

Implementing IBM FlashSystem 840

Karen Orlando

Detlef Helmbrecht

Jon Herd

Carsten Larsen

Matt Levan



Storage



International Technical Support Organization

Implementing IBM FlashSystem 840

July 2015

Note: Before using this information and the product it supports, read the information in “Notices” on page ix.

Third Edition (July 2015)

This edition applies to the IBM FlashSystem 840, Release 1.3.

© Copyright International Business Machines Corporation 2014, 2015. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Notices	ix
Trademarks	x
IBM Redbooks promotions	xi
Preface	xiii
Authors	xv
Now you can become a published author, too!	xvi
Comments welcome	xvii
Stay connected to IBM Redbooks	xvii
Summary of changes	xix
July 2015, Third Edition	xix
September 2014, Second Edition	xx
Chapter 1. FlashSystem storage introduction	1
1.1 FlashSystem storage overview	2
1.2 Why Flash matters	2
1.3 IBM FlashSystem family: Product differentiation	4
1.4 Technology and architectural design overview	5
1.4.1 IBM Variable Stripe RAID and two-dimensional flash RAID overview	7
1.5 Variable Stripe RAID	8
1.6 How VSR works	9
1.7 Two-dimensional (2D) Flash RAID	10
Chapter 2. IBM FlashSystem 840 architecture	13
2.1 Introduction to the IBM FlashSystem 840 architecture	14
2.1.1 IBM FlashSystem 840 capacity	14
2.1.2 IBM FlashSystem 840 performance and latency	16
2.1.3 IBM FlashSystem 840 power requirements	17
2.1.4 IBM FlashSystem 840 physical specifications	17
2.1.5 IBM FlashSystem 840 reliability and serviceability	18
2.1.6 IBM FlashSystem 840 scalability	18
2.1.7 IBM FlashSystem 840 protocol support	19
2.1.8 IBM FlashSystem 840 encryption support	20
2.1.9 IBM FlashSystem models 820 and 840 comparison	21
2.1.10 IBM FlashSystem 840 management	22
2.2 IBM FlashSystem 840 architecture	23
2.2.1 IBM FlashSystem 840 architecture overview	23
2.2.2 IBM FlashSystem 840 hardware components	25
2.2.3 IBM FlashSystem 840 canisters	25
2.2.4 IBM FlashSystem 840 interface cards	26
2.2.5 IBM FlashSystem 840 flash modules	30
2.2.6 IBM FlashSystem 840 battery modules	33
2.3 IBM FlashSystem 840 administration and maintenance	35
2.3.1 IBM FlashSystem 840 serviceability and software enhancements	35
2.3.2 IBM FlashSystem 840 system management	35
2.4 IBM FlashSystem 840 support matrix	41
2.5 IBM FlashSystem 840 IBM product integration overview	42

2.5.1 IBM Spectrum Virtualize: SAN Volume Controller	42
2.5.2 IBM Storwize V7000 storage array	43
2.5.3 IBM PureFlex System and IBM PureSystems.	44
2.5.4 IBM DB2 database environments	44
2.5.5 IBM Spectrum Scale	44
2.5.6 IBM TS7650G ProtecTIER	45
Chapter 3. Planning	47
3.1 Installation prerequisites	48
3.1.1 General information	48
3.1.2 Completing the hardware location chart	49
3.2 Planning cable connections	50
3.2.1 Management port connections	50
3.2.2 Interface card connections	51
3.3 Planning for power	56
3.4 Planning for configuration	57
3.5 Call Home option.	58
3.6 TCP/IP requirements.	58
3.7 Planning for encryption	60
3.8 Checking your web browser settings for the management GUI	61
3.9 Licensing.	63
3.10 Supported hosts and operating system considerations	63
Chapter 4. Installation and configuration	65
4.1 First-time installation	66
4.1.1 Installing the hardware	66
4.2 Cabling the system	69
4.2.1 Cabling for Fibre Channel.	69
4.2.2 Cabling for FCoE	71
4.2.3 Cabling for iSCSI	71
4.2.4 Cabling for QDR InfiniBand.	72
4.2.5 FC cable type	72
4.2.6 Ethernet management cabling	72
4.2.7 Power requirements	73
4.2.8 Cooling requirements	73
4.2.9 Cable connector locations.	73
4.3 Initializing the system	74
4.3.1 About encryption.	74
4.3.2 Prepare for initialization using InitTool	76
4.3.3 Initializing the system through the web management interface	91
4.4 RAID storage modes.	104
4.4.1 Changing RAID modes	104
4.5 Connectivity guidelines for improved performance	107
4.5.1 Interface card configuration guidelines	107
4.5.2 Host adapter guidelines	108
4.5.3 Cabling guidelines.	108
4.5.4 Zoning guidelines	109
Chapter 5. IBM FlashSystem 840 client host attachment and implementation.	111
5.1 Host implementation and procedures	112
5.2 Host connectivity	112
5.2.1 Fibre Channel SAN attachment	112
5.2.2 Fibre Channel direct attachment.	113
5.2.3 General Fibre Channel attachment rules	114

5.3 Operating system connectivity and preferred practices	114
5.3.1 FlashSystem 840 sector size	114
5.3.2 File alignment for the best RAID performance	114
5.3.3 IBM AIX and FlashSystem 840	115
5.3.4 FlashSystem 840 and Linux client hosts	119
5.3.5 FlashSystem 840 and Microsoft Windows client hosts	121
5.3.6 FlashSystem 840 and client VMware ESX hosts	125
5.3.7 FlashSystem 840 and IBM SAN Volume Controller or Storwize V7000	126
5.3.8 FlashSystem iSCSI host attachment	126
5.3.9 FlashSystem iSCSI configuration	126
5.3.10 Windows 2008 R2 and Windows 2012 iSCSI attachment	127
5.3.11 Linux iSCSI attachment	131
5.4 Miscellaneous host attachment	134
5.4.1 FlashSystem 840 and Solaris client hosts	134
5.4.2 FlashSystem 840 and HP-UX client hosts	142
5.5 FlashSystem 840 preferred read and configuration examples	143
5.5.1 FlashSystem 840 deployment scenario with preferred read	143
5.5.2 Implementing preferred read	145
5.5.3 Linux configuration file multipath.conf example	155
5.5.4 Example of a VMWare configuration	160
5.6 FlashSystem 840 and Easy Tier	160
5.7 Troubleshooting	161
5.7.1 Troubleshooting Linux InfiniBand configuration issues	161
5.7.2 Linux fdisk error message	162
5.7.3 Changing FC port properties	163
5.7.4 Changing iSCSI port properties	163
Chapter 6. Using the IBM FlashSystem 840	165
6.1 Overview of IBM FlashSystem 840 management tools	166
6.1.1 Access to the graphical user interface	166
6.1.2 Graphical user interface layout	167
6.1.3 Navigation	168
6.1.4 Multiple selections	170
6.1.5 Status indicators	171
6.2 Monitoring menu	172
6.2.1 Monitoring System menu	172
6.2.2 Monitoring events	185
6.2.3 Monitoring performance menu	196
6.3 Volumes	204
6.3.1 Navigating to the Volumes menu	204
6.3.2 Volumes menu	205
6.3.3 Volumes by Host menu	211
6.4 Hosts	216
6.4.1 Navigating to the Hosts menu	216
6.4.2 Volumes by Host	224
6.5 Access menu	225
6.5.1 Navigating to the Access menu	225
6.5.2 Users menu	226
6.5.3 Access CLI by using PuTTY	229
6.5.4 User groups	232
6.5.5 Audit log menu	236
Chapter 7. Configuring settings	239

7.1 Settings menu	240
7.1.1 Navigating to the Settings menu	240
7.1.2 Notifications menu	241
7.1.3 Security menu	243
7.1.4 Network menu	254
7.1.5 Support menu	256
7.1.6 System menu	259
7.2 Service Assistant Tool	271
7.2.1 Accessing Service Assistant Tool	271
7.2.2 Log in to Service Assistant Tool	272
Chapter 8. Product integration	275
8.1 Running the FlashSystem 840 with Spectrum Virtualize - SAN Volume Controller	276
8.1.1 IBM System Storage SAN Volume Controller introduction	276
8.1.2 SAN Volume Controller architecture and components	279
8.1.3 SAN Volume Controller hardware options	281
8.1.4 IBM Spectrum Virtualize - SAN Volume Controller advanced functionality.	284
8.2 SAN Volume Controller connectivity to FlashSystem 840.	286
8.2.1 SAN Volume Controller FC cabling to SAN	287
8.2.2 SAN zoning and port designations	288
8.2.3 Port designation recommendations.	289
8.2.4 Verifying FlashSystem 840 connectivity in SAN Volume Controller	291
8.2.5 Import/export	303
8.3 Integrating FlashSystem 840 and SAN Volume Controller considerations	303
8.4 Integrating FlashSystem 840 and IBM Storwize V7000 considerations	304
Chapter 9. Use cases and solutions	305
9.1 Introduction to the usage cases	306
9.2 Tiering	307
9.2.1 Easy Tier or block-level tiering	307
9.2.2 Information Life Management or file-level tiering	309
9.3 Preferred read	309
9.3.1 Implementing preferred read	313
9.4 Flash only	315
9.5 Comparison	316
Chapter 10. Hints and tips	319
10.1 Encryption hints	320
10.2 System check	320
10.2.1 Checking the Fibre Channel connections	320
10.3 Host attachment hints	321
10.3.1 Fibre Channel link speed	321
10.3.2 Host is in a degraded state	322
10.3.3 FlashSystem port status	322
10.3.4 AIX multipathing	322
10.3.5 Direct attach hints	323
10.4 General guidelines for testing a specific configuration	323
10.4.1 Save the default configuration	324
10.4.2 Test scenarios	324
10.4.3 Data center environment	325
10.4.4 Secure erase of data	325
10.4.5 Performance data gathering basics	325
10.5 Troubleshooting	329
10.5.1 Troubleshooting prerequisites	329

10.5.2 User interfaces for servicing your system	331
10.5.3 Event reporting	334
10.5.4 Resolving a problem	337
10.6 IBM System Storage Interoperation Center (SSIC)	338
Appendix A. SAN preferred practices for	
16 Gbps	339
Sixteen Gbps Fibre Channel benefits	340
IBM System Storage b-type Gen 5 SAN product overview	341
SAN design basics	347
Topologies	347
Inter-switch link	348
Intercluster links	349
Device placement	349
Fan-in ratios and oversubscription	350
FCoE as a top of rack (ToR) solution	352
Data flow considerations	352
Redundancy and resiliency	354
Distance and Fibre Channel over IP (FCIP) design preferred practices	356
Monitoring	359
Scalability and supportability	361
Implementation	363
Initial setup	363
Related publications	369
IBM Redbooks	369
Other publications	369
Online resources	370
Help from IBM	370

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.


COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®	IBM FlashCore™	PureFlex®
BigInsights™	IBM FlashSystem®	PureSystems®
DB2®	IBM Flex System®	Real-time Compression™
developerWorks®	IBM SmartCloud®	Redbooks®
DS8000®	IBM Spectrum™	Redbooks (logo)  ®
Easy Tier®	IBM Spectrum Storage™	Storwize®
FICON®	MicroLatency®	System Storage®
FlashCopy®	NetView®	Tivoli®
FlashSystem™	Power Systems™	Tivoli Enterprise Console®
GPFS™	PowerPC®	Variable Stripe RAID™
IBM®	ProtecTIER®	XIV®

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.

Find and read thousands of IBM Redbooks publications

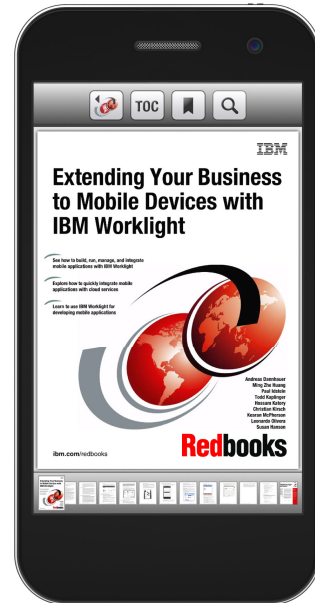
- ▶ Search, bookmark, save and organize favorites
- ▶ Get up-to-the-minute Redbooks news and announcements
- ▶ Link to the latest Redbooks blogs and videos

Get the latest version of the Redbooks Mobile App



Download
Now

iOS



Promote your business in an IBM Redbooks publication

Place a Sponsorship Promotion in an IBM® Redbooks® publication, featuring your business or solution with a link to your web site.

Qualified IBM Business Partners may place a full page promotion in the most popular Redbooks publications. Imagine the power of being seen by users who download millions of Redbooks publications each year!



ibm.com/Redbooks

About Redbooks → Business Partner Programs

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

Almost all technological components in the data center are getting faster: central processing units, networks, storage area networks (SANs), and memory. All of them have improved their speed by a minimum of 10X; some of them by 100X, for example, data networks. However, spinning disk performance has only increased by 1.2 times.

IBM® FlashSystem™ 840 version 1.3 closes this gap. The FlashSystem 840 is optimized for the data center to enable organizations of all sizes to strategically harness the value of stored data. It provides flexible capacity and extreme performance for the most demanding applications, including virtualized or bare-metal online transaction processing (OLTP) and online analytical processing (OLAP) databases, virtual desktop infrastructures (VDI), technical computing applications, and cloud environments. The system accelerates response times with IBM MicroLatency® access times as low as 90 µs write latency and 135 µs read latency to enable faster decision making.

The introduction of a low capacity 1 TB flash module allows the FlashSystem 840 to be configured in capacity points as low as 2 TB in protected RAID 5 mode. Coupled with 10 GB iSCSI, the FlashSystem is positioned to bring extreme performance to small and medium-sized businesses (SMB) and growth markets.

Implementing the IBM FlashSystem® 840 provides value that goes beyond those benefits that are seen on disk-based arrays. These benefits include better user experience, server and application consolidation, development cycle reduction, application scalability, data center footprint savings, and improved price performance economics.

This IBM Redbooks® publication discusses IBM FlashSystem 840 version 1.3. It provides in-depth knowledge of the product architecture, software and hardware, its implementation, and hints and tips. Also illustrated are use cases that show real-world solutions for tiering, flash-only, and preferred read, as well as examples of the benefits gained by integrating the FlashSystem storage into business environments.

Also described are product integration scenarios running the IBM FlashSystem 840 with the IBM SAN Volume Controller, and the IBM Storwize® family of products such as V7000, V5000, and the V3700, as well as considerations when integrating with the IBM FlashSystem 840. The preferred practice guidance is provided for your FlashSystem environment with IBM 16 Gbps b-type products and features, focusing on Fibre Channel design.

This book is intended for pre-sales and post-sales technical support professionals and storage administrators, and for anyone who wants to understand and learn how to implement this exciting technology.

The following IBM Spectrum™ Storage family of offerings is discussed and referenced in this Redbooks publication:

IBM Spectrum Storage

The IBM Spectrum Storage™ family is based on proven technologies and designed specifically to simplify storage management, scale to keep up with data growth, and optimize data economics. It represents a new, more agile way of storing data, and helps organizations prepare themselves for new storage demands and workloads. The software defined storage solutions included in the IBM Spectrum Storage family can help organizations simplify their storage infrastructures, cut costs, and start gaining more business value from their data.

For details about the entire IBM Spectrum Storage family, see the following website:

<http://www.ibm.com/systems/storage/spectrum>

IBM Spectrum Control

Provides efficient infrastructure management for virtualized, cloud, and software-defined storage to simplify and automate storage provisioning, capacity management, availability monitoring, and reporting.

The functionality of IBM Spectrum Control is provided by IBM Data and Storage Management Solutions and includes functionality delivered by IBM SmartCloud® Virtual Storage Center, IBM Tivoli® Storage Productivity Center, IBM Storage Integration Server, and others.

For more information, see the IBM Data Management and Storage Management website:

<http://www.ibm.com/software/tivoli/csi/cloud-storage>

IBM Spectrum Virtualize

IBM Spectrum Virtualize is industry-leading storage virtualization that enhances existing storage to improve resource utilization and productivity in order to achieve a simpler, more scalable, and cost-efficient IT infrastructure.

The functionality of IBM Spectrum Virtualize is provided by IBM SAN Volume Controller.

For details about IBM Spectrum Virtualize: SAN Volume Controller, see the following website:

<http://www.ibm.com/systems/storage/software/virtualization/svc>

IBM Spectrum Scale

IBM Spectrum Scale is a proven, scalable, high-performance data and file management solution, based on IBM General Parallel File System or IBM GPFS™. IBM Spectrum Scale technology is a high-performance enterprise file management platform, and it can help you move beyond simply adding storage to optimize data management.

For more information, see the following IBM Spectrum Scale website:

<http://www.ibm.com/systems/storage/spectrum/scale>

Authors

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.



Karen Orlando is a Project Leader at the International Technical Support Organization, Tucson Arizona Center. Karen has over 25 years in the IT industry with extensive experience in open systems management, and Information and Software development of IBM hardware and software storage. She holds a degree in Business Information Systems from the University of Phoenix and is Project Management Professional (PMP) certified since 2005.



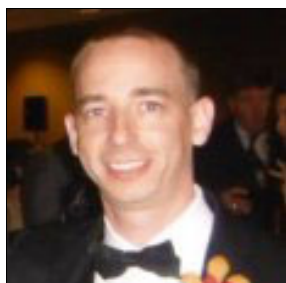
Detlef Helmbrecht is an Advanced Technical Skills (ATS) IT Specialist working for the IBM Systems & Technology Group. He is located in the European Storage Competence Center (ESCC), Germany. Detlef has over 25 years of experience in IT, performing numerous different roles, including software design, sales, and solution architect. His areas of expertise include high-performance computing (HPC), disaster recovery, archiving, application tuning, and FlashSystem.



Jon Herd is an IBM Storage Technical Advisor working for the European Storage Competence Center (ESCC), Germany. He covers UK and Ireland, advising clients on a portfolio of IBM storage products including FlashSystem products. Jon has been in IBM for more than 40 years and has held various technical roles, including EMEA level support on mainframe servers and technical education development. He holds IBM certifications in Supporting IT Solutions at expert level and Actualizing IT Solutions experienced level. He is also a certified member of the British Computer Society (MBCS CITP) and the Institute of Engineering and Technology (MIET).



Carsten Larsen is an IBM Certified Senior IT Specialist working for the Technical Services Support organization in IBM Denmark, delivering consulting services to IBM clients within the storage arena. Carsten joined IBM in 2007, leaving behind a job at HP where he worked with storage arrays and UNIX for 10 years, holding, among others, the HP certification: Master Certified Systems Engineer (MASE). While working for IBM, Carsten obtained Brocade BCFP and BCSD certifications and NetApp NCDA and NCIE certifications. Carsten is the author of a number of IBM Redbooks publications related to these product certifications.



Matt Levan is an IBM FlashSystem Corporate Solutions Architect. Matt has been involved in IT and Storage Technologies for over 15 years. He has performed innumerable roles, including storage administration, technical support, and solutions architect. Matt spent many years at technology-leading companies, such as Novus Consulting Group, VeriSign, EMC, and Innovative Data Solutions. Matt's current role is to help worldwide field technical sellers for IBM Flash Storage products. His primary responsibilities are to provide technical sales engineering, competitive winning strategies, and cohesive technical sales solutions to a multitude of IBM internal organizations and IBM clients worldwide.

The following authors wrote the second edition, *Implementing IBM FlashSystem 840*, SG24-8189-01, which was published on 16 September 2014:

- ▶ Chip Elmlad
- ▶ Detlef Helmbrecht
- ▶ Carsten Larsen
- ▶ Matt Levan
- ▶ Karen Orlando

Thanks to the following people for their contributions to this project:

Jim Cioffi, Mark Flemming, Brian Groff, Kim Miller
IBM Systems

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at this website:

<http://www.ibm.com/redbooks/residencies.html>

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- ▶ Use the online **Contact us** review Redbooks form found at this website:

<http://www.ibm.com/redbooks>

- ▶ Send your comments in an email to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- ▶ Follow us on Twitter:

<https://twitter.com/ibmredbooks>

- ▶ Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- ▶ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>

- ▶ Stay current on recent Redbooks publications with RSS Feeds:

<http://www.redbooks.ibm.com/rss.html>

Summary of changes

This section describes the technical changes that were made in this edition of the book and in previous editions. This edition might include minor corrections and editorial changes that are not identified.

Summary of Changes
for SG24-8189-02
for Implementing IBM FlashSystem 840
as created or updated on July 1, 2015.

July 2015, Third Edition

This revision reflects the addition, deletion, or modification of new and changed information described below for FlashSystem 840 release 1.3.

New information

This revision includes the following new information:

1300-watt power supply

The new 1300 W power supply for high-line voltage provides IBM FlashSystem 840 and FlashSystem V840 systems (as well as the new IBM FlashSystem 900 and FlashSystem v9000 systems) with a high-power alternative to run at maximum performance for longer durations during power supply servicing, resulting in more predictable performance under unexpected failure conditions. While FlashSystem 840 already offers improved data economics, upgrading to a 1300 W power supply facilitates performance at a higher level, thereby increasing return on investment.

FlashSystem 840 GUI management

Additional FlashSystem 840 management software improvements make system management and performance health monitoring even more effective in FlashSystem's already intuitive GUI. Additional features include:

- ▶ 300 days of performance data with pan and zoom capability to better identify and research trends
- ▶ Five predefined graphs showing System IOPS, Latency, Bandwidth, Total Port IOPS, and queue depth

Encryption

FlashSystem 840 added new encryption functions:

- ▶ Hot Encryption Activation: Adding an encryption license to a previously initialized system
- ▶ Encryption Rekey: Changing the encryption key on a previously initialized system

Battery reconditioning

This release enables a battery reconditioning feature that calibrates the gauge that reports the amount of charge on the batteries. This feature ensures that the batteries are kept at the optimal level to support the IBM FlashSystem 840, and to conduct a controlled and orderly shutdown if there is an external power outage.

Changed information

This revision also includes additional feedback from the field.

September 2014, Second Edition

This revision reflects the addition, deletion, or modification of new and changed information described below for FlashSystem 840 release 1.2.

New information

IBM enhanced the flexibility of the IBM FlashSystem 840 with new, more granular capacity offerings starting as low as 2 TB, providing an attractive entry point so that you can start with smaller, application-specific flash deployments and scale up to Tier 1 disk replacement as business needs grow. The IBM FlashSystem 840 is also cloud-optimized with support for 10 Gbps iSCSI interfaces, allowing private cloud environments and Managed Service Providers (MSPs) to take advantage of high-performance storage inside existing Ethernet-based infrastructures.

This book includes the following new information based on Release 1.2:

- ▶ One TB flash modules
- ▶ Capacity points:
 - RAID 0: 2, 4, 6, 8, 10, 12, 16, 24, 32, and 48 TB
 - RAID 5: 2, 4, 6, 8, 10, 12, 16, 20, 24, 32, and 40 TB
- ▶ Available in multiple new interface choices:
 - Ten Gb Fibre Channel over Ethernet (FCoE)
 - Ten Gb Internet Small Computer System Interface (iSCSI)

Changed information

This revision also includes the following updates:

- ▶ Added additional host connectivity guidelines
- ▶ Added feedback from the field



FlashSystem storage introduction

Flash technology in the data center is too relevant to be ignored for a few simple reasons:

- ▶ Since its introduction, flash storage has improved across all metrics: higher performance, density, and reliability, all of which translate to improved business efficiency.
- ▶ Flash cost per capacity and cost per transaction relative to hard disk-based storage make it extremely attractive to businesses that are attempting to maintain pace in a 24x7 competitive marketplace.
- ▶ Flash is easily integrated into existing data center environments and provides an instant boost to the mission critical applications.

Although flash in storage is pervasive in the data center, its implementation varies considerably among competitors and technologies. Some use it as a simple cache accelerator while others implement it as yet another permanent data tier. The reality is that flash only matters when two conditions in the data center are met:

- ▶ Flash eliminates I/O bottlenecks while generating higher levels of application efficiency (improved performance).
- ▶ Storage economics are improved by its use. That is, it provides lower total cost of ownership (TCO) and faster return on investment (ROI) to the existing environment (enables new business opportunities).

The IBM FlashSystem storage delivers high performance, efficiency, and reliability for shared enterprise storage environments. It helps clients address performance issues with their most important applications and infrastructure.

This chapter provides an introduction to the IBM FlashSystem storage system and its core value, benefits, and technological advantages.

1.1 FlashSystem storage overview

Flash technology fundamentally changed the paradigm for IT systems, enabling new use cases and unlocking the scale of enterprise applications. Flash technology enhances the performance, efficiency, reliability, and design of essential enterprise applications and solutions by addressing the bottleneck in the IT process (data storage), enabling truly optimized information infrastructure.

The IBM FlashSystem shared flash storage systems offer affordable, high-density, ultra low-latency, highly reliable and scalable performance in a storage device that is both space efficient and power efficient. IBM Flash products, which can either augment or replace traditional hard disk drive (HDD) storage systems in enterprise environments, empower applications to work faster and scale further.

In addition to optimizing performance, the IBM FlashSystem family helps bring enterprise reliability and macro efficiency to the most demanding data centers so that businesses can see the following benefits:

- ▶ Reduce customer complaints by improving application response time
- ▶ Service more users with less hardware
- ▶ Reduce I/O wait and response times of critical applications
- ▶ Simplify solutions
- ▶ Reduce power and floor space requirements
- ▶ Speed up applications, therefore enhancing the pace of business
- ▶ Improve the utilization of the existing infrastructure
- ▶ Extend the existing infrastructure
- ▶ Mitigate risk

From the client business perspective, an IBM FlashSystem provides benefits and value in four essential areas:

Extreme performance

Enables businesses to unleash the power of performance, scale, and insight to drive services and products to market faster

MicroLatency

Achieves competitive advantage through applications that enable faster decision making due to microsecond response times

Macro efficiency

Decreases costs by getting more from the efficient use of the IT staff, IT applications, and IT equipment due to the efficiencies flash brings to the data center

Enterprise reliability

Enhances customer experience through durable and reliable designs that use enterprise class flash and patented data protection technology

1.2 Why Flash matters

Flash is a vibrant and fast growing technology. Clients are looking to solve data center problems, optimize applications, reduce costs, and grow their businesses.

Here are several reasons why Flash is a *must* in every data center, and why an IBM FlashSystem changes the storage economics:

- ▶ Reduces application and server licensing costs, especially those related to databases and virtualization solutions.
- ▶ Improves application efficiency, that is, an application's ability to process, analyze, and manipulate more information, faster.
- ▶ Improves server efficiency. Helps you get more out of your existing processors, use less RAM per server, and consolidate operations by having server resources spend more time processing data as opposed to waiting for data.
- ▶ Improves storage operations. Helps eliminate costly application tuning, wasted developer cycles, storage array hot spots, array tuning, and complex troubleshooting. Decreases floor space usage and energy consumption by improving overall storage environment performance.
- ▶ Enhances performance for critical applications by providing the lowest latency in the market.

Almost all technological components in the data center are getting faster: central processing units, network, storage area networks (SANs), and memory. All of them have improved their speeds by a minimum of 10X; some of them by 100X, such as data networks. However, spinning disk has only increased its performance 1.2 times.

The IBM FlashSystem 840 provides benefits that include a better user experience, server and application consolidation, development cycle reduction, application scalability, data center footprint savings, and improved price performance economics.

Flash improves the performance of applications that are critical to the *user experience*, such as market analytics and research applications, trading and data analysis interfaces, simulation, modeling, rendering, and so on. Server and application consolidation is possible due to the increased process utilization resulting from the low latency of flash memory, which enables a server to load more users, more databases, and more applications. Flash provides or gives back *time* for further processing within the existing resources of such servers. Clients soon realize that there is no need to acquire or expand server resources as often or as soon as was previously expected.

Development cycle reduction is possible because developers spend less time designing an application to work around the inefficiencies of HDDs and less time tuning for performance.

Data center footprint savings are due to the high density and high performance per density flash solutions that are replacing racks of spinning HDDs. Reducing the data center footprint also translates into power and cooling savings, making flash one of the greenest technologies for the data center.

Improved price: Performance economics are due to the low cost for performance from the IBM FlashSystem. The cost savings result from deploying fewer storage enclosures, fewer disk drives, fewer servers with fewer processors, and less RAM while using less power, space, and cooling. Flash is one of the best tools for the data center manager for improving data center economics.

1.3 IBM FlashSystem family: Product differentiation

Flash is used widely in the data center, either within a server (Peripheral Component Interconnect Express (PCIe) cards) or internal solid-state drives (SSDs), in storage arrays (hybrid or all-flash), appliances, or platform solutions (hardware/software/network). Flash can be used as cache or as a data tier. Due to the vast and wide adoption of flash, there are a number of different flash architectures and, therefore, criteria that can be applied to compare flash options. See Figure 1-1.

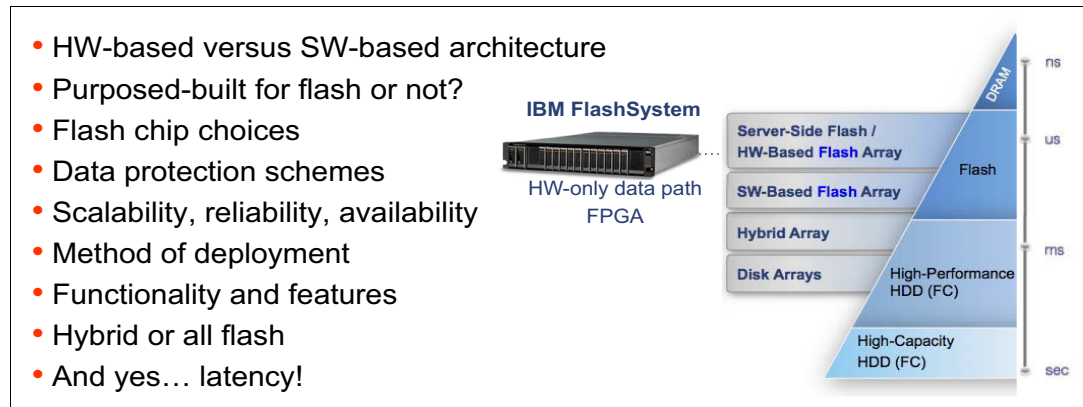


Figure 1-1 The different deployments of Flash

Most storage vendors use and promote flash. The difference is how it is implemented, and the impact that such implementation has on the economics (cost reduction and revenue generation) for clients.

Flash technology is used to eliminate the storage *performance bottleneck*. The IBM FlashSystem family is a key shared-storage market leader and it provides extremely low latency and consistent response times. It is purpose-built and designed from the ground up for flash.

Some other vendors create flash appliances based on commodity server platforms and use software-heavy stacks. Some suppliers use hardware technologies designed and created for disk, not flash. Some hybrid arrays combine existing storage designs, spinning HDDs, and solid-state disk (SSD). The IBM storage portfolio includes SSD and flash on various storage platforms; however, these alternate solutions do not have the same low latency (MicroLatency) as the hardware-accelerated FlashSystem.

IBM FlashSystem family versus SSD-based storage arrays

Flash memory technologies appeared in the traditional storage systems some time ago. These SSD-based storage arrays help to successfully address the challenge of increasing I/Os per second needed by applications, and the demand for lower response times in particular tasks. An implementation example is the IBM Easy Tier® technology. For an overview of this technology, see “Easy Tier” on page 284.

However, these technologies typically rely on flash in the format of Fibre Channel (FC), serial-attached SCSI (SAS), or Serial Advanced Technology Attachment (SATA) disks, placed in the same storage system as traditional spinning disks, and using the same resources and data paths. This approach can limit the advantages of flash technology due to the limitations of traditional disk storage systems.

For more information about IBM Easy Tier, see the following publications:

- ▶ Chapter 7, “Advanced features for storage efficiency” in *Implementing the IBM System Storage SAN Volume Controller V7.4*, SG24-7933
- ▶ Chapter 11, “IBM System Storage Easy Tier function” in *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- ▶ *IBM System Storage DS8000 Easy Tier*, REDP-4667

The IBM FlashSystem storage provides a hardware-only data path that realizes all of the potential of flash memory. These systems are different from traditional storage systems, both in the technology and usage.

An SSD device with an HDD disk form factor has flash memory that is put into a carrier or tray. This carrier is inserted into an array, such as an HDD. The speed of storage access is limited by the following technology because it adds latency and cannot keep pace with flash technology:

- ▶ Array controllers and software layers
- ▶ SAS controllers and shared bus
- ▶ Tiering and shared data path
- ▶ Form factor enclosure

The IBM FlashSystem products are fast and efficient. The hardware-only data path has a minimum number of software layers, which are mostly firmware components, and management software that is separated from the data path (out-of-band). The only other family of products with hardware-only access to flash technology is the PCI Express (PCIe) flash product family, where products are installed into a dedicated server. With the appearance of the IBM FlashSystem, the benefits of PCIe flash products to a single server can now be shared by many servers.

1.4 Technology and architectural design overview

The IBM FlashSystem, with an all-hardware data path using field programmable-gate array (FPGA) modules, is engineered to deliver the lowest possible latency. The modules incorporate proprietary flash controllers and use numerous patented technologies. The FlashSystem controllers have a proprietary logic design, firmware, and system software.

There are no commodity 2.5-inch SSDs, PCIe cards, or any other significant non IBM assemblies within the system. The flash chips, FPGA chips, processors, and other semiconductors in the system are carefully selected to be consistent with the purpose-built design, which is designed from the ground up for high performance, reliability, and efficiency.

The IBM FlashSystem storage systems offer the following notable architectural concepts:

- ▶ Hardware-only data path.
- ▶ Use of FPGAs extensively.
- ▶ Field-upgradable hardware logic.
- ▶ Less expensive design cycle.
- ▶ Extremely high degree of parallelism.
- ▶ Intelligent flash modules.
- ▶ Distributed computing model.

- ▶ Low-power IBM PowerPC® processors (PPCs).
- ▶ Interface and flash processors run thin real-time operating systems.
- ▶ The management processor communicates with the interface and flash processors through an internal network.
- ▶ Minimal management communication.

Hardware-only data path

The hardware-only data path design of the IBM FlashSystem eliminates software layer latency. To achieve extremely low latencies, the IBM FlashSystem advanced software functions are carefully assessed and implemented on a limited basis. For environments requiring advanced storage services, implementing the IBM FlashSystem with the IBM SAN Volume Controller, which delivers the function of IBM Spectrum Virtualize technology, can offer an unmatched combination of performance, low latency, and rich software functionality.

Notes:

IBM SAN Volume Controller delivers the functions of IBM Spectrum Virtualize, part of the IBM Spectrum Storage family, and has been improving infrastructure flexibility and data economics for more than 10 years. Its innovative data virtualization capabilities provide the foundation for the entire IBM Storwize family. SAN Volume Controller provides the latest storage technologies for unlocking the business value of stored data, including virtualization and IBM Real-time Compression™.

In addition, the latest system includes the new SAN Volume Controller Data Engine to help support the massive volumes of data created by today's demanding enterprise applications. SAN Volume Controller is designed to deliver unprecedented levels of efficiency, ease of use and dependability for organizations of all sizes.

In the IBM FlashSystem, data traverses the array controllers through FPGAs and dedicated, low-power CPUs. There are no wasted cycles on *interface* translation, protocol control, or tiering.

The IBM FlashSystem, with an all-hardware data path design, has an internal architecture that is different from other hybrid (SSD + HDD) or SSD-only-based disk systems.

Flash chips

The *flash chip* is the basic storage component of the flash module. A maximum of 80 enterprise multi-level cell (eMLC) flash chips can exist for each flash module. Combining flash chips of different flash technologies is not supported in the same flash module or storage system to maintain consistent wearing and reliability.

Gateway interface FPGA

The gateway interface FPGA is responsible for providing I/O to the flash module and direct memory access (DMA) path. It is on the flash module and has two connections to the backplane.

Flash controller FPGA

The flash controller FGPA of the flash module is used to provide access to the flash chips and is responsible for the following functions:

- ▶ Provides data path and hardware I/O logic
- ▶ Uses lookup tables and a write buffer

- ▶ Controls 20 flash chips
- ▶ Operates independently of other controllers
- ▶ Maintains write ordering and layout
- ▶ Provides write setup
- ▶ Maintains garbage collection
- ▶ Provides error handling

Figure 1-2 shows the flash controller design details.

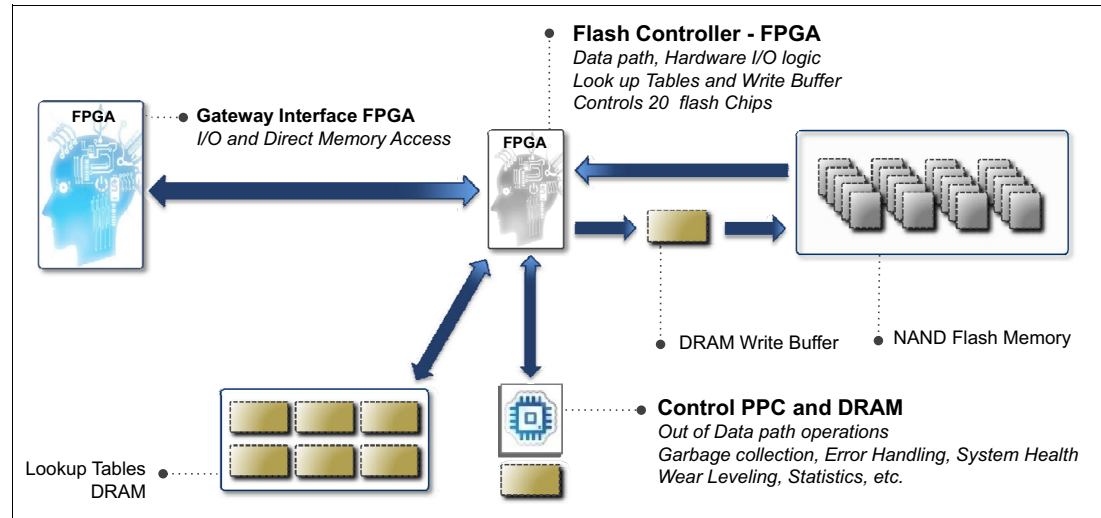


Figure 1-2 IBM FlashSystem controller details

The concurrent operations performed on the flash chips include moving data in and out of the chip through DMA, and by internally moving data and performing erasures. While actively transferring user data in the service of host-initiated I/O, the system can simultaneously run garbage collection activities without affecting the I/O. The ratio of transparent background commands running concurrent to active data transfer commands is 7:1.

There are a maximum of four flash controllers per flash module: two for each primary board and two for each expansion board.

1.4.1 IBM Variable Stripe RAID and two-dimensional flash RAID overview

Storage systems of any kind are typically designed to perform two main functions: to store and protect data. The IBM FlashSystem includes the following options for data protection. Table 1-1 on page 8 shows the various methods of protection.

- ▶ RAID data protection:
 - IBM Variable Stripe RAID™
 - Two-dimensional (2D) Flash RAID
- ▶ Flash memory protection methods
- ▶ Optimized RAID rebuild times

Table 1-1 Various types of IBM FlashSystem protection

Layer	Managed by	Protection
System-level RAID 5	Centralized RAID controllers	Module failure
Module-level RAID 5	Each module across the chips	Chip failure and page failure
Module-level Variable Stripe RAID	Each module across the chips	Subchip, chip, or multi-chip failure
Chip-level error correction code (ECC)	Each module using the chips	Bit and block error

Note: The proprietary two-dimensional (2D) Flash RAID data protection scheme of the IBM FlashSystem 840 storage system combines system-level RAID 5 and module-level Variable Stripe RAID (not just module-level RAID).

1.5 Variable Stripe RAID

Variable Stripe RAID (VSR) is a unique IBM technology that provides data protection of the memory page, block, or whole chip, which eliminates the necessity to replace a whole flash module in a single memory chip failure or plane failures. This, in turn, expands the life and endurance of flash modules and reduces considerably maintenance events throughout the life of the system.

VSR provides high redundancy across chips within a flash module. RAID is implemented at multiple addressable segments within chips, in a 9+1 RAID 5 fashion, and it is controlled at the flash controller level (four in each flash module). Due to the massive parallelism of DMA operations controlled by each FPGA and parallel access to chip sets, dies, planes, blocks, and pages, the implementation of VSR has minimal impact on performance.

The following information describes some of the most important aspects of VSR implementation:

- ▶ VSR is managed and controlled by each of the four flash controllers within a single module.
- ▶ A flash controller is in charge of only 20 flash chips.
- ▶ Data is written on flash pages of 8 KB and erased in 1 MB flash blocks.
- ▶ VSR is implemented and managed at flash chip *plane* levels.
- ▶ There are 16 planes per chip.
- ▶ Before a plane fails, at least 256 flash blocks within a plane must be deemed *failed*.
- ▶ A plane can also fail in its entirety.
- ▶ Up to 64 planes can fail before a whole module is considered failed.
- ▶ Up to four chips can fail before a whole module is considered failed.
- ▶ When a flash module is considered failed, 2D Flash RAID takes control of data protection and recovery.
- ▶ When a plane or a chip fails, VSR activates to protect data while maintaining system-level performance and capacity.

1.6 How VSR works

Variable Stripe RAID is an IBM patented technology. It includes but is more advanced than a simple RAID of flash chips. Variable Stripe RAID introduces two key concepts:

- ▶ The RAID stripe is not solely across chips; it spans across flash layers.
- ▶ The RAID stripe can automatically vary based on observed flash plane failures within a flash module. For example, stripes are not fixed at 9+1 RAID 5 stripe members, but they can go down to 8+1, 7+1, or even 6+1 based on plane failures.

This ability to protect the data at *variable* stripes effectively maximizes flash capacity even after flash component failures. Figure 1-3 shows an overview of the IBM FlashSystem Variable Stripe RAID (VSR).

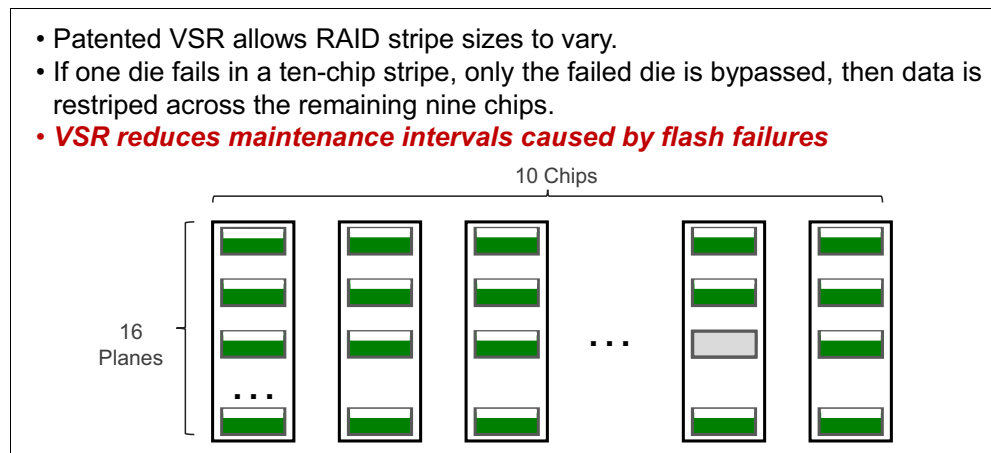


Figure 1-3 IBM FlashSystem Variable Stripe RAID (VSR)

Figure 1-4 shows the benefits of IBM VSR.

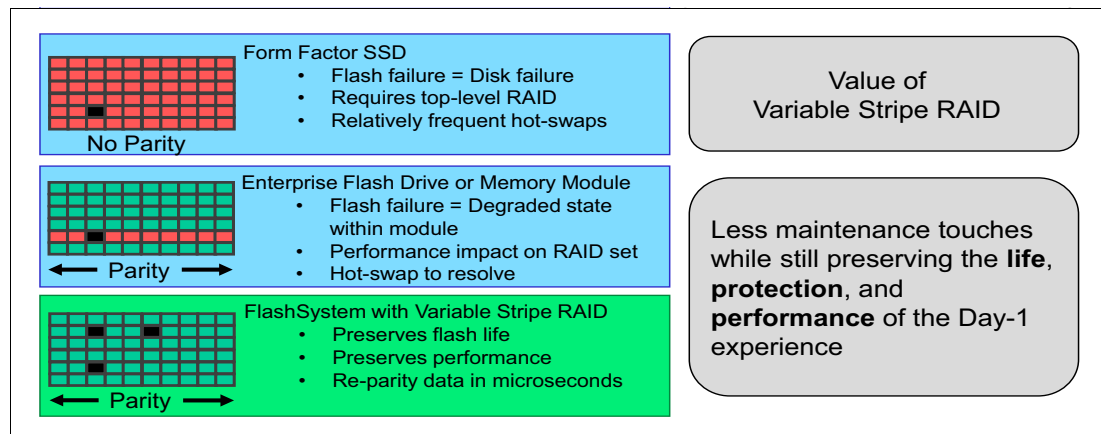


Figure 1-4 The value of the IBM FlashSystem Variable Stripe RAID

It is important to emphasize that VSR only has an effect at the *plane* level. Therefore, *only* the affected planes within a plane failure are converted to (N-1). VSR maintains the current stripe member count (9+1) layout through the rest of the areas of all other planes and chips that are not involved in the plane failure.

To illustrate how VSR functions, assume that a plane fails within a flash chip and is no longer available to store data. This might occur as a result of a physical failure within the chip, or some damage is inflicted on the address or power lines to the chip. The plane failure is detected and the system changes the format of the page stripes that are used. The data that was previously stored in physical locations across chips in all 10 lanes using a page stripe format with ten pages is now stored across chips in only nine lanes using a page stripe format with nine pages. Therefore, no data stored in the memory system was lost, and the memory system can self-adapt to the failure and continue to perform and operate by processing read and write requests from host devices.

This ability of the system to automatically self-adapt, when needed, to chip and intra-chip failures makes the FlashSystem flash module extremely rugged and robust, and capable of operating despite the failure of one or more chips or intra-chip regions. It also makes the system easier to use because the failure of one, two, or even more individual memory chips or devices does not require the removal and potential disposal of previously used memory storage components.

The reconfiguration or reformatting of the data to change the page stripe formatting to account for chip or intra-chip failures might reduce the amount of physical memory space that is held in reserve by the system and available for the system for background operation. Note that in all but the most extreme circumstances (in which case the system creates alerts), it does not affect usable capacity or performance.

Reliability, availability, and serviceability

The previous explanation points out an increase in reliability, availability, and serviceability (RAS) levels and the IBM FlashSystem RAS levels over other technologies.

In summary, VSR has these capabilities:

- ▶ Patented Variable Stripe RAID allows RAID stripe sizes to vary.
- ▶ If one plane fails in a 10-chip stripe, only the failed plane is bypassed, and then data is restriped across the remaining nine chips. No system rebuild is needed.
- ▶ VSR reduces maintenance intervals caused by flash failures.

1.7 Two-dimensional (2D) Flash RAID

Two-dimensional (2D) Flash RAID refers to the combination of Variable Stripe RAID (at the flash module level) and system-level RAID 5.

The second dimension of data protection is implemented across flash modules of RAID 5 protection. This system-level RAID 5 is striped across the appropriate number of flash modules in the system based on the selected configuration. System-level RAID-5 can stripe across four (2D+1P+1S), eight (6D+1P+1S), or 12 flash modules (10D+1P+1S).

The architecture allows you to designate a dynamic flash module hot spare. Figure 1-5 on page 11 shows the IBM FlashSystem 2D RAID.

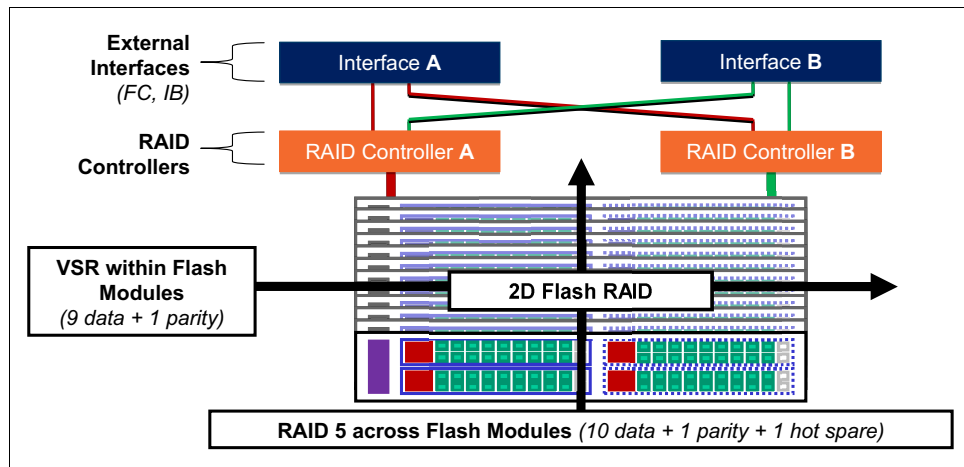


Figure 1-5 IBM FlashSystem 2D RAID

Two-dimensional (2D) Flash RAID technology within the IBM FlashSystem provides two independent layers of RAID 5 data protection within each system: the module-level Variable Stripe RAID technology and an additional system-level RAID 5 across flash modules. When operating in system-level RAID 5 mode, redundant centralized RAID controllers create a stripe arrangement across the 4, 8, or 12 flash modules in the system.

The system-level RAID 5 complements the Variable Stripe RAID technology implemented within each flash module, and it provides protection against data loss and data unavailability resulting from flash module failures. It also allows data to be rebuilt onto a hot spare flash module, so that flash modules can be replaced without data disruption.

In addition to 2D Flash RAID and Variable Stripe RAID data protection, the IBM FlashSystem family storage systems incorporate other reliability features:

- ▶ Error-correcting codes to provide bit-level reconstruction of data from flash chips.
- ▶ Checksums and data integrity fields designed to protect all internal data transfers within the system.
- ▶ Overprovisioning to enhance write endurance and decrease write amplification.
- ▶ Wear-leveling algorithms balance the number of writes among flash chips throughout the system.
- ▶ Sweeper algorithms help ensure that all data within the system is read periodically to avoid data fade issues.

Understanding 2D Flash RAID allows you to visualize the advantage over other flash storage solutions. Both VSR and 2D Flash RAID are implemented and controlled at FPGA hardware-based levels. Two-dimensional flash RAID eliminates single points of failure and provides enhanced system-level reliability.



IBM FlashSystem 840 architecture

The IBM FlashSystem 840 architecture is described in detail. An introduction to the IBM FlashSystem 840, product features, a comparison to the IBM FlashSystem 820, and an overview of the architecture and hardware are included. An overview is provided of the administration and serviceability of the IBM FlashSystem 840, interoperability, and integration with other IBM products.

For more information about the IBM FlashSystem architecture, see the FlashSystem 840 IBM Knowledge Center website:

http://www.ibm.com/support/knowledgecenter/ST2NVR_1.3.0

2.1 Introduction to the IBM FlashSystem 840 architecture

The IBM FlashSystem 840 is an all-flash storage array that provides extreme performance and large capacity while also providing enterprise class reliability and “green” data center power and cooling requirements. The IBM FlashSystem 840 holds up to twelve 4 TB flash modules in only 2U of rack space, making it an extremely dense all-flash storage array solution.

The IBM FlashSystem 840 also provides up to 1,100,000 I/O per second (IOPS) performance, up to 8 GBps bandwidth, and latency as low as 90 microseconds. This high capacity and extreme performance are also protected by the IBM FlashSystem patented reliability technologies. The IBM FlashSystem 840 supports several protocols, including FC (16, 8, and 4 Gbps), Fibre Channel over Ethernet (FCoE), iSCSI, and InfiniBand, enabling connections to high performance servers and storage area networks.

The IBM FlashSystem 840 core attributes are described next. Figure 2-1 shows the front view of the IBM FlashSystem 840.



Figure 2-1 IBM FlashSystem 840 front view

2.1.1 IBM FlashSystem 840 capacity

The IBM FlashSystem 840 supports a maximum of twelve 4 TB flash modules, which provide a maximum capacity of 48 TB (RAID 0). The IBM FlashSystem 840 can be ordered with 2, 4, 6, 8, 10, or 12 flash modules. The flash modules available are either 1 TB, 2 TB, or 4 TB storage capacity.

Important: One TB, 2 TB, and 4 TB flash modules cannot be intermixed in the same IBM FlashSystem 840 chassis.

The IBM FlashSystem 840 supports both RAID 0 and RAID 5 configurations. However, RAID 0 flash arrays have no redundancy for flash module failure, and they do not support hot spare takeover.

Notes:

- ▶ The maximum usable capacity of the IBM FlashSystem 840 in RAID 0 mode is 45 TiB. The maximum usable capacity of the IBM FlashSystem 840 in RAID 5 mode is 37.5 TiB.
- ▶ As of FlashSystem 840 release 1.3, before enabling RAID 0, you must submit a SCORE/RPQ to IBM. To submit a SCORE/RPQ, contact your IBM representative.

The IBM FlashSystem 840 supports the creation of up to 2,048 logical unit numbers (LUNs). The size of the LUNs can be 1 MiB - 45 TiB (not to exceed the total system capacity). The IBM FlashSystem 840 supports up to 2,084 host connections and up to 256 host connections for each interface port. The IBM FlashSystem 840 allows the mapping of multiple LUNs to each host for Fibre Channel, Fibre Channel over Ethernet (FCoE), iSCSI, and InfiniBand protocols.

The IBM FlashSystem 840 supports up to 256 host connections for the iSCSI protocol. Table 2-1 lists all the combinations of storage capacities for various configurations of the IBM FlashSystem 840.

Table 2-1 IBM FlashSystem 840 capacity in TB and TiB for RAID 0 and RAID 5

IBM FlashSystem 840 configuration	RAID 0 TB	RAID 5 TB	RAID 0 TiB	RAID 5 TiB
Two 1 TB flash modules	2	N/A	1.88	N/A
Four 1 TB flash modules	4	2	3.75	1.88
Six 1 TB flash modules	6	4	5.63	3.75
Eight 1 TB flash modules	8	6	7.5	5.63
Ten 1 TB flash modules	10	8	9.38	7.5
Twelve 1 TB flash modules	12	10	11.25	9.38
Two 2 TB flash modules	4	N/A	3.75	N/A
Four 2 TB flash modules	8	4	7.5	3.75
Six 2 TB flash modules	12	8	11.25	7.5
Eight 2 TB flash modules	16	12	15	11.25
Ten 2 TB flash modules	20	16	18.75	15
Twelve 2 TB flash modules	24	20	22.5	18.75
Two 4 TB flash modules	8	N/A	7.5	N/A
Four 4 TB flash modules	16	8	15	7.5
Six 4 TB flash modules	24	16	22.5	15
Eight 4 TB flash modules	32	24	30	22.5
Ten 4 TB flash modules	40	32	37.5	30
Twelve 4 TB flash modules	48	40	45	37.5

Note: The following exact byte counts are for the flash modules that are used in the IBM FlashSystem 840:

- ▶ Four TB module:
 - Presented: 4123162312704 bytes (after Variable Stripe RAID (VSR) and reserve).
 - Raw flash: 5497558138880 bytes (80 chips x 64 GiB).
 - The IBM FlashSystem 840 GUI reports 3.75 TiB usable.
 - Therefore, the actual maximum “raw” flash capacity of an IBM FlashSystem 840 storage system is 5497558138880 bytes x 12 or approximately 66 TB.
- ▶ Two TB module:
 - Presented: 2061581156352 bytes (after VSR and reserve).
 - Raw flash: 2748779069440 bytes (40 chips x 64 GiB).
 - The IBM FlashSystem 840 GUI reports 1.875 TiB usable.
- ▶ One TB module:
 - Presented: 1030790578176 bytes (after VSR and reserve).
 - Raw flash: 1374389534720 bytes (20 chips x 64 GiB).
 - The IBM FlashSystem 840 GUI reports 0.938 TiB usable.

2.1.2 IBM FlashSystem 840 performance and latency

The IBM FlashSystem 840 uses all hardware field-programmable gateway array (FPGA) components in the data path, which enables fast I/O rates and low latency. The IBM FlashSystem 840 provides extreme performance of up to 1,100,000 IOPS and up to 8 GBps in bandwidth. The IBM FlashSystem 840 provides write latency as low as 90 μ s and read latency as low as 135 μ s.

Table 2-2 on page 17 illustrates the IBM FlashSystem performance at various I/O patterns.

Table 2-2 IBM FlashSystem 840 performance at various I/O patterns

Performance criteria ^{ab}	Maximum capacity (12 flash modules)	Middle capacity (8 flash modules)	Minimum capacity (4 flash modules)
100% Read IOPS	1.1 M	1.1 M	1.0 M
100% Write IOPS	600 K	400 K	225 K
70/30 IOPS	750 K	500 K	225 K
100% large block sequential read	8 GBps	8 GBps	4 GBps
100% large block sequential write	4 GBps	2.5 GBps	1 GBps
Read latency	135 µs	135 µs	135 µs
Write latency	90 µs	90 µs	90 µs

a. Data gathered using an Oakgate storage test appliance and an FC protocol analyzer.

b. All measurements are made in a RAID 5 configuration, 4 TB cards, and 90% of usable capacity.

2.1.3 IBM FlashSystem 840 power requirements

The IBM FlashSystem 840 is *green data center friendly*. The IBM FlashSystem 840 only consumes 625 W of power (steady state RAID 5 configuration for a 70/30 read/write workload on an eight module 2 TB flashcard system) and uses two standard single phase (100v - 240v) electrical outlets.

Notes:

Plan to attach each of the two power supplies in the enclosure to separate main power supply lines.

The new 1300 W power supply, feature AF1H, for high-line voltage provides IBM FlashSystem 840 with a high-power alternative. Optimal operation is achieved when operating between 200 V - 240 V (Nominal). The maximum and minimum voltage ranges (Vrms) and associated high line AC Ranges are specified below:

- ▶ Minimum-180V, Nominal-200 V - 240 V, Maximum-265 V
- ▶ Using two power sources provides power redundancy. It is recommended that the two power supplies are placed on different circuits.

Important: The power cord is the main power disconnect. Ensure that the socket outlets are located near the equipment and are easily accessible.

2.1.4 IBM FlashSystem 840 physical specifications

The IBM FlashSystem 840 installs in a standard 19-inch equipment rack. The IBM FlashSystem 840 is 2U high and 19 inches wide. A standard data 42U 19-inch data center rack can be fully populated with 21 IBM FlashSystem 840 storage systems.

The IBM FlashSystem 840 has the following physical dimensions:

- ▶ Height: 8.90 mm (3.5 inches)
- ▶ Width: 48 mm (19 inches)
- ▶ Length: 79 mm (31.4 inches)
- ▶ Weight (maximum configuration - 12 flash modules): 34.02 kg (75 lbs)

- ▶ Airflow path: Cool air flows into the front of the unit (intake) to the rear of the unit (exhaust)
- ▶ Heat: 2133 BTU (standard configuration); 3753 BTU (maximum configuration RAID 5)

2.1.5 IBM FlashSystem 840 reliability and serviceability

Similar to all IBM FlashSystem products, the IBM FlashSystem 840 provides enterprise class reliability and serviceability that are unique for all-flash storage arrays. The IBM FlashSystem 840 uses the following technologies for data protection and maximum system uptime:

- ▶ Block remapping: Assures that flash cells are protected from adjacent activity.
- ▶ Flash cell leveling: A technology that reduces flash cell wear due to electrical programming.
- ▶ Variable Stripe RAID (VSR): A patented IBM technology that provides an intramodule RAID stripe on each flash module.
- ▶ Two-dimensional RAID (2D) Flash: System-wide RAID 5 along with VSR helps reduce downtime and maintains performance and allows the provisioning of an entire flash module as a spare to be used in another flash module failure.

New reliability and serviceability features of IBM FlashSystem 840

In addition to the standard features, the IBM FlashSystem 840 includes the following new reliability and serviceability features:

- ▶ Hot-swappable flash modules through the front panel. In a flash module failure, critical client applications can remain online while the defective module is replaced.

Because client application downtime does not need to be scheduled, you can typically perform this service immediately versus having to wait days for a service window. The “directed maintenance procedure”, accessible from the GUI, can be used to prepare the IBM FlashSystem 840 for a flash module replacement. You can remove the flash modules easily from the front of the IBM FlashSystem 840 unit without needing to remove the top access panels or extend cabling.
- ▶ Concurrent code loads. The IBM FlashSystem 840 supports concurrent code load, enabling client applications to remain online during firmware upgrades to all components, including the flash modules.
- ▶ Redundant hot-swappable components. RAID controllers called *canisters*, management modules, and interface cards (all contained in the canister), batteries, fans, and power supplies are all redundant and hot swappable. All components are easily accessible via the front or rear of the unit so the IBM FlashSystem 840 does not need to be moved in the rack, and top access panels or cables do not need to be extended. This makes servicing the unit easy.

Tip: Concurrent code loads require that all connected hosts have at least two connections, at least one to each canister, to the FlashSystem 840. For more information, see 10.1, “Encryption hints” on page 320.

2.1.6 IBM FlashSystem 840 scalability

The IBM FlashSystem 840 supports the ability to grow the storage capacity after deployment. The IBM FlashSystem 840 supports a maximum configuration of twelve 1 TB, 2 TB, or 4 TB

flash modules. The IBM FlashSystem 840 can be purchased with 2, 4, 6, 8, 10, or twelve 1 TB, 2 TB, or 4 TB modules.

Tip: It is possible to buy an entry-level IBM FlashSystem 840 unit with only two flash modules. This unit can only be configured as RAID 0. It can be expanded to 4, 6, 8, 10, or 12 modules and can then be reconfigured as RAID 5.

The IBM FlashSystem 840 offers these upgrade options:

- ▶ Systems that are purchased with four flash modules can be expanded to 6, 8, 10, or 12 of the same capacity flash modules.
- ▶ Systems that are purchased with six flash modules can be expanded to 8, 10, or 12 of the same capacity flash modules.
- ▶ Systems that are purchased with eight flash modules can be expanded to 10 or 12 of the same capacity flash modules.
- ▶ Systems that are purchased with ten flash modules can be expanded to 12 of the same capacity flash modules.

Notes: Remember these important considerations:

- ▶ Mixing different capacity flash modules (1 TB, 2 TB, or 4 TB) in any configuration on the IBM FlashSystem 840 is not supported.
- ▶ If an IBM FlashSystem 840 is purchased with 1 TB flash modules, all system expansions must be with 1 TB flash modules.
- ▶ If an IBM FlashSystem 840 is purchased with 2 TB flash modules, all system expansions must be with 2 TB flash modules.
- ▶ If an IBM FlashSystem 840 is purchased with 4 TB flash modules, all system expansions must be with 4 TB flash modules.
- ▶ Expanding an IBM FlashSystem 840 unit with 2, 4, 6, or 8 additional flash modules requires that the system is reconfigured. A backup of the system configuration and data migration, if needed, must be planned before the expansion.

2.1.7 IBM FlashSystem 840 protocol support

The IBM FlashSystem 840 supports the following interface protocols and number of connections:

- ▶ Fibre Channel (16 ports of 4 Gbps or 8 Gbps)
- ▶ Fibre Channel (8 ports of 16 Gbps (these ports also support 8 Gbps and 4 Gbps))
- ▶ Fibre Channel over Ethernet (FCoE) (16 ports of 10 Gbps FCoE)
- ▶ InfiniBand (8 ports of Quad Data Rate (QDR) 40 Gbps)
- ▶ iSCSI (16 ports of 10 Gbps Ethernet)

Notes: Remember these important considerations:

- ▶ The IBM FlashSystem 840 only supports one interface type per system. For example, it is not possible to have two FC interface cards and two InfiniBand interface cards in the same IBM FlashSystem 840 storage system.
- ▶ The IBM FlashSystem 840 supports eight active ports across the entire system if 16 Gbps FC is enabled. These eight ports can operate at 16, 8, or 4 Gbps.

2.1.8 IBM FlashSystem 840 encryption support

The IBM FlashSystem 840 provides optional encryption of data at rest, which protects against the potential exposure of sensitive user data and user metadata that are stored on discarded or stolen flash modules. Encryption of system data and metadata is not required, so system data and metadata are not encrypted.

Note: Some IBM products, which implement encryption of data at rest stored on a fixed block storage device, implement encryption using self-encrypting disk drives (SEDs). The IBM FlashSystem 840 flash module chips do not use SEDs. The IBM FlashSystem 840 data encryption and decryption are performed by the flash modules, which can be thought of as the functional equivalent of Self-Encrypting Flash Controller (SEFC) cards.

The following list describes general encryption concepts and terms for the IBM FlashSystem 840:

- ▶ *Encryption-capable* refers to the ability of the IBM FlashSystem 840 to optionally encrypt user data and metadata by using a secret key.
- ▶ *Encryption-disabled* describes a system where no secret key is configured. The secret key is neither required, or used, to encrypt or decrypt user data. Encryption logic is actually still implemented by the IBM FlashSystem 840 while in the encryption-disabled state, but uses a default, or well-known, key. Therefore, in terms of security, encryption-disabled is effectively the same as not encrypting at all.
- ▶ *Encryption-enabled* describes a system where a secret key is configured and used. This does not necessarily mean that any access control was configured to ensure that the system is operating securely. Encryption-enabled only means that the system is encrypting user data and metadata using the secret key.
- ▶ *Access-control-enabled* describes an encryption-enabled system that is configured so that an access key must be provided to authenticate with an encrypted entity, such as a secret key or flash module, to unlock and operate that entity. The IBM FlashSystem 840 permits access control enablement only when it is encryption-enabled. A system that is encryption-enabled can optionally also be access-control-enabled to provide functional security.
- ▶ *Protection-enabled* describes a system that is both encryption-enabled and access-control-enabled. An access key must be provided to unlock the IBM FlashSystem 840 so that it can transparently perform all required encryption-related functionality, such as encrypt on write and decrypt on read.
- ▶ The *Protection Enablement Process* (PEP) transitions the IBM FlashSystem 840 from a state that is not protection-enabled to a state that is protection-enabled. The PEP requires that the client provide a secret key to access the system, and the secret key must be resiliently stored and backed up externally to the system, for example, on a USB flash drive.

PEP is not merely activating a feature via the GUI or CLI. To avoid the loss of data that was written to the system before the PEP occurs, the client must move all of the data to be retained off the system before the PEP is initiated, and then must move the data back onto the system after the PEP completes. The PEP is performed during the system initialization process, if encryption is activated.

- ▶ *Application-transparent encryption* is an attribute of the IBM FlashSystem 840 encryption architecture, referring to the fact that applications are not aware that encryption and protection are occurring. This is in contrast to Application-Managed Encryption (AME), where an application must serve keys to a storage device.

- *Hot Key Activation* is the process of changing an *encryption-disabled* FlashSystem 840 to *encryption-enabled* while the system is running beginning with Version 1, Service Pack 3.
- *Non-Disruptive Rekey* is the process of creating a new encryption key that supersedes the existing key on a running FlashSystem 840 beginning with Version 1, Service Pack 3.

Note: The IBM FlashSystem 840 requires a license for encryption. If encryption is required, validate with IBM marketing or your IBM Business Partner that the license is ordered with the equipment.

Configuring encryption

You can activate encryption with the easy setup wizard during initialization or the Hot Key Activation process after the FlashSystem 840 has already been initialized, when an encryption feature code is purchased. If encryption is activated, an encryption key is generated by the system to be used for access to the system. The process invokes a wizard that guides the user through the process of copying the encryption key to multiple USB keys.

The IBM FlashSystem 840 provides support for Encryption Rekey in order to create new encryption keys that supersede the existing encryption keys.

Note: It is recommended that if you are planning to implement either Hot Key Activation or Encryption Rekey, that you inform IBM Support so they can monitor the operation.

Accessing an encrypted system

At system start (power on) or to access an encrypted system, the encryption key must be provided by an outside source so that the IBM FlashSystem 840 can be accessed. The encryption key is provided by inserting the USB flash drives that were created during system initialization into a canister.

Encryption technology

Key encryption is protected by an Advanced Encryption Standard (XTS-AES) algorithm key wrap using the 256-bit symmetric option in XTS mode, as defined in the IEEE1619-2007 standard. An HMAC-SHA256 algorithm is used to create a hash message authentication code (HMAC) for corruption detection, and it is additionally protected by a system-generated cyclic redundancy check (CRC).

2.1.9 IBM FlashSystem models 820 and 840 comparison

Table 2-3 on page 22 lists the differences between the IBM FlashSystem 820 and the IBM FlashSystem 840 modules.

Table 2-3 IBM FlashSystem 820 and 840 comparison

Feature to compare	IBM FlashSystem 820	IBM FlashSystem 840
Storage capacity options (TB)	Ten, 12, 20, and 24	Two, 4, 6, 8, 10, 12, 16, 20, 24, 32, 40, and 48
Form factor	1U	2U
Performance (IOPS)	525,000	1,100,000
Bandwidth	5 GBps	8 GBps
Latency (read/write)	110 μ s/25 μ s	135 μ s/90 μ s
Available interfaces	Eight and 4 Gbps FC Forty Gbps QDR InfiniBand	Sixteen, 8, and 4 Gbps FC Forty Gbps QDR InfiniBand Ten Gbps FCoE Ten Gbps iSCSI
Chip type	eMLC	eMLC
Chip RAID	Yes	Yes
System RAID	Yes	Yes
Power consumption (steady-state RAID 5)	300 W	625 W ^a
LUN masking	Yes	Yes
Management	CLI HTML 3.0 c2001 GUI Interface Simple Network Management Protocol (SNMP)	IBM SAN Volume Controller CLI HTML 5.0 c2011 IBM Storage GUI SNMP Email alerts Syslog redirect

a. 625 W is for a 70/30 read/write workload on an eight module 2 TB flashcard system.

2.1.10 IBM FlashSystem 840 management

The IBM FlashSystem 840 includes state-of-the-art IBM storage management interfaces. The IBM FlashSystem 840 graphical user interface (GUI) and command-line interface (CLI) are updated from previous versions of the IBM FlashSystem products to include the IBM SAN Volume Controller CLI and the new IBM SAN Volume Controller GUI, which delivers the functionality of IBM Spectrum Virtualize.

The IBM FlashSystem 840 also uses a USB key for system initialization, similar to the IBM V7000 disk system. The IBM FlashSystem 840 also supports Simple Network Management Protocol (SNMP), email notification (Simple Mail Transfer Protocol (SMTP)), and syslog redirection.

Figure 2-2 on page 23 shows the IBM FlashSystem 840 GUI. For more details about the use of the FlashSystem 840 GUI and CLI, see 2.3.2, “IBM FlashSystem 840 system management” on page 35.

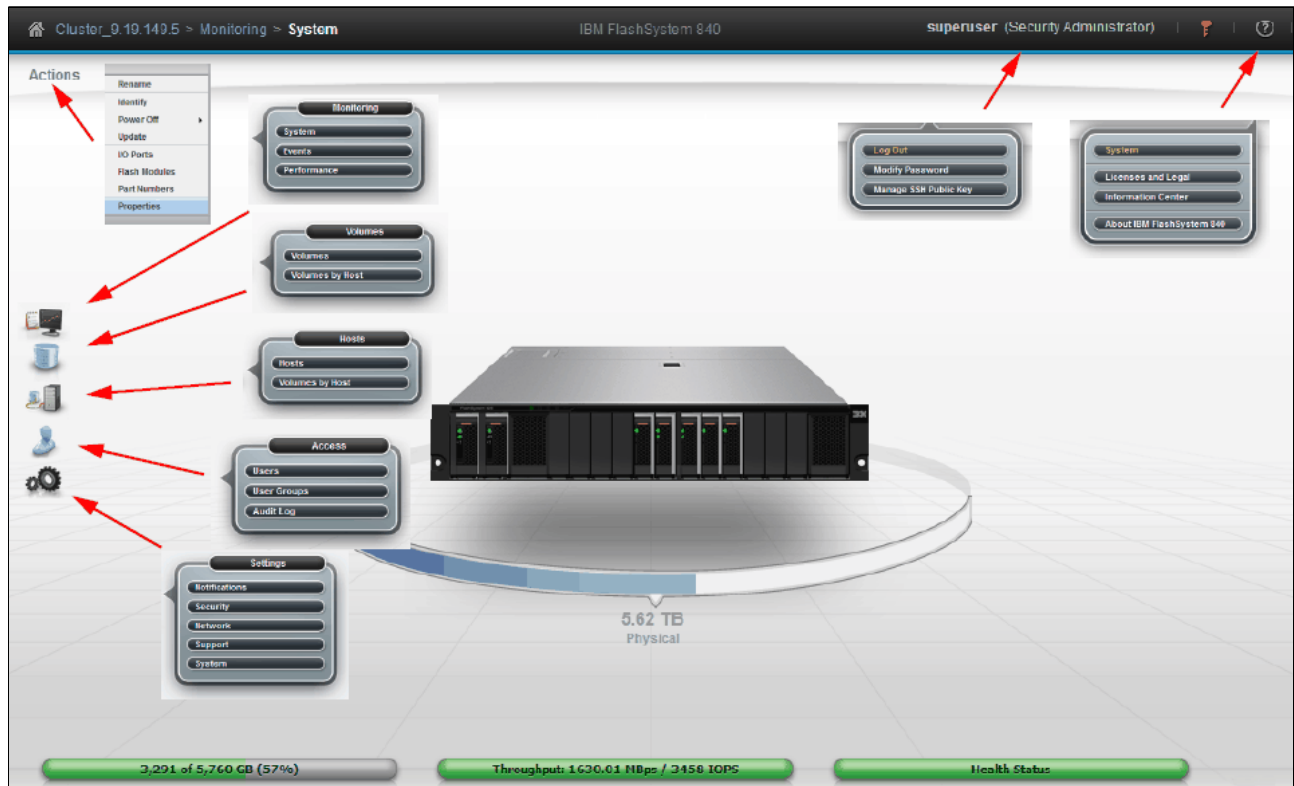


Figure 2-2 IBM FlashSystem 840 GUI

2.2 IBM FlashSystem 840 architecture

The IBM FlashSystem architecture is explained. Key product design characteristics, including performance, reliability, and serviceability, are described. Hardware components are also described.

2.2.1 IBM FlashSystem 840 architecture overview

The design goals for the IBM FlashSystem 840 are to provide the client with the fastest and most reliable all-flash storage array on the market, while making it simple to service and support with as little downtime as possible. The IBM FlashSystem 840 uses many FPGA components and as little software as possible, keeping I/O latency to a minimum and I/O performance to a maximum.

Figure 2-3 on page 24 illustrates the IBM FlashSystem 840 design. At the core of the system are the two high-speed non-blocking crossbar buses. The crossbar buses provide two high-speed paths, which carry the data traffic, and they can be used by any host entry path into the system. There is also a slower speed bus for management traffic.

Connected to the crossbar buses are high-speed non-blocking RAID modules and flash modules. There is also a main system board (midplane) to which both the RAID canisters and all the flash modules connect, as well as connections to battery modules, fan modules, and power supply units. The two RAID canisters contain crossbar controllers, management modules, interface controllers and interface adapters, and fan modules. The two RAID canisters form a logical cluster, and there is no single point of failure in the design (assuming that all host connections have at least one path to each canister).

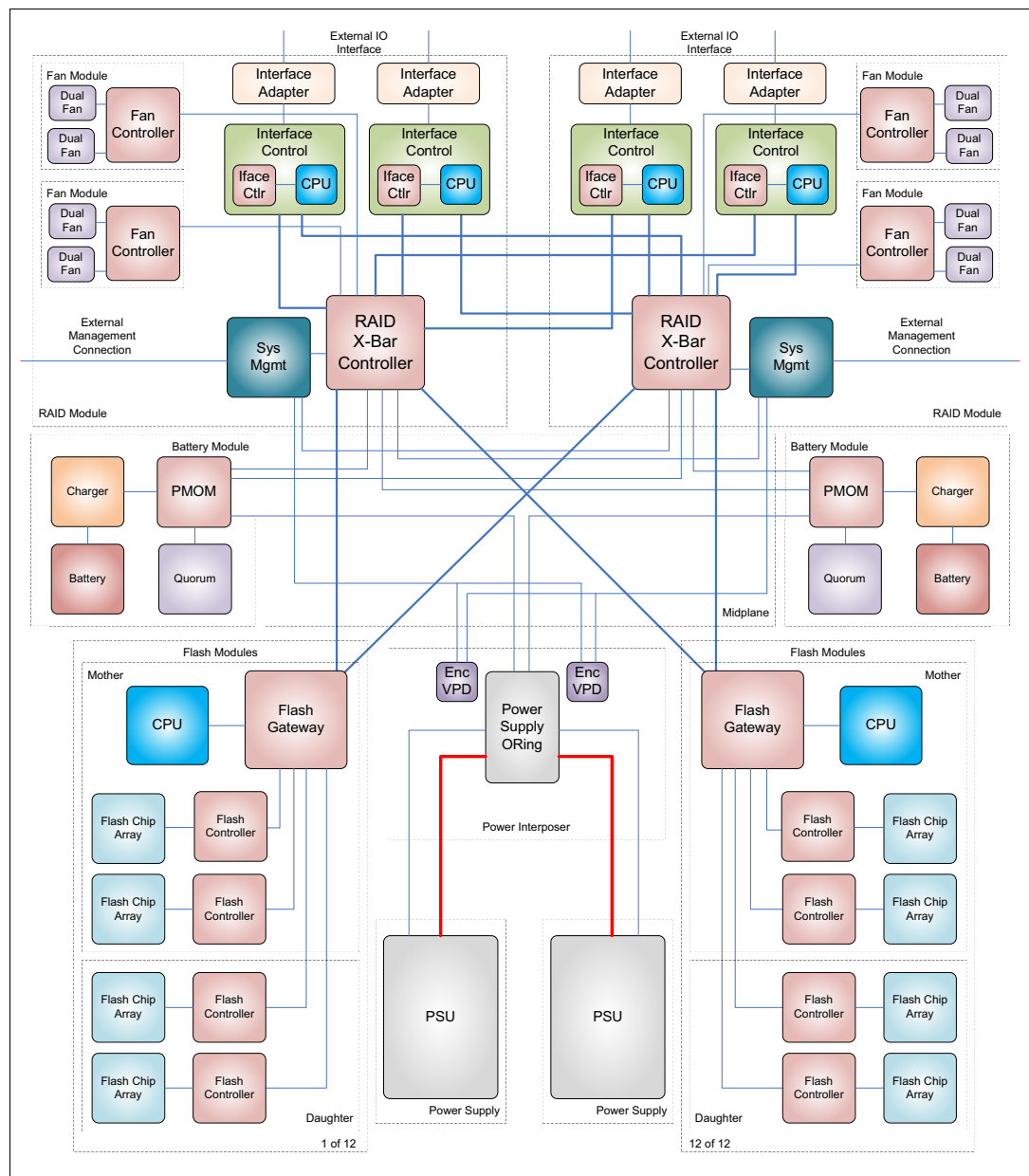


Figure 2-3 IBM FlashSystem 840 architecture

2.2.2 IBM FlashSystem 840 hardware components

The following list shows the core IBM FlashSystem 840 components:

- ▶ Canisters
- ▶ Interface cards
- ▶ Flash modules
- ▶ Battery modules
- ▶ Power supply units
- ▶ Fan modules

Figure 2-4 shows the IBM FlashSystem 840 front view. The two battery modules are to the left and the twelve flash modules are to the right.



Figure 2-4 IBM FlashSystem 840 front view

Figure 2-5 shows the IBM FlashSystem 840 rear view. The canisters are to the left (large units) and the two power supply units are to the right (small units).



Figure 2-5 IBM FlashSystem 840 rear view

2.2.3 IBM FlashSystem 840 canisters

Each IBM FlashSystem 840 storage system contains two fully redundant canisters. The fan modules are at the bottom and the interface cards are at the top. Each canister contains a RAID controller, two interface cards, and a management controller with an associated 1 Gbps Ethernet port. Each canister also has a USB port and two hot swappable fan modules.

Figure 2-6 on page 26 shows the components of the IBM FlashSystem 840 from the rear. One of the two canisters was removed, and you see two interface cards and two fan modules. The power supply unit to the right of the fans provides redundant power to the system. All components are concurrently maintainable except the midplane and the power interposer, which has no active components. All external connections are from the rear of the system.



Figure 2-6 Rear view of the FlashSystem 840 with canister removed

To maintain redundancy, the canisters are hot swappable. If any of the components (except the fans) within a canister fail, the entire canister is replaced as a unit. Both fan modules in each canister are hot swappable.

Notes: If either interface card in a canister fails, the entire canister (minus the fans) must be replaced as an entire unit. When replacing hardware in the IBM FlashSystem 840, follow the “directed maintenance procedure” that is accessible via the GUI. For more information, see Chapter 6, “Using the IBM FlashSystem 840” on page 165.

For more information about the IBM FlashSystem canisters, including canister state LEDs, see the FlashSystem 840 IBM Knowledge Center website:

<http://ibm.co/1o0Z8br>

2.2.4 IBM FlashSystem 840 interface cards

The IBM FlashSystem 840 supports three different protocol interface cards:

- ▶ Fibre Channel (16 Gbps, 8 Gbps, and 4 Gbps)
- ▶ Fibre Channel over Ethernet (FCoE) (10 Gbps)
- ▶ InfiniBand QDR (40 Gbps)
- ▶ iSCSI (10 Gbps)

Note: Although FC and FCoE protocols use the same interface card, only one protocol is supported per IBM FlashSystem 840 unit.

In the IBM FlashSystem 840, both the FC and FCoE protocols use the same interface card type. The InfiniBand and iSCSI protocols require different module types. Figure 2-7 shows a four port FC interface card, which is used for 16 Gbps FC (two ports only used), 8 Gbps and 4 Gbps FC (four ports used), and FCoE (four ports used at 10 Gbps each).

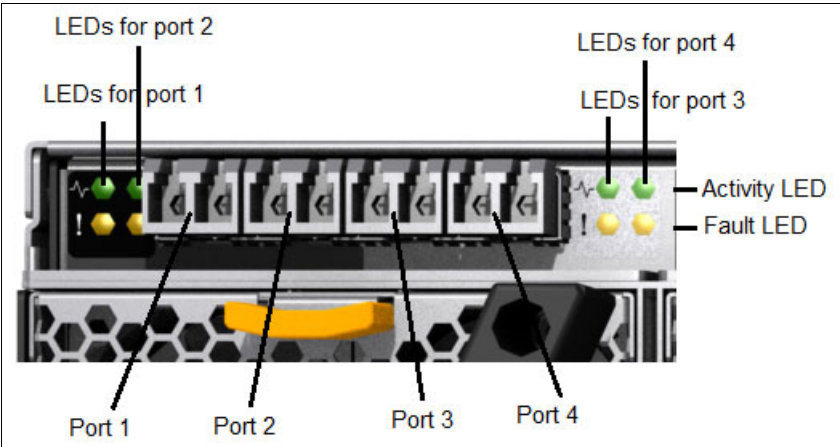


Figure 2-7 IBM FlashSystem 840 FC and FCoE interface card

The two LEDs per port (Figure 2-7) have different status meanings, depending on whether the interface is FC or FCoE:

- ▶ Fibre Channel LED description:
 - A: Link state
 - B: Link speed
- ▶ Fibre Channel over Ethernet LED description:
 - A: Activity
 - B: Link state

Fibre Channel and FCoE interface card ports and indicators

The FC and FCoE ports on each interface card are numbered 1 - 4, starting from the left.

There are two LED indicators for each FC or FCoE port, or a total of four pairs per interface card.

Fibre Channel and FCoE port LED descriptions

Each FC or FCoE interface port in the IBM FlashSystem 840 has a set of LEDs to indicate status. Figure 2-7 shows the locations of the port LED. Table 2-4 shows the LED states and descriptions for FC ports.

Table 2-4 FC LED port descriptions

LED name	Color	States
Link state	Green	OFF - NO SFP transceiver installed. SLOW flash - SFP transceiver installed, no link. SOLID - Link connected.
Link speed	Amber	OFF - No link. Two fast flashes - 4 Gb FC connection. Three fast flashes - 8 Gb FC connection. Four fast flashes - 16 Gb FC connection.

Table 2-5 shows the LED states and descriptions for FCoE ports.

Table 2-5 FCoE LED port descriptions

LED name	Color	States
Activity	Green	FLASHING - Activity on the link OFF - No activity
Link state	Amber	OFF - No link established SOLID - Link established

IBM FlashSystem 840 16 Gbps Fibre Channel support

The IBM FlashSystem 840 supports the new 16 Gbps FC connection speed via the standard FC interface card. The following rules apply to supporting 16 Gbps FC on the IBM FlashSystem 840:

- ▶ If using 16 Gbps FC, only two (of the four) ports on the FC modules can be used. The two leftmost ports (1 and 2) on each interface card are used for 16 Gbps support. The two rightmost ports (3 and 4) are disabled when 16 Gbps is sensed on any port in the IBM FlashSystem 840.
- ▶ If using 16 Gbps FC, the interface is configured as either 16 Gb FC (only two ports active), 8 Gb FC (4 ports active), or 10 Gb FCoE (4 ports). This is configured at the factory and is not changeable by the client. Direct-attach can be supported via Point-to-Point topology in 16 Gb FC, if the host supports direct attach.
- ▶ Four Gbps and 8 Gbps FC connections are supported on the same system connecting to 16 Gbps devices, but there will still only be a total of eight available active ports (ports 1 and 2 on each interface card).

For example, an IBM FlashSystem 840 storage system can have four FC connections at 16 Gbps and four FC connections at 8 Gbps.

- ▶ FC interfaces support Fibre Channel Protocol (FCP) only, with point-to-point (FC-P2P), arbitrated loop (FC-AL), and switched fabric (FC-SW) topologies. FC interfaces can be configured as N_port or NL_port types.
- ▶ Sixteen Gbps FC ports do not work in FC-AL mode, and need to be connected to a storage area network (SAN) fabric.
- ▶ Two Gbps FC ports are not supported directly by the IBM FlashSystem 840; a SAN fabric must be used to support these older hosts.

For details about a high-level design for your SAN environment and preferred practices guidance based on IBM 16 Gbps b-type products and features, focusing on FC SAN design details, see Appendix A, “SAN preferred practices for 16 Gbps” on page 339.

IBM FlashSystem 840 InfiniBand interface card

The IBM FlashSystem 840 supports four 2-port InfiniBand 40 Gbps interface cards. A total of eight ports of 40 Gbps InfiniBand connections are supported per IBM FlashSystem 840. Figure 2-8 on page 29 shows a two-port IBM FlashSystem 840 module.

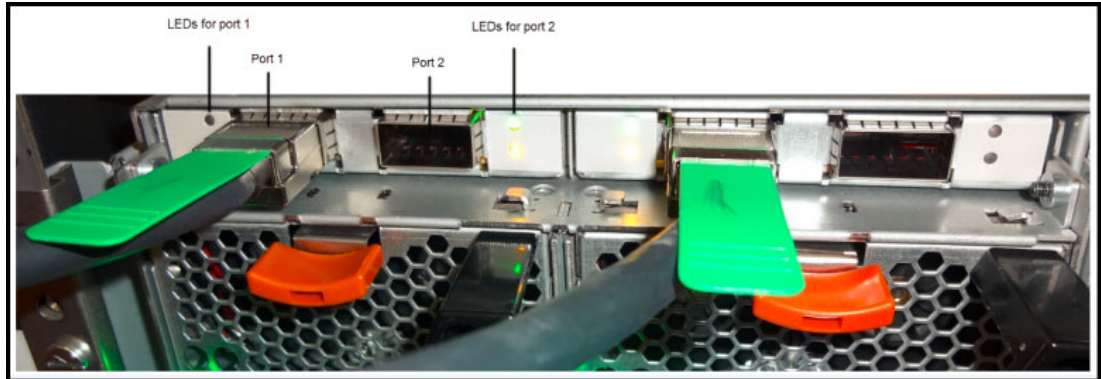


Figure 2-8 Two-port IBM FlashSystem module with InfiniBand

IBM FlashSystem 840 InfiniBand support

The IBM FlashSystem 840 InfiniBand interface cards have two 4X QDR ports each. The InfiniBand interface card ports are capable of connecting to Quad Data Rate (QDR), Double Data Rate (DDR), or Single Data Rate (SDR) InfiniBand host channel adapters (HCAs) using the SCSI Remote Direct Memory Access (RDMA) Protocol Secure Remote Password (SRP). The IBM FlashSystem 840 InfiniBand interfaces support SCSI RDMA Protocol (SRP) only.

InfiniBand interface card port LED descriptions

Each InfiniBand interface port in the IBM FlashSystem 840 has a set of LEDs to indicate the status. Table 2-6 shows the InfiniBand LED port descriptions.

Table 2-6 InfiniBand LED port descriptions

LED name	Color	States
Link state	Green	OFF - No link established. SOLID - Link is established.
Activity	Amber	OFF - No physical link. SOLID - Link is established, no activity. FLASHING - Activity on the link.

IBM FlashSystem 840 iSCSI interface card

The IBM FlashSystem 840 supports four of the 4-port iSCSI 10 Gbps interface cards. A total of 16 ports of 10 Gbps iSCSI connections are supported per IBM FlashSystem 840.

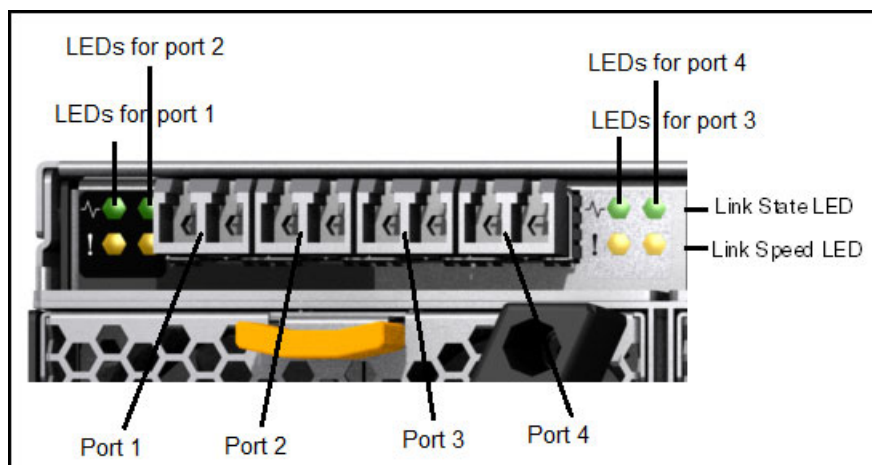


Figure 2-9 IBM FlashSystem 840 iSCSI interface card

IBM FlashSystem 840 iSCSI support

The IBM FlashSystem 840 iSCSI interface cards have four ports each. The iSCSI interface card ports are capable of connecting to Ethernet switches using 10 Gbps with either optical or copper interconnects. The IBM FlashSystem 840 supports 256 hosts connections using iSCSI. The IBM FlashSystem 840 assigns a T11 Network Address Authority (NAA) name to the system, which, by default, is in this form:

naa.<WWNN>

iSCSI interface card port LED descriptions

Each iSCSI interface port in the IBM FlashSystem 840 has a set of LEDs to indicate the status. Table 2-7 shows the iSCSI LED port descriptions.

Table 2-7 iSCSI LED port descriptions

LED name	Color	States
Link state	Green	FLASHING - Activity on the link. OFF - No activity.
Activity	Amber	OFF - No link established. SOLID - Link established.

2.2.5 IBM FlashSystem 840 flash modules

The IBM FlashSystem 840 supports up to 12 flash modules, accessible from the enclosure front panel. Each flash module has a usable capacity of either 0.938 TiB (1 TB), 1.875 TiB (2 TB), or 3.75 TiB (4 TB) of flash storage. Flash modules without the daughter board are either half-populated with 0.938 TiB (1 TB) or fully populated with 1.875 TiB (2 TB). The optional daughter board adds another 1.875 TB (2 TB) for a total of 3.75 TiB (4 TB). Figure 2-10 on page 31 illustrates an IBM FlashSystem 840 flash module (base unit and optional daughter board).



Figure 2-10 IBM FlashSystem 840 flash module

Note: All flash modules in the IBM FlashSystem 840 must be ordered as 1 TB, 2 TB, or 4 TB. Flash modules types cannot be mixed. The daughter board *cannot* be added after deployment.

The maximum storage capacity of the IBM FlashSystem 840 is based on the following factors:

- ▶ In a RAID 5 configuration, one flash module is reserved as an active spare, and capacity equivalent to one module is used to implement a distributed parity algorithm. Therefore, the maximum usable capacity of a RAID 5 configuration is 40 TB (37.5 TiB) (10 flash modules x 4 TB (3.75 TiB)).
- ▶ The maximum capacity of a RAID 0 configuration is 48 TB (45 TiB) because there are no spare flash modules (12 flash modules x 4 TB (3.75 TiB)).

Modules are installed in the IBM FlashSystem 840 based on the following configuration guidelines:

- ▶ A minimum of two flash modules must be installed in the system. RAID 0 is the only supported configuration of the IBM FlashSystem 840 with two flash modules.
- ▶ The system only supports configurations of 2, 4, 6, 8, 10, and 12 flash modules. RAID 5 can be used for systems with 4, 6, 8, 10, and 12 flash modules.
- ▶ The default configuration for the IBM FlashSystem 840 is RAID 5, unless ordered with only two flash modules.
- ▶ All flash modules that are installed in the enclosure must be identical in capacity and type.
- ▶ For optimal airflow and cooling, if fewer than 12 flash modules are installed in the enclosure, populate the flash module bays beginning in the center of the slots and adding on either side until all 12 slots are populated.

See Table 2-8 on page 32 for suggestions to populate flash module bays.

Table 2-8 Supported flash module configurations

No. of installed flash modules ^a	Flash mod. slot 1	Flash mod. slot 2	Flash mod. slot 3	Flash mod. slot 4	Flash mod. slot 5	Flash mod. slot 6	Flash mod. slot 7	Flash mod. slot 8	Flash mod. slot 9	Flash mod. slot 10	Flash mod. slot 11	Flash mod. slot 12
Two						X	X					
Four					X	X	X	X				
Six				X	X	X	X	X	X			
Eight			X	X	X	X	X	X	X	X		
Ten		X	X	X	X	X	X	X	X	X	X	
Twelve	X	X	X	X	X	X	X	X	X	X	X	X

a. RAID 5 is supported by configurations of 4, 6, 8, 10, and 12 flash modules. RAID 0 is supported by configurations of 2, 4, 6, 8, 10, and 12 flash modules.

Notes:

If fewer than 12 modules are installed, flash module blanks must be installed in the empty bays to maintain cooling airflow in the system enclosure.

During system setup in the storage enclosure management GUI, the system automatically configures RAID settings based on the number of flash modules in the system. For systems with four or more flash modules, RAID 5 is used. If the storage enclosure management GUI is used, the system automatically configures RAID 0 for systems with two flash modules. The system supports RAID 0 for a larger number of flash modules, but the configuration must be completed using the command-line interface (CLI).

All flash modules installed in the enclosure must be identical in capacity and type.

Important: Flash modules are hot swappable. However, to replace a module, you must power down the flash module by using the management GUI before you remove and replace the module. This service action does not affect the active logical unit numbers (LUNs), and I/O to the connected hosts can continue while the flash module is replaced. Be sure to follow the “directed maintenance procedure” from the IBM FlashSystem 840 GUI before any hardware replacement. See Chapter 6, “Using the IBM FlashSystem 840” on page 165.

Important: It is suggested that the storage enclosure remain powered on, or be powered on periodically, to retain array consistency. The storage enclosure can be safely powered down for up to 90 days, in temperatures up to 40-degrees C. Although the flash modules retain data if the enclosure is temporarily disconnected from power, if the system is powered off for a period exceeding 90 days, data might be lost.

When replacing this part, you must follow the recommended procedures for handling electrostatic discharge (ESD)-sensitive devices.

2.2.6 IBM FlashSystem 840 battery modules

The IBM FlashSystem 840 contains two hot swappable battery modules. The function of the battery modules is to ensure that the system is gracefully shut down (write cache fully flushed and synchronized) when AC power is lost to the unit. The IBM FlashSystem 840 battery modules are hot-swappable. Figure 2-11 shows Battery Module # 1, which is in the leftmost front of the IBM FlashSystem 840. An IBM FlashSystem 840 battery module can be hot-swapped without software intervention; however, be sure to follow the “directed maintenance procedure” from the IBM FlashSystem 840 GUI before any hardware replacement.



Figure 2-11 IBM FlashSystem 840 Battery Module # 1

IBM FlashSystem 840 power supply units

The IBM FlashSystem 840 contains two hot swappable power supply units. The system can remain fully online if one of the power supply units fails. The IBM FlashSystem 840 power supply units are accessible from the rear of the unit and are fully hot swappable. Figure 2-12 on page 34 shows the two IBM FlashSystem 840 hot swappable power supply units. The IBM FlashSystem 840 GUI and alerting systems (SNMP, and so on) report a power supply fault. The power supply can be hot-swapped without software intervention; however, be sure to follow the “directed maintenance procedure” from the IBM FlashSystem 840 GUI before any hardware replacement.

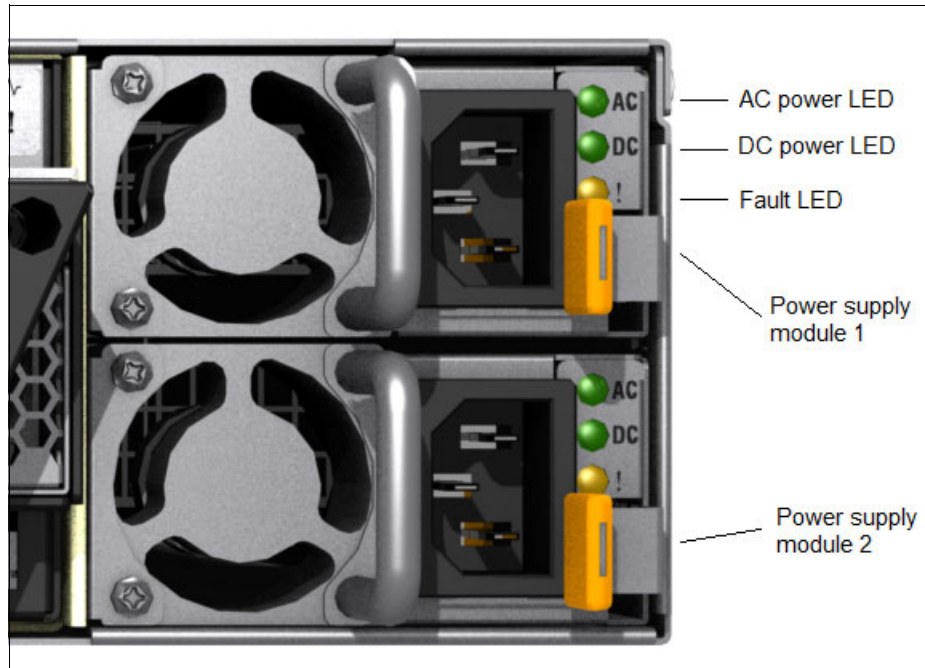


Figure 2-12 IBM FlashSystem 840 hot swappable power supply units

IBM FlashSystem 840 fan modules

The IBM FlashSystem 840 contains four hot swappable fan modules. Each IBM FlashSystem 840 canister holds two hot swappable fan modules. Each fan module contains two fans. The system can remain fully online if one of the fan modules fails. The IBM FlashSystem 840 fan modules are accessible from the rear of the unit (in each canister) and are fully hot swappable. Figure 2-13 shows an IBM FlashSystem 840 hot swappable fan module. The IBM FlashSystem 840 GUI and alerting systems (SNMP, and so on) report a fan module fault. The fan module can be hot-swapped without software intervention; however, be sure to follow the “directed maintenance procedure” from the IBM FlashSystem 840 GUI before any hardware replacement.

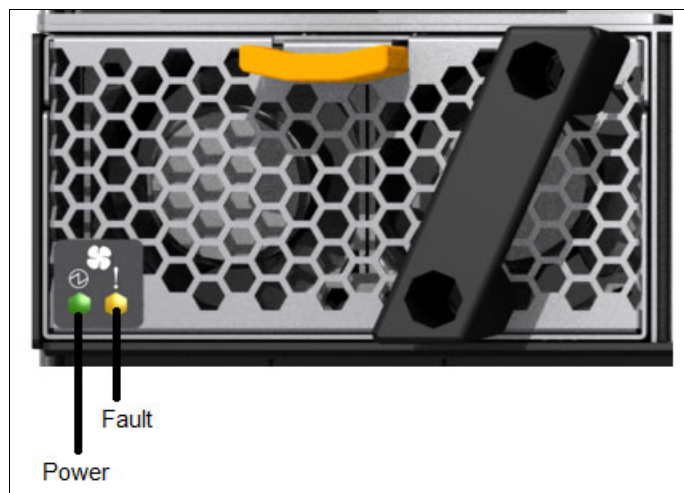


Figure 2-13 IBM FlashSystem 840 fan module

2.3 IBM FlashSystem 840 administration and maintenance

The IBM FlashSystem 840 storage system capabilities for administration, maintenance, and serviceability are described.

2.3.1 IBM FlashSystem 840 serviceability and software enhancements

The IBM FlashSystem 840 includes several design enhancements for the administration, management, connectivity, and serviceability of the system:

- ▶ Concurrent code load: The IBM FlashSystem 840 supports the ability to upgrade the system firmware on the canister (RAID controllers, management modules, and interface cards) and flash modules without affecting the connected hosts or their applications.
- ▶ Easily accessible hot swappable modules with no single point of failure: The IBM FlashSystem 840 design enables the easy replacement of any hardware module via the front or rear of the unit. The IBM FlashSystem 840 does not require the top panel to be removed nor does it need to be moved in the rack to replace any component.
- ▶ Standard IBM SAN Volume Controller CLI and GUI: The IBM FlashSystem 840 uses the latest IBM SAN Volume Controller CLI and GUI for simple and familiar management of the unit.
- ▶ Encryption support: The IBM FlashSystem 840 supports hardware encryption of the flash modules to meet the audit requirements of enterprise, financial, and government clients.
- ▶ Sixteen Gbps FC and FCoE support: The IBM FlashSystem 840 supports 16 Gbps FC and 10 Gbps FCoE, enabling clients to take advantage of the latest high-speed networking equipment while increasing performance.

2.3.2 IBM FlashSystem 840 system management

The IBM FlashSystem 840 includes the use of the common IBM SAN Volume Controller CLI and the popular IBM SAN Volume Controller GUI, which is based on the user-friendly IBM XIV® GUI. The IBM FlashSystem 840 supports SNMP, email forwarding (SMTP), and syslog redirection for complete enterprise management access.

IBM FlashSystem 840 USB key initialization process

The IBM FlashSystem 840 uses a new Universal Serial Bus (USB) key initialization process, which is similar to the IBM V7000 disk systems initialization. A USB key has an initialization file on it and is placed in a Microsoft Windows or Linux workstation to program the initial IP address information into a utility. The USB key is then placed into the IBM FlashSystem 840 on the first start and the initialization file is read and applied. The IBM FlashSystem 840 can then be managed via the GUI and CLI. For more information about the USB key initialization process, see Chapter 4, “Installation and configuration” on page 65.

IBM FlashSystem 840 graphical user interface

The IBM FlashSystem 840 includes the use of the standard IBM SAN Volume Controller GUI. This GUI is simple to use and based on the popular IBM XIV GUI.

The IBM FlashSystem 840 GUI is launched from a supported Internet browser by simply entering the systems management IP address. The system then presents the login window that is shown in Figure 2-14 on page 36.



Figure 2-14 IBM FlashSystem 840 GUI login window

After you enter a valid user name and password, the IBM FlashSystem 840 GUI presents the system overview window shown in Figure 2-15. The system overview window shows a real-time graphic depicting the IBM FlashSystem 840 in the middle. Five function icons are on the left of the window. Three dashboard icons that represent capacity, performance, and system status are at the bottom of the window.



Figure 2-15 System overview window

The IBM FlashSystem 840 GUI has five function icons:

- ▶ Monitoring element
- ▶ Volumes element
- ▶ Hosts element
- ▶ Access element
- ▶ Settings element

The following windows provide an overview of the five function icons and a brief description of their use.

Figure 2-16 shows the Monitoring icon and the associated branch-out menu. By clicking the Monitoring icon, you can perform the following actions:

- ▶ System: Monitor the system health of the IBM FlashSystem 840 hardware
- ▶ Events: View the events log of the IBM FlashSystem 840
- ▶ Performance: Launch the system I/O performance graphs



Figure 2-16 IBM FlashSystem 840 GUI Monitoring icon

Figure 2-17 on page 38 shows the Volumes icon and the associated branch-out menu. By clicking the IBM FlashSystem 840 GUI Volumes icon, the following actions can be performed:

- ▶ Volumes: View a list of all system storage volumes (LUNs), create new volumes, edit existing volumes, and delete volumes
- ▶ Volumes by Host: View a list of volumes that are associated with hosts, create new associations, or delete associations



Figure 2-17 IBM FlashSystem 840 GUI Volumes icon

Figure 2-18 shows the Hosts icon and the associated branch-out menu. By clicking the IBM FlashSystem 840 GUI Hosts icon, the following actions can be performed:

- ▶ Hosts: View a list of all hosts, create new hosts, edit existing hosts, and delete hosts
- ▶ Volumes by Host: View a list of volumes that are associated with hosts, create new associations, or delete associations



Figure 2-18 IBM FlashSystem 840 GUI Hosts icon

Figure 2-19 on page 39 shows the Access icon and associated branch-out menu. From the IBM FlashSystem 840 Access icon, the following actions are possible:

- ▶ Users: View a list of current users, create new users, edit existing users, and delete users
- ▶ User Groups: Create user groups (based on access rights) and associate users with groups
- ▶ Audit Log: View the system access log and view actions by individual users



Figure 2-19 IBM FlashSystem 840 GUI Access icon

The IBM FlashSystem 840 GUI Settings icon is used to configure system parameters, including alerting, open access, GUI settings, and other system-wide configuration. Figure 2-20 shows the Settings icon and associated branch-out menu. For more information about these parameters, see Chapter 6, “Using the IBM FlashSystem 840” on page 165.

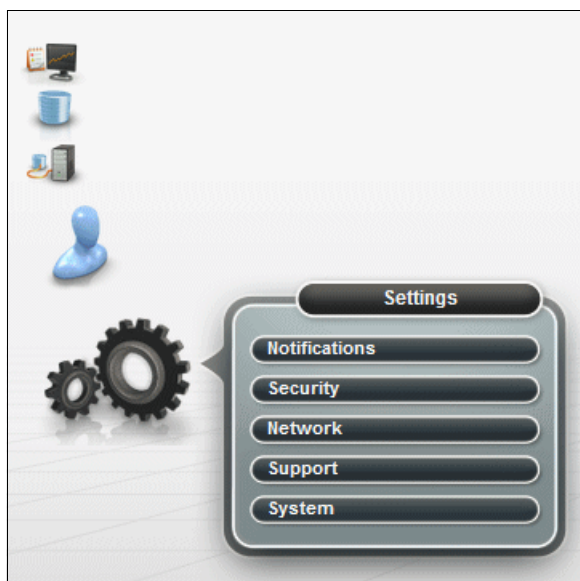


Figure 2-20 IBM FlashSystem 840 Settings GUI branch-out menu

For more detailed information about how to use the FlashSystem 840 GUI, see Chapter 6, “Using the IBM FlashSystem 840” on page 165.

IBM FlashSystem 840 command-line interface

The IBM FlashSystem 840 uses the standard IBM SAN Volume Controller storage CLI. This CLI is common among several IBM storage products, including the IBM SAN Volume Controller and the IBM Storwize family of products: the V7000, IBM V5000, IBM V3700, and IBM V3500 disk systems. IBM SAN Volume Controller CLI is easy to use with built-in help and hint menus.

To access the IBM FlashSystem 840 SAN Volume Controller CLI, a Secure Shell (SSH) session to the management IP address must be established (Telnet is not enabled on the IBM FlashSystem 840). The client is then prompted for a user name and password.

IBM FlashSystem 840 Call Home email SMTP support

The IBM FlashSystem 840 supports the ability to set up a Simple Mail Transfer Protocol (SMTP) mail server for alerting the IBM Support Center of system incidents that might require a service event. These emails can also be sent within the client's enterprise to other email accounts that are specified. After it is set up, system events that might require service will be emailed automatically to an IBM Service account automatically specified in the IBM FlashSystem 840 SAN Volume Controller code. The email alerting can be set up as part of the system initialization process or added or edited at anytime via the IBM FlashSystem 840 GUI. Also, a test email can be generated at anytime to test the connections. Figure 2-21 shows the IBM FlashSystem 840 Email setup window.

Event Notifications			
Email			
SNMP			
Syslog			

Email

Use this panel to configure email servers to send alerts to specified users.

[Edit](#) [Disable Email Event Notification](#)

Email Servers

IP Address	Server Port
9.9.9.9	25

Email Users

User Type	Email Address	Event Type	Inventory
Support	flash-sc1@vnet.ibm.com	Alerts	<input type="checkbox"/>

[Test](#)

Email Contact

* Contact Name	* Email Reply Address		
joe smith	jsmith@ibm.com		
* Machine Location	* Telephone (Primary)	Telephone (Alternate)	
test	7134567866		

Figure 2-21 IBM FlashSystem 840 Email alerting setup window

IBM FlashSystem 840 SNMP support

The IBM FlashSystem 840 supports SNMP versions 1 and 2. The GUI is used to set up SNMP support on the IBM FlashSystem 840.

To set up SNMP support on the IBM FlashSystem 840, go to the Settings icon on the lower left of the window, click the **Event Notifications** tab, and click the **SNMP** tab to enter the SNMP trap receiver IP address and community access information. Figure 2-22 on page 41 shows the IBM FlashSystem 840 SNMP setup window.

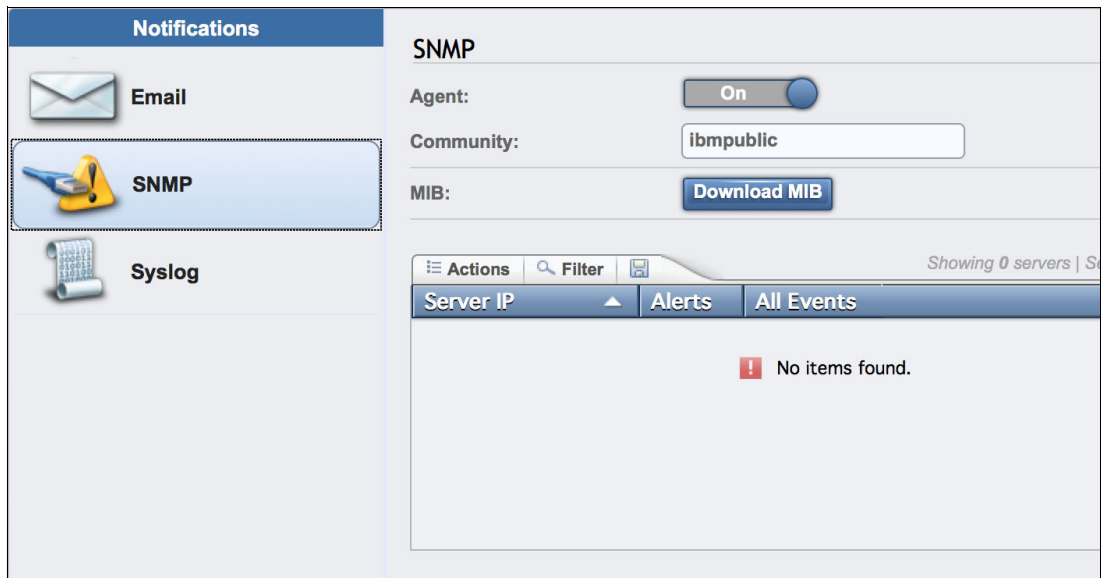


Figure 2-22 IBM FlashSystem 840 SNMP setup window

Note: The IBM FlashSystem 840 SAN Volume Controller CLI can also be used to program the SNMP settings.

IBM FlashSystem 840 syslog redirection

The IBM FlashSystem 840 allows the redirection of syslog messages to another host for system monitoring. The GUI is used to set up syslog redirection on the IBM FlashSystem 840.

To set up syslog redirection support on the IBM FlashSystem 840, go to the Settings icon on the lower left of the window, then click the **Event Notifications** tab, and then click the **Syslog** tab to enter the remote host trap IP address and directory information. Figure 2-23 shows the IBM FlashSystem 840 GUI Syslog redirection setup window.



Figure 2-23 IBM FlashSystem 840 Syslog redirection setup window

Note: The IBM FlashSystem 840 CLI can also be used to set up syslog redirection.

2.4 IBM FlashSystem 840 support matrix

The IBM FlashSystem 840 supports a wide range of operating systems (Windows Server 2008 and 2012, Linux, and IBM AIX®), hardware platforms (IBM System x, IBM Power Systems™, and x86 servers not from IBM), host bus adapters (HBAs), and SAN fabrics.

For specific information, see the IBM System Storage Interoperation Center (SSIC):

<http://ibm.com/systems/support/storage/ssic>

Also, consider using the IBM SAN Volume Controller as a front-end host-facing interface for the IBM FlashSystem 840. Therefore, if the IBM FlashSystem 840 is used with the IBM SAN Volume Controller, the host interoperability matrix for the IBM SAN Volume Controller is relevant. Obtain the IBM SAN Volume Controller interoperability information at this website:

<http://ibm.com/systems/support/storage/ssic>

Contact your IBM sales representative or IBM Business Partner for assistance or questions about the IBM FlashSystem 840 or IBM SAN Volume Controller interoperability.

2.5 IBM FlashSystem 840 IBM product integration overview

The IBM FlashSystem 840 is an all-flash storage system that can enhance the performance of almost any application. A high-level overview of how the IBM FlashSystem 840 works with a list of IBM products and applications is described. In addition to the products described here, the IBM FlashSystem 840 also works with a wide variety of other IBM software applications and hardware products, as well as products from third-party vendors. For more detail, consult your IBM salesperson or IBM Business Partner for advice about incorporating the IBM FlashSystem 840 into any of these or other applications.

Chapter 8, “Product integration” on page 275 also has more information about deploying the IBM FlashSystem 840 with the IBM products described here.

For more information about using the IBM FlashSystem with other IBM and third-party solutions and applications, see this website:

<http://www.ibm.com/storage/flash>

2.5.1 IBM Spectrum Virtualize: SAN Volume Controller

IBM SAN Volume Controller delivers the functions of IBM Spectrum Virtualize, part of the IBM Spectrum Storage family, and has been improving infrastructure flexibility and data economics for more than 10 years. Its innovative data virtualization capabilities provide the foundation for the entire IBM Storwize family. SAN Volume Controller provides the latest storage technologies for unlocking the business value of stored data, including virtualization and Real-time Compression. IBM SAN Volume Controller enriches any storage environment by adding storage management functions and a wide variety of features

- ▶ IBM FlashCopy® point-in-time copies
- ▶ Local and remote mirroring
- ▶ Thin provisioning
- ▶ Real-time Compression (RTC)
- ▶ EasyTier support (automatically directs “hot I/O” to fastest storage)
- ▶ Support for virtual environment APIs
- ▶ Support for OpenStack Cloud environments
- ▶ Support for IBM Storage Integration Server
- ▶ Storage consolidation

The IBM SAN Volume Controller product provides all of these features with minimal delay or latency in the I/O path. The combination of the IBM FlashSystem 840 and IBM SAN Volume Controller enables clients to take advantage of the speed of the IBM FlashSystem 840 and the robust storage management capabilities of the IBM SAN Volume Controller.

Note: Clients who want the advanced software features and low latency of the IBM FlashSystem 840 combined with SAN Volume Controller functions and services, such as mirroring, IBM FlashCopy, thin provisioning, IBM Real-time Compression Copy Services, and broader host support can purchase the IBM FlashSystem V9000. For product details, see the IBM Redbooks Product Guide, IBM FlashSystem V9000, at this website:

<http://www.redbooks.ibm.com/abstracts/tips1281.html?Open>

For more information about the IBM SAN Volume Controller and the IBM FlashSystem Redbooks publications, see this website:

<http://www.ibm.com/storage/flash>

For more information about the FlashSystem running in an IBM SAN Volume Controller environment, see the following publications:

- *Implementing the IBM SAN Volume Controller and FlashSystem 820*, SG24-8172
- *IBM SAN Volume Controller and IBM FlashSystem 820: Best Practices and Performance Capabilities*, REDP-5027

For more information about the IBM SAN Volume Controller, see the following books:

- *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- *Implementing the IBM System Storage SAN Volume Controller V7.4*, SG24-7933

For an in-depth description of combining the IBM FlashSystem 840 and IBM SAN Volume Controller, see 8.1, “Running the FlashSystem 840 with Spectrum Virtualize - SAN Volume Controller” on page 276.

2.5.2 IBM Storwize V7000 storage array

Similar to the IBM Spectrum Virtualize: IBM SAN Volume Controller product, the IBM Storwize V7000 storage array can provide storage management services (FlashCopy, thin-provisioning, mirroring, replication, Real-time Compression, support for virtual environments, and Easy Tier support) to externally connected storage systems.

The IBM FlashSystem 840 can be connected to the IBM V7000 storage array and provide high performance and low latency to connected hosts, while taking advantage of the IBM V7000 storage management services.

The IBM Storwize V7000 FlashSystem Edition and FlashSystem Enterprise Edition enable you to accelerate your mid-range storage solution by taking advantage of the extreme performance and low latency of the FlashSystem. For more information, see the following website:

<http://www.ibm.com/systems/storage/flash>

For more information about deploying the IBM V7000 and IBM FlashSystem 840 together, contact your IBM sales representative or IBM Business Partner.

For more information about the IBM V7000 and IBM FlashSystem, see the following website:

<http://www.ibm.com/storage/flash>

For more information about the IBM Storwize V7000, see the following publications:

- *IBM Storwize V7000 and SANSlide Implementation*, REDP-5023

- *IBM Flex System V7000 Storage Node Introduction and Implementation Guide*, SG24-8068

For an in-depth description of combining the IBM FlashSystem 840 and the IBM V7000, see 8.3, “Integrating FlashSystem 840 and SAN Volume Controller considerations” on page 303.

2.5.3 IBM PureFlex System and IBM PureSystems

IBM PureFlex® System and IBM PureSystems® products provide IBM clients with powerful hypervisor and database hardware and applications that are needed to perform mission-critical functions via the integration of the V7000 storage node.

As with all virtual environments and databases, the IBM FlashSystem 840 provides extreme I/O rates and significantly low latency, which enables the fastest possible response times. The combination of the IBM FlashSystem 840, IBM PureFlex System, and IBM PureSystems provides the fastest applications and data processing possible.

For more information about using the IBM FlashSystem 840, IBM PureFlex System, and IBM PureSystems products together, contact your IBM sales representative or IBM Business Partner.

You also can obtain information at this website:

<http://www.ibm.com/storage/flash>

Also, see the IBM Redbooks Solution Guide, *IBM FlashSystem in IBM PureFlex System Environments*, TIPS1042:

<http://www.redbooks.ibm.com/abstracts/tips1042.html?Open>

2.5.4 IBM DB2 database environments

The IBM FlashSystem 840 enables clients to speed up their databases dramatically. IBM DB2® is a high-performance, enterprise-scale database that is used by several of the largest IBM clients worldwide. Moving some or all of an IBM DB2 database onto the IBM FlashSystem 840 accelerates performance and increases CPU utilization at the same time. Also, moving a small portion of an IBM DB2 database onto the IBM FlashSystem 840 can have dramatic results.

For more information about using the IBM FlashSystem 840 and IBM DB2 products together, contact your IBM sales representative or IBM Business Partner.

You also can obtain information at this website:

<http://www.ibm.com/storage/flash>

Also, see the IBM Redbooks Solution Guide, *Faster DB2 Performance with IBM FlashSystem*, TIPS1041:

<http://www.redbooks.ibm.com/abstracts/tips1041.html?Open>

2.5.5 IBM Spectrum Scale

IBM Spectrum Scale is a proven, scalable, high-performance data and file management solution, based on IBM General Parallel File System or GPFS. IBM Spectrum Scale technology is a high-performance enterprise file management platform, and it can help you move beyond simply adding storage to optimize data management. With IBM Spectrum

Scale, businesses can achieve higher performance while reducing the footprint by placing *elastic* storage metadata on the FlashSystem 840. Running Spectrum Scale *virtualizes* IBM FlashSystem for file-based access in much the same way that SAN Volume Controller and the V7000 virtualize it for block-based access.

Spectrum Scale is also a key component for many big data products. IBM BigInsights™, DB2 PureScale, and SAP HANA all use Spectrum Scale. In addition, Spectrum Scale is used for many high-performance storage applications (media, life sciences, High Performance Computing (HPC) and so on). Some applications need fast I/O for data and some applications need fast I/O for metadata. There is a mix of random and streaming I/O for data, depending on the type of workload.

Spectrum Scale integrates with the IBM FlashSystem 840 and offers your business environment the following potential benefits:

- ▶ Spectrum Scale enables the IBM FlashSystem to be used as a storage *tier* under user control, scheduled control, or dynamically, where files are moved to and from flash and disk (and even tape) under policy control or when they are used.
- ▶ You can use the IBM FlashSystem to support either data, metadata, or both, and it can support millions of file creations and deletions per minute.
- ▶ A small part of an IBM FlashSystem can be partitioned for metadata use and the rest can be used for hot data.
- ▶ This integration provides the capability to replicate data sync/async to another site, and it supports full active-active two-site configurations (sync replication only).
- ▶ Spectrum Scale can perform mirroring of two FlashSystem 840s, either at a single site, or across multiple sites.
- ▶ Spectrum Scale is used at many sites where InfiniBand is used. Spectrum Scale and InfiniBand are a good match for the IBM FlashSystem 840.

In this scenario, Spectrum Scale and DataDirect Networks (DDNs) communicate by using the same SCSI Remote Direct Memory Access (RDMA) Protocol (SRP) supported by the IBM FlashSystem 840. SRP is similar to iSCSI protocol. SRP provides access to LUNs across networks, and SRP protocol is used across InfiniBand networks. GPFS was historically implemented by using SRP protocols to access DDN disk subsystems from GPFS Network Shared Disks (NSDs) servers.

2.5.6 IBM TS7650G ProtecTIER

IBM TS7650G ProtecTIER® is an enterprise class deduplication system that is used by IBM clients to make their backup environments more efficient. IBM TS7650G ProtecTIER uses IBM System x servers, IBM TS7650G ProtecTIER software, and a Fibre Channel-attached storage array to create a deduplication storage unit. This deduplication storage unit can be used by backup applications, such as IBM Tivoli Storage Manager, to back up and restore data. IBM TS7650G ProtecTIER also provides disaster recovery by replicating backup data to another location for safekeeping. The IBM TS7650G ProtecTIER back-end storage array is configured with two types of LUNs: Metadata and user data. Metadata LUNs are used to record where data is kept. User data LUNs are used to store the actual data. Metadata LUN performance is critical and 15 K RPM HDD spindles in a RAID 10 4+4 configuration are commonly used.

The IBM FlashSystem 840 can be used by IBM TS7650G ProtecTIER as the back-end storage device for metadata and user data LUNs. A common use case for the IBM FlashSystem 840 is for IBM TS7650G ProtecTIER metadata LUNs. Compared to the cost of dozens of 15 K or 10 Kb HDD spindles, it can be more cost-effective to use the IBM

FlashSystem 840 for IBM TS7650G ProtecTIER metadata LUNs. It might also be more cost-effective to use the IBM FlashSystem 840 for the entire IBM TS7650G ProtecTIER repository if high performance, but small capacity is needed.

Consult your IBM sales representative or IBM Business Partner on more information about deploying the IBM FlashSystem 840 with IBM TS7650G ProtecTIER.

For more information about IBM TS7650G ProtecTIER running with the IBM FlashSystem 840, see the following documents:

- ▶ *Implementing ProtecTIER 3.3 and IBM FlashSystem*, TIPS1140:
<http://www.redbooks.ibm.com/abstracts/tips1140.html?open>
- ▶ Chapter 9 “FlashSystem” of the *IBM ProtecTIER Implementation and Best Practices Guide*, SG24-8025:
<http://www.redbooks.ibm.com/redpieces/abstracts/sg248025.html?open>

For more information about IBM TS7650G ProtecTIER, see this website:

<http://www.ibm.com/systems/storage/tape/ts7650g>

Notes: IBM TS7650G ProtecTIER only supports the IBM FlashSystem 840 in RAID 5 mode. If the FlashSystem 840 is used for the IBM TS7650G ProtecTIER metadata LUNs, all future expansions must be with the IBM FlashSystem 840 storage.



Planning

This chapter provides planning information and general considerations for you to review before you perform the IBM FlashSystem 840 (FlashSystem 840) installation. This information includes connectivity, supported host environments, and IP addresses.

3.1 Installation prerequisites

This chapter provides the information needed to plan for your FlashSystem 840 environment. Plan to provide the required network infrastructure and the storage network infrastructure as described in the following sections.

3.1.1 General information

Before installing the system, you need to make sure to collect the necessary information used during the installation and setup of the system.

Contact information

The FlashSystem 840 is customer installable. Enabling the Call Home feature allows IBM personnel to be notified of any critical hardware problems.

Important: It is strongly suggested to enable the Call Home feature. Enabling Call Home allows IBM personnel to be notified 24 x 7 for any critical hardware problems. Not enabling the Call Home feature can result in long delays for any necessary service actions.

When the IBM Support Center receives a Call Home report, an IBM service representative contacts your company to work on resolving the problem. You need basic information, such as an email address and a Simple Mail Transfer Protocol (SMTP) gateway address to set up Call Home. You can complete the necessary information in Table 3-13 on page 58, section 3.5, “Call Home option” on page 58 to set up Call Home before you install the system.

Checklist before you start

Review this checklist to be sure that you have the latest information for planning the installation:

- Check for the latest firmware and **InitTool** on the IBM fix central website:

<http://www.ibm.com/support/fixcentral>

The **InitTool** uses features of Microsoft Internet Explorer (IE). Make sure to use the latest IE version if you encounter problems using the **InitTool**.

- If the FlashSystem 840 is used with the IBM SAN Volume Controller, the SAN Volume Controller version must support the FlashSystem 840. You have to check the SAN Volume Controller supported hardware list for the SAN Volume Controller version that you want to use.

SAN Volume Controller 7.4 supports the FlashSystem 840. For more details, see *V7.4.x Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller*.

<http://www.ibm.com/support/docview.wss?uid=ssg1S1004946>

You can also find references to older SAN Volume Controller versions at this site.

- If the FlashSystem 840 is used with Easy Tier, the FlashSystem 840 must be configured in RAID 5 mode.
- The FlashSystem 840 uses 2U of rack space. You must check for sufficient free space in the rack and for round or square holes in the rack to attach the enclosure rails.

- ▶ Always check the IBM System Storage Interoperation Center (SSIC) to get the latest information about supported operating systems, hosts, switches, and so on:
<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>
- ▶ If the host uses clustering software, for example, Microsoft Cluster Server (MSCS) or Oracle Automatic Storage Management (ASM), make sure that it is supported by the FlashSystem 840.
- ▶ The preferred RAID mode is RAID 5. RAID 5 provides redundancy for failed flash modules and keeps one flash module as a *spare*. For more details about the storage modes, see “RAID storage modes” on page 104.
- ▶ Set up a risk mitigation plan that lists all unresolved action items.
- ▶ Connect the Ethernet ports to a switch with offline alert notification enabled.
- ▶ Collect all necessary information, for example IP addresses, port information, user and email names.

3.1.2 Completing the hardware location chart

Planning for the physical location includes documenting the rack locations of the IBM FlashSystem 840 enclosure and other devices, based on the requirements of each device.

The hardware location chart in Table 3-1 represents the rack into which the enclosure is installed. Each row of the chart represents one Electronic Industries Alliance (EIA) 19-inch wide by 1.75-inch tall rack space or unit, each of which is commonly referred to as *1U of the rack*. As you design your rack, use the hardware location chart to record the physical configuration of the 2U enclosure and other devices in your system.

Note: Install the enclosure where it can be easily serviced, but ensure that the rack is kept stable, for example, by installing enclosures beginning from the bottom.

Complete Table 3-1 for the hardware locations of the FlashSystem 840 enclosure and other devices.

Table 3-1 Hardware location of the FlashSystem 840 enclosure and other devices

Rack unit	Component
EIA 36	
EIA 35	
EIA 34	
EIA 33	
EIA 32	
EIA 31	
EIA 30	
EIA 29	
EIA 28	
EIA 27	
EIA 26	

Rack unit	Component
EIA 25	
EIA 24	
EIA 23	
EIA 22	
EIA 21	
EIA 20	
EIA 19	
EIA 18	
EIA 17	
EIA 16	
EIA 15	
EIA 14	
EIA 13	
EIA 12	
EIA 11	
EIA 10	
EIA 9	
EIA 8	
EIA 7	
EIA 6	
EIA 5	
EIA 4	
EIA 3	
EIA 2	
EIA 1	

3.2 Planning cable connections

This section provides the necessary steps to plan for setting up cable connections for the enclosure. You can use the following tables to record the details of the cable connections for your system.

3.2.1 Management port connections

Each of the two canisters contained in the FlashSystem 840 includes an Ethernet port for accessing the management GUI.

Table 3-2 shows the management port connections.

Table 3-2 Enclosure management port Ethernet connections

Canister		Management port
Canister 1 Ethernet Management Port	Switch:	
	Port:	
	Speed:	
Canister 2 Ethernet Management Port	Switch:	
	Port:	
	Speed:	

Important: Three IP addresses are required for managing the storage enclosure: a cluster IP address and two service IP addresses. Each of the three IP addresses must be a unique value.

The left canister, when viewed from the back, is the first node. If the left canister is not available at system start, the right canister defaults to be the first node.

You use the USB-Key **InitTool** to set up the cluster IP address. You have to set the two service IP addresses after the installation of the system. Use Table 3-3 to record the cluster and service IP address settings for the storage enclosure.

Table 3-3 IP addresses for the storage enclosure

Cluster name:	
Cluster IP address:	
IP:	
Subnet mask:	
Gateway:	
Service IP address 1:	
IP:	
Subnet mask:	
Gateway:	
Service IP address 2:	
IP:	
Subnet mask:	
Gateway:	

3.2.2 Interface card connections

This section describes the information needed for various interface card connections and for the connection switches.

Fibre Channel

Each canister supports two optional Fibre Channel (FC) interface cards. Eight Gb FC cards support four ports. Sixteen Gb cards support the two leftmost ports on the card.

Important: Each host must be connected to both canisters to add redundancy and improve performance. If a host is not connected to both canisters, its state shows as *degraded*.

Small form-factor pluggable (SFP) transceivers must be obtained and deployed in pairs to support multipathing.

Table 3-4 shows the FC port connections.

Table 3-4 Fibre Channel port connections

Location	Item	Fibre Channel port 1	Fibre Channel port 2	Fibre Channel port 3 (8 Gb FC only)	Fibre Channel port 4 (8 Gb FC only)
Canister 1 Fibre Channel card 1 (left)	Switch or host:				
	Port:				
	Speed:				
Canister 1 Fibre Channel card 1 (right)	Switch or host:				
	Port:				
	Speed:				
Canister 2 Fibre Channel card 2 (left)	Switch or host:				
	Port:				
	Speed:				
Canister 2 Fibre Channel card 2 (right)	Switch or host:				
	Port:				
	Speed:				

Fibre Channel over Ethernet

Each canister supports two optional Fibre Channel over Ethernet (FCoE) interface cards. There are four ports on each card.

Note: SFP transceivers must be obtained and deployed in pairs to support multipathing.

Table 3-5 on page 53 shows the FCoE port connections.

Table 3-5 FCoE port connections

Location	Item	FCoE port 1	FCoE port 2	FCoE port 3	FCoE port 4
Canister 1 FCoE card 1 (left)	Switch or host:				
	Port:				
	Speed:				
Canister 1 FCoE card 1 (right)	Switch or host:				
	Port:				
	Speed:				
Canister 2 FCoE card 2 (left)	Switch or host:				
	Port:				
	Speed:				
Canister 2 FCoE card 2 (right)	Switch or host:				
	Port:				
	Speed:				

Quad Data Rate InfiniBand

Each canister supports two optional InfiniBand interface cards. There are two ports on each card.

Complete Table 3-6 on page 54 for the InfiniBand port connections.

Table 3-6 InfiniBand port connections

Location	Fibre Channel port 1	InfiniBand port 1	InfiniBand port 2
Canister 1 InfiniBand card 1 (left)	Switch or hosts:		
	Port:		
	Speed:		
Canister 1 InfiniBand card 1 (right)	Switch or hosts:		
	Port:		
	Speed:		
Canister 2 InfiniBand card 2 (left)	Switch or hosts:		
	Port:		
	Speed:		
Canister 2 InfiniBand card 2 (right)	Switch or hosts:		
	Port:		
	Speed:		

Internet Small Computer System Interface (iSCSI)

iSCSI is an IP-based standard for transferring data that supports host access by carrying SCSI commands over IP networks. The iSCSI standard is defined by Request for Comments memorandum (RFC) 3720. For the FlashSystem 840, connections from iSCSI-attached hosts to nodes are supported. The FlashSystem 840 supports up to sixteen 10 Gb Ethernet ports, using small form-factor pluggable transceiver (SFPs) with optical cabling or direct attach copper (DAC) cabling. iSCSI uses iSCSI qualified name (IQN), extended unique identifier (EUI), or T10 Network Address Authority (NAA) identifiers.

Table 3-7 shows that iSCSI and Fibre Channel terms have analogous components.

Table 3-7 Comparison of iSCSI and Fibre Channel components

iSCSI components	Fibre Channel components
iSCSI host bus adapter	Fibre Channel host bus adapter
Network interface controller (NIC) and iSCSI software initiator	Fibre Channel host bus adapter
IP switch	Fibre Channel switch
IP router	N/A
iSCSI name, such as iSCSI qualified name (IQN), extended unique identifier (EUI), or T11 Network Address Authority (NAA) identifiers	Worldwide node name (WWNN)

iSCSI initiators and targets

In an iSCSI configuration, the iSCSI host or server sends requests to a node. The host contains one or more initiators that attach to an IP network to initiate requests to send and receive responses from an iSCSI target, for example, the FlashSystem 840. Each initiator and target are given a unique iSCSI name, such as an IQN, EUI, or NAA identifier. An IQN is a 223-byte ASCII name. An EUI is a 64-bit identifier. A NAA is a 64 or 128-bit identifier. An iSCSI name represents a worldwide unique naming scheme that is used to identify each initiator or target in the same way that worldwide node names (WWNNs) are used to identify devices in a Fibre Channel fabric.

iSCSI targets are devices that respond to iSCSI commands. An iSCSI device can be an end node, such as a storage device, or it can be an intermediate device, such as a bridge between IP and Fibre Channel devices. Each iSCSI target is identified by a unique iSCSI name. Table 3-8 shows the iSCSI port connections.

Table 3-8 iSCSI port connections

Location	Item	iSCSI port 1	iSCSI port 2	iSCSI port 3	iSCSI port 4
Canister 1 iSCSI card 1 (left)	Switch or host: Port:				
	IP address:				
	Subnet:				
	Gateway:				
Canister 1 iSCSI card 1 (right)	Switch or host: Port:				
	IP address:				
	Subnet:				
	Gateway:				
Canister 2 iSCSI card 1 (left)	Switch or host: Port:				
	IP address:				
	Subnet:				
	Gateway:				
Canister 2 iSCSI card 2 (right)	Switch or host: Port:				
	IP address:				
	Subnet:				
	Gateway:				

Switch information

Each of the two canisters includes an Ethernet port that is attached to an Ethernet switch. These two ports are used for the management of the system.

Complete Table 3-9 on page 56 for the switch information.

Table 3-9 Switch information

	IPv4 address	IPv6 address	Media Access Control (MAC) address	Physical location
Switch 1				
Switch 2				

3.3 Planning for power

This section describes necessary information for planning to attach each of the two power supplies in the enclosure to separate main power supply lines.

Plan to connect the power cords on the right side of the rack (when viewed from the rear) to power sources that provide power in the 100 - 127 V/200 - 240 V ac range at 10.0/5.0A 50/60 z.

Two power supplies are available for use:

► 900 W power supply

Using either a 200 - 240 V ac power source at 5.0A, 50/60 Hz, or a 100 - 127 V ac power source at 10.0A, 50/60 Hz, the maximum output power is 900 W

► 1300 W power supply

– Using a 100 - 127 V ac power source at 10.0A, 50/60 Hz, the maximum output power of each power supply is 900 W

– Using a 200 - 240 V ac power source at 6.9A, 50/60 Hz, the maximum output power of each power supply is 1300 W (this is the recommended power configuration)

Using two power sources provide power redundancy.

The power supplies must be matched. You cannot have one 900 W and one 1300 W power supply.

Figure 3-1 shows two power connections.



Figure 3-1 FlashSystem 840 rear view and power connections

Note: We suggest that you place the two power supplies on different circuits.

The power cables are specific to the power requirements of your country or region.

Important: The power cord is the main power disconnect. Ensure that the socket outlets are near the equipment and easily accessible.

3.4 Planning for configuration

This section describes necessary information to plan for the management address and service address data before the system is installed.

A management address must be allocated for the system. The *management address* provides access to system configuration and administration functions, such as the management GUI.

Important: The management and service addresses for the enclosure must be allocated within the same network.

The management IP address is required when the system is initialized. The system initialization tool (**InitTool**) that is provided on a USB flash drive provides a convenient wizard for initializing the system and configuring service IP addresses.

Use Table 3-10 to record the management IP address that is allocated for use by the system.

Table 3-10 Management IP address configuration

Configuration item	Value
Management IP address:	
Subnet mask:	
Gateway address:	

Two service addresses must be allocated to the enclosure. The enclosure canisters retain the service IP addresses, allowing convenient access to node configuration and service functions, such as the service assistant GUI and CLI for that node. IPv4 and IPv6 protocols can be used.

Use Table 3-11 to plan the service IP addresses required to perform service actions.

Table 3-11 Service IP address configuration

Configuration item	Value
Service address 1	
Management IP address:	
Subnet mask:	
Gateway address:	
Service address 2	
Management IP address:	
Subnet mask:	
Gateway address:	

Use Table 3-12 to configure the system for event notification.

Table 3-12 Event notification settings

Configuration item	Value
Email server address	
Simple Network Management Protocol (SNMP) server address	
SNMP community strings	
Syslog servers	

3.5 Call Home option

The FlashSystem 840 supports the ability to set up a Simple Mail Transfer Protocol (SMTP) mail server for alerting the IBM Support Center of system incidents that might require a service event. This is the Call Home option. You can enable this option during the setup. Table 3-13 lists the necessary items.

Table 3-13 Call Home option

Configuration item	Value
Primary Domain Name System (DNS) server	
SMTP gateway address	
SMTP gateway name	
SMTP "From" address	Example: Flashsystem_name@customer_domain.com
Optional: Customer email alert group name	Example: group_name@customer_domain.com
Network Time Protocol (NTP) manager	
Time zone	

3.6 TCP/IP requirements

To plan your installation, consider the TCP/IP address requirements of the system and the requirements to access other services. You must also plan for the Ethernet address allocation and for the configuration of the Ethernet router, gateway, and firewall.

Table 3-14 on page 59 lists the TCP/IP ports and services that are used.

Table 3-14 TCP/IP ports and services

Service	Traffic direction	Protocol	Port	Service type
Email (SMTP) notification and inventory reporting	Outbound	Transmission Control Protocol (TCP)	25	Optional
SNMP event notification	Outbound	User Datagram Protocol (UDP)	162	Optional
Syslog event notification	Outbound	UDP	514	Optional
IPv4 Dynamic Host Configuration Protocol (DHCP) (Node service address)	Outbound	UDP	68	Optional
IPv6 DHCP (Node service address)	Outbound	UDP	547	Optional
Network Time Protocol (NTP) server	Outbound	UDP	123	Optional
Secure Shell (SSH) for CLI access	Inbound	TCP	22	Mandatory
HTTPS for GUI access	Inbound	TCP	443	Mandatory
HTTPS for new firmware check by the GUI	Outbound	TCP	443	Optional
Remote user authentication service - HTTP	Outbound	TCP	16310	Optional
Remote user authentication service - HTTPS	Outbound	TCP	16311	Optional
Remote user authentication service - Lightweight	Outbound	TCP	389	Optional
Wake On LAN	Inbound	N/A	N/A	Mandatory

Note: Both IPv4 and IPv6 addresses are supported.

For configuration and management, you must allocate an IP address to the Ethernet management port on each canister, which is referred to as the *management IP address*. If both IPv4 and IPv6 operate concurrently, an address is required for each protocol.

You can configure the enclosure for event notification by SNMP, syslog, or email. To configure notification, you must ensure that the SNMP agent, syslog IP addresses, or SMTP email server IP addresses can be accessed from all management addresses.

The system does not use name servers to locate other devices. You must supply the numeric IP address of the device. To locate a device, the device must have a fixed IP address.

When you click **GUI** → **Settings** → **General** → **Upgrade Software**, the software level is checked. This check is done by the GUI using the specific URL:

<https://public.dhe.ibm.com/storage/flash/9840.js>

To be able to perform this check, the system that runs the GUI must have access to this URL using port 443.

3.7 Planning for encryption

Planning for encryption involves purchasing a licensed function and then activating and enabling the function on the system.

To encrypt data that is stored on drives, the control enclosure on which they are connected must contain an active license and be configured to use encryption. When encryption is activated and enabled on the system, valid encryption keys must be present on the system when the system unlocks the drives or the user generates a new key. The encryption key must be stored on USB flash drives that contain a copy of the key that was generated when encryption was enabled. Without these keys, user data on the drives cannot be accessed. The encryption key is read from the USB flash drives that were created during system initialization.

Before activating and enabling encryption, you must determine the method of accessing key information during times when the system requires an encryption key to be present. The system requires an encryption key to be present during the following operations:

- ▶ System power-on
- ▶ System reboot
- ▶ User initiated rekey operations
- ▶ Firmware update

Several factors must be considered when planning for encryption:

- ▶ Physical security of the system
- ▶ Need and benefit of manually providing encryption keys when the system requires
- ▶ Availability of key data
- ▶ Encryption license is purchased, activated, and enabled on the system

Two options are available for accessing key information on USB flash drives:

1. USB flash drives are inserted in the system at all times

If you want the system to unlock the drives automatically when the system requires an encryption key to be present, a USB flash drive must be left inserted in both of the two canisters. This way both canisters have access to the encryption key. This method requires that the physical environment where the system is located is secure. If the location is secure, it prevents an unauthorized person from making copies of the encryption keys, stealing the system, or accessing data that is stored on the system. If a USB flash drive containing valid encryption keys is left inserted in both of the two canisters, the system will always have access to the encryption keys and the user data on the drives will always be accessible.

Note: This method requires that the physical environment where the system is located is secure. If the location is secure, it prevents an unauthorized person from making copies of the encryption keys, stealing the system, or accessing data that is stored on the system.

2. USB flash drives are never inserted into the system except as required

For the most secure operation, do not keep the USB flash drives inserted into the canisters on the system. However, this method requires that you manually insert the USB flash drives that contain copies of the encryption key in the canisters during operations that the system requires an encryption key to be present. USB flash drives that contain the keys must be stored securely to prevent theft or loss. During operations that the system requires an encryption key to be present, the USB flash drives must be inserted manually into each canister so data can be accessed. After the system has completed unlocking the drives, the USB flash drives must be removed and stored securely to prevent theft or loss.

Note: You can encrypt an existing FlashSystem 840. For assistance or questions about purchasing this licensed function, see your IBM sales representative or IBM Business Partner.

3.8 Checking your web browser settings for the management GUI

To access the management GUI, you must ensure that your web browser is supported and that the correct settings are enabled.

At the time of writing this book, the management GUI supports the following web browsers with the following versions:

- ▶ Mozilla Firefox 32
- ▶ Mozilla Firefox Extended Support Release (ESR) 31
- ▶ Microsoft Internet Explorer (IE) 10 and 11
- ▶ Google Chrome 37

For a list of supported web browsers, see *Checking your web browser settings for the management GUI* in the FlashSystem 840 IBM Knowledge Center:

<http://ibm.co/1o0Z8br>

IBM supports higher versions of the browsers as long as the vendors do not remove or disable functionality that the product relies upon. For browser levels higher than the versions that are certified with the product, IBM customer support accepts usage-related and defect-related service requests. As with operating system and virtualization environments, if the support center cannot re-create the issue in our lab, we might ask the client to re-create the problem on a certified browser version to determine whether a product defect exists. Defects are not accepted for cosmetic differences between browsers or browser versions that do not affect the functional behavior of the product. If a problem is identified in the product, defects are accepted. If a problem is identified with the browser, IBM might investigate potential solutions or workarounds that the client can implement until a permanent solution becomes available.

Procedure to configure your web browser

To configure your web browser, follow these steps:

1. Enable JavaScript for your web browser.

For Mozilla Firefox:

- a. On the menu bar in the Firefox browser window, click **Tools** → **Options**.
- b. On the Options window, select **Content**.
- c. Select **Enable JavaScript**.

- d. Click **OK**.
- e. Refresh your browser.

For Microsoft Internet Explorer (IE) running on Microsoft Windows 7:

- a. In Internet Explorer, click **Tools** → **Internet Options**.
- b. Click **Security Settings**.
- c. Click **Internet** to choose the Internet zone.
- d. Click **Custom Level**.
- e. Scroll down to the Scripting section, and then in Active Scripting, click **Enable**.
- f. Click **OK** to close Security Settings.
- g. Click **Yes** to confirm the change for the zone.
- h. Click **OK** to close Internet Options.
- i. Refresh your browser.

For Microsoft Internet Explorer (IE) running on Microsoft Windows Server 2008:

- a. In Internet Explorer, click **Tools** → **Internet Options**.
- b. Click **Security**.
- c. Click **Trusted sites**.
- d. On the **Trusted sites** dialog, verify that the web address for the management GUI is correct and click **Add**.
- e. Verify that the correct web address was added to the Trusted sites window.
- f. Click **Close** on the Trusted sites window.
- g. Click **OK**.
- h. Refresh your browser.

For Google Chrome:

- a. On the menu bar in the Google Chrome browser window, click **Settings**.
- b. Click **Show advanced settings**.
- c. In the **Privacy** section, click **Content settings**.
- d. In the JavaScript section, select **Allow all sites to run JavaScript**.
- e. Click **OK**.
- f. Refresh your browser.

2. Enable cookies in your web browser.

For Mozilla Firefox:

- a. On the menu bar in the Firefox browser window, click **Tools** → **Options**.
- b. On the Options window, select **Privacy**.
- c. Set "Firefox will" to **Use custom settings for history**.
- d. Select **Accept cookies from sites** to enable cookies.
- e. Click **OK**.
- f. Refresh the browser.

For Microsoft Internet Explorer:

- a. In Internet Explorer, click **Tools** → **Internet Options**.
- b. Click **Privacy**. Under Settings, move the slider to the bottom to allow all cookies.
- c. Click **OK**.
- d. Refresh your browser.

For Google Chrome:

- a. On the menu bar in the Google Chrome browser window, click **Settings**.
- b. Click **Show advanced settings**.
- c. In the **Privacy** section, click **Content settings**.
- d. In the **Cookies** section, select **Allow local data to be set**.

- e. Click **OK**.
 - f. Refresh your browser.
3. Enable scripts to disable or replace context menus (Mozilla Firefox only).
- For Mozilla Firefox:
- a. On the menu bar in the Firefox browser window, click **Tools** → **Options**.
 - b. On the Options window, select **Content**.
 - c. Click **Advanced** by the Enable JavaScript setting.
 - d. Select **Disable or replace context menus**.
 - e. Click **OK** to close the Advanced window.
 - f. Click **OK** to close the Options window.
 - g. Refresh your browser.
4. Enable TLS 1.1/1.2 (Microsoft Internet Explorer 10 only, IE 11 and later enable TLS 1.1/1.2 by default).
- For Mozilla Firefox:
- a. Open Internet Explorer.
 - b. Select **Tools** → **Internet Options**.
 - c. Select the **Advanced** tab.
 - d. Scroll to the **Security** section.
 - e. Check **Use TLS 1.1** and **Use TLS 1.2**.

3.9 Licensing

There is only one license with the FlashSystem 840. The FlashSystem 840 storage system provides support for AES XTS 256 data-at-rest encryption when the Encryption Enablement Pack, Feature Code AF14 is ordered.

For assistance or questions about the Encryption Enablement Pack for the FlashSystem 840, see your IBM sales representative or IBM Business Partner.

3.10 Supported hosts and operating system considerations

Always check the IBM System Storage Interoperation Center (SSIC) to get the latest information about supported operating systems, hosts, switches, and so on:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

If a configuration that you want is not available at the SSIC, a Solution for Compliance in a Regulated Environment (SCORE)/request for price quotation (RPQ) must be submitted to IBM requesting approval. To submit a SCORE/RPQ, contact your IBM representative.



Installation and configuration

In this chapter, you learn how to install and configure the IBM FlashSystem 840. The system cabling and management are described. How the initial setup procedure prepares the system for use is demonstrated.

The following topics are covered:

- ▶ First-time installation
- ▶ Initial setup wizard
- ▶ Call Home
- ▶ Performance guidelines
- ▶ RAID storage modes

4.1 First-time installation

The initial installation of the IBM FlashSystem 840 includes unpacking the system and installing it in a rack. When the system is physically installed in rack, it must be connected to power and cabled for management. Also, it must be cabled for communication with hosts.

This chapter describes unpacking the system to getting it operating in a functional state so it is ready for use.

4.1.1 Installing the hardware

Unlike previous models of the IBM FlashSystem family of products, which are serviced from the top of the unit, the IBM FlashSystem 840 has its replaceable components installed from either the front or the back side of the enclosure.

Installation poster

The installation poster, which comes with the system when it is delivered from the factory, gives a quick overview of how to prepare the system for first-time use as shown in Figure 4-1 on page 67.

IBM FlashSystem 840

Installation Poster

① Prepare for setup

You will need:



Additional resources:



Installation Guide
<https://bit.ly/BdRcQ>

Be aware that:



Before installing your
FlexSystem 840
compose site, read the *IBM
Systems Safety Notices* in
the publication package.

Important:

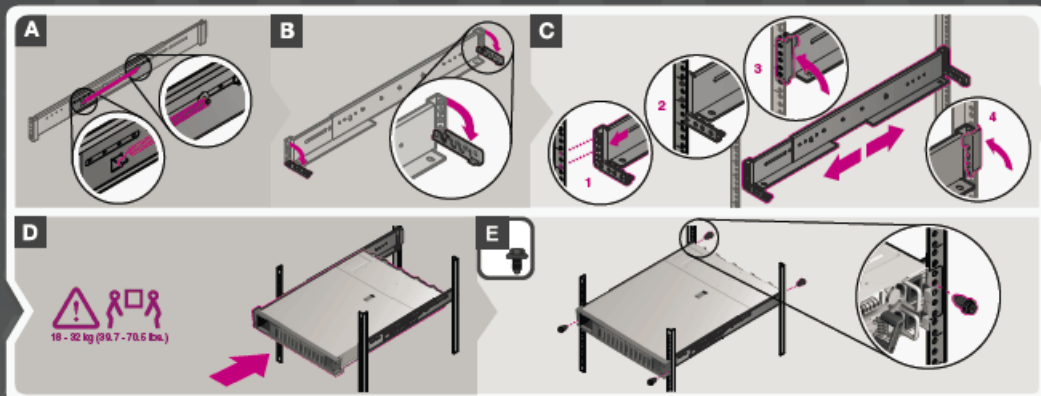


If this system will use encryption, ensure that you always know where the USB flash drives are located. During initialization, encryption keys are copied on the three USB flash drives, and these drives should be stored securely.

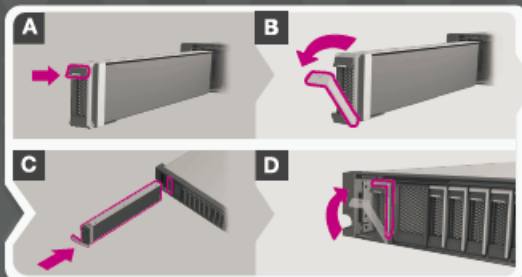
② Verify the enclosure package



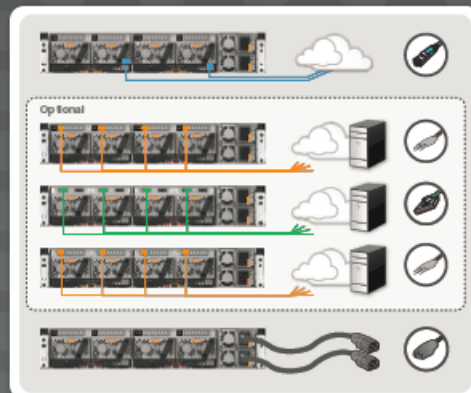
③ Install enclosures in a rack



④ Install batteries (x2)



⑤ Connect cables and power on



⑥ Set up the system (You will need: IP address / Subnet mask / Gateway)



© Copyright IBM Corporation 2014. Printed in USA

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available at www.ibm.com/legal/copytrade.shtml.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Windows is a trademark of Microsoft Corporation in the United States, other countries, or both.

IBM FlashSystem 840 Installation Poster



Figure 4-1 Installation poster

Rack installation

To install an enclosure in a rack, complete the following steps:

1. Install the rack mount rails.
2. To reduce the weight of the enclosure before lifting it, temporarily remove the two battery modules from the left side of the enclosure front panel.
3. Align the enclosure with the front of the rack cabinet.
4. Carefully slide the enclosure into the rack along the rails until the enclosure is fully inserted.
5. Secure the enclosure to the rack with a screw in the rack mounting screw hole on each side of the enclosure.
6. Pull off the front bezel to reveal the holes for the screws.
7. Insert the screws
8. Replace the bezel by snapping it in place.
9. Install the two battery modules.

The rails are not designed to hold an enclosure that is partially inserted. The enclosure must therefore always be in a fully inserted position. To reduce the weight of the enclosure before lifting it, you can temporarily remove the two battery modules and the flash modules from the front of the enclosure.

Note: Install the enclosure where it can be easily serviced, but ensure that the rack is kept stable, for example, by installing enclosures beginning from the bottom.

Installing batteries

The IBM FlashSystem 840 storage system contains two redundant batteries. In a loss of power to the enclosure, the batteries supply power so that any volatile data is written to the flash modules and the system will shut down in an orderly manner.

The IBM FlashSystem 840 batteries install in the left front side of the enclosure and plug into the midplane as shown in Figure 4-2.

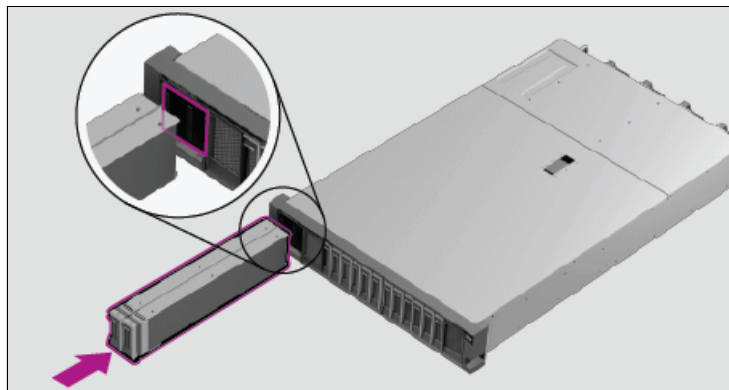


Figure 4-2 Install the batteries

Note: The IBM FlashSystem 840 does not have power switches, so whenever power is applied to the power inlets, the system is powered on. Batteries and flash cards must be installed before power cables are connected to the power inlets.

4.2 Cabling the system

This chapter details the considerations when planning and executing the physical installation of the IBM FlashSystem 840. It also includes several notes and preferred practice considerations for Fibre Channel cabling, network cabling, physical installation, and configuration.

4.2.1 Cabling for Fibre Channel

The IBM FlashSystem 840 supports either 16 Gbps Fibre Channel (FC), 8 Gbps FC, 10 Gbps Fibre Channel over Ethernet (FCoE), 10 Gbps Internet Small Computer System Interface (iSCSI) or Quadruple Data Rate (QDR) InfiniBand interfaces.

In our example environment, our FlashSystem 840 storage system is configured with four dual port 16 Gbps FC cards, for a total of eight 16 Gbps FC ports. *Only two ports for each card are active for 16 Gbps.* The optimal cabling scenario with this hardware configuration is to cable Port 1 (P1) on each interface card to storage area network (SAN) Fabric A and Port 2 (P2) on each card to SAN Fabric B, as shown in Figure 4-3.

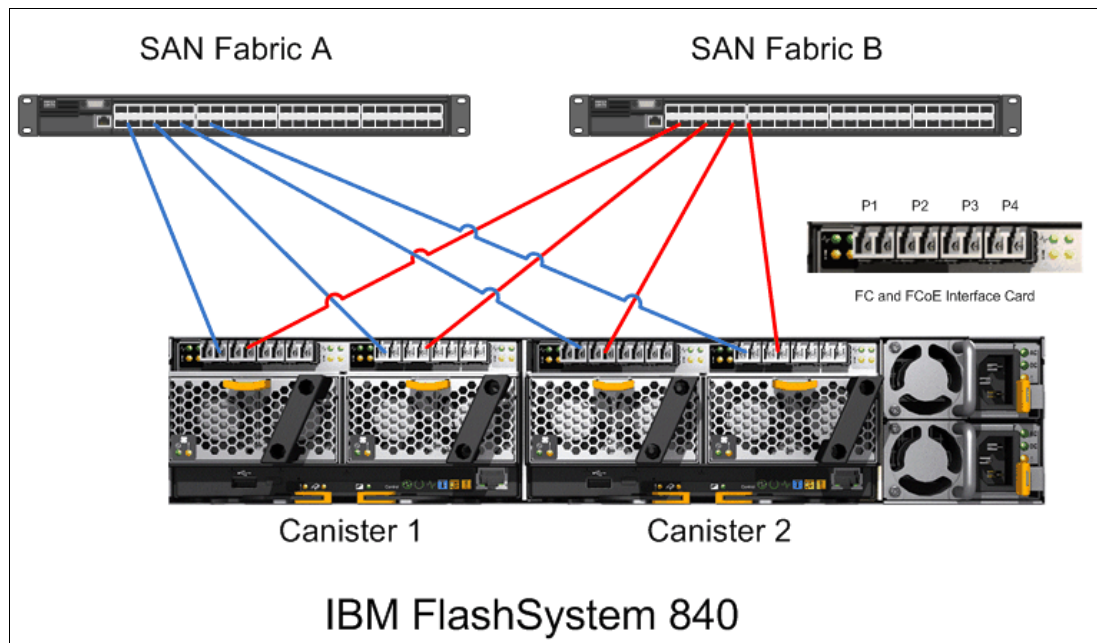


Figure 4-3 Cabling for 16 Gbps FC connectivity

The FC interface cards physically have four ports, but only port P1 and P2 are enabled in the 16 Gbps configuration. The same type of interface card hardware is used for 8 Gbps FC connectivity and for 10 Gbps FCoE connectivity. In the 8 Gbps FC and the 10 Gbps FCoE configuration, all four ports on each interface card are enabled. This configuration achieves SAN fabric switch-level redundancy and IBM FlashSystem 840 interface card-level redundancy, providing protection in an issue or failure.

A further consideration is to distribute the IBM FlashSystem 840 FC ports onto different SAN switch port groups to distribute the workload and provide the best available bandwidth. For more information about SAN switch port groups for IBM/Brocade switches, see the book about the IBM b-type Gen 5 16 Gbps switches and Network Advisor, *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116.

If the environment only contains a single SAN switch, all FC ports from the FlashSystem 840 are cabled to this switch. Although this configuration is technically possible, this configuration does not provide any redundancy in a SAN switch failure.

Note: The 16 Gbps FC, 8 Gbps FC, 10 Gbps FCoE, and iSCSI interface cards look similar and are all equipped with four physical ports. However, these interface cards are not interchangeable and must be ordered with the correct IBM feature code to provide the needed functionality.

FC port speed settings

When running FC, the IBM FlashSystem 840 can be configured with up to eight ports at 16 Gbps or 16 ports at 8 Gbps.

Depending on the configuration, the connected SAN switches can be 16 Gbps or 8 Gbps capable. However, the IBM FlashSystem 840 FC ports also run at 4 Gbps, if required. The preferred practice is to manually configure and fix the speed of the SAN switch ports to the highest mutually available speed, rather than using auto negotiate. This is done purely for consistency and stability.

Example 4-1 shows a truncated `switchshow` command output from an IBM SAN switch, showing ports connected to the IBM FlashSystem 840 set to 16 Gbps.

Example 4-1 The switchshow output from an IBM SAN switch that shows ports connected at 16 Gbps

IBM_2498_F48:FID128:admin> switchshow

Index	Port	Address	Media	Speed	State	Proto			
=====									
0	0	010000	id	N8	Online	FC	F-Port	21:01:00:1b:32:2a:23:b1	
1	1	010100	id	N16	Online	FC	F-Port	50:05:07:60:5e:fe:0a:dd	
2	2	010200	--	N16	No_Module	FC			
3	3	010300	--	N16	No_Module	FC			
4	4	010400	id	N8	Online	FC	F-Port	21:00:00:24:ff:22:f9:ea	
5	5	010500	id	N16	Online	FC	F-Port	50:05:07:60:5e:fe:0a:d9	
6	6	010600	--	N16	No_Module	FC			
7	7	010700	--	N16	No_Module	FC			
8	8	010800	id	N16	Online	FC	F-Port	10:00:8c:7c:ff:0b:0f:00	
9	9	010900	id	N16	Online	FC	F-Port	10:00:8c:7c:ff:0b:78:81	
10	10	010a00	--	N16	No_Module	FC			
11	11	010b00	id	N8	Online	FC	F-Port	10:00:00:00:c9:d4:94:11	

The **N16** identifier located under **Speed** in Example 4-1 indicates that the system negotiates to 16 Gbps. To fix the ports at 16 Gbps, use the command shown in Example 4-2 on page 71.

Example 4-2 Fixing port speed at 16 Gbps on the IBM/Brocade SAN switch

```
IBM_2498_F48:FID128:admin> portcfgspeed 1 16;  
IBM_2498_F48:FID128:admin> portcfgspeed 5 16;
```

```
IBM_2498_F48:FID128:admin> switchshow
```

Index	Port	Address	Media	Speed	State	Proto		
=====								
0	0	010000	id	N8	Online	FC	F-Port	21:01:00:1b:32:2a:23:b1
1	1	010100	id	16G	Online	FC	F-Port	50:05:07:60:5e:fe:0a:dd
2	2	010200	--	N16	No_Module	FC		
3	3	010300	--	N16	No_Module	FC		
4	4	010400	id	N8	Online	FC	F-Port	21:00:00:24:ff:22:f9:ea
5	5	010500	id	16G	Online	FC	F-Port	50:05:07:60:5e:fe:0a:d9
6	6	010600	--	N16	No_Module	FC		
7	7	010700	--	N16	No_Module	FC		
8	8	010800	id	16G	Online	FC	F-Port	10:00:8c:7c:ff:0b:0f:00
9	9	010900	id	16G	Online	FC	F-Port	10:00:8c:7c:ff:0b:78:81
10	10	010a00	--	N16	No_Module	FC		
11	11	010b00	id	N8	Online	FC	F-Port	10:00:00:00:c9:d4:94:11

The ports are no longer negotiating speed but are fixed to 16 Gbps.

For more details about how to configure the IBM/Brocade SAN switches for the correct and optimal interconnection with the IBM FlashSystem 840, see Appendix A, “SAN preferred practices for 16 Gbps” on page 339.

4.2.2 Cabling for FCoE

The IBM FlashSystem 840 supports Fibre Channel over Ethernet (FCoE) when ordered with the correct interface cards for FCoE. Cables used for FCoE are the Fibre Optical Multi Mode type that is also used for FC and iSCSI connectivity. The IBM FlashSystem 840 FCoE ports connect to switches, such as the Cisco Nexus 5000 Series Switches and the Cisco MDS 9000 Family configured with Fibre Optical ports and connectors. The operating speed for FCoE is 10 Gbps. The IBM FlashSystem 840 supports 16 FCoE ports.

For optimal redundancy and performance, the IBM FlashSystem 840 FCoE ports must connect to different and redundant FCoE switches in the same way that is illustrated in Figure 4-3 on page 69, except that the switches are FCoE-capable switches and not SAN switches.

4.2.3 Cabling for iSCSI

The IBM FlashSystem 840 supports iSCSI when ordered with the correct interface cards for iSCSI. Cables used for iSCSI are the same Fibre Optical Multi Mode type that is also used for FC and FCoE connectivity. The IBM FlashSystem 840 iSCSI ports connect to switches, such as the Cisco Nexus 5000 Series Switches equipped with Fibre Optical ports. The operating speed for iSCSI is 10 Gbps. The IBM FlashSystem 840 supports 16 iSCSI ports and IPV4 connectivity.

For optimal redundancy and performance, the FlashSystem 840 iSCSI ports must connect to different and redundant LAN switches. The hosts that connect to the FlashSystem 840 storage must access iSCSI ports on different iSCSI I/O cards on both of the FlashSystem 840 controller canisters.

4.2.4 Cabling for QDR InfiniBand

The IBM FlashSystem 840 supports Quadruple Data Rate (QDR) InfiniBand when ordered with the correct interface cards for this protocol. Each interface card has two ports operating at 40 Gbps. The IBM FlashSystem 840 supports four QDR InfiniBand adapters for a total of eight ports, each operating at 40 Gbps.

QDR InfiniBand is used in situations where powerful, high-demand servers need high-bandwidth access to the IBM FlashSystem 840.

The 40 Gbps QDR InfiniBand host bus adapters (HBAs) can be ordered from IBM or from Mellanox Technologies. Mellanox Technologies also offers InfiniBand switches with which you can connect multiple hosts, via a redundant switch configuration, to a single IBM FlashSystem 840.

Cables used for QDR InfiniBand can be ordered from IBM and can be up to 10 meters (32.8 feet) long.

Figure 4-4 shows a FlashSystem 840 canister mounted with a four port QDR InfiniBand interface card.

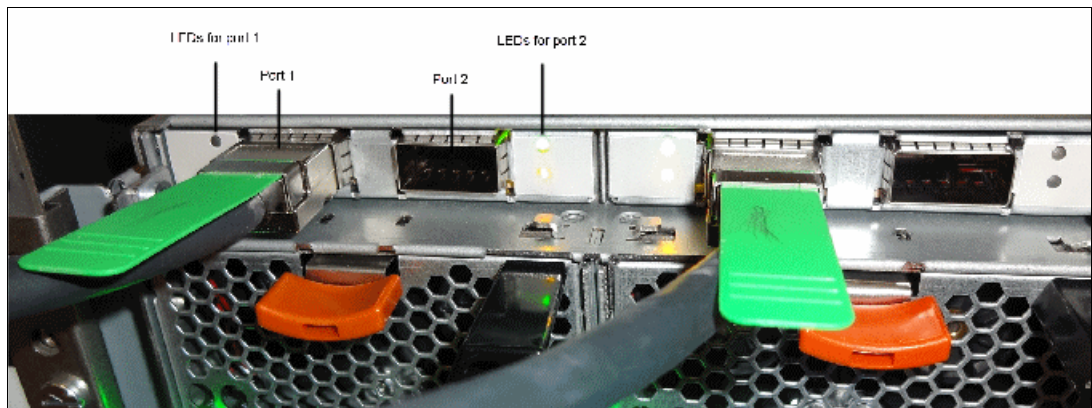


Figure 4-4 QDR InfiniBand interface card

4.2.5 FC cable type

OM4 standard cabling must be used where possible to provide the clearest connection. All of the connectors need to be the LC-LC connector standard.

4.2.6 Ethernet management cabling

The IBM FlashSystem 840 contains dual management control processors, each with its own Ethernet management port. These two Ethernet management ports operate with a single clustered IP address for system management. If one control canister is taken out for service or for any other reason inoperable, the other control canister will service the clustered IP address.

Individual canister service IP addresses are configurable via the Ethernet management ports. These IP addresses provide access to each of the canister modules and are, among other features, capable of setting a canister into a service state or rebooting a specified canister.

See Figure 4-5 on page 74 for a diagram that identifies the Ethernet management ports on the system.

The default speed setting of the IBM FlashSystem 840 Ethernet management network interface is auto, allowing the port to negotiate speed and duplex settings with the switch. The maximum configurable speed of the interface is 1 Gbps full duplex.

4.2.7 Power requirements

The IBM FlashSystem 840 comes with dual, redundant AC power modules. Plan to attach each of the two power supplies in the enclosure to separate main power supply lines and to power sources that provide power in the 100V - 240V AC range, depending on the country.

A single IBM FlashSystem 840 has a power consumption rating of 1300 watts (W) maximum with 625 watts typically seen operating in RAID 5 (625 W for a 70/30 read/write workload on a FlashSystem 900 with eight 2 TB flash cards).

Although the system operates when only one power supply is connected, this configuration is not advised. Using the power cords that are provided, connect each IBM FlashSystem 840 power inlet to an uninterruptible power supply (UPS) battery-backed power source. If possible, always connect each of the power cords to separate circuits.

Notes:

- ▶ The IBM FlashSystem 840 device must be connected to a UPS-protected power source and each power supply must be on a different phase of power to provide power redundancy.
- ▶ If one power supply fails, the performance is limited. Write performance is reduced by 40% and read performance is not affected.
- ▶ In order for a 1300-watt power supply to supply the full 1300 watts for which it is rated, it must be attached to a high-line voltage (220 - 240 volts).

4.2.8 Cooling requirements

The IBM FlashSystem 840 storage system has a British Thermal Unit (BTU) per-hour rating of approximately 2133 BTU. For maximum configurations, it can be as high as 3753 BTU.

It is suggested that the cooling vents in the room be at the front of the rack because the air flows from the front of the rack to the back.

4.2.9 Cable connector locations

The IBM FlashSystem 840 has cable connections for power and management and optional connections for either FC, FCoE, iSCSI, or QDR InfiniBand interface cards.

The FC, FCoE, and iSCSI interface cards are typically used for connecting to switches where multiple servers are able to connect to the IBM FlashSystem 840. The QDR InfiniBand interfaces can be connected directly to a host, or they can be connected through QDR InfiniBand switches.

The various options for host connectivity are described in more detail in Chapter 2, "IBM FlashSystem 840 architecture" on page 13.

Planning for Ethernet connections and connectivity is described in Chapter 3, "Planning" on page 47.

Figure 4-5 shows the rear side of the IBM FlashSystem 840 and its connectors.

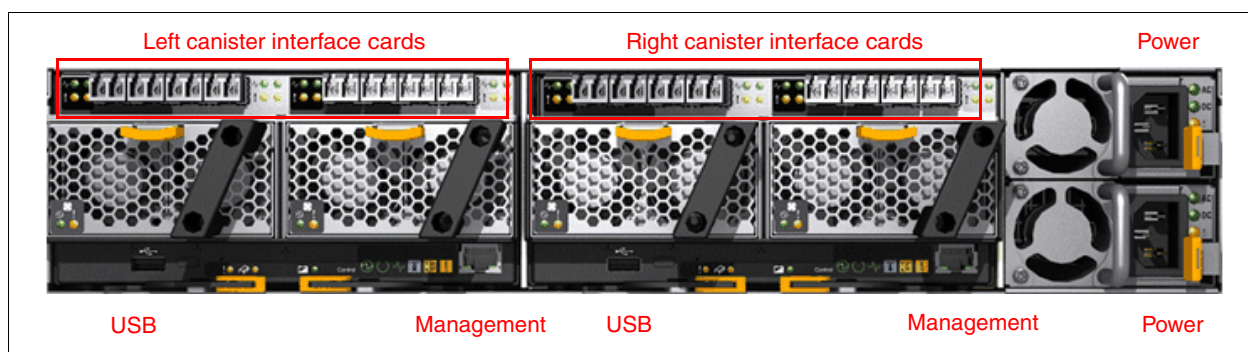


Figure 4-5 Rear view of the IBM FlashSystem 840

The IBM FlashSystem 840 is configured with either 16 Gbps FC, 8 Gbps FC, 10 Gbps FCoE, 10 Gbps iSCSI, or 40 Gbps QDR InfiniBand interface cards. Combinations of interface card types in a single system are not supported. The FlashSystem 840 depicted in Figure 4-5 is configured with FC/FCoE interface card types.

4.3 Initializing the system

After installing and powering on the new system, you must initialize it to be able to manage it. These steps are included:

1. Prepare a USB flash drive with the correct IP address using **InitTool** by inserting the provided USB flash drive into a personal computer.
2. Execute **InitTool** and follow the instructions.
3. Initialize the system by inserting the USB flash drive into the left canister USB port.
4. Log on to the system and continue with the System Setup wizard.
5. Upon completion of the wizard, insert the USB flash drive back into the same computer used in step 1.
6. Use the **InitTool** final window to review the results of the initialization.

4.3.1 About encryption

The IBM FlashSystem 840 provides optional encryption of data at rest, which protects against the potential exposure of sensitive user data and user metadata that are stored on discarded, lost, or stolen flash modules. Encryption of system data and system metadata is not required, so system data and metadata are not encrypted.

The FlashSystem 840 storage system provides support for AES-XTS 256-bit data-at-rest encryption when the Encryption Enablement Pack, feature AF14, is ordered. Starting with FlashSystem 840 1.3, the following functions are available:

- ▶ Hot Encryption Activation
- ▶ Non-Disruptive Rekey

Hot Key Activation allows a decrypted FlashSystem to be encryption-enabled while the system is running, without impacting client data in any way.

Non-Disruptive Rekey permits creating a new encryption access key that supersedes the existing key on a running FlashSystem without impacting client data.

Note: It is recommended that if you are planning to implement either Hot Encryption Activation or Encryption Rekey, that you inform IBM Support so they can monitor the operation.

Both of these operations can be done concurrently and do not cause loss of access to data. Both operations do require that you purchase the Feature Code AF14: Encryption Enablement Pack.

Data Encryption Mythology

The IBM FlashSystem 900 data encryption uses the Advanced Encryption Standard (AES) algorithm, with a 256-bit symmetric encryption key in XTS mode. This encryption mode is known as XTS–AES–256, which is described in the IEEE 1619–2007 data encryption standard. The data encryption key itself is protected by a 256-bit AES key wrap when it is stored in non-volatile form. There are two layers of encryption used with stored data, first on the data being protected, and second on the data encryption key itself.

Protection Enablement Process

The Protection Enablement Process (PEP) transforms a system from a state that is not protection-enabled to a state that is protection-enabled.

The PEP establishes a secret *encryption access key* to access the system, which must be stored and made available for use later, whenever the system needs to be unlocked. The secret encryption access key must be stored outside the system on a USB drive, which the system reads to obtain the key. The encryption access key should also be backed up to other forms of storage.

Note: For FlashSystem 840, hot encryption activation and rekey of an already initialized system is provided only via the command-line interface (CLI).

About USB flash drives and encryption keys

When the encryption Feature Code AF14 is purchased, IBM sends a total of three USB flash drives: One USB flash drive for the system, and two additional USB flash drives for the encryption feature code.

When encryption is activated, an encryption key is generated by the system to be used for access to encrypted data that is stored on the system. The GUI launches a wizard that guides you through the process of copying the encryption key to multiple USB flash drives. The following actions are considered preferred practices for copying and storing encryption keys:

1. Make copies of the encryption key on at least three USB flash drives to access the system.
2. In addition, copy the encryption keys to other forms of storage to provide resiliency and to mitigate risk, if, for example, the three USB flash drives are from a faulty batch of drives.
3. Test each copy of the encryption key before writing any user data to the initialized system.
4. Securely store all copies of the encryption key. As an example, any USB flash drives that are not left inserted into the system can be locked in a safe. Take comparable precautions to securely protect any other copies of the encryption key stored to other forms of storage.

Initializing a new system with encryption

When the IBM FlashSystem 840 initializes, it provides the option to encrypt the system and generate encryption keys. Users must purchase the Feature Code AF14 to enable encryption. Two USB flash drives are required and the encryption enablement procedure creates and store these keys on two USB flash drives, which the user must install at the rear side into the FlashSystem 840 controllers.

If encryption is activated, an encryption key is generated by the system to be used for access to encrypted data that is stored on the system. The initialization tool launches a wizard that guides you through the process of copying the encryption key to multiple USB flash drives.

Note: During the initialization, the wizard prompts you to insert two of the USB flash drives that contain the encryption keys, one into each canister. This action assumes that the physical environment where the system is located is secure.

If the system is encrypted, the system can function without the USB flash drives containing the encryption key, but it cannot be rebooted, repaired, or upgraded and the internal flash disks cannot be formatted or erased without access to the USB flash drives holding the encryption key.

FlashSystem 840 encryption can also be enabled after the system is initialized and while it is running full production. Starting with FlashSystem 840 release 1.3, it is not a disruptive operation to encrypt a FlashSystem 840.

Warning: At system start (power on) or to access an encrypted system, the encryption key must be provided by an outside source so that the system can be accessed. The encryption key is read from the USB flash drives that store copies of the keys that were created during system initialization. If you want the system to reboot automatically, a USB flash drive with the encryption keys must be left inserted in each of the canisters, so that both canisters have access to the encryption key when they power on.

This method requires that the physical environment where the system is located is secure, so that no unauthorized person can make copies of encryption keys on the USB flash drives and gain access to data stored on the system. For the most secure operation, do not keep the USB flash drives inserted into the canisters on the system. However, this method requires that you manually insert the USB flash drives that contain copies of the encryption key in both canisters before rebooting the system.

The encryption key is required to access encrypted data, and it resides only on the USB flash drive copies and on any additional copies made on other forms of storage. The encryption key cannot be recovered or regenerated by IBM if all user-maintained copies are lost or unrecoverable.

4.3.2 Prepare for initialization using InitTool

The supplied USB flash drive contains an initialization tool called **InitTool**, which is used to initialize the system. After initializing, you will be able to access the management GUI to complete the configuration procedures.

InitTool can be used with the following systems:

- ▶ Microsoft Windows computer
- ▶ Linux computer
- ▶ Apple Macintosh computer

Before you begin, ensure that the physical installation of the enclosure is complete. You need a computer to complete the initialization procedure. The computer must have a USB 2.0 port.

To initialize the system, complete the following steps:

1. Gather the information that you will use to configure the system:
 - a. You must have the IP network address that you will use to manage the system:
 - IP address
 - Subnet mask
 - Gateway
 - b. Other information is optional but useful for enabling additional capabilities:
 - The IP address of a Network Time Protocol (NTP) server for automatically setting date and time
 - The IP address of a Simple Mail Transfer Protocol (SMTP) server for sending alert notifications
 - c. Locate the USB flash drive that was shipped with your order in the documentation package.
 - d. Insert the USB flash drive into a USB port on the personal computer.
 - e. To launch the tool, open the USB flash drive, and double-click **InitTool.bat**. The initialization tool wizard starts.
 - f. In the wizard, click **Next** and select **Create a new system**.
 - g. Follow the instructions on the window that are provided by the initialization tool. You will be instructed to complete the following actions:
 - i. Enter the details of the system management address that you want to use.
 - ii. Take the USB flash drive and insert it into the IBM FlashSystem 840 and allow it to initialize.
 - iii. Plug in both power cables into the power supply units.
Wait for the status LED to flash.
This process can take up to 10 minutes.
 - iv. Return the flash drive to the workstation to check that the initialization completed.
2. Make backup copies of the encryption key (if encryption was selected).
3. If the system initialization completed successfully, click **Finish**. If you have a network connection to the system, the system management GUI is displayed. If the workstation does not have a network connection to the system, go to the workstation that you will use to manage the system and start a supported browser. Direct the browser to the management address that you specified for the system.
4. Log in with the user name superuser and password passw0rd.

Note: The 0 character in the password is a zero.

5. Follow the instructions on the window to begin setting up your system.

Initializing the system with a Microsoft Windows computer

The following example shows the steps involved in initializing the system using a Microsoft Windows computer to prepare the supplied USB flash drive using **InitTool**.

The first window in the System Initialization wizard shows that the tool can be used for different purposes:

- ▶ Install a new system
- ▶ Reset the superuser password
- ▶ Edit the service IP address

Start the InitTool process

Figure 4-6 shows the initial step in using **InitTool**.



Figure 4-6 *InitTool Welcome*

Tasks

In the second step of the System Initialization wizard, select **Yes** to configure a new system as shown in Figure 4-7 on page 79.

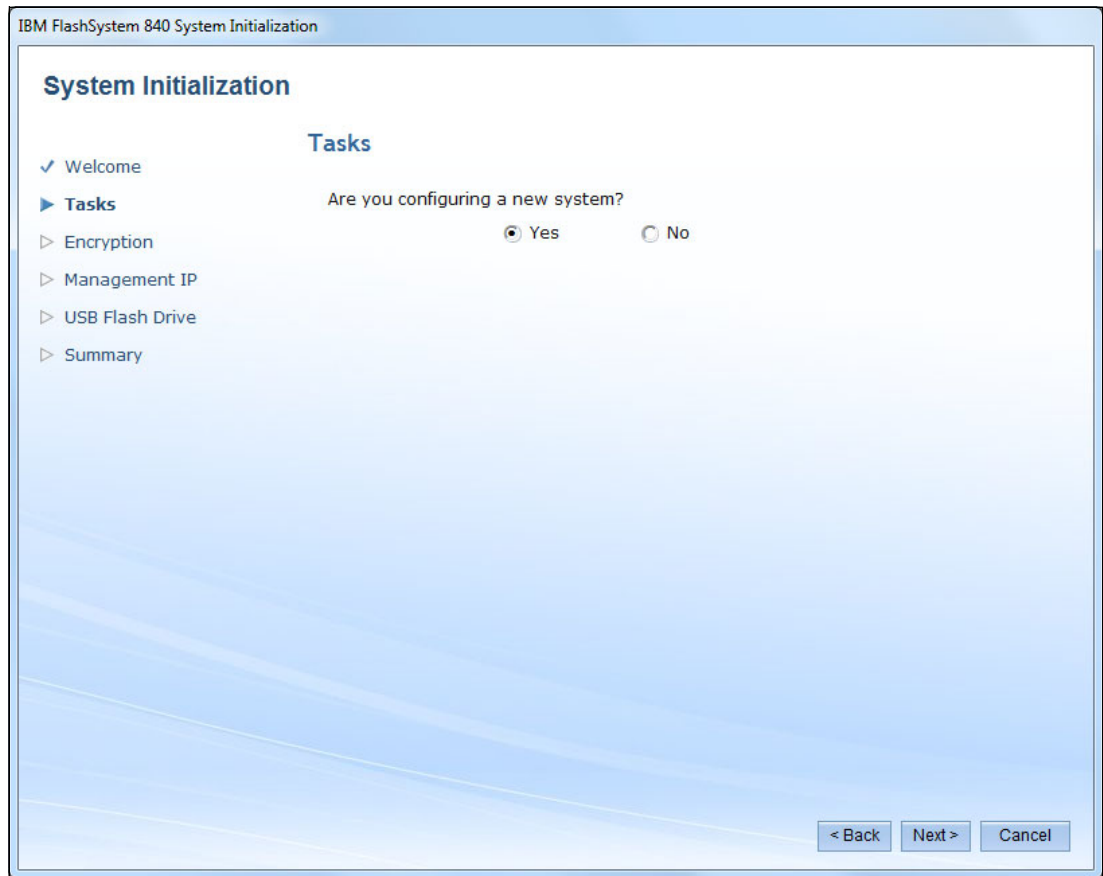


Figure 4-7 *InitTool Tasks*

If you selected No at the Tasks window, the **InitTool** assumes that you are unable to access your system and gives you two options:

- ▶ Reset the superuser password
- ▶ Set the service IP address

Encryption

In the next window, select whether encryption is needed. The system provides optional encryption of data at rest, which protects against the potential exposure of sensitive user data and user metadata that are stored on discarded, lost, or stolen flash modules. Encryption of system data and system metadata is not required, so system data and metadata are not encrypted.

If you want to use encryption, ensure that you purchased Feature Code AF14: Encryption Enablement Pack (Plant).

Assuming that a license for encryption was purchased, click **Yes** at the Encryption window as shown in Figure 4-8 on page 80.

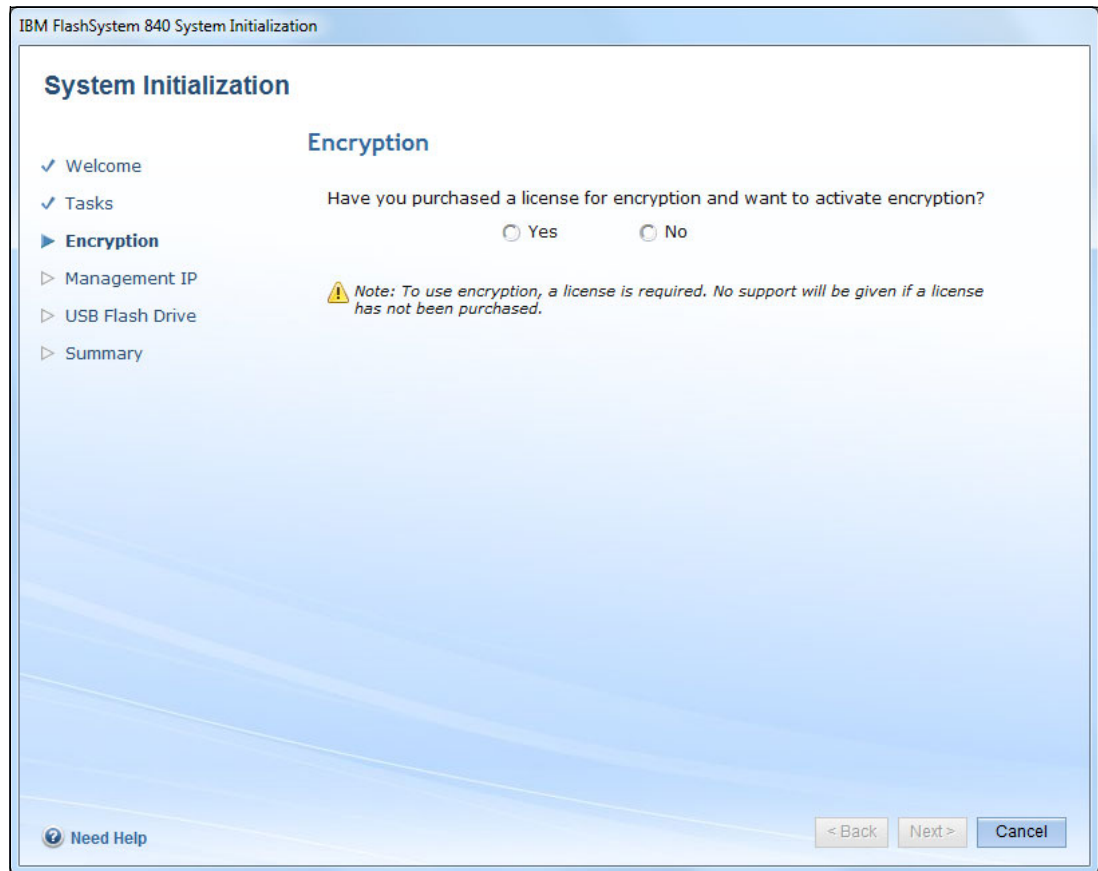


Figure 4-8 InitTool Encryption

If you activate encryption by selecting **Yes**, to obtain further IBM support, you must provide proof of purchase of the encryption feature code (FC) AF14: Encryption Enablement Pack (Plant).

Management IP address

In the next step of the System Initialization wizard, type the IP address for managing the IBM FlashSystem 840 as shown in Figure 4-9 on page 81.

The screenshot shows a window titled "IBM FlashSystem 840 System Initialization". Inside, there's a section titled "System Initialization" with a list of steps on the left: "Welcome", "Tasks", "Encryption", "Management IP" (which is highlighted with a blue arrow), "USB Flash Drive", and "Summary". To the right of this list, under the heading "Management IP Address", is the instruction "Select the Internet Protocol (IP) address to use on your system." Below this, there are two radio buttons: "IPv4" (which is selected) and "IPv6". Further down, there are three input fields labeled "IP address:", "Subnet mask:", and "Gateway:". At the bottom right of the window, there are three buttons: "< Back", "Apply and Next >", and "Cancel".

Figure 4-9 InitTool Management IP Address window

Power on

In the next step, the operator is instructed to power on the IBM FlashSystem 840 and wait for the status LED to flash as shown in Figure 4-10 on page 82. This power-on process can take up to 10 minutes.

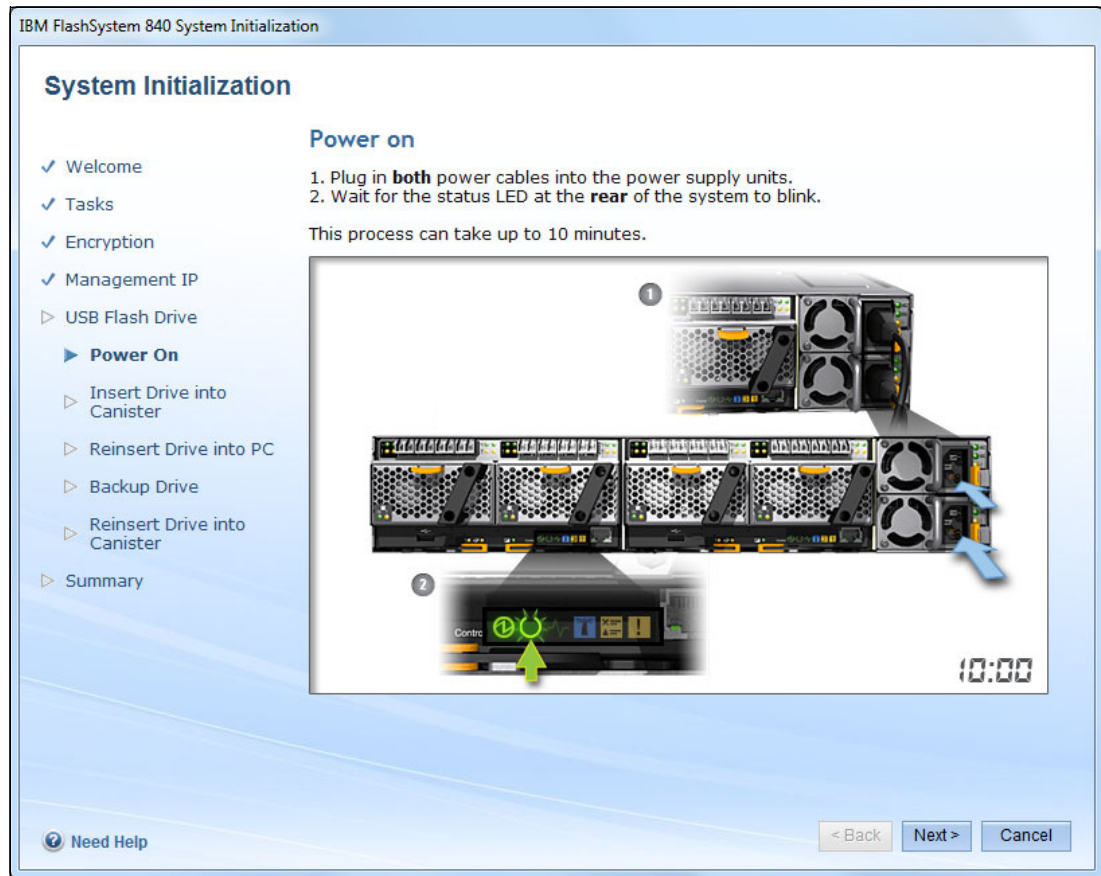


Figure 4-10 InitTool Power on

Insert USB into canister

In the next step, the operator is instructed to remove the USB flash drive from the PC and insert it into the USB port on the *left* IBM FlashSystem 840 canister (controller) as shown in Figure 4-11 on page 83. At the end of the process, the Identify LED turns on and then off.

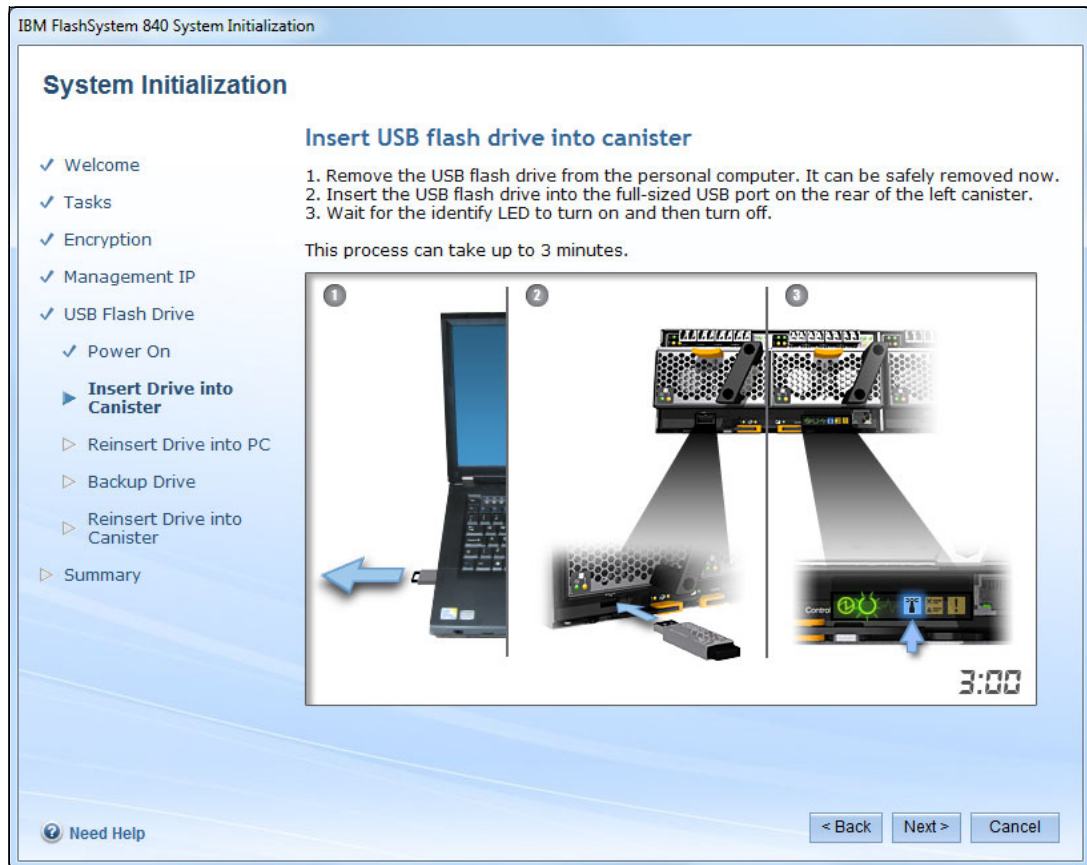


Figure 4-11 InitTool Insert USB flash drive into canister

Note: The USB flash drive must be inserted into the left canister to correctly initialize the IBM FlashSystem 840.

When the system initialization process finishes, a results file named `satask_result.html` is written to the USB flash drive by the IBM FlashSystem 840 canister. The results file indicates success or failure of the process. **InitTool** can be used to verify this result.

Reinsert USB flash drive into personal computer

InitTool instructs the operator to reinsert the USB flash drive into the Microsoft Windows computer as shown in Figure 4-12 on page 84.

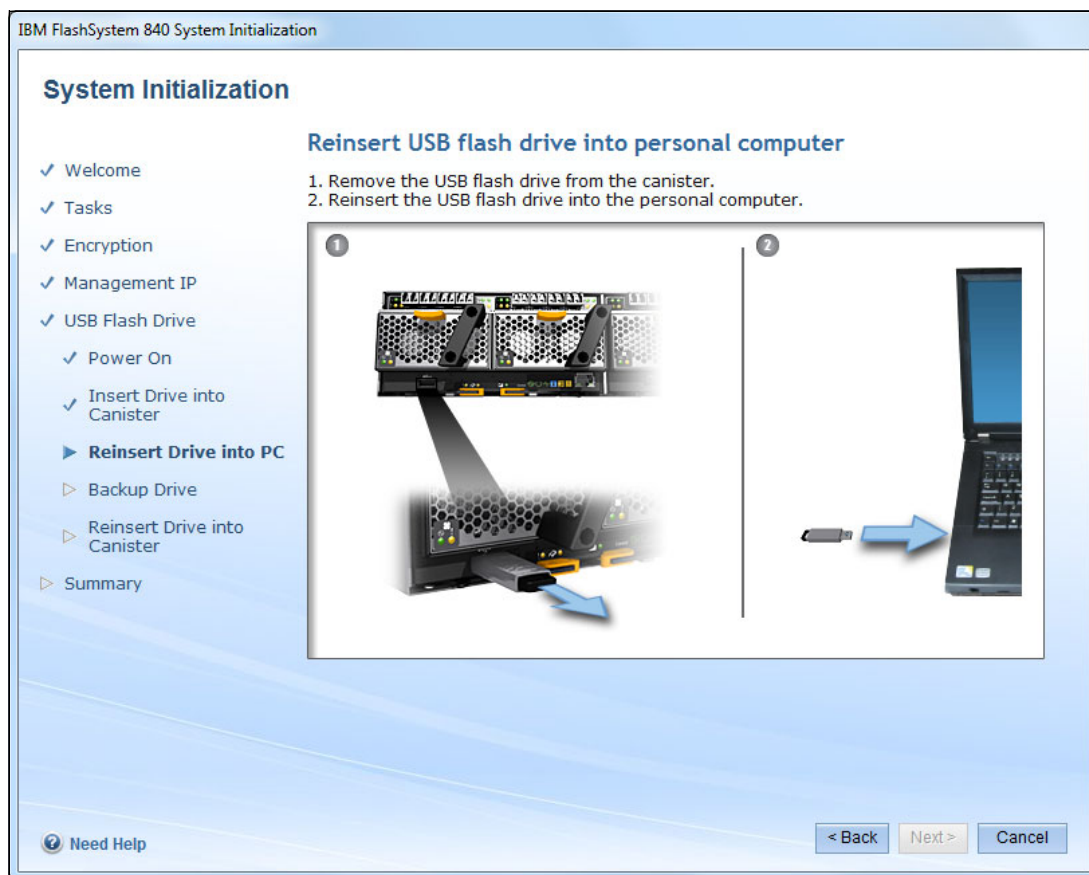


Figure 4-12 InitTool Reinsert USB flash drive into personal computer

System initialization failed

If the system initialization failed, you get the message from **InitTool** that is shown in Figure 4-13.

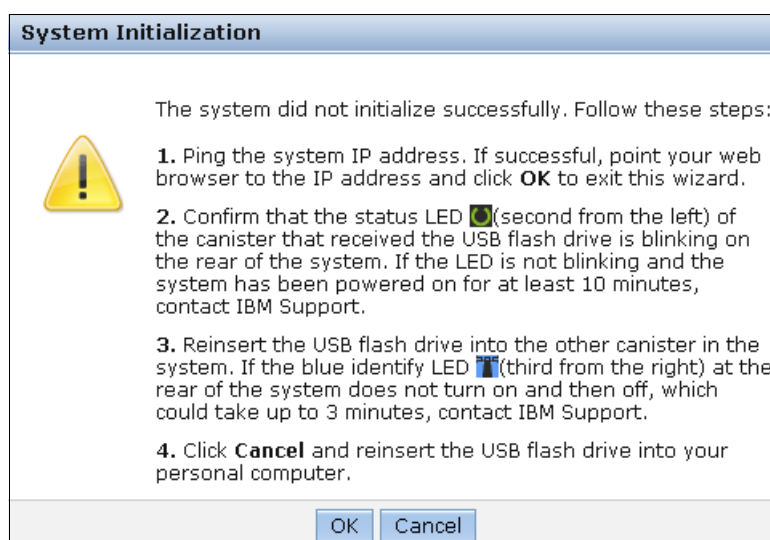


Figure 4-13 System initialization failed

System initialization can fail for multiple reasons. Follow the steps shown to resolve the situation. **InitTool** checks the content of the `satask_result.html` file. Only if a cluster was successfully created, the Next option is enabled for the initialization process to continue.

Backup USB flash drive with encryption key

If “Encryption Yes” was selected, **InitTool** prompts the operator to make at least two backup copies of the encryption key.

IBM provides two additional USB flash drives when the encryption feature code is purchased. You can make as many copies of the USB flash drive as you want, but you must make a minimum of two copies so that you have at least three USB flash drives in total.

At this point, you can either take out the original USB flash drive where you initiated **InitTool**, or you can insert an empty USB flash drive in another USB port of the personal computer where **InitTool** is executed.

Figure 4-14 shows that an empty USB flash drive is recognized, and we clicked **Next** to back up the encryption key.

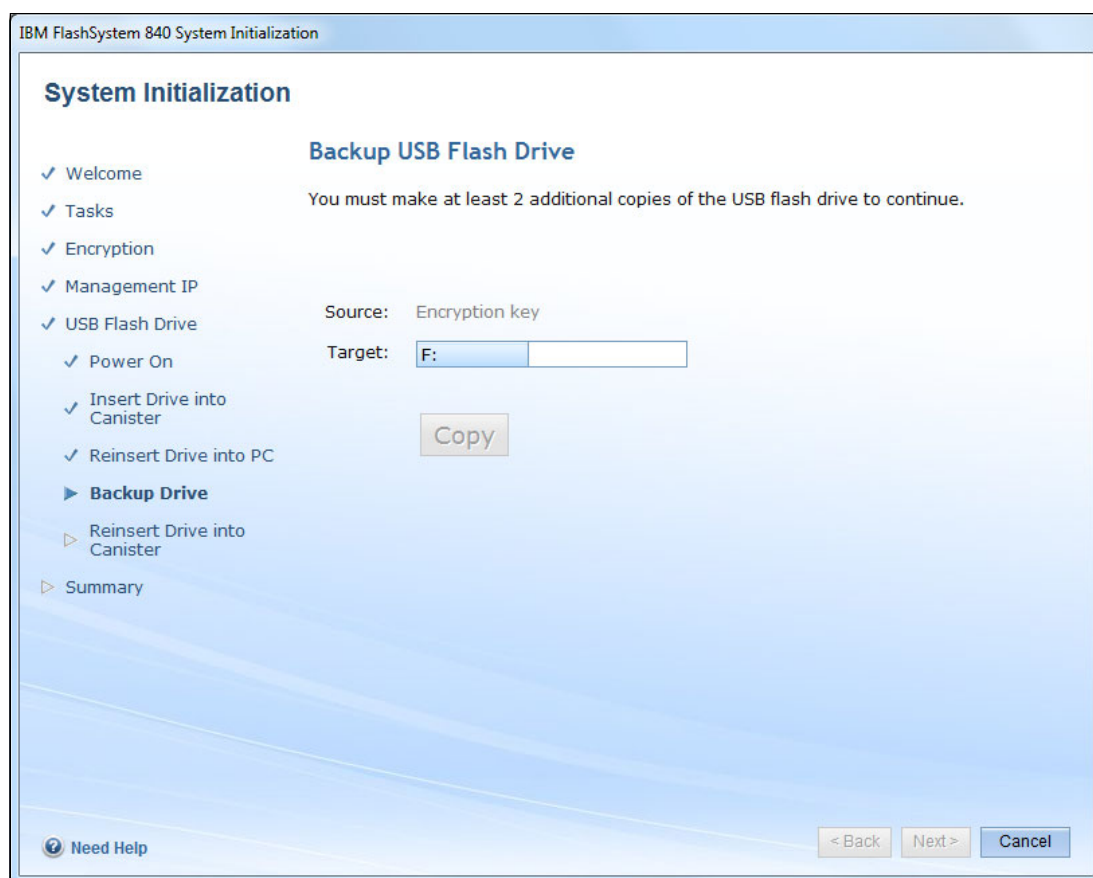


Figure 4-14 *InitTool Backup USB Flash Drive*

The previous step must be repeated for at least two backup copies of the encryption key.

Reinsert the USB flash drives into the canisters

Because we are installing a system with encryption key protection, **InitTool** now prompts us to insert the USB flash drives into the canisters. We insert the USB flash drives into the FlashSystem 840 canisters and click **Next** as shown in Figure 4-15 on page 86.

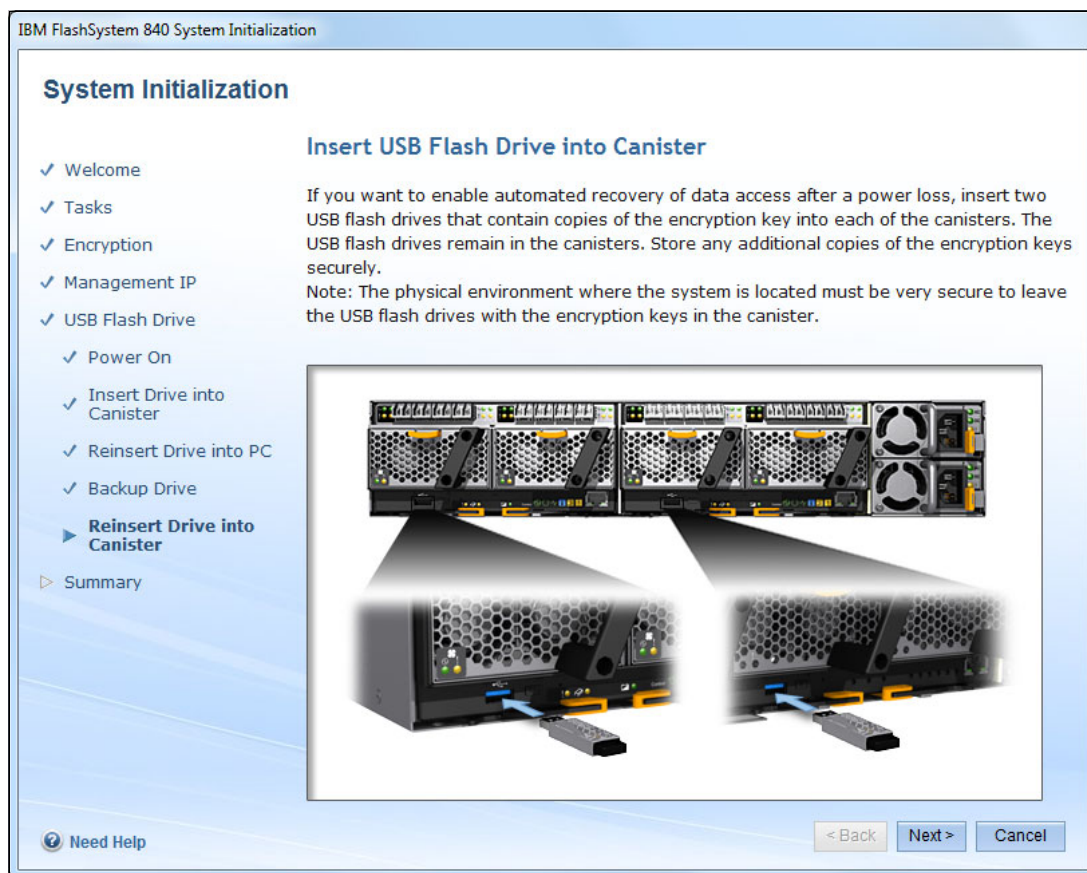


Figure 4-15 InitTool: Insert USB with encryption keys into canister

Note: In the latest version of FlashSystem 840, the encryption enablement procedure has moved from **InitTool** to the System Setup wizard similar to the encryption enablement procedure for FlashSystem 900. For more information about FlashSystem 900 System Setup wizard, see *Introducing and Implementing IBM FlashSystem V9000*, SG24-8273.

Check connectivity

InitTool now checks connectivity to the IBM FlashSystem 840, as shown in Figure 4-16 on page 87.

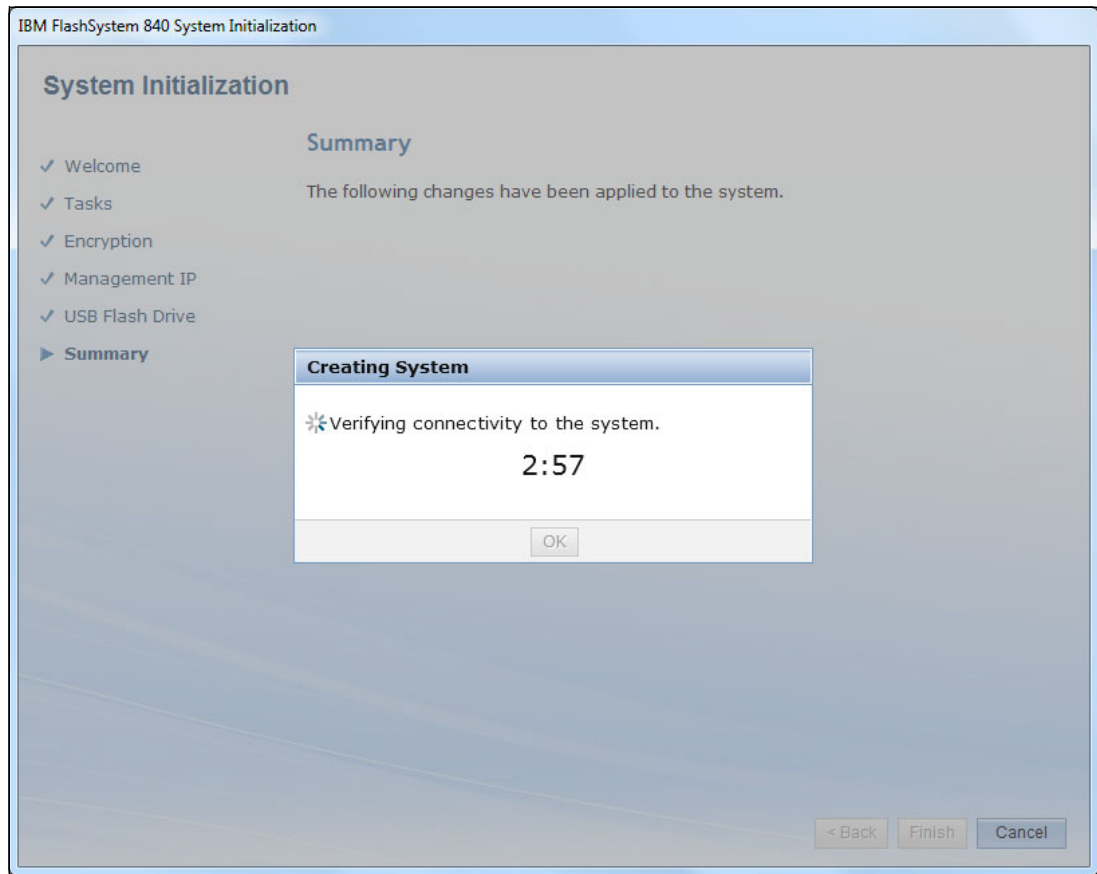


Figure 4-16 *InitTool: Verify connectivity*

Summary

If the Ethernet ports of the newly initialized FlashSystem 840 are attached to the same network as the personal computer where **InitTool** was executed, **InitTool** checks connectivity to the system and displays the result of the system initialization process.

If connectivity was successful, the Summary window is displayed as shown in Figure 4-17 on page 88.

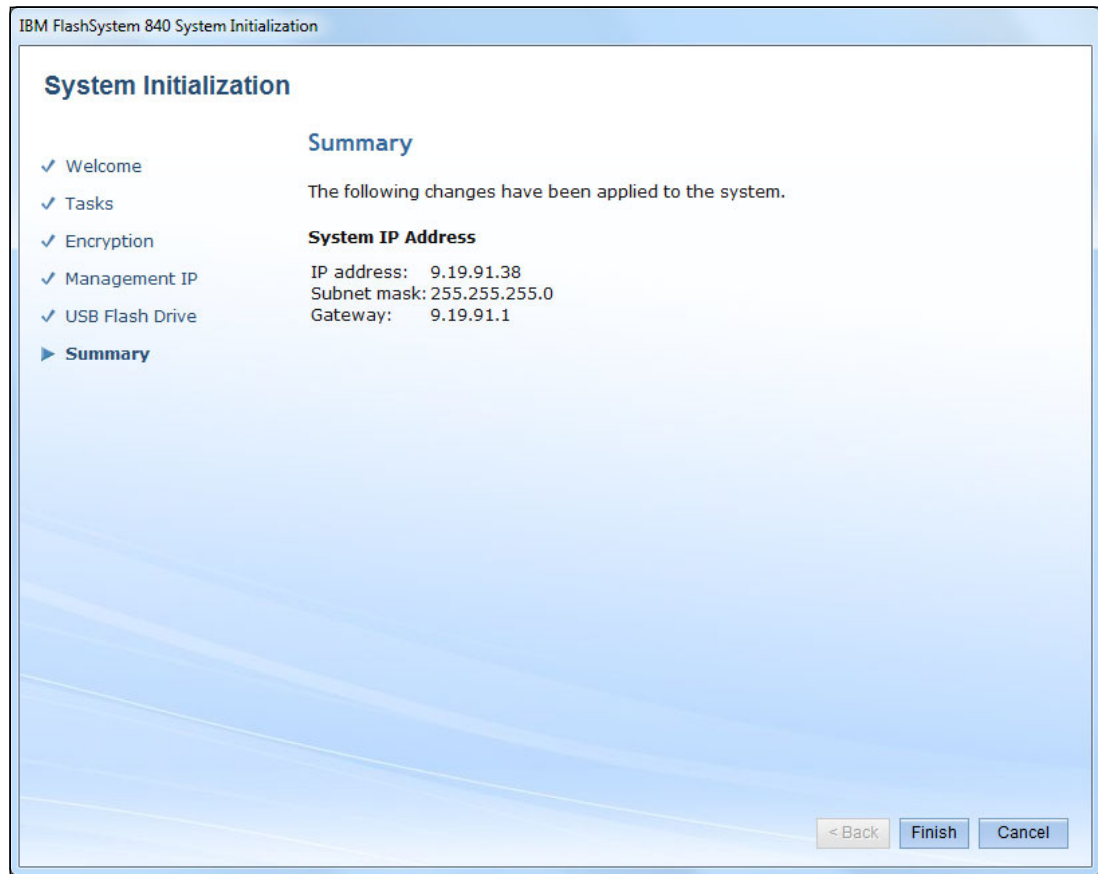


Figure 4-17 InitTool Summary

System initialization completed successfully

Figure 4-18 on page 89 shows the final window of the system initialization process that completed successfully.

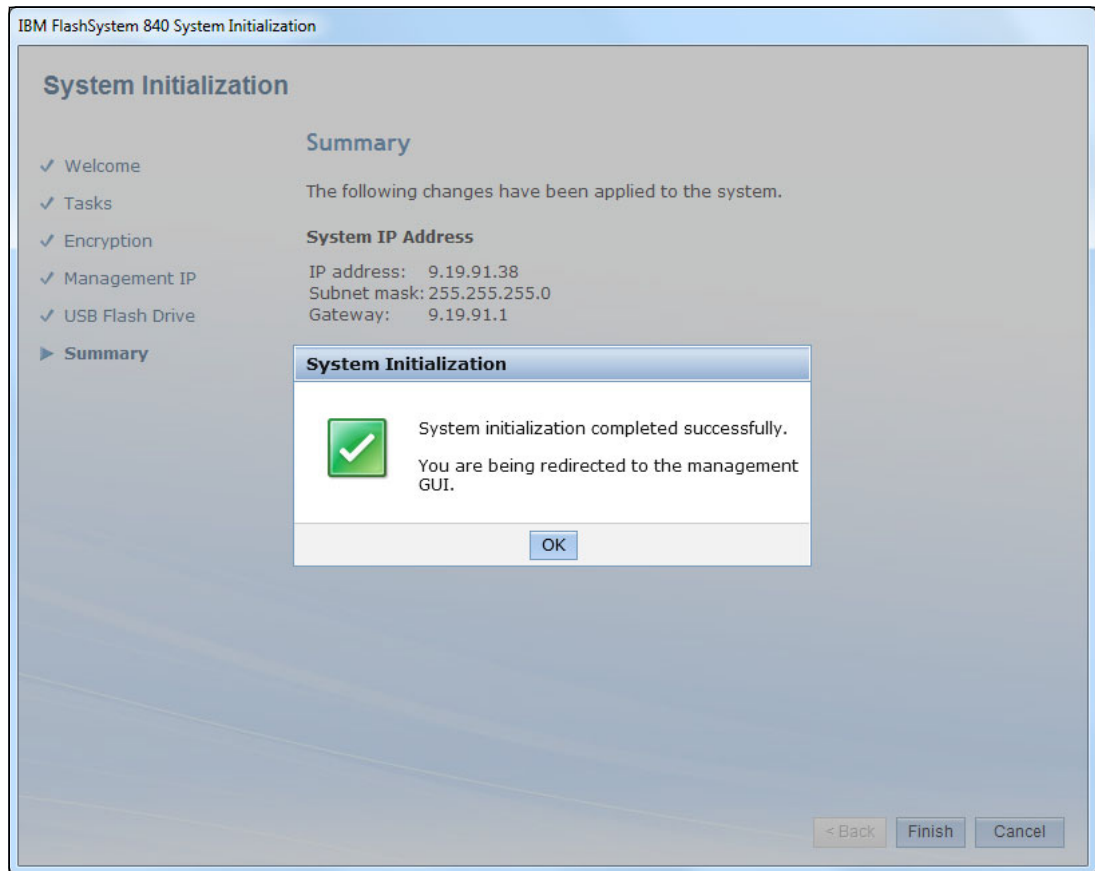


Figure 4-18 InitTool completion

The IBM FlashSystem 840 completed the first part of the initialization process and is ready for the next step.

The next step in the IBM FlashSystem 840 installation process is to open a web browser and log on to the system at the selected IP address, which we demonstrate in 4.3.3, “Initializing the system through the web management interface” on page 91.

Other purposes of InitTool

InitTool can be used for other purposes in addition to initializing the IBM FlashSystem 840. **InitTool** can also be used if you are unable to access the system. **InitTool** provides two options:

- ▶ Reset the superuser password:
 - You get to the login prompt, but you do not know the superuser password.
 - **InitTool** can, in this case, modify the superuser password.
- ▶ Set the service IP address:
 - You want to access a specific canister but you do not know the service IP addresses.
 - Specifying /service in the URL does not open the IBM FlashSystem 840 Service Assistant Tool.
 - **InitTool** can, in this case, modify the service IP addresses.

Just like with the IBM FlashSystem 840 initialization process, **InitTool** creates a file on the USB flash drive called `satask.txt`. This file contains a command that is read and executed by the IBM FlashSystem 840 canister when the USB flash drive is inserted into the USB port of the system.

Service Assistant Tool

IBM FlashSystem 840 Service Assistant Tool can be used in various service event cases. Service Assistant Tool is normally only used in cases where the client is instructed to use it by IBM Support because Service Assistant Tool contains destructive and disruptive functions.

Note: Only the superuser account is allowed to log on to Service Assistant Tool.

The following examples show the functions of IBM FlashSystem 840 Service Assistant Tool:

- ▶ Review installed hardware and firmware
- ▶ Review Ethernet ports and IP addresses
- ▶ Review worldwide names (WWNs)
- ▶ Change WWNs
- ▶ Canister enters the Service state
- ▶ Canister reboot
- ▶ Collect logs
- ▶ Reinstall software
- ▶ Configure CLI access
- ▶ Restart web service
- ▶ Recover system

There are two ways to open Service Assistant Tool. One is to point your web browser directly to the Service IP address of each canister. Another way is to point your web browser to the management IP address of your IBM FlashSystem 840 and to specify service in the URL, for example:

`https://192.168.10.10/service`

Important: Service Assistant Tool has destructive and disruptive functions. *Only open the Service Assistant Tool when you are instructed to do so by IBM Support.*

Supported web browsers

The web-based GUI is designed to simplify storage management and provide a fast and efficient management tool. It is loosely based on the IBM System Storage XIV software and has a similar look and feel.

The management GUI requires a supported web browser. At the time of writing this book, the management GUI supports the following web browsers with equal or higher versions:

- ▶ Mozilla Firefox 32
- ▶ Mozilla Firefox Extended Support Release (ESR) 31
- ▶ Microsoft Internet Explorer (IE) 10 and 11
- ▶ Google Chrome 37

IBM supports higher versions of the browsers if the vendors do not remove or disable functionality that the product relies on.

Additionally, the following are requirements for web browsers:

- ▶ JavaScript must be enabled
- ▶ Cookies must be allowed

- ▶ Enable scripts to disable or replace context menus. (Mozilla Firefox only)
- ▶ Enable TLS 1.1/1.2 (Microsoft Internet Explorer 9 and 10 only)

For a list of supported web browsers, and how to configure these, see *Web Browser Requirements* in the FlashSystem 840 IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/ST2NVR_1.3.0

4.3.3 Initializing the system through the web management interface

After the IBM FlashSystem 840 is initialized using the USB flash drive, you use a supported web browser to point it to the selected address.

Log on and change password

In this step of the IBM FlashSystem 840 initialization procedure, you log on to the web management GUI using the password `passw0rd` (with a zero) as shown in Figure 4-19.



Figure 4-19 Initialization procedure login as superuser

The IBM FlashSystem 840 at this point only allows the user `superuser` to access the system and therefore does not prompt for user name.

The next step is to change the password as shown in Figure 4-20 on page 92.



Figure 4-20 Initialization procedure to change the password

The IBM FlashSystem 840 administrators are encouraged not to leave the superuser password as the default but instead to create individual users with their own passwords for security reasons. You can also configure authentication and authorization for users of the clustered system as described in “Configure remote authentication” on page 246.

System Setup wizard

After the password is changed, the System Setup wizard starts. During the System Setup wizard, the administrator is prompted for the following information:

- ▶ Provide a system name.
- ▶ Configure the date and time in one of these ways:
 - NTP server (preferred)
 - Manual
- ▶ Confirm the number of flash modules.
- ▶ Configure the access type:
 - Open Access Yes: All hosts have access to all volumes.
 - Open Access No: Volumes must be mapped to hosts.
- ▶ Set up Call Home.
- ▶ Confirm the summary of changes.

All of these configuration steps can also be configured after the System Setup wizard is finished. The options for changing any of these configuration settings are to use either the GUI or CLI.

Figure 4-21 on page 93 shows the first step of the System Setup wizard, which is invoked automatically after the change password step.

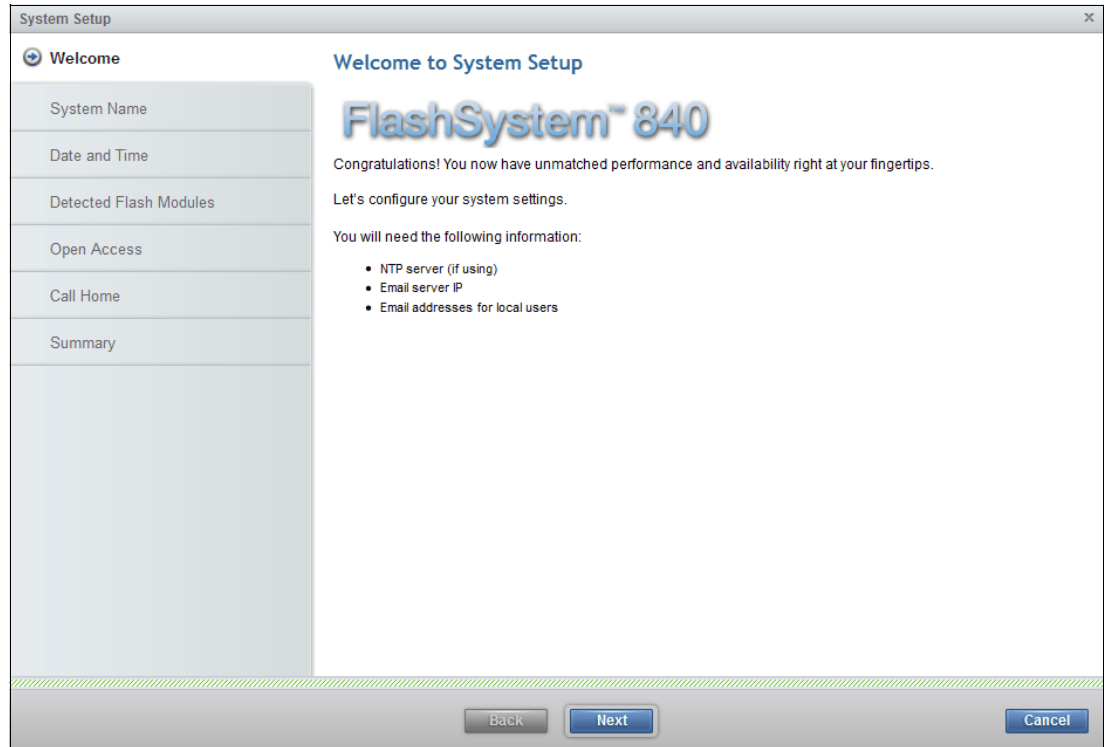


Figure 4-21 Initialization procedure welcome window

The Welcome window of the System Setup wizard tells you that the following information is needed:

- ▶ NTP server IP address for automatic and accurate system time
- ▶ SMTP server IP address for sending emails with warnings and alerts
- ▶ Email addresses for local users who will receive warnings and alerts

Configure a system name

Next, provide a host name for the system. In our example, we typed the system name, FlashSystem_840, as shown in Figure 4-22 on page 94. The host name can be changed later at any time from the main window at the FlashSystem 840 GUI.

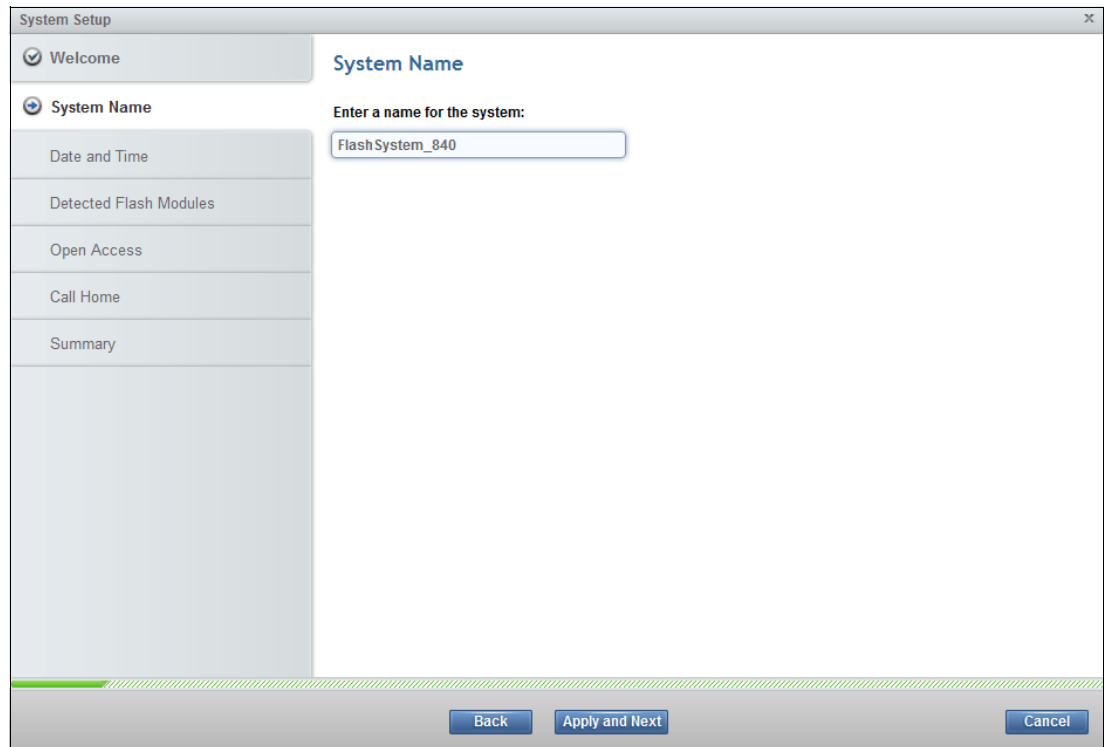


Figure 4-22 Initialization procedure to name the system

Each time that a new command executes, the Task window opens. At first, the Task window shows collapsed output. By clicking **Details**, the expanded output shows the command that is executed on the IBM FlashSystem 840, as shown in Figure 4-23.

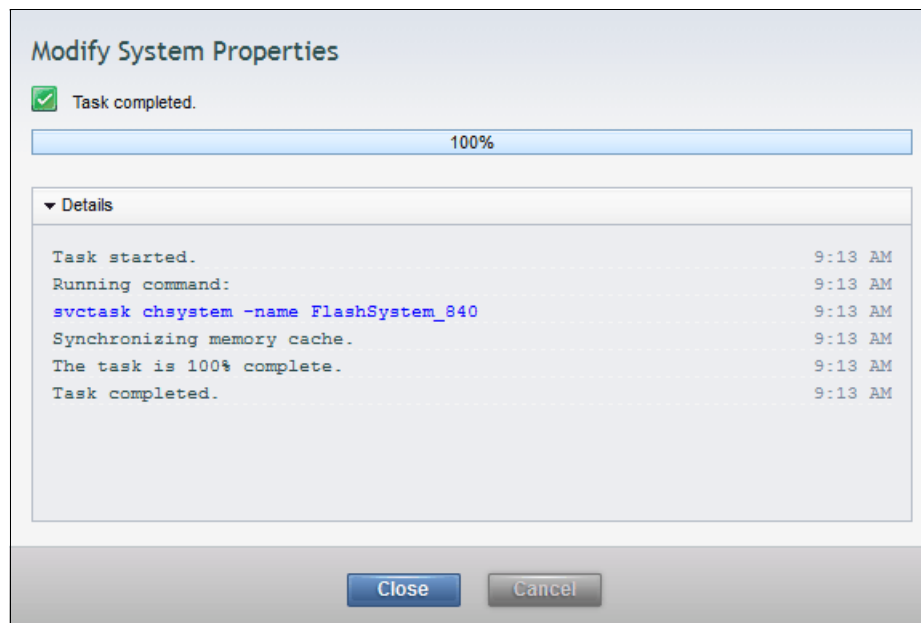


Figure 4-23 Initialization procedure task executing

Configure date and time

Next, configure the system date and time. The preferred practice is to configure the system with an NTP server. By using an NTP server, the date and time settings are always correct, which is useful in log analysis and troubleshooting. If an NTP server is not available at the time of installation, it can be added later. The date and time can then be set manually as shown in Figure 4-24.

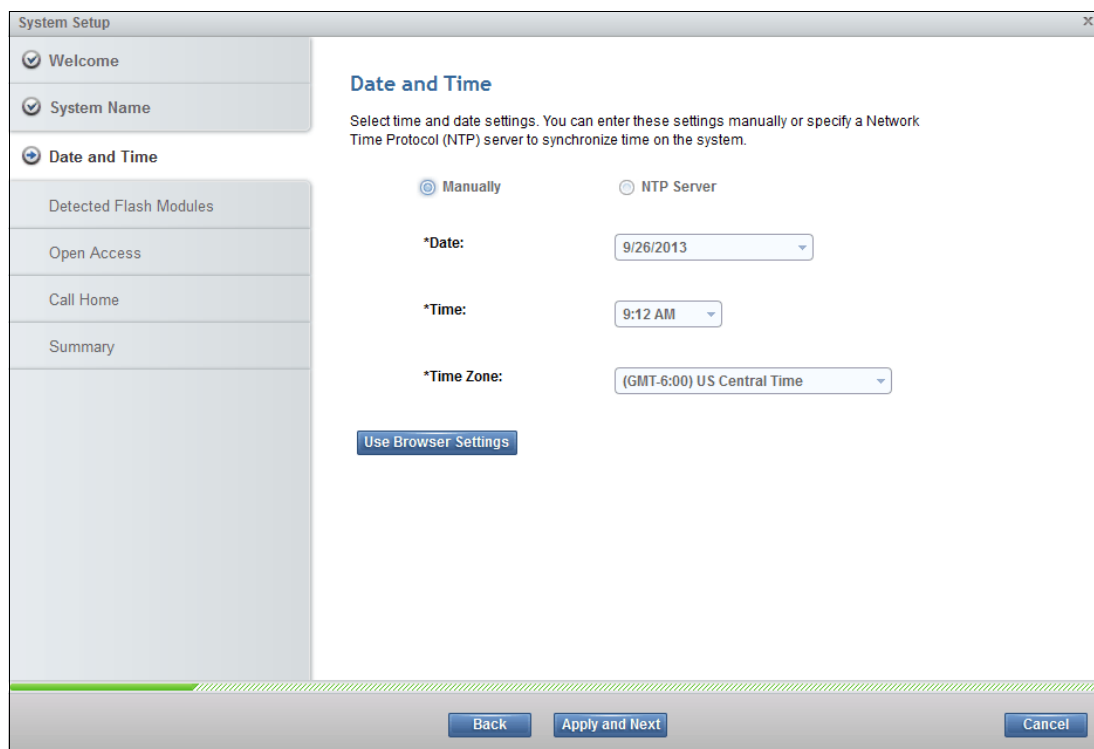
The screenshot shows the 'System Setup' window with a sidebar on the left containing 'Welcome', 'System Name', 'Date and Time' (selected), 'Detected Flash Modules', 'Open Access', 'Call Home', and 'Summary'. The main area is titled 'Date and Time' and contains the text: 'Select time and date settings. You can enter these settings manually or specify a Network Time Protocol (NTP) server to synchronize time on the system.' There are two radio buttons: 'Manually' (selected) and 'NTP Server'. Below them are three fields: '*Date:' with a dropdown showing '9/26/2013', '*Time:' with a dropdown showing '9:12 AM', and '*Time Zone:' with a dropdown showing '(GMT-6:00) US Central Time'. A 'Use Browser Settings' button is located below these fields. At the bottom of the window are 'Back', 'Apply and Next', and 'Cancel' buttons.

Figure 4-24 Initialization procedure date and time

A shortcut to setting the date and time manually is to click **Use Browser Settings**, which then inherits the date and time from the web browser in use.

Note: In the latest version of FlashSystem 840, the encryption enablement procedure has moved from **InitTool** to the System Setup wizard, and is the next step after setting date and time. The new System Setup wizard is similar to the encryption enablement procedure for FlashSystem 900 when enabling encryption on a new system.

For more information about FlashSystem 900 System Setup wizard, see *Introducing and Implementing IBM FlashSystem V9000*, SG24-8273, section 4.3.2, *Initializing the system through the web management interface*.

Verify detected flash modules

When the IBM FlashSystem 840 initializes, it detects the number of flash modules in the system. You must verify and confirm to the System Setup wizard that the size and quantity of the detected flash modules are correct.

By default, RAID 5 is the only redundancy option that can be configured. Changing to RAID 0 can be invoked after the System Setup wizard finishes and then only by using the CLI. For information about how to change the configuration to RAID 0, see Example 4-3 on page 106.

Note: As of FlashSystem 840 release 1.3, before enabling RAID 0, you must submit a SCORE/RPQ to IBM. To submit a SCORE/RPQ, contact your IBM representative.

Figure 4-25 shows the detected flash modules in the system. Our system has four flash modules configured in RAID 5.

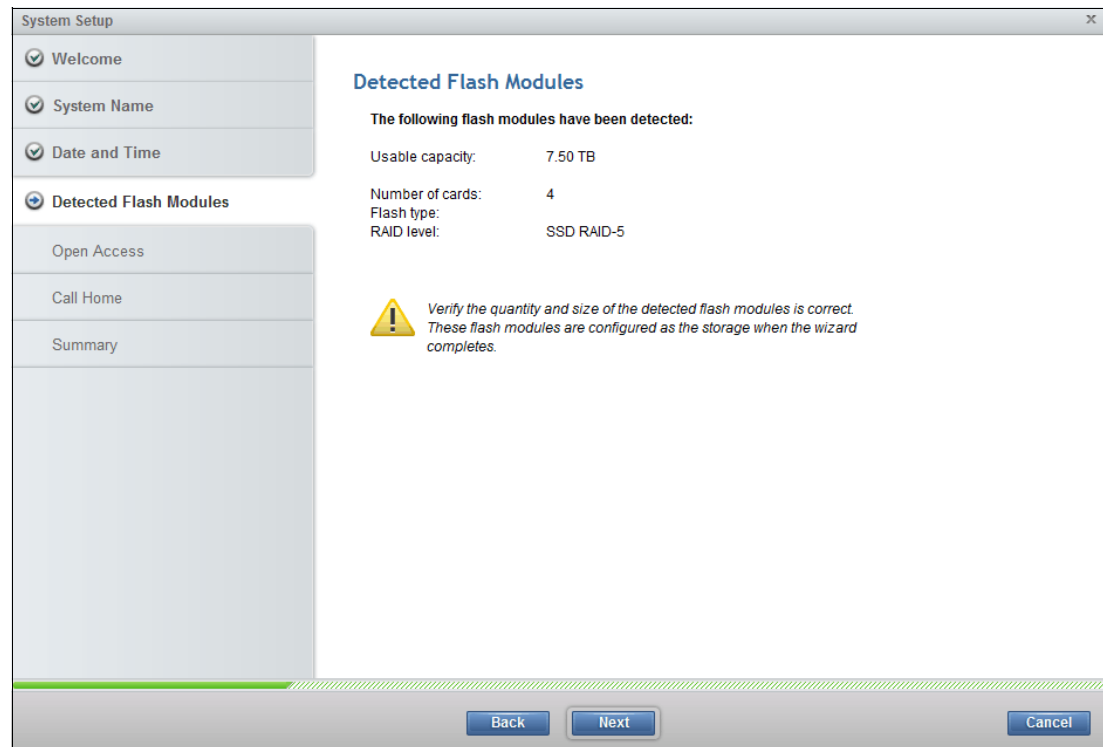


Figure 4-25 Initialization procedure detects flash modules

Allow or disallow open access

The IBM FlashSystem 840 can be configured to allow open access or to disallow open access. Open access is feasible when the system is connected to a group of hosts that all need the same access to the FlashSystem 840 volumes.

An example is a FlashSystem 840 that functions as a fast data store for a number of VMware servers in a cluster. By zoning the SAN switches to connect the IBM FlashSystem 840 to the VMware servers, they are all able to access all volumes.

Allowing open access is used where the IBM FlashSystem 840 is connected to correctly zoned SAN switches. However, disallowing open access and forcing the system to map its volumes only to selected hosts provide an extra layer of security.

When the system is configured to disallow open access, a few extra buttons appear in the GUI of the managed system. These buttons include the Hosts group of buttons and the Volumes by Host from the Volumes group of buttons. These buttons allow for selective mapping of volumes to hosts.

Figure 4-26 on page 97 shows the System Setup wizard when open access is allowed.

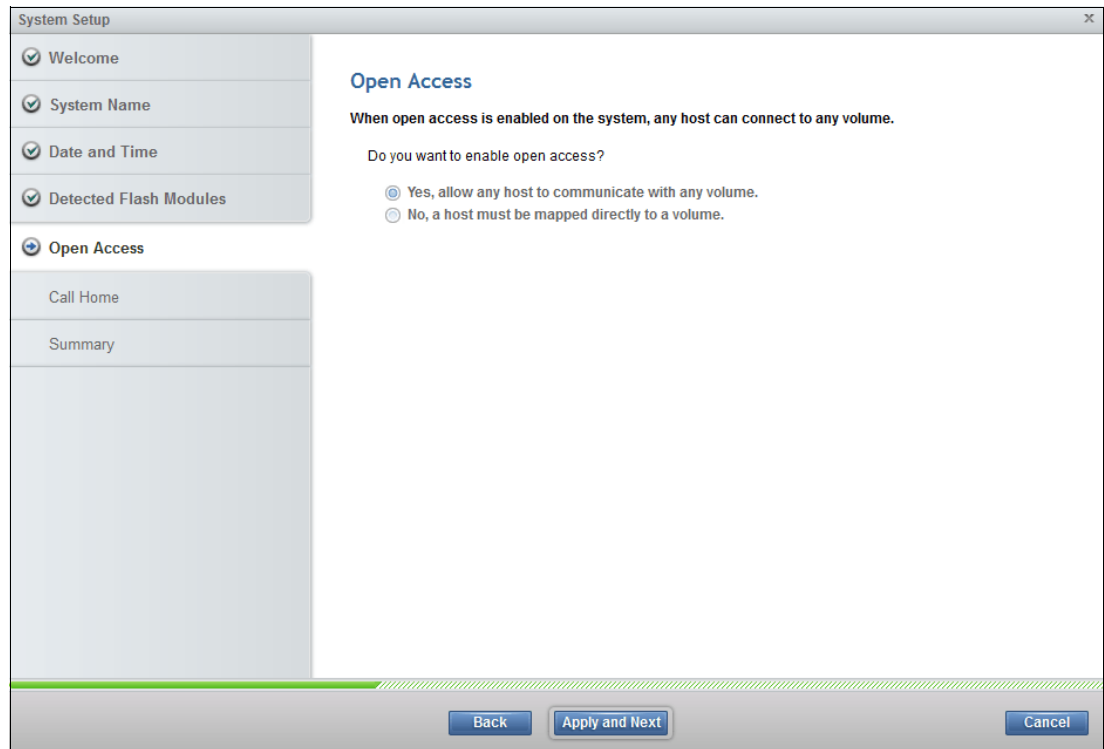


Figure 4-26 Initialization procedure to configure open access

Figure 4-27 shows the System Setup wizard where open access is disallowed, which is what we selected in our example.

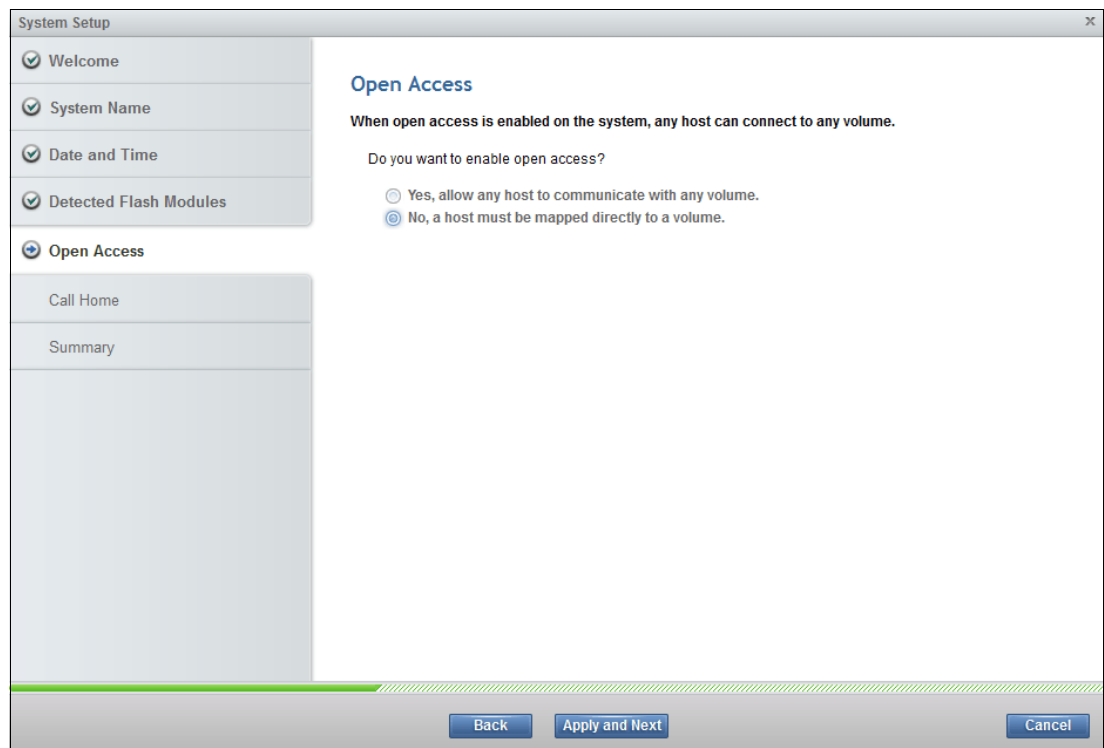


Figure 4-27 Initialization procedure denying open access

Configure Call Home

Next, configure Call Home. Call Home from the IBM FlashSystem 840 is invoked by sending an email with messages and alerts to IBM Support.

The following information must be provided to set up Call Home correctly:

- ▶ System Location: Where is the system physically installed?
- ▶ Contact Details: Who is the IBM Support contact for issues that require attention:
 - Name
 - Phone
 - Email
- ▶ Email server or servers (SMTP server):
 - IP address
 - Port (defaults to port 25)
- ▶ Event notification: Who else is to be notified of alerts, warnings, and errors?
- ▶ Summary: Confirm the settings.

Figure 4-28 shows that we set up Call Home by selecting **Yes**.

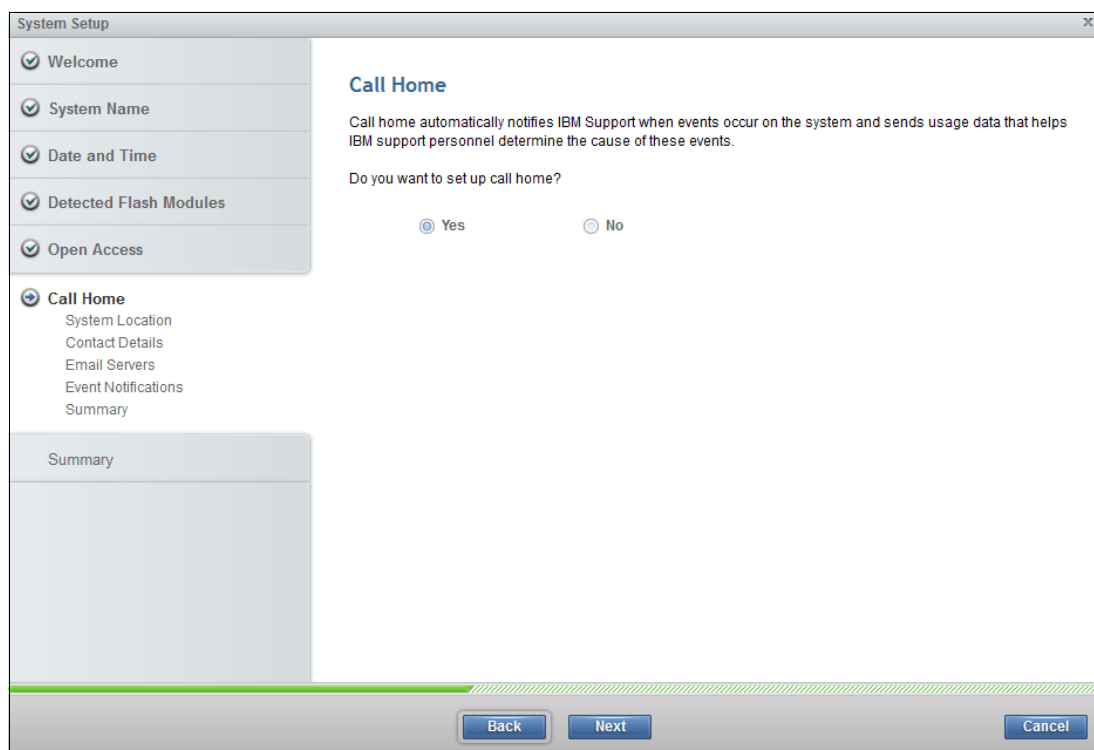


Figure 4-28 Initialization procedure: Call Home

If you select **No** for “Do you want to set up Call Home?”, you get a warning as shown in Figure 4-29 on page 99.

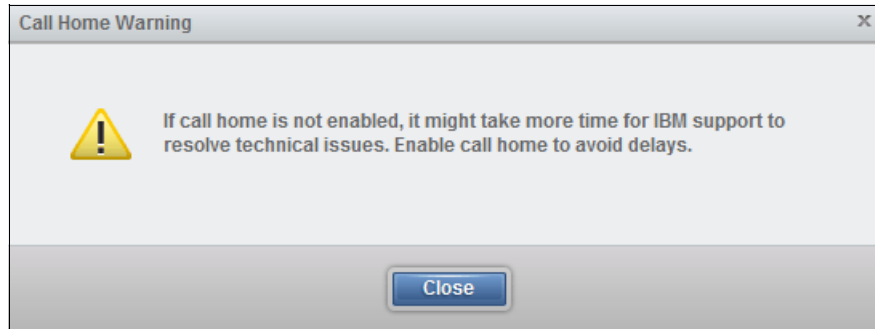


Figure 4-29 Initialization procedure: Call Home warning

In our example, we chose to enable Call Home and continued with the configuration of Call Home. The first window in the Call Home wizard is to configure the system location as shown in Figure 4-30.

 A screenshot of the "System Setup" window, specifically the "System Location" step. On the left is a sidebar with a list of steps: Welcome, System Name, Date and Time, Detected Flash Modules, Open Access, Call Home (expanded), System Location (selected), Contact Details, Email Servers, Event Notifications, and Summary. Below the sidebar is a "Summary" section. The main area is titled "System Location" and contains the instruction "Enter the company name and address to ship parts." Below this are several labeled input fields:

- *Company name: IBM_ITSO
- *Street address: 10777 Westheimer
- *City: Houston
- *State or province: TX
- *Postal code: 77077
- *Country or region: US (dropdown menu)

 At the bottom of the window are three buttons: "Back", "Next", and "Cancel".

Figure 4-30 Initialization procedure: Call Home System Location

The next step is to configure contact details. These reflect the client contact person that IBM Support will contact if issues arise that need attention and that create a service incident.

Figure 4-31 on page 100 shows where the client contact person is entered.

System Setup

- ✓ Welcome
- ✓ System Name
- ✓ Date and Time
- ✓ Detected Flash Modules
- ✓ Open Access
- Call Home
 - ✓ System Location
 - ➔ **Contact Details**
 - Email Servers
 - Event Notifications
 - Summary
- Summary

Contact Details

Enter the name and contact information for the person in your organization that IBM Support can contact to help resolve problems on the system.

*Contact name:

*Email address:

*Telephone (primary):

Telephone (alternate):

Comment:

Figure 4-31 Initialization procedure: Call Home Contact Details

Call Home from the IBM FlashSystem 840 is used by sending an email to IBM at a fixed email destination address that cannot be altered.

To send emails, the IBM FlashSystem 840 needs to know which mail server can transport the email. It also needs to know the TCP port through which the emails are sent. Figure 4-32 on page 101 demonstrates how to configure an email server. Multiple email servers can be configured by clicking the plus sign (+) to the right of the Server Port box.

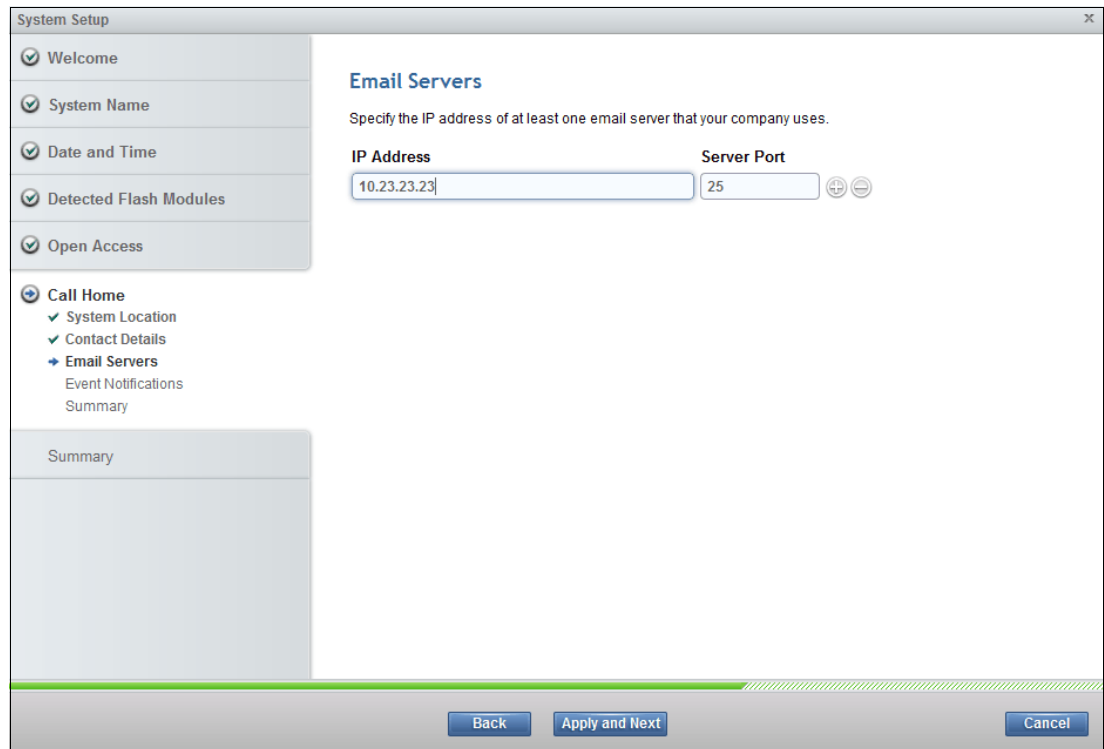


Figure 4-32 Initialization procedure: Configure the email servers to use Call Home

Next, optionally, you can configure event notifications. Email notification to IBM Support is automatically configured, but typically the client prefers to be notified if any issues occur that need attention. Client event notification is valuable if email transport to IBM fails. An email transport error can occur in an SMTP server outage, or if the SMTP server IP address is changed without the Call Home function of the IBM FlashSystem 840 being updated correctly.

Figure 4-33 on page 102 shows where client event notification is configured. Multiple client contacts can be configured by clicking the plus sign (+) to the right of the Event Type box.

Figure 4-33 Initialization procedure: Call Home Event Notifications

In Figure 4-34, a summary of the configured Call Home settings is displayed.

Figure 4-34 Initialization procedure: Call Home Summary

Summary

Finally, when all of the System Setup and Call Home details are configured, the System Setup wizard shows a summary of the configuration as shown in Figure 4-35.

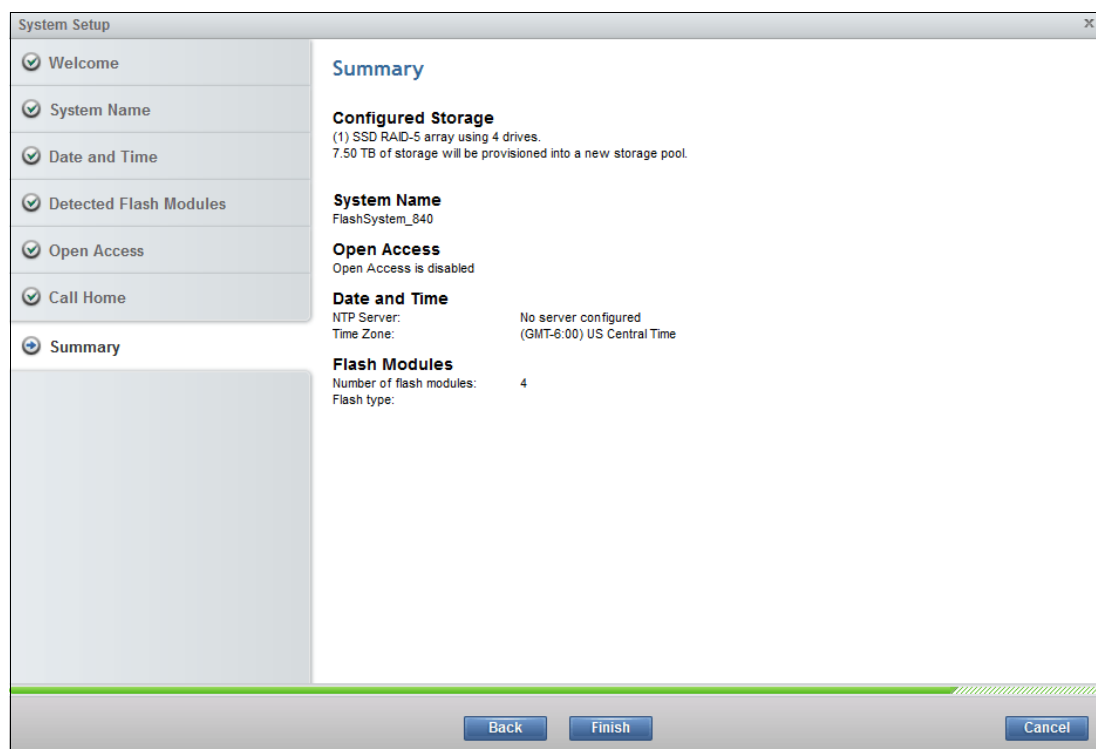


Figure 4-35 Initialization procedure: Summary

Initialization complete

The IBM FlashSystem 840 is now initialized and ready to use. The web browser automatically points to the main page of the management GUI as shown in Figure 4-36 on page 104.



Figure 4-36 IBM FlashSystem 840 GUI main page

For more information about how to use the FlashSystem 840 GUI, see Chapter 6, “Using the IBM FlashSystem 840” on page 165.

4.4 RAID storage modes

The process to implement the various RAID protection technologies that are used in the IBM FlashSystem 840 is described in more detail in Chapter 2, “IBM FlashSystem 840 architecture” on page 13.

4.4.1 Changing RAID modes

Every flash module in the IBM FlashSystem 840 implements patented IBM Variable Stripe RAID (VSR) within the flash module, as described in “Variable Stripe RAID” on page 107. This capability is in place whether the system is set up as a RAID 0 or a RAID 5 system. The IBM FlashSystem 840 supports the following system-level RAID protection configurations:

- ▶ RAID 0:
 - With only two flash modules in the system, RAID 0 is the only supported configuration.
 - All other flash module configurations of 4, 6, 8, 10, and 12 modules also support RAID 0.
 - RAID 0 configurations are protected against chip failures by VSR.
 - RAID 0 is not protected against flash module failure.
 - RAID 0 is not protected with a spare flash module.

- ▶ RAID 5:
 - Flash module configurations of 4, 6, 8, 10, and 12 modules support RAID 5.
 - RAID 5 configurations are protected against chip failures by VSR.
 - RAID 5 configurations are protected against entire flash module failures.
 - RAID 5 configurations have a spare flash module that becomes an active RAID 5 member immediately after a flash module failure.

The RAID controller in each of the two canisters is directly connected to the two interface cards in that canister, and to all the flash modules in the enclosure.

Note: RAID 0 flash arrays have no redundancy for flash module failure, and they do not support hot spare takeover.

Implementing RAID 5

RAID 5 is the default RAID protection mode. RAID 5 is the only RAID protection mode that is configured using **InitTool** and the subsequent IBM FlashSystem 840 initialization process. Also, RAID protection mode cannot be changed from within the GUI after the system is initialized. Therefore, implementing RAID 5 occurs when the system is initialized as demonstrated in 4.3, “Initializing the system” on page 74.

Implementing RAID 0

Configure the IBM FlashSystem 840 with RAID 5, which provides higher availability than RAID 0. RAID 0 does not provide redundancy for the loss of flash modules. You might have reasons, however, to configure RAID 0. One reason to configure RAID 0 is if more capacity is required than available with RAID 5. Another reason to configure RAID 0 is that a connected host or the IBM SAN Volume Controller mirrors data between two IBM FlashSystem 840 storage systems. In these cases, the same requirements of high availability do not apply.

RAID 0 can be implemented only after the IBM FlashSystem 840 initialization process is complete. This IBM FlashSystem 840 initialization process configures the system for the RAID 5 level of protection.

Note: Using the **rmarray** CLI command destroys the array that is configured. Any data existing on the array before you use this command is lost.

Use the CLI as shown in Example 4-3 on page 106 (output shortened for clarity) to change from RAID 5 to RAID 0.

Example 4-3 Remove RAID 5 and create RAID 0 flash array

```
IBM_9840:ITS0 FS840:superuser> rmarrray
```

```
IBM_9840:ITS0 FS840:superuser> lsarray
```

```
IBM_9840:ITS0 FS840:superuser> svctask mkarray -level raid0
```

```
MDisk, id [0], successfully created
```

```
You have new mail in /var/spool/mail/root
```

```
IBM_9840:ITS0 FS840:superuser> lsarrayinitprogress
```

```
mdisk_id mdisk_name progress estimated_completion_time
```

```
0          array0      100
```

```
IBM_9840:ITS0 FS840:superuser> lsarray
```

mdisk_id	mdisk_name	status	mdisk_grp_id	mdisk_grp_name	capacity	raid_status	raid_level	redundancy	strip_size
0	array0	online	0	mdiskgrp0	7.5TB	online	raid0	0	4

After implementing RAID 0 from the IBM FlashSystem 840 CLI, you can change it back to RAID 5 by deleting the created RAID 0 flash array and then creating a new RAID 5 flash array as shown in Example 4-4 (output shortened for clarity).

Example 4-4 Remove RAID 0 and create RAID 5 flash array

```
IBM_9840:ITS0 FS840:superuser> rmarrray
```

```
IBM_9840:ITS0 FS840:superuser> lsarray
```

```
IBM_9840:ITS0 FS840:superuser> svctask mkarray -level raid5
```

```
MDisk, id [0], successfully created
```

```
You have new mail in /var/spool/mail/root
```

```
[09:43:50] 825_sys41_1:~ # lsarray
```

mdisk_id	mdisk_name	status	mdisk_grp_id	mdisk_grp_name	capacity	raid_status	raid_level	redundancy	strip_size
0	array0	offline	0	mdiskgrp0	3.7TB	initting	raid5	0	4

```
[09:43:56] 825_sys41_1:~ # lsarrayinitprogress
```

```
mdisk_id mdisk_name progress estimated_completion_time
```

```
0          array0      8          131008094452
```

```
[09:45:27] 825_sys41_1:~ # lsarrayinitprogress
```

```
mdisk_id mdisk_name progress estimated_completion_time
```

```
0          array0      100
```

```
[09:45:36] 825_sys41_1:~ # lsarray
```

mdisk_id	mdisk_name	status	mdisk_grp_id	mdisk_grp_name	capacity	raid_status	raid_level	redundancy	strip_size
0	array0	online	0	mdiskgrp0	3.7TB	online	raid5	1	4

Note: Any volumes with data that existed on the IBM FlashSystem 840 before the RAID 5 array was removed are lost. Therefore, CLI commands, such as **svctask rmarrray** and **svctask mkarray**, must only be used on a system that has no data that needs to be preserved.

Variable Stripe RAID

In addition to RAID 0 and RAID 5, which protect the IBM FlashSystem 840 against an entire flash module failure, the IBM FlashSystem 840 also supports Variable Stripe RAID (VSR). VSR is a built-in data protection technology that protects data from subcomponent failure on a flash module.

The subcomponent failure VSR protects against a *plane*. When a bad plane occurs, VSR technology allows a plane to be removed from use without affecting the available capacity of other devices within the RAID stripe. Upon detection of a failure, the failing plane is removed from use (no further writes are allowed) and all used pages within the affected stripe are marked as “critical to move”. Information from the affected stripe is then gradually relocated to stripes that are known to be good, which is a process that is performed as a background task to minimize the required processing power.

Note: Even in the case of the IBM FlashSystem 840 operating in RAID 0 mode, data is still protected by VSR.

VSR is described in greater detail in Chapter 1, “FlashSystem storage introduction” on page 1.

4.5 Connectivity guidelines for improved performance

You can configure the various network connections to improve the overall performance of the IBM FlashSystem 840. Next, we describe considerations for planning the installation of the storage system.

4.5.1 Interface card configuration guidelines

You can improve reliability and performance by following specific network connection guidelines for the interface cards in the enclosure.

Figure 4-37 shows the rear side of the IBM FlashSystem 840 when installed with four FC or FCoE interface cards.

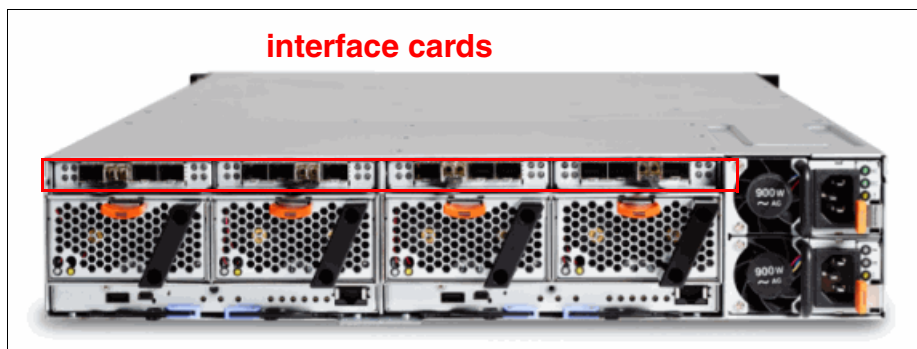


Figure 4-37 Interface cards

The FlashSystem 840 enclosure includes two canisters, each containing two FC, FCoE, iSCSI, or QDR InfiniBand interface cards. To eliminate any single failure point, use multipathing to span across interface cards in both canisters. Most major operating systems incorporate a multipathing option.

The storage system allows the flexible presentation of volumes, which are also called *logical unit numbers* (LUNs), through the interface cards. A volume or LUN can be presented and accessed through every port simultaneously. The loss of one interface card does not affect the I/O performance of the other interface cards. This configuration is commonly referred to as *active/active* or *active/active symmetric multipathing*.

Note: The IBM FlashSystem 840 interface cards are equipped with either two 16 Gbps FC, four 8 Gbps FC, four 10 Gbps FCoE, four 10 Gbps iSCSI, or four QDR InfiniBand ports.

Each canister mounts with two interface cards. Although having up to 16 ports on a single IBM FlashSystem 840 is possible, it is not a cabling or zoning requirement that host communication is established through all available ports. Therefore, an attached host can connect to fewer than all four ports on the interface cards.

4.5.2 Host adapter guidelines

To improve bandwidth, install dual-port host bus adapters and host controller adapters (HCAs) in the servers that are used with the storage system.

Dual-port HBAs and HCAs provide more ports for aggregating bandwidth. In an ideal case, multiple dual-port HBAs and HCAs are installed in each server for redundancy.

Quad Data Rate (QDR) InfiniBand HCAs maximize the bandwidth to the storage system. Most QDR InfiniBand HCAs use PCI Express 2.0 instead of version 1.0 to allow for the higher-rated bandwidth. Servers with PCI Express 2.0, or later, expansion slots provide the highest bandwidth.

The storage system is tested for interoperability against all major HBA and HCA vendors.

To verify valid and supported configurations, see this website:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

4.5.3 Cabling guidelines

As described in 4.2, “Cabling the system” on page 69, several considerations exist related to the correct cabling.

In addition, design cabling to provide resiliency across storage system management ports and FC, FCoE, iSCSI, or QDR InfiniBand interface card ports. Also, consider the related network switches and server HBA and HCA ports and create a design that provides redundancy and has enough paths for optimal performance.

All high-availability concepts of a dual-switched fabric setup apply to the storage system. One key element to recognize when cabling the storage system is the use of available paths. The storage system is designed to deliver I/O throughput through all connected ports. Use all ports if the application can benefit from more bandwidth. To take advantage of these ports on the storage system, there must be an equal number of server ports. Otherwise, there are under-utilized ports on one side of the fabric.

4.5.4 Zoning guidelines

In a switched fabric environment, implementing zoning when you connect the storage system to the network can improve performance and simplify upgrades.

In a switched fabric deployment, it is common to isolate one application's server storage devices from other applications' server storage devices. This practice prevents cross-traffic and helps ease maintenance. Therefore, employ zoning in all multiple server environments.

Zoning a server's HBA or HCA is best deployed when only a single HBA or HCA is contained in a zone. This approach is also referred to as *Single Initiator Zoning*. Having more than one HBA or HCA in a zone can lead to unexpected errors and can cause unplanned downtime. HBAs or HCAs are called *initiators*, and these initiators can be zoned to a single or multiple *target* ports. A target port is a storage controller port, such as the ports in the IBM FlashSystem 840.

The host initiators must be zoned to sufficient storage target ports to carry the expected workload. Host initiator ports must however not be zoned to more than the necessary number of storage target ports. Zoning host initiator ports to more than the necessary number of storage target ports can result in excessive paths to storage, which can decrease performance.

Note: When you connect the storage system to a dual-switched fabric architecture, ensure that each zoned server uses interface cards in both canisters in the storage system. Ensure that each zoned server does not depend on a single interface card or a single canister for availability.

For information about how to zone an IBM FlashSystem to a SAN Volume Controller, see Chapter 8, "Product integration" on page 275.

For detailed information about how to zone storage devices in general to the IBM SAN Volume Controller, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.



IBM FlashSystem 840 client host attachment and implementation

This chapter provides installation, implementation, and other general information and guidelines for connecting client host systems to the IBM FlashSystem 840 (FlashSystem 840).

This chapter describes the following topics:

- ▶ FlashSystem 840 sector size and host block size considerations
- ▶ Partition and file alignment for best performance
- ▶ Multipath support implementation for various host operating systems
- ▶ Necessary drivers for several operating systems
- ▶ Host integration for various operating systems

Note: Some of the following sections mention IBM SAN Volume Controller, which delivers the functions of IBM Spectrum Virtualize, part of the IBM Spectrum Storage family.

For the purposes of this chapter, the name IBM SAN Volume Controller has been retained with the intent to update it to IBM Spectrum Virtualize as appropriate for publication.

5.1 Host implementation and procedures

The necessary procedures to connect the IBM FlashSystem 840 to client hosts using various operating systems are described in the following sections.

5.2 Host connectivity

The IBM FlashSystem 840 can be attached to a client host by four methods:

- ▶ Fibre Channel (FC)
- ▶ Fibre Channel over Ethernet (FCoE)
- ▶ InfiniBand
- ▶ Internet Small Computer System Interface (iSCSI)

Always check the IBM System Storage Interoperation Center (SSIC) to get the latest information about supported operating systems, hosts, switches, and so on:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

If a configuration that you want is not available on the SSIC, a Solution for Compliance in a Regulated Environment (SCORE)/request for price quotation (RPQ) must be submitted to IBM requesting approval. To submit a SCORE/RPQ, contact your IBM FlashSystem marketing representative or IBM Business Partner.

The IBM FlashSystem 840 can be storage area network (SAN)-attached using a switch or directly attached to a client host. Check the IBM SSIC for specific details. Several operating system and FC driver combinations allow point-to-point direct access with 16 Gbps FC. Check your environment and the SSIC to use 16 Gbps direct attachment to the host.

Note: The FlashSystem 840 16 Gbps FC attachment does not support arbitrated loop topology (direct connection to client hosts). The IBM FlashSystem 840 must be connected to a SAN switch when using 16 Gbps FC if the host operating system does not support point-to-point FC direct connections. At the time that this book was written, IBM AIX did not support point-to-point FC direct connections.

5.2.1 Fibre Channel SAN attachment

If you attach a host using a SAN switch to the FlashSystem 840, ensure that each host port is connected and zoned to both canisters of the FlashSystem 840. If only one FlashSystem 840 canister is connected to a host port, the host state is shown as *degraded*. This is referred to in the remainder of this chapter as the *switch rule*.

Note: When you use a switch, you must zone host ports according to the switch rule. A host port has to be connected to each FlashSystem 840 canister.

Figure 5-1 on page 113 shows the correct SAN connection of an AIX server with two ports to the FlashSystem 840. In this example, four zones are set up:

- ▶ AIX port 8a FlashSystem 840 port 41
- ▶ AIX port 8a FlashSystem 840 port 61
- ▶ AIX port 27 FlashSystem 840 port 51
- ▶ AIX port 27 FlashSystem 840 port 71

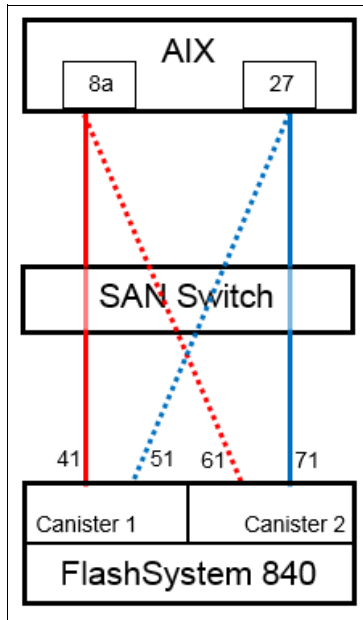


Figure 5-1 SAN attachment

5.2.2 Fibre Channel direct attachment

If you attach the FlashSystem 840 directly to a host, the host must be attached to both canisters. If the host is not attached to both canisters, the host is shown as *degraded*.

Figure 5-2 shows the correct direct attachment of an AIX server with two ports to the FlashSystem 840. The two connections are shown in this example:

- ▶ AIX port 8a directly attached to FlashSystem 840 port 41
- ▶ AIX port 27 directly attached to FlashSystem 840 port 71

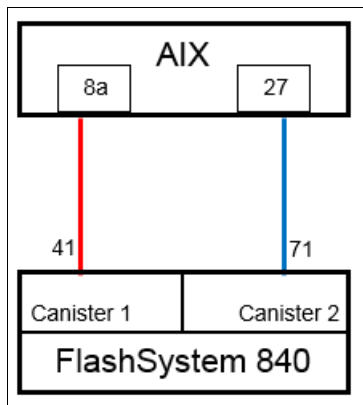


Figure 5-2 Direct attachment

If you use SAN attachment and direct attachment simultaneously on a FlashSystem 840, the direct-attached host state is *degraded*. Using a switch enforces the *switch rule* for all attached hosts, which means that a host port has to be connected to both FlashSystem canisters. Because a direct-attached host cannot connect one port to both canisters, it does not meet the **switch rule** and its state is *degraded*.

Note: You can attach a host through a switch and simultaneously attach a host directly to the FlashSystem 840. But then, the direct-attached host is shown as *degraded*.

5.2.3 General Fibre Channel attachment rules

These rules apply to FC connections:

- ▶ If direct attached, a host must have ports connected to both canisters.
- ▶ If connected to a switch, all host ports must have paths to both canisters. This is the switch rule.
- ▶ If any port is connected to a switch, the switch rule applies to all hosts, regardless of whether that host is connected through a switch.

5.3 Operating system connectivity and preferred practices

Detailed information about the IBM FlashSystem 840 client host connections using various operating systems is described in the following sections.

5.3.1 FlashSystem 840 sector size

With traditional spinning disk, a *sector* refers to a physical part of a disk. The size of the sector is defined by the disk manufacturer and most often set to 512 bytes. The 512-byte sector size is supported by most operating systems.

The FlashSystem 840 does not have fixed physical sectors like spinning disks. Data is written in the most effective way on flash. The sector size however is the same as the sector size of most traditional spinning disk sectors to maintain compatibility. Therefore, the default sector size that is presented to the host by the FlashSystem 840 is 512 bytes. Starting with firmware version 1.1.3.0, you can create volumes with a sector size of 4096 bytes by using the command-line interface (CLI) command the **mkvdisk** and the new parameter **-blocksize**. For details about this command, check the FlashSystem 840 IBM Knowledge Center at:

http://www.ibm.com/support/knowledgecenter/ST2NVR_1.3.0

The **mkvdisk** parameter **-blocksize** specifies the Small Computer System Interface (SCSI) logical unit sector size. The only two possible values are 512 (the default) and 4096.

- ▶ 512 is the default. It is supported by all operating systems.
- ▶ 4096 provides better performance but it might not be supported by your host operating system or application.

Note: Format all client host file systems on the storage system at 4 KB or at a multiple of 4 KB. This is a recommendation for a used sector size of 512 and 4096 bytes. For example, file systems that are formatted at an 8 KB allocation size or a 64 KB allocation size are satisfactory because they are a multiple of 4 KB.

5.3.2 File alignment for the best RAID performance

File system alignment can improve performance for storage systems using a RAID storage mode. File system alignment is a technique that matches file system I/O requests with important block boundaries in the physical storage system. Alignment is important in any

system that implements a RAID layout. I/O requests that fall within the boundaries of a single stripe have better performance than an I/O request that affects multiple stripes. When an I/O request crosses the endpoint of one stripe and into another stripe, the controller must then modify both stripes to maintain their consistency.

Unaligned accesses include those requests that start at an address that is not divisible by 4 KB, or are not a multiple of 4 KB. These unaligned accesses are serviced at much higher response times, and they can also significantly reduce the performance of aligned accesses that were issued in parallel.

The IBM FlashSystem 840 provides 512-byte sector size support that greatly improves response times for I/O requests that cannot be forcibly aligned. However, alignment to 4 KB must be maintained whenever possible.

5.3.3 IBM AIX and FlashSystem 840

The IBM FlashSystem 840 can be attached to AIX client hosts using the following FC method.

The IBM FlashSystem 840 connects to AIX through Node Port Identifier Virtualization (NPIV) and Virtual I/O Server (VIOS) modes.

Directly attached Fibre Channel topology for AIX

Configure the FlashSystem 840 FC controllers to arbitrated loop topology when the controllers are directly attached to the AIX hosts. You have to check the SSIC for supported configurations. For more details about the SSIC, see 5.2, “Host connectivity” on page 112.

Note: The FlashSystem 840 16 Gbps FC ports do not support direct connection to client hosts. A SAN switch must be placed between the IBM FlashSystem 840 and any 16 Gbps-attached client host. If arbitrated loop is required by the client host, connect at 8 Gbps FC to the IBM FlashSystem 840.

Optimal logical unit number configurations for AIX

The number of logical unit numbers (LUNs) that you create on the IBM FlashSystem 840 can affect the overall performance of AIX.

Applications perform optimally if at least 32 LUNs are used in a volume group. If fewer volumes are required by an application, use the Logical Volume Manager (LVM) to map fewer logical volumes to 32 logical units. This does not affect performance in any significant manner (LVM overhead is small).

Sector size restrictions for AIX

The IBM FlashSystem 840 supports only a 512-byte sector size. The AIX operating system supports the 512-byte sector size.

Auto Contingent Allegiance support

Certain host systems require the Auto Contingent Allegiance (ACA) support to run multiple concurrent commands. When using the round-robin multipathing algorithm, IBM AIX sends out extraneous ACA task management commands. ACA support on logical units is always enabled on the IBM FlashSystem 840.

Volume alignment

The IBM AIX operating system volumes align to 4 KB boundaries.

Implementing multipathing for IBM AIX hosts

How to set up multipathing with AIX is described. *Multipathing* enables the host to access the FlashSystem 840 LUNs through different paths. This architecture helps to protect against I/O failures, such as port, cable, or other path issues.

Important: For the latest updates for multipathing support on the IBM AIX operating system, go to IBM Fix Central:

<http://www.ibm.com/support/fixcentral>

Resetting the host bus adapter and disk configuration

In the following sections, we describe how to reconfigure the host bus adapters (HBAs) to implement multipathing. After you install the latest IBM AIX updates for support of the IBM FlashSystem 840, AIX must rescan the SCSI bus for the LUNs to recognize them as devices that support multipathing. Begin by reconfiguring the HBA and its attached disks.

Important: If other disks are attached to any HBA devices, the following commands remove the configuration for those disks and the HBA. If you are attempting to save the current configuration, skip these steps.

1. To determine the device names of the HBAs to which the storage system is connected, enter the following command:

```
lsdev -t efscsi
```
2. For each HBA device name, enter the following command to remove the HBA and the disk configuration that is associated with it:

```
rmdev -l <device name> -R
```
3. To determine whether any disks are already defined that must be removed before rescanning, enter the following command:

```
lsdev -C -c disk
```
4. If any LUNs are already defined as Other FC SCSI Disk Drive, remove the old definitions. For each disk name, enter the following command:

```
rmdev -l <disk name> -R
```

Setting the fast fail recovery flag for the host bus adapter

You can set the fast fail recovery flag for the HBA to improve the failover response.

For the multipath I/O (MPIO) driver to fail over to an available path in a timely manner after a path failure, set the **fast_fail** recovery flag for the HBA devices to which the storage system is connected.

At a command prompt, enter this command:

```
chdev -a fc_err_recov=fast_fail -l <device name>
```

The **<device name>** is the device name of the HBA that is connected to the system.

Rescanning for the storage system logical unit numbers

After the host system is configured to recognize that the storage device supports multipathing, you must rescan for the LUNs.

At a command prompt, enter this command:

```
cfgmgr -vl <device name>
```

The **<device name>** is the device name of the HBA connected to the system.

Checking the configuration

After you change the configuration to support multipathing, confirm that the configuration is working correctly.

To check the new configuration, complete the following steps:

1. To ensure that the configuration is successful, enter the following command to list all disks available to the system:

```
lsdev -C -c disk
```

All LUNs must use MPIO. They must show as an MPIO IBM FlashSystem Disk.

The following command gives you detailed information about a LUN:

```
lscfg -vl <LUN>
```

Example 5-1 shows you the output of those two commands. This AIX system has four disks attached:

- FlashSystem 820 LUN using MPIO
- A LUN without multipathing
- Another LUN without multipathing
- FlashSystem 840 LUN using MPIO

Both IBM FlashSystem units are shown as MPIO IBM FlashSystem Disks. But, you see the different models when you look at the Machine Type and Model attribute of the **lscfg** command output, for example, the fourth LUN is a FlashSystem 840.

Some output lines are left out for clarity (Example 5-1).

Example 5-1 Check the AIX MPIO configuration

```
# lsdev -C -c disk
hdisk0 Available          Virtual SCSI Disk Drive
hdisk1 Available 00-00-02 MPIO IBM FlashSystem Disk
hdisk2 Available 00-01-02 Other FC SCSI Disk Drive
hdisk3 Available 01-01-02 Other FC SCSI Disk Drive
hdisk4 Available 00-01-02 MPIO IBM FlashSystem Disk

# lscfg -vl hdisk1
hdisk1      U78C0.001.DBJ2497-P2-C1-T1-W20040020C2117377-L0  MPIO IBM FlashSystem Disk

      Manufacturer.....IBM
      Machine Type and Model.....FlashSystem
      ...

# lscfg -vl hdisk4
hdisk4      U78C0.001.DBJ2497-P2-C1-T2-W500507605EFE0AD1-L0  MPIO IBM FlashSystem Disk

      Manufacturer.....IBM
      Machine Type and Model.....FlashSystem-9840
      ...
```

2. If there are missing disks, or extra disks, or the LUNs do not show as an *MPIO IBM FlashSystem Disk*, check that the connections and the storage system configuration are correct. You must then remove the configuration for the HBAs and complete the rescan again. For more information, see “Resetting the host bus adapter and disk configuration” on page 116.

3. To ensure that all the connected paths are visible, enter the following command:

```
lspath
```

Example 5-2 shows the paths for the IBM FlashSystem 840 used in Example 5-1 on page 117.

Example 5-2 AIX lspath output

```
# lspath -l hdisk4
Enabled hdisk4 fscsi1
Enabled hdisk4 fscsi1
Enabled hdisk4 fscsi3
Enabled hdisk4 fscsi3
```

4. If paths are missing, check that the connections and the storage system configuration are correct. You must then remove the configuration for the HBAs and perform the rescan again. For more information, see “Resetting the host bus adapter and disk configuration” on page 116.

Configuring path settings

All paths on the IBM FlashSystem 840 are equal. All ports have access to the LUNs, and there is no prioritized port. Therefore, you use them all at the same time. You have to set the distribution of the I/O load at the operating system level. The round-robin distribution is the ideal way to use all of the ports equally.

Set the algorithm attribute to `round_robin` before you add the hdisk to any volume group. All outgoing traffic is then spread evenly across all of the ports, as shown in the following example:

```
chdev -l <LUN> -a algorithm=round_robin
```

The `shortest_queue` algorithm is available in the latest technology levels of AIX for some devices. The algorithm behaves similarly to `round_robin` when the load is light. When the load increases, this algorithm favors the path that has the fewest active I/O operations. Therefore, if one path is slow due to congestion in the SAN, the other less congested paths are used for more of the I/O operations. `Shortest_queue` (if available) or `round_robin` enables the maximum use of the SAN resources. You can use the `load_balance` algorithm to spread the load equally across the paths.

To list the attributes of the LUN, enter the following command:

```
lsattr -El <LUN>
```

Example 5-3 on page 119 shows the output of the `chdev` and the `lsattr` commands of the FlashSystem 840 used in Example 5-1 on page 117. The number of spaces in the output is changed for clarity.

Example 5-3 AIX chdev and lsattr commands

```
# chdev -l hdisk4 -a algorithm=round_robin
hdisk4 changed

# lsattr -El hdisk4
PCM                PCM/friend/fcpothe  Path Control Module      False
PR_key_value       none                Persistent Reserve Key Value True+
algorithm          round_robin        Algorithm                 True+
clr_q              no                 Device CLEARS its Queue on error True
dist_err_pcnt      0                  Distributed Error Percentage True
dist_tw_width      50                 Distributed Error Sample Time True
hcheck_cmd         test_unit_rdy      Health Check Command      True+
hcheck_interval    60                 Health Check Interval     True+
hcheck_mode        nonactive          Health Check Mode         True+
location           Location            Label                     True+
lun_id             0x0                Logical Unit Number ID    False
lun_reset_spt      yes                LUN Reset Supported       True
max_coalesce       0x40000            Maximum Coalesce Size     True
max_retry_delay    60                 Maximum Quiesce Time      True
max_transfer       0x80000            Maximum TRANSFER Size     True
node_name          0x500507605efe0ad0 FC Node Name              False
pvid              none                Physical volume identifier False
q_err              yes                Use QERR bit              True
q_type             simple             Queuing TYPE              True
queue_depth        64                 Queue DEPTH               True
reassign_to        120                REASSIGN time out value   True
reserve_policy     no_reserve          Reserve Policy             True+
rw_timeout         30                 READ/WRITE time out value True
scsi_id            0x10100            SCSI ID                   False
start_timeout      60                 START unit time out value True
timeout_policy     fail_path           Timeout Policy             True+
unique_id          54361IBM          FlashSystem-9840041263a20412-0000-0004-00006410FlashSystem-984003IBMfcp
                    Unique device identifier  False
ww_name            0x500507605e800e41 FC World Wide Name        False
```

5.3.4 FlashSystem 840 and Linux client hosts

The FlashSystem 840 can be attached to Linux client hosts by using the following methods:

- ▶ FC
- ▶ InfiniBand
- ▶ FCoE
- ▶ iSCSI

The FlashSystem 840 benefits the most from operating systems where multipathing and logical volumes are supported. Most Linux distributions have the same optimum configurations. Specific Linux configuration settings are shown.

Network topology guidelines

You can use an arbitrated loop or point-to-point topology on FC configurations for Linux hosts.

Note: The FlashSystem 840 16 Gbps FC ports do not support direct connection to client hosts. A SAN switch must be placed between the IBM FlashSystem 840 and any 16 Gbps-attached client host.

Aligning a partition using Linux

Use this procedure to improve performance by aligning a partition in the Linux operating system.

The Linux operating system defaults to a 63-sector offset. To align a partition in Linux using **fdisk**, complete the following steps:

1. At the command prompt (#), enter the **fdisk /dev/mapper/<device>** command
2. To change the listing of the partition size to sectors, enter **u**.
3. To create a partition, enter **n**.
4. To create a primary partition, enter **p**.
5. To specify the partition number, enter **1**.
6. To set the base sector value, enter **128**.
7. Press Enter to use the default last sector value.
8. To write the changes to the partition table, enter **w**.

Note: The **<device>** is the FlashSystem 840 volume. Example 5-23 on page 156 shows how to create device names for a FlashSystem 840 volume.

The newly created partition now has an offset of 64 KB and works optimally with an aligned application.

If you are installing the Linux operating system on the storage system, create the partition scheme before the installation process. For most Linux distributions, this process requires starting at the text-based installer and switching consoles (press Alt+F2) to get the command prompt before you continue.

Multipathing information for Linux

You can use MPIO to improve the performance of the Linux operating system. Linux kernels of 2.6, and later, support multipathing through device-mapper-multipath. This package can coexist with other multipathing solutions if the other storage devices are excluded from device-mapper.

For a template for the `multipath.conf` file, see Example 5-22 on page 155.

Because the storage system controllers provide true active/active I/O, the `rr_min_io` field in the `multipath.conf` file is set to 4. This results in the best distribution of I/O activity across all available paths. You can set it to 1 for a round-robin distribution or if the I/O activity is more sequential in nature, you can increase the `rr_min_io` field by factors of 2 for a performance gain by using buffered I/O (non-DIRECT).

Integrating InfiniBand controllers

To integrate with InfiniBand technology, the storage system provides block storage by using the SCSI Remote Direct Memory Access (RDMA) Protocol (SRP).

The Linux operating system requires several software modules to connect to the storage system through InfiniBand technology and SRP. In particular, make sure that you install the *srp* and *srptools* modules, and install drivers for the server's host channel adapter (HCA). Use the OpenFabrics Enterprise Distribution (OFED) package from the following website to install these modules, either individually or by using the Install All option:

<http://www.openfabrics.org>

Figure 5-3 on page 121 shows the setting of the InfiniBand `/etc/infiniband/openib.conf` configuration file for SRP.


```
# Load SRP module
SRP_LOAD=yes #
Enable SRP High Availability daemon
SRPHA_ENABLE=yes
SRP_DAEMON_ENABLE=yes
```

Figure 5-3 InfiniBand configuration file

These settings cause the SRP and the SRP daemons to load automatically when the InfiniBand driver starts. The SRP daemon automatically discovers and connects to InfiniBand SRP disks.

Use the `SRPHA_ENABLE=yes` setting. This setting triggers the multipath daemon to create a multipath disk target when a new disk is detected.

InfiniBand technology also requires a Subnet Manager (SM). An existing InfiniBand network already has an SM. In many cases, an InfiniBand switch acts as the SM. If an SM is needed, install OpenSM, which is included with the OFED package, and start it on a single server in the network by entering the following command:

```
# /etc/init.d/opensmd start
```

This script opens an SM on a single port only. If multiple ports directly connect to the storage system, a custom script is needed to start the SM on all ports.

5.3.5 FlashSystem 840 and Microsoft Windows client hosts

The FlashSystem 840 can be attached to Windows client hosts by using the following methods:

- ▶ FC
- ▶ FCoE

The IBM FlashSystem 840 sees the most benefit from operating systems where multipathing and logical volumes are supported. However, certain applications depend on operating systems that are designed for workstations, and they can still benefit from the storage system performance.

Network topologies for Windows hosts

Arbitrated loop or point-to-point topology can be used on the FC configuration for Windows hosts.

Note: The FlashSystem 840 16 Gbps FC ports do not support direct connection to client hosts. A SAN switch must be placed between the IBM FlashSystem 840 and any 16 Gbps-attached client host.

Implementing 4 KB alignment for Windows Server 2003

Use this procedure to improve performance by establishing a 4 KB alignment on a Windows operating system.

Before Windows Vista and Windows Server 2008, systems running the Windows operating systems offset the partition by 63 sectors, or 31.5 KB.

To align to the preferred 4 KB sector size, implement the offset by using the **diskpart.exe** utility:

1. Start the Windows **diskpart.exe** utility to view the list of available LUNs. Enter the following command:

```
DISKPART> list disk
```

2. To select the LUN that holds the file system, enter the following command:

```
DISKPART> select disk <disk number>
```

3. To create a partition on the selected LUN, enter the following command:

```
DISKPART> create partition primary align=64
```

4. Use the Microsoft Management Console (MMC) or other method to assign a file system or drive letter (raw access) to the partition.

If you are installing the Windows XP or Windows Server 2003 operating system, create the partition on the LUN before the installation of the operating system. You can create the partition by using a Linux Live CD, or by presenting the LUN to another Windows host and disconnecting the drive after the partitioning is complete.

Windows Server 2003 multipathing

Configuring MPIO on a Windows Server 2003 operating system can improve reliability. Windows Server 2003 has a built-in MPIO driver that is provided by Microsoft. This driver is responsible for aggregating the links of storage systems and reporting the addition or removal of links to the kernel while online. To use this feature, a Device Specific Module (DSM) is provided to identify the storage system to the MPIO driver. All major storage vendors have support for the multipathing function and coexist safely because of the common driver in the Windows operating system.

To obtain a copy of the FlashSystem 840 driver for Windows Server 2003, a SCORE/RPQ must be submitted to IBM requesting approval. To submit a SCORE/RPQ, contact your IBM representative or IBM Business Partner.

If you are using EMC PowerPath data path management software, it must be at version 4.6, or later, before the storage system DSM can be used.

Windows Server 2008 and Windows Server 2012 multipathing

Windows Server operating system versions that begin with Windows Server 2008 no longer require a separate DSM. Instead, the MPIO function must be installed on the server. For more information, see the Microsoft TechNet website:

[http://technet.microsoft.com/en-us/library/ee619752\(WS.10\).aspx](http://technet.microsoft.com/en-us/library/ee619752(WS.10).aspx)

You can enable multipathing by selecting the **Server Manager** → **Features** option. See Figure 5-4 on page 123.

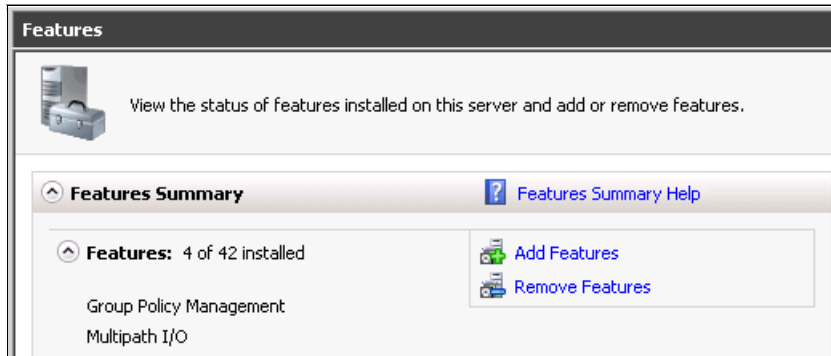


Figure 5-4 Windows 2008 example of activated multipathing

You can set vendor ID (IBM) and product ID (FlashSystem-9840) using **Administrative Tools** → **MPIO**. You enter the eight-character vendor ID and the 16-character product ID by using the MPIO Devices pane. See Figure 5-5.

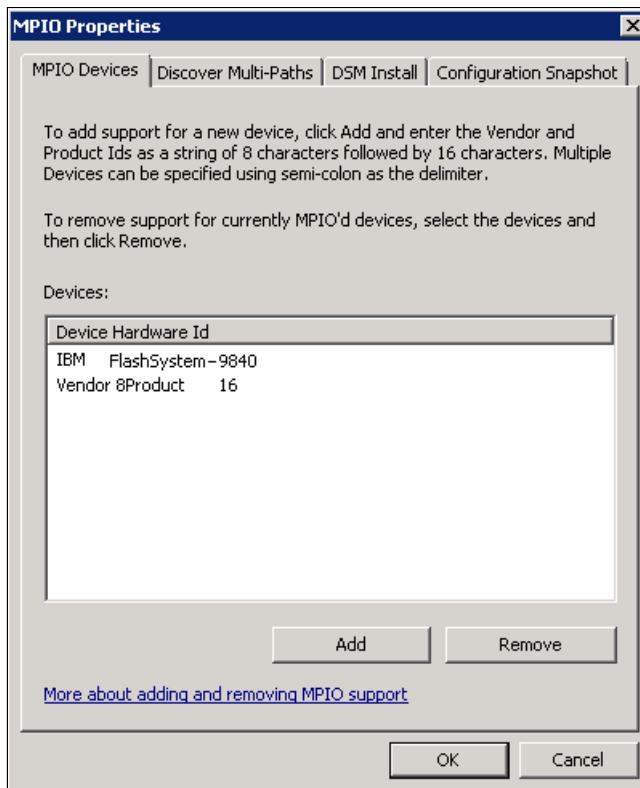


Figure 5-5 Windows MPIO vendor ID and product ID for the FlashSystem 840

Note: The vendor ID must be eight characters in length, including spaces. The product ID must be 16 characters in length, including spaces.

Table 5-1 on page 124 shows the correct vendor ID and product ID for different IBM FlashSystem products.

Table 5-1 IBM FlashSystem SCSI standard inquiry data

IBM FlashSystem	Vendor identification	Product identification
IBM FlashSystem 840	IBM	FlashSystem-9840
IBM FlashSystem 820	IBM	FlashSystem
IBM FlashSystem 720	IBM	FlashSystem
IBM FlashSystem 810	IBM	FlashSystem
IBM FlashSystem 710	IBM	FlashSystem

After you install the MPIO function, set the load balance policy on all storage system LUNs to **Least Queue Depth**. All available paths to the LUNs are then used to aggregate bandwidth. The load balance policy is set through the Properties pane of each multipath disk device in the Windows Device Manager. See Figure 5-6.

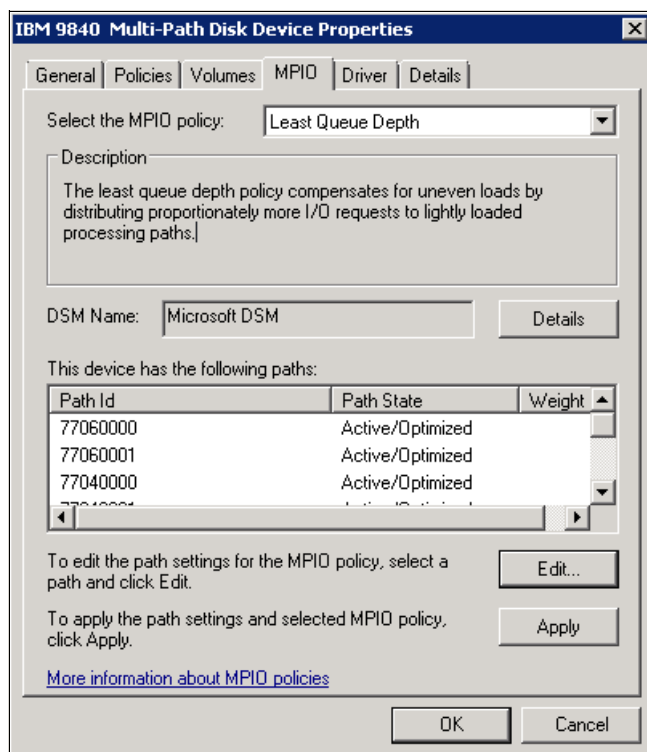


Figure 5-6 Windows MPIO queue configuration

Power option setting for the highest performance

Set the Windows power plan to **High performance** by selecting the **Control Panel** → **Hardware** → **Power Options** menu. See Figure 5-7 on page 125.

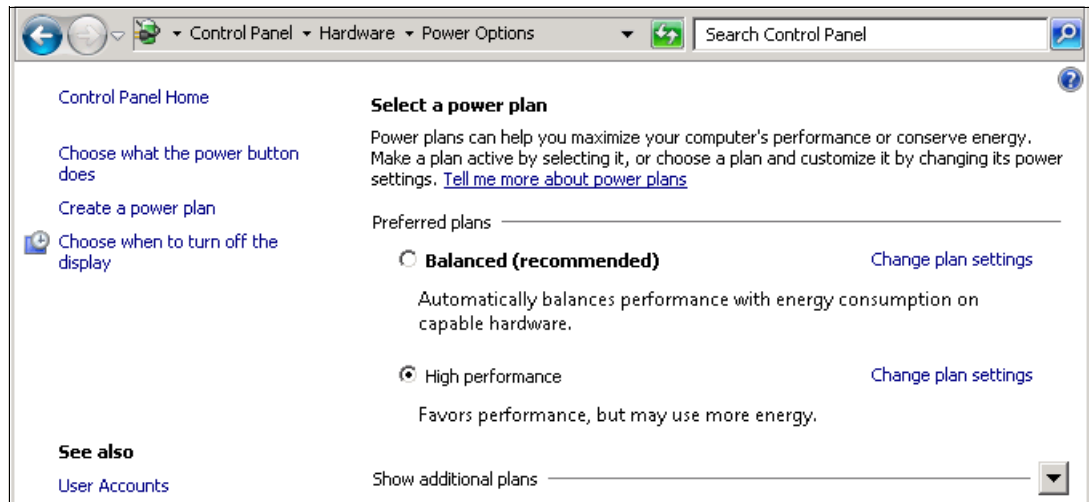


Figure 5-7 Windows Power Options

Optimum disk command timeout settings

Adjust the disk **TimeOutValue** parameter on the Windows operating system for more reliable multipath access.

Windows operating systems have a default disk command **TimeOutValue** of 60 seconds. If a SCSI command does not complete, the application waits 60 seconds before an I/O request is tried again. This behavior can create issues with most applications, so you must adjust the disk **TimeOutValue** in the registry key to a lower value. For more information about setting this value, see the Microsoft TechNet website.

Adjust this key to your needs:

HKLM\System\CurrentControlSet\Services\Disk\TimeOutValue

For a disk in a *nonclustered* configuration, set **TimeOutValue** to 10.

For a disk in a *clustered* configuration, set **TimeOutValue** to 20.

5.3.6 FlashSystem 840 and client VMware ESX hosts

The FlashSystem 840 can be attached to VMware ESX client hosts by using the following methods:

- ▶ FC
- ▶ FCoE

Arbitrated loop topology is required when you attach the IBM FlashSystem 840 directly to the client VMware ESX hosts.

Note: The FlashSystem 840 16 Gbps FC ports do not support direct connection to client hosts. A SAN switch must be placed between the IBM FlashSystem 840 and any 16 Gbps-attached client host. If arbitrated loop connection is required by the host, 8 Gbps FC must be used.

To configure round-robin multipathing in a VMware ESX environment, complete the following steps:

1. In the vSphere client, select the **Configuration** tab.
2. In the Devices view, select each disk on which you want to change the path selection.
3. In the Manage Paths pane, change the Path Selection setting to **Round Robin** (VMware).

You have to set the alignment in the guest operating system. VMware aligns its data stores to 64 KB, but guest virtual machines must still align their own presentation of the storage. Before you continue the installation of a guest Linux operating system or a guest Windows Server 2003 operating system, partition the storage to the aligned accesses.

5.3.7 FlashSystem 840 and IBM SAN Volume Controller or Storwize V7000

For details about IBM SAN Volume Controller or Storwize V7000 product integration, considerations, and configuration with the IBM FlashSystem 840, see Chapter 8, “Product integration” on page 275.

5.3.8 FlashSystem iSCSI host attachment

Support for iSCSI hosts is planned for these operating systems:

- ▶ Red Hat Enterprise Linux 6.5
- ▶ SUSE Linux Enterprise Server 11 Service Pack 3
- ▶ Microsoft Windows 2012 R2

You can find an excellent description about the host site for iSCSI connection in *Implementing the IBM Storwize V7000 V7.4*, SG24-7938.

Note: Always check the IBM System Storage Interoperation Center (SSIC) to get the latest information about supported operating systems, hosts, switches, and so on:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

5.3.9 FlashSystem iSCSI configuration

Figure 5-8 on page 127 shows the GUI I/O port information. In this picture, ports 3 and 11 are online.

Note: The FlashSystem 840 iSCSI port ID numbering starts with ID 3. The two ports with ID 1 and 2 are the management ports of the two canisters as shown in Example 5-4.

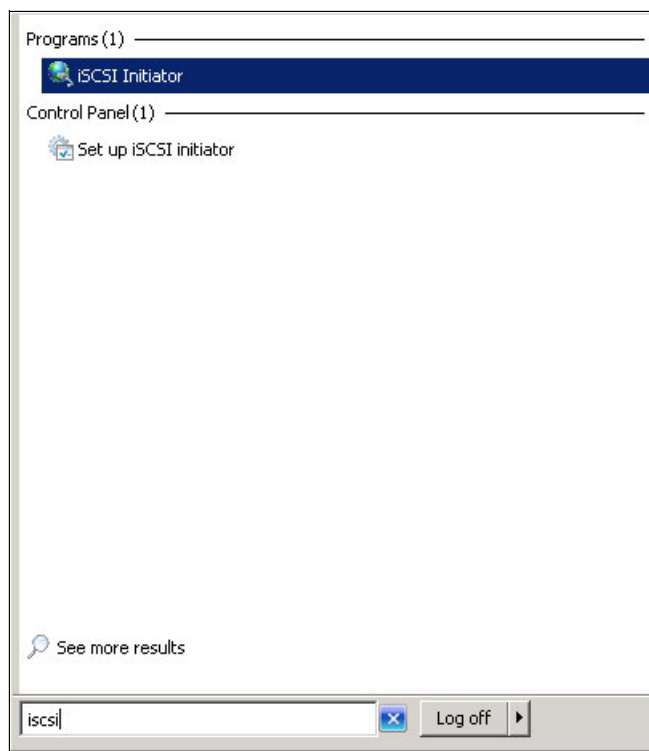


Figure 5-9 Windows iSCSI initiator

You have to confirm the automatic start of the iSCSI service (Figure 5-10).

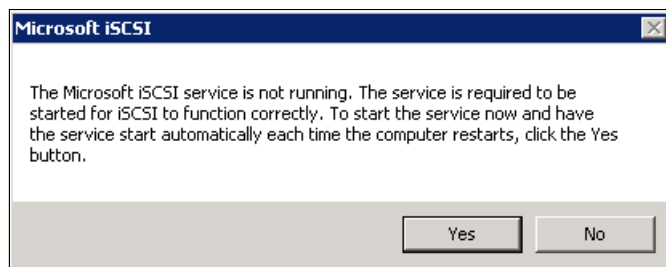


Figure 5-10 Automatic start of the iSCSI service

The iSCSI Configuration window opens. Select the **Configuration** tab (Figure 5-11 on page 129). Write down the initiator name of your Windows host. You use it later on the host configuration part when configuring FlashSystem 840.

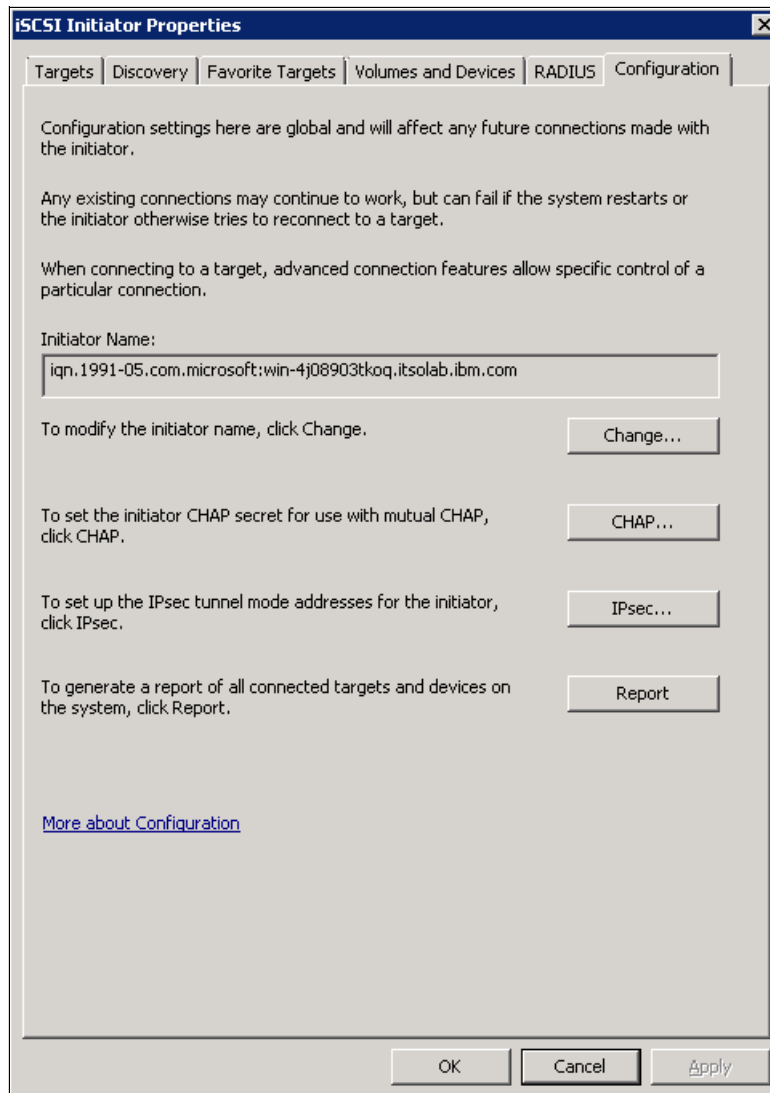


Figure 5-11 iSCSI Initiator Properties window

Open the FlashSystem 840 GUI and open the **Add Host** dialog. Enter a name for the windows host and add the initiator name of the windows host shown in Figure 5-12 on page 130.

Figure 5-12 on page 130 shows the dialog with values.

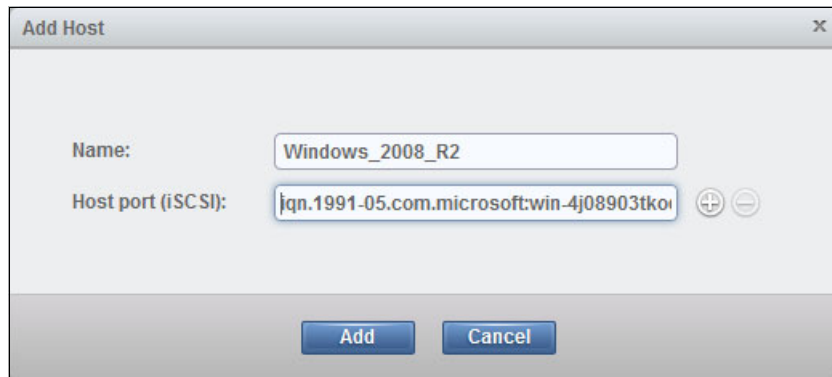


Figure 5-12 iSCSI add host dialog

Create the FlashSystem 840 LUNs for this host and map them to this host as described in 6.3, “Volumes” on page 204.

Go back to the windows system’s iSCSI Initiator Properties window and select the **Targets** tab. Click the **refresh** button and the FlashSystem 840 is listed in the **Discovered Targets** section (Figure 5-13). If FlashSystem 840 is not listed, you can add its IP addresses in the Target field and click **Quick Connect**.

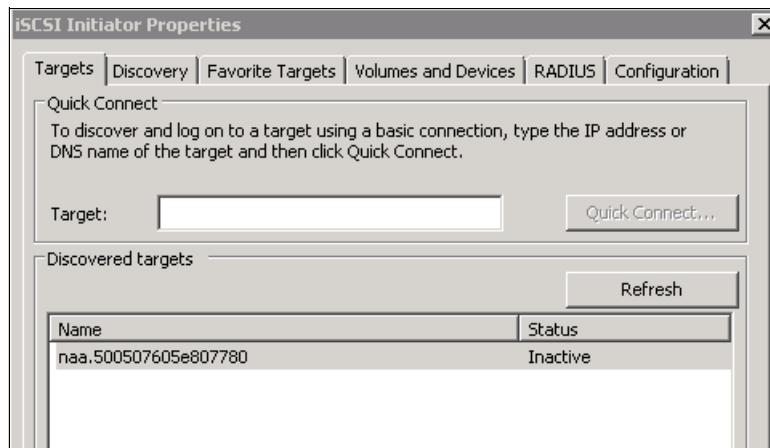


Figure 5-13 iSCSI discovered targets

Click the **connect** button and select **Enable multi-path** if the host is connected with multiple Ethernet ports to FlashSystem 840 (Figure 5-14).

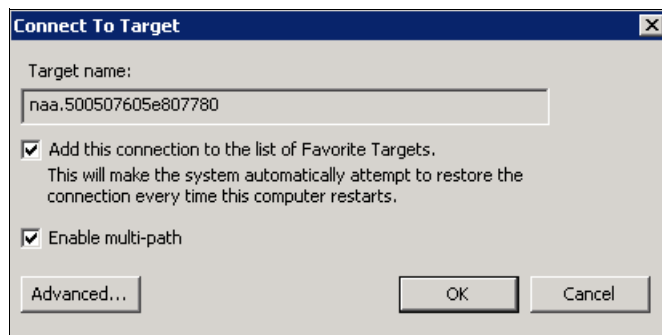


Figure 5-14 Connect to target dialog

5.3.11 Linux iSCSI attachment

You have to install or set up the iSCSI initiator software according to the instructions of your Linux distribution and version.

To configure the FlashSystem 840, you need the iSCSI initiator name. Example 5-5 shows how to find the initiator name on Red Hat Enterprise Linux 6.2.

Example 5-5 Red Hat 6.2 iSCSI initiator name

```
# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1994-05.com.redhat:21774cad37da
```

Open the FlashSystem 840 GUI and open the **Add Host** dialog. Enter a name for the Linux host and add the initiator name of the host as shown in Example 5-5.

Figure 5-15 shows the dialog with values.

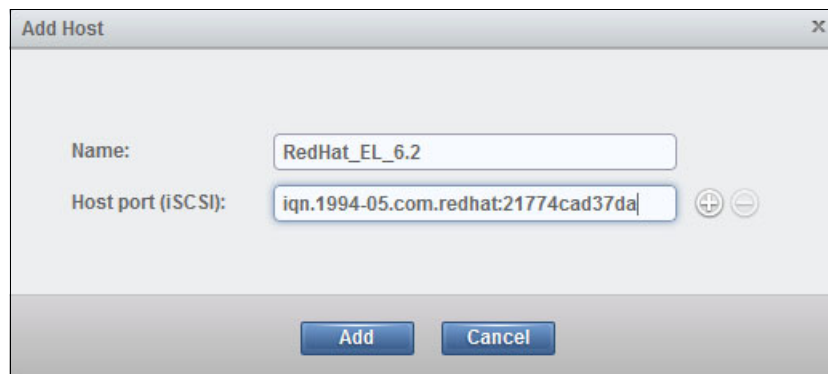


Figure 5-15 iSCSI Add Host dialog

Create the FlashSystem 840 LUNs for this host and map them to this host as described in 6.3, “Volumes” on page 204.

Go back to the Linux host and start the iSCSI disk discovery. Example 5-6 shows the command that is used to discover the LUNs created on the FlashSystem 840 for this host. In this example, two FlashSystem LUNs are attached to the Linux host. The Linux host has one path to each canister. Multipathing is set up.

Example 5-6 Discover iSCSI targets and devices

```
#
# # use iscsiadm --mode discoverydb to detect iSCSI targets
# # detect a connection to each FlashSystem 840 to be able to set up multipathing

# # first FlashSystem canister
# iscsiadm --mode discoverydb --type sendtargets --portal 192.168.61.215 --discover
192.168.61.215:3260,-1 naa.500507605e807780

# # second FlashSystem canister
# iscsiadm --mode discoverydb --type sendtargets --portal 192.168.61.216 --discover
192.168.61.216:3260,-1 naa.500507605e807780

# # login to first target to detect iSCSI disks
# iscsiadm --mode node --targetname naa.500507605e807780 --portal 192.168.61.215 --login
Logging in to [iface: default, target: naa.500507605e807780, portal: 192.168.61.215,3260] (multiple)
Login to [iface: default, target: naa.500507605e807780, portal: 192.168.61.215,3260] successful.
```



```

# # login to second target to detect iSCSI disks
# iscsiadm --mode node --targetname naa.500507605e807780 --portal 192.168.61.216 --login
Logging in to [iface: default, target: naa.500507605e807780, portal: 192.168.61.216,3260] (multiple)
Login to [iface: default, target: naa.500507605e807780, portal: 192.168.61.216,3260] successful.

# # use iscsiadm to get detailed information about the target
# # both targets will only differ in the node addresses
# iscsiadm -m node --targetname=naa.500507605e807780 --portal=192.168.61.215 --op=show
# BEGIN RECORD 6.2.0-873.2.e16
node.name = naa.500507605e807780
... (lines left out)
node.discovery_address = 192.168.61.215
... (lines left out)
node.conn[0].address = 192.168.61.215
... (lines left out)

# # use iscsiadm -m session to get the attached iSCSI disk
# # they are listed at the end of the command output
# # the two flashsystem LUNs are seen by each portal
# iscsiadm -m session -P 3
iSCSI Transport Class version 2.0-870
version 6.2.0-873.2.e16
Target: naa.500507605e807780
    Current Portal: 192.168.61.215:3260,50
    Persistent Portal: 192.168.61.215:3260,50

... (lines left out)

*****
Attached SCSI devices:
*****
Host Number: 16 State: running
scsi16 Channel 00 Id 0 Lun: 0
scsi16 Channel 00 Id 0 Lun: 2
    Attached scsi disk sds                State: running
scsi16 Channel 00 Id 0 Lun: 3
    Attached scsi disk sdt                State: running
Current Portal: 192.168.61.216:3260,18
Persistent Portal: 192.168.61.216:3260,18

... (lines left out)

*****
Attached SCSI devices:
*****
Host Number: 17 State: running
scsi17 Channel 00 Id 0 Lun: 0
scsi17 Channel 00 Id 0 Lun: 2
    Attached scsi disk sdu                State: running
scsi17 Channel 00 Id 0 Lun: 3
    Attached scsi disk sdv                State: running

```

You can set up Linux multipathing by using the configuration file `multipath.conf` and the udev rules as described in 5.5.3, “Linux configuration file `multipath.conf` example” on page 155 and Example 5-26 on page 159 by adding the FlashSystem 840 LUN Volume Unique Identifier to the `blacklist_exceptions` in the `multipath.conf` file. Figure 5-16 on page 133 shows this value in the properties of the FlashSystem 840 LUN using the GUI.

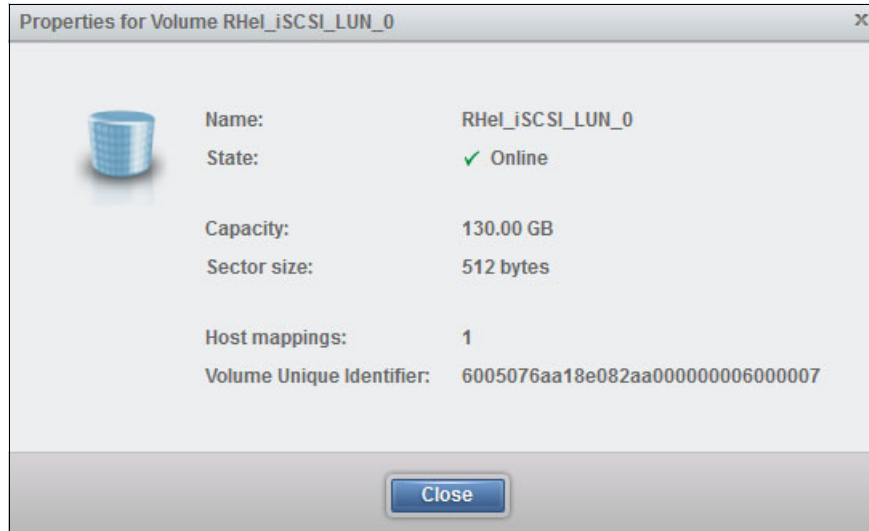


Figure 5-16 FlashSystem 840 LUN Volume Unique Identifier

When using the CLI, you get this information from the `lsvdisk` command. Example 5-7 shows this value in the `multipath.conf` file.

Example 5-7 Blacklist exception for iSCSI

```
blacklist_exceptions {
    wwid                "36005076*"
}
```

This entry allows all LUNs starting with "36005076" to be used by the Linux multipath daemon. This also includes the FlashSystem 840 FC LUNs. You must restart the Linux multipath daemon for this change to take effect, and then check for new devices as shown in Example 5-8.

Example 5-8 Restarting the Linux multipath daemon

```
# # add the Volume Unique Identifier to /multipath.conf
# vi /etc/multipath.conf

# # restart multipathd
# service multipathd restart
ok
Stopping multipathd daemon:      [ OK ]
Starting multipathd daemon:      [ OK ]

# # check for new devices
# multipath -v 2
create: mpathd (36005076aa18e082aa000000006000007) undef IBM,FlashSystem-9840
size=130G features='0' hwhandler='0' wp=undef
~+- policy='queue-length 0' prio=1 status=undef
  |- 16:0:0:2 sds 65:32 undef ready running
  ~- 17:0:0:2 sdu 65:64 undef ready running
create: mpathe (36005076aa18e082aa000000007000008) undef IBM,FlashSystem-9840
size=131G features='0' hwhandler='0' wp=undef
~+- policy='queue-length 0' prio=1 status=undef
  |- 16:0:0:3 sdt 65:48 undef ready running
```



```

^- 17:0:0:3 sdv 65:80 undef ready running

# # list devices
# multipath -ll
mpathe (36005076aa18e082aa000000007000008) dm-5 IBM,FlashSystem-9840
size=131G features='0' hwhandler='0' wp=rw
^-+- policy='queue-length 0' prio=1 status=active
  |- 16:0:0:3 sdt 65:48 active ready running
  ^- 17:0:0:3 sdv 65:80 active ready running
mpathd (36005076aa18e082aa000000006000007) dm-4 IBM,FlashSystem-9840
size=130G features='0' hwhandler='0' wp=rw
^-+- policy='queue-length 0' prio=1 status=active
  |- 16:0:0:2 sds 65:32 active ready running
  ^- 17:0:0:2 sdu 65:64 active ready running

#

```

5.4 Miscellaneous host attachment

This section provides implementation and other general information for connecting client host systems to IBM FlashSystem 840.

Notes: Always check the IBM System Storage Interoperation Center (SSIC) to get the latest information about supported operating systems, hosts, switches, adapters, and so on:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

If the IBM SSIC does not list support, a SCORE/RPQ must be submitted to IBM requesting approval. To submit a SCORE/RPQ, contact your IBM FlashSystem representative or IBM Business Partner.

5.4.1 FlashSystem 840 and Solaris client hosts

The Oracle Solaris operating system has some slight differences between x86 and SPARC support when the disks are partitioned. The Solaris Multiplexed I/O (MPxIO) multipathing setups, however, are nearly identical.

Network topology guidelines

Configure the settings of the FC ports in the system for the network topology that is used with the storage system.

If directly attached to a server, configure the FC ports in the storage system to an arbitrated loop topology. If configured through a switch, set the FC ports to point-to-point to correctly negotiate with the switch.

Note: The FlashSystem 840 16 Gbps FC ports do not support direct connection to client hosts. A SAN switch must be placed between the FlashSystem 840 and any 16 GBps-attached client host.

Sector size information for Solaris

The Solaris operating system supports a 512-byte sector size. The FlashSystem 840 uses a 512-byte sector size.

Aligning the partition for Solaris

Aligning the partition to a 4 KB boundary improves performance of the storage system.

Solaris SPARC aligns slices on LUNs when the LUN is using an Oracle SUN label for a LUN smaller than 2 TB. If the 2 TB capacity is exceeded, the operating system must use an Extensible Firmware Interface (EFI) label. If you use the EFI label, the partitions that are created do not start on a 4 KB boundary. EFI disks default to 34-sector offsets that have the default partition table. To align the partition to a 4 KB boundary for optimal performance, complete the following steps:

1. After formatting the disk, select the partition option.
2. Choose the All Free Hog partitioning base option:

```
partition> 0
Select partitioning base:
0. Current partition table (unnamed)
1. All Free Hog
Choose base (enter number) [0]? 1
```

3. Change the base sector to 40.
4. Press Enter to use the default values for the remaining options:

```
partition> 0
Enter partition id tag[usr]:
Enter partition permission flags[wm]:
Enter new starting Sector[34]: 40
Enter partition size[10066067388b, 10066067427e, 4915071mb, 4799gb, 4tb]:
Part Tag Flag First Sector Size Last Sector
0 usr wm 40 4.69TB 10066067427
```

5. After the first sector is changed, save the configuration and then continue by using the newly created partition.

Multipathing information for Solaris 11 hosts

The method of implementing multipathing depends on the HBA used in the server. MPxIO is the built-in multipathing mechanism for Solaris and no longer requires an HBA that is branded Oracle SUN.

To enable MPxIO support, copy the `/kernel/drv/scsi_vhci.conf` file to the `/etc/driver/drv/scsi_vhci.conf` file and modify the `/etc/driver/drv/scsi_vhci.conf` file.

Example 5-9 shows the FlashSystem entry for Solaris 11.

Example 5-9 Example of `/etc/driver/drv/scsi_vhci.conf` for Solaris 11 that is reduced for clarity

```
name="scsi_vhci" class="root";
load-balance="round-robin";
auto-failback="enable";
ddi-forceload =
    "misc/scsi_vhci/scsi_vhci_f_asym_sun",
    "misc/scsi_vhci/scsi_vhci_f_asym_lsi",
    "misc/scsi_vhci/scsi_vhci_f_asym_emc",
    "misc/scsi_vhci/scsi_vhci_f_sym_emc",
```



```

        "misc/scsi_vhci/scsi_vhci_f_sym_hds",
        "misc/scsi_vhci/scsi_vhci_f_sym",
        "misc/scsi_vhci/scsi_vhci_f_tpgs";

scsi-vhci-failover-override =
    "IBM    FlashSystem-9840", "f_sym";

spread-iptort-reservation = "yes";
iptort-rlentime-snapshot-interval = 30;

```

Note: The entry "IBM FlashSystem-9840" in Example 5-9 contains exactly five spaces.

Preferred read with Solaris

You can speed up an application by accelerating the read I/Os. Implementing preferred read with the FlashSystem 840 gives you an easy way to deploy the FlashSystem 840 in an existing environment. This section describes how to set up preferred read with Solaris.

Use Solaris Volume Manager (SVM) to create mirrored volumes and to set up preferred read to the first disk in the mirrored volume. You use the following command:

```
metaparam -r first <device>
```

This command modifies the read option for a mirror. The option *first* specifies reading only from the first submirror.

FlashSystem must be the first disk in the mirrored volume. If you create a new volume, you have to select the FlashSystem 840 metadvice as the first metadvice for the volume. If you have an existing mirrored volume, you can use these steps:

1. Add FlashSystem metadvice to the existing volume.
2. Sync the data.
3. Destroy the mirrored volume without losing data.
4. Re-create the mirrored volume and use the FlashSystem 840 metadvice as the first entry.

The next example shows the different steps in creating a Solaris file system with mirrored devices and preferred read on the FlashSystem 840. Here are the steps:

1. Check for the attached FlashSystem 840.
2. Create the partition on the attached FlashSystem 840.
3. Create a mirrored file system.
4. Set preferred read on the FlashSystem 840.

Check for attached FlashSystem 840

With the **fcinfo** command, you can get information about the local FC ports and the devices attached to this port. Example 5-10 shows these two use cases.

Example 5-10 Solaris FC port information

```

#
# # list local FC ports
# fcinfo hba-port
HBA Port WWN: 2100001b320f324e
    Port Mode: Initiator
    Port ID: 10800
    OS Device Name: /dev/cfg/c9
    Manufacturer: QLogic Corp.

```



```

Model: QLE2462
Firmware Version: 5.6.4
FCode/BIOS Version: BIOS: 1.29; fcode: 1.27; EFI: 1.09;
Serial Number: RFC0802K03058
Driver Name: qlc
Driver Version: 20120717-4.01
Type: N-port
State: online
Supported Speeds: 1Gb 2Gb 4Gb
Current Speed: 4Gb
Node WWN: 2000001b320f324e
Max NPIV Ports: 127
NPIV port list:

```

```

# # list systems attached to local FC -ports
# fcinfo remote-port -s -p 2100001b320f324e
Remote Port WWN: 500507605efe0ac2
Active FC4 Types: SCSI
SCSI Target: yes
Port Symbolic Name: IBM      FlashSystem-9840 0020
Node WWN: 500507605efe0ad0
LUN: 0
Vendor: IBM
Product: FlashSystem-9840
OS Device Name: /dev/rdisk/c10t500507605EFE0AC2d0s2

```

Create partition on attached FlashSystem 840

The Solaris Server in this example is based on the x86 architecture. Before creating Solaris partitions, the disk must be partitioned for possible operating systems on the x86 server. In Example 5-11, the disk is only used for Solaris.

Some lines are left out for clarity.

Example 5-11 Create Solaris partition

```

#
# # Format disk
# format
Searching for disks...done

AVAILABLE DISK SELECTIONS:
  0. c7d0 ...
    /pci@0,0/pci-ide@1f,2/ide@0/cmdk@0,0
  1. c9t0d1 ...
    /pci@0,0/pci8086,d13a@5/pci1077,138@0/fp@0,0/disk@w20080020c2117377,1
  2. c9t0d2 ...
    /pci@0,0/pci8086,d13a@5/pci1077,138@0/fp@0,0/disk@w20080020c2117377,2
  3. c10t500507605EFE0AC2d0 <IBM-9840-0020 cyl 16716 alt 2 hd 224 sec 56>
    /pci@0,0/pci8086,d13a@5/pci1077,138@0,1/fp@0,0/disk@w500507605efe0ac2,0
Specify disk (enter its number): 3
selecting c10t500507605EFE0AC2d0
[disk formatted]
No Solaris fdisk partition found.

```


FORMAT MENU:

- disk - select a disk
- type - select (define) a disk type
- partition - select (define) a partition table
- current - describe the current disk
- format - format and analyze the disk
- fdisk - run the fdisk program
- repair - repair a defective sector
- label - write label to the disk
- analyze - surface analysis
- defect - defect list management
- backup - search for backup labels
- verify - read and display labels
- save - save new disk/partition definitions
- inquiry - show disk ID
- volname - set 8-character volume name
- !<cmd> - execute <cmd>, then return
- quit

format> fdisk

No fdisk table exists. The default partition for the disk is:

a 100% "SOLARIS System" partition

Type "y" to accept the default partition, otherwise type "n" to edit the partition table.

y

format> partition

PARTITION MENU:

- 0 - change `0' partition
- 1 - change `1' partition
- 2 - change `2' partition
- 3 - change `3' partition
- 4 - change `4' partition
- 5 - change `5' partition
- 6 - change `6' partition
- 7 - change `7' partition
- select - select a predefined table
- modify - modify a predefined partition table
- name - name the current table
- print - display the current table
- label - write partition map and label to the disk
- !<cmd> - execute <cmd>, then return
- quit

partition> print

Current partition table (default):

Total disk cylinders available: 16715 + 2 (reserved cylinders)

Part	Tag	Flag	Cylinders	Size	Blocks	
0	unassigned	wm	0	0	(0/0/0)	0
1	unassigned	wm	0	0	(0/0/0)	0
2	backup	wu	0 - 16714	99.98GB	(16715/0/0)	209672960
3	unassigned	wm	0	0	(0/0/0)	0

4	unassigned	wm	0	0	(0/0/0)	0
5	unassigned	wm	0	0	(0/0/0)	0
6	unassigned	wm	0	0	(0/0/0)	0
7	unassigned	wm	0	0	(0/0/0)	0
8	boot	wu	0 - 0	6.12MB	(1/0/0)	12544
9	unassigned	wm	0	0	(0/0/0)	0

```
partition> 0
Part      Tag      Flag      Cylinders      Size      Blocks
  0 unassigned      wm          0          0      (0/0/0)          0
```

```
Enter partition id tag[unassigned]:
Enter partition permission flags[w]:
Enter new starting cyl[0]:
Enter partition size[0b, 0c, 0e, 0.00mb, 0.00gb]: 16713e
partition> print
Current partition table (unnamed):
Total disk cylinders available: 16715 + 2 (reserved cylinders)
```

Part	Tag	Flag	Cylinders	Size	Blocks
0	unassigned	wm	1 - 16713	99.97GB	(16713/0/0) 209647872
1	unassigned	wm	0	0	(0/0/0) 0
2	backup	wu	0 - 16714	99.98GB	(16715/0/0) 209672960
3	unassigned	wm	0	0	(0/0/0) 0
4	unassigned	wm	0	0	(0/0/0) 0
5	unassigned	wm	0	0	(0/0/0) 0
6	unassigned	wm	0	0	(0/0/0) 0
7	unassigned	wm	0	0	(0/0/0) 0
8	boot	wu	0 - 0	6.12MB	(1/0/0) 12544
9	unassigned	wm	0	0	(0/0/0) 0

```
... < create a partition on slice 7, size 1 cylinder, to be used for metadb >
```

```
partition> label
Ready to label disk, continue? yes
```

```
... < partition written, quit format >
```

Create a mirrored file system

You can easily set up a mirrored file system using the Solaris Volume Manager (SVM) Soft Partitioning. Example 5-12 shows all the steps to create SVM metadevices and use them to set up a mirrored file system. One mirror is on spinning disk and the other mirror is on the FlashSystem 840.

Some lines are left out for clarity.

Example 5-12 Solaris mirrored file system

```
#
# # create an SVM database with the attached devices
# metadb -f -a c9t0d0s7 c10t500507605EFE0AC2d0s7
# metadb -i
      flags      first blk      block count
a      u      16      8192      /dev/dsk/c9t0d0s7
a      u      16      8192      /dev/dsk/c10t500507605EFE0AC2d0s7
...
```



```

u - replica is up to date
...
a - replica is active, commits are occurring to this replica
...

# # create a one-on-one concatenation on spinning disk
# metainit d_disk 1 1 c9t0d0s0
d_disk: Concat/Stripe is setup

# # create a one-on-one concatenation mn flashsystem
# metainit d_FlashSystem 1 1 c10t500507605EFE0AC2d0s0

# # mirror setup. first on spinning disk
# metainit d_mirror -m d_disk
d_mirror: Mirror is setup

# # attach second disk
# metattach d_mirror d_FlashSystem
d_mirror: submirror d_FlashSystem is attached

# # check
# metastat -p d_mirror
d_mirror -m /dev/md/rdsk/d_disk /dev/md/rdsk/d_FlashSystem 1
d_disk 1 1 /dev/rdsk/c9t0d0s0
d_FlashSystem 1 1 /dev/rdsk/c10t500507605EFE0AC2d0s0

# # create FS
# newfs /dev/md/rdsk/d_mirror
newfs: construct a new file system /dev/md/rdsk/d_mirror: (y/n)? y
Warning: 512 sector(s) in last cylinder unallocated
/dev/md/rdsk/d_mirror: 115279360 sectors in 18763 cylinders of 48 tracks, 128
sectors
        56288.8MB in 1173 cyl groups (16 c/g, 48.00MB/g, 5824 i/g)
super-block backups (for fsck -F ufs -o b=#) at:
    32, 98464, 196896, 295328, 393760, 492192, 590624, 689056, 787488, 885920,
Initializing cylinder groups:
.....
super-block backups for last 10 cylinder groups at:
    114328992, 114427424, 114525856, 114624288, 114722720, 114821152, 114919584,
    115018016, 115116448, 115214880

# # mount
# mount /dev/md/dsk/d_mirror /mnt

# # check
# df -h | grep mirror
/dev/md/dsk/d_mirror      54G      55M      54G      1%      /mnt

```

The file system is now on two mirrored devices.

Set preferred read on FlashSystem 840

Solaris Volume Manager uses the round-robin read algorithm as a default. To read from only one device, you can choose the first option. Then, the reads are performed only from the first

device. In this example, the first device that was used was on spinning disk. To use the FlashSystem 840 as the first device in the mirror, you must destroy the mirror and re-create the mirror with the FlashSystem 840 as the first disk.

Important: Be extremely careful while performing the steps described in Example 5-13 because the data can be at risk. Ensure that you back up any critical data before performing any file system management activities.

Example 5-13 Solaris preferred read setup

```
#
# # destroy mirror, recreate mirror
# # first umount
# umount /dev/md/dsk/d_mirror

# # delete the mirror
# metaclear d_mirror
d_mirror: Mirror is cleared

# # recreate mirror, FlashSystem as first entry
# metainit d_mirror -m d_FlashSystem
d_mirror: Mirror is setup

# # attach spinning disk
# metattach d_mirror d_disk
d_mirror: submirror d_disk is attached

# # check
# metastat
d_mirror: Mirror
    Submirror 0: d_FlashSystem
        State: Okay
    Submirror 1: d_disk
        State: Resyncing
    Resync in progress: 20 % done
    Pass: 1
    Read option: roundrobin (default)
    Write option: parallel (default)
    Size: 115279360 blocks (54 GB)

d_FlashSystem: Submirror of d_mirror
    State: Okay
    Size: 115279360 blocks (54 GB)
    Stripe 0:
        Device          Start Block  Dbase   State Reloc Hot Spare
        c10t500507605EFE0AC2d0s0 0      No      Okay   Yes

d_disk: Submirror of d_mirror
    State: Resyncing
    Size: 115279360 blocks (54 GB)
    Stripe 0:
        Device      Start Block  Dbase   State Reloc Hot Spare
        c9t0d0s0      0          No      Okay   Yes

# # first disk preferred read
```



```
# # metaparam -r first d_mirror

# # check changed read algorithm
# metaparam d_mirror
d_mirror: Mirror current parameters are:
    Pass: 1
    Read option: first (-r)
    Write option: parallel (default)

# # mount
# mount /dev/md/dsk/d_mirror /mnt

# # check
# df -h | grep mirror
/dev/md/dsk/d_mirror    54G    10G        44G    19%    /mnt

# # create a read process and check preferred read with iostat:
# # for example: iostat -xMnz 10
# # only md/d_FlashSystem and md/d_mirror will show activity.
```

Note: You must destroy and re-create a metadevice mirror to change the device that is used as the first disk for the preferred read.

5.4.2 FlashSystem 840 and HP-UX client hosts

This section discusses HP-UX client host with FlashSystem 840 configurations.

Note: Always check the IBM SSIC to ensure that the IBM FlashSystem 840 supports your required client host and version required.

HP-UX operating system configurations

Here is a checklist when you use HP-UX client hosts:

- ▶ If using Veritas File System (VxFS), select a block size of 4 Kb or higher.
- ▶ If using HP physical volume links (PVLlinks), add each disk and path into the same Volume Group under Logical Volume Manager (LVM).
- ▶ For direct-attaching FlashSystem, FC ports on the FlashSystem must be set to arbitrated loop (AL) topology (not for 16 Gb attachment).
- ▶ For fabric-attached, use Point-to-Point topology or Auto.
- ▶ LUNs greater than 2 TB have not been tested.
- ▶ For LUNs that require sequential detection, you can use the **mkvdiskhostmap** command to explicitly assign LUNs:
LUNs on a bus (FC port) start at LUN 0, LUN 1, and so on.
- ▶ LUNs are historically limited to 0 - 7 on a bus:
When HP-UX is on traditional SCSI-2 storage architecture, you will not see LUN 8+ unless the SCSI driver in HP-UX is updated to SCSI-3 compliance.
- ▶ No virtual bus or virtual target architecture in the FlashSystem:
More than eight LUNs require LUN masking to overlapping I/O paths.

Detecting LUNs on HP-UX server

Changes in the FlashSystem LUN configuration do not generate a CHECK_CONDITION to the host. You have to rescan the SCSI bus manually:

```
# ioscan -fnC disk
```

Or, use:

```
# ioinit -i (HP-UX 11v1, 11v2)
```

You might need to force CHECK_CONDITION by a link reset. You can use the **chportfc** command and its **-reset** option to perform a link reset on a dedicated port.

Alignment

HP-UX volumes align to 4 KB boundaries. The VxFS file system must be set to a block size of 4096 or higher to keep alignment:

```
# mkfs -F vxfs -o bsize=4096 <disk>
```

5.5 FlashSystem 840 preferred read and configuration examples

Examples of implementing preferred read in different environments and of the Linux `multipath.conf` configuration file are shown in the following sections.

5.5.1 FlashSystem 840 deployment scenario with preferred read

Implementing preferred read with the IBM FlashSystem 840 gives you an easy way to deploy the IBM FlashSystem 840 in an existing environment. The data is secured by writing it to two different storage systems. Data is read at the FlashSystem 840 speed, because it is always read from the FlashSystem 840. This implementation does not change the existing infrastructure concepts, for example, data security, replication, backup, disaster recovery, and so on. Preferred read can be implemented by using the following techniques:

- ▶ IBM SAN Volume Controller/V7000:
 - Virtual Disk Mirroring (also known as *Volume Mirroring*)
- ▶ At the volume manager/operating system level:
 - IBM AIX
 - Linux LVM (native least queue read)
- ▶ At the application level:
 - Oracle Automatic Storage Management (ASM)
 - Standby or Reporting Instance
 - SQL Server: AlwaysOn Availability Groups maximizes the availability of a set of user databases for an enterprise. An availability group supports a failover environment for a discrete set of user databases, known as *availability databases*, that fail over together.

The following examples are schemas that show the logical setup. You have to plan the FC, FCoE, or InfiniBand cabling and SAN setup, depending on your environment and needs.

An example of implementing preferred read with the operating system volume manager is shown in Figure 5-17. This picture shows a schema of IBM AIX LVM mirroring.

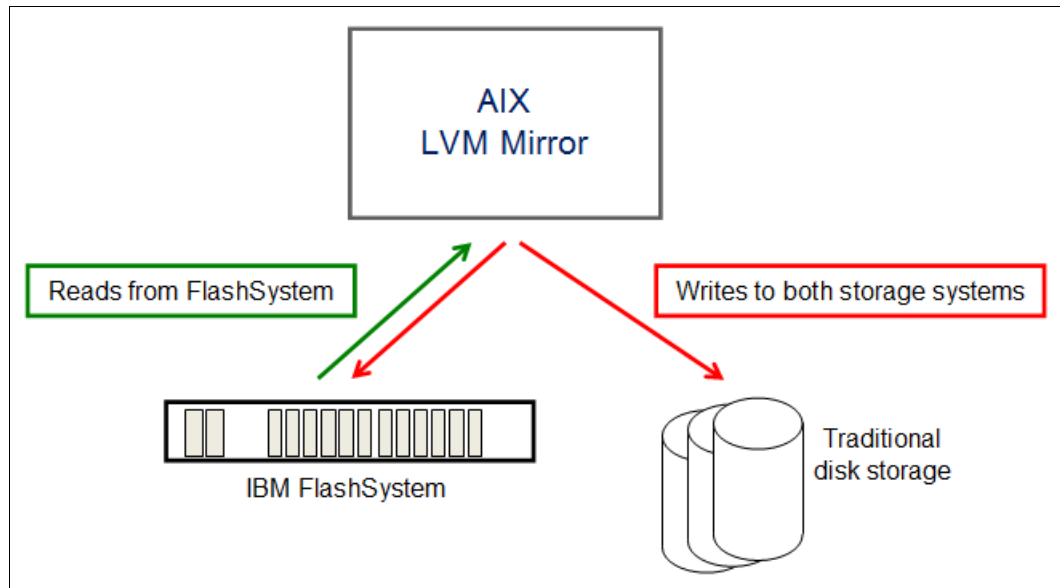


Figure 5-17 Preferred read with AIX

An example of implementing preferred read on the application level is shown in Figure 5-18. This picture shows a schema of Oracle Automatic Storage Management (ASM) mirroring.

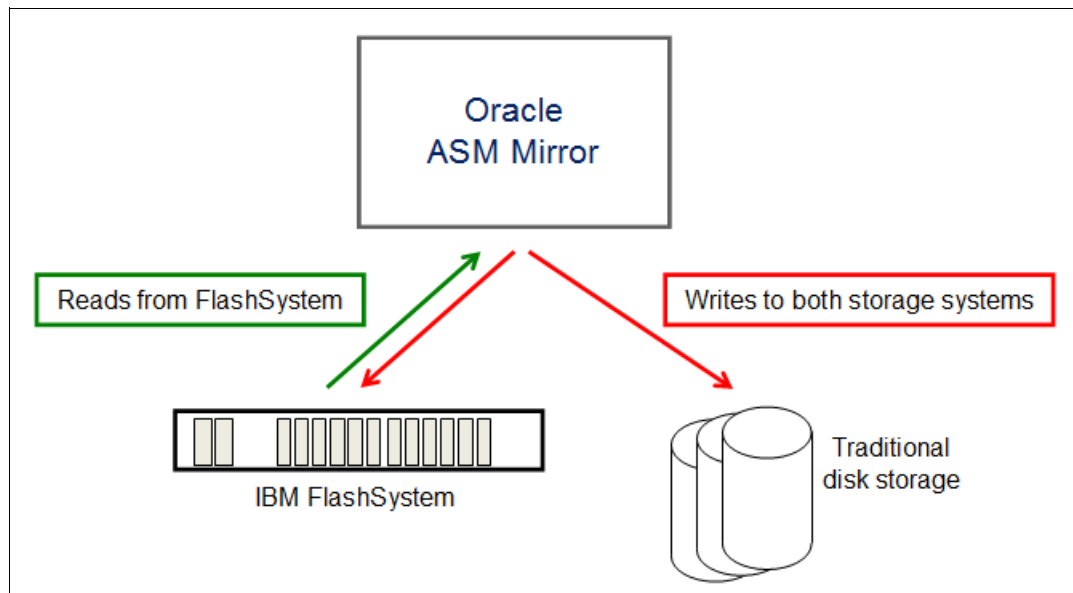


Figure 5-18 Preferred read with Oracle ASM

An example of implementing preferred read on a virtualization layer is shown in Figure 5-19 on page 145. This picture shows a schema of the IBM SAN Volume Controller.

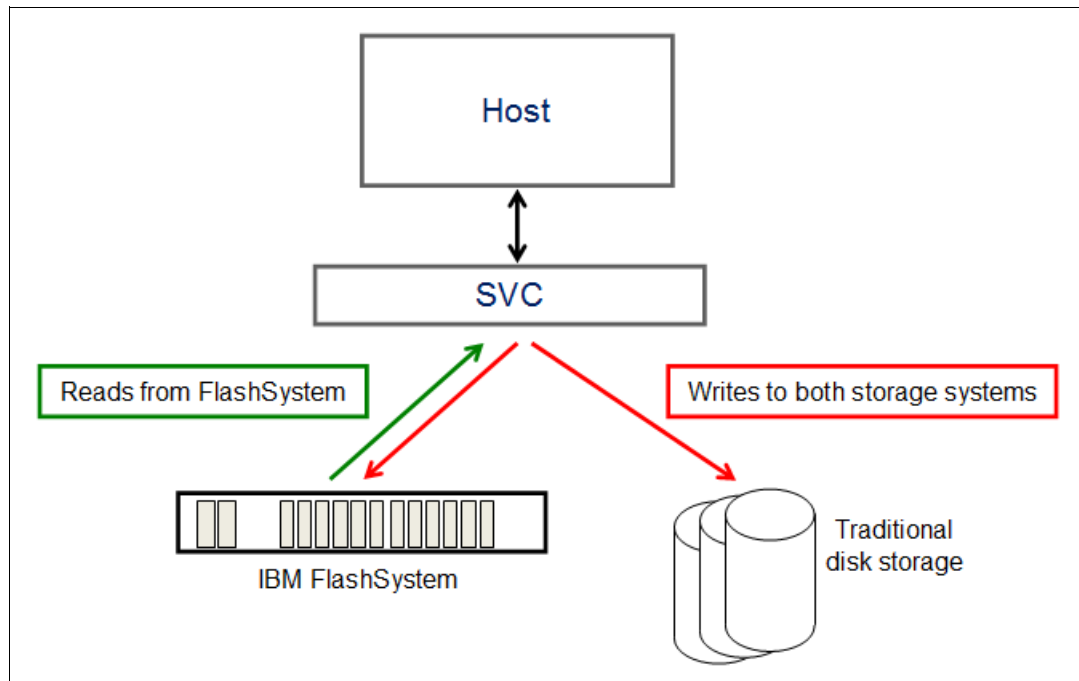


Figure 5-19 Preferred read with the IBM SAN Volume Controller

5.5.2 Implementing preferred read

You can speed up an application by accelerating the read I/Os. Implementing preferred read with the IBM FlashSystem 840 gives you an easy way to deploy the IBM FlashSystem 840 in an existing environment. The data is secured by writing it to two different storage systems. Data is read at the FlashSystem 840 speed, because it is always read from the IBM FlashSystem 840. This implementation does not change the existing infrastructure concepts, for example, data security, replication, backup, disaster recovery, and so on. Preferred read is described in 5.5.1, “FlashSystem 840 deployment scenario with preferred read” on page 143.

Preferred read with AIX

On AIX, preferred read is implemented by the AIX Logical Volume Manager (LVM).

The following steps are illustrated in Example 5-14 on page 146 through Example 5-20 on page 153. The examples walk you through the process of creating a preferred read configuration with the FlashSystem 840. The steps assume that the AIX server is cabled and zoned correctly.

- ▶ Create a file system on spinning disk.
- ▶ Add the IBM FlashSystem 840 as a mirrored copy to this file system.
- ▶ Set the correct read and write policy.
- ▶ Set preferred read to the IBM FlashSystem 840.

In the following steps (Example 5-14 on page 146 through Example 5-20 on page 153), two systems, which are attached through a SAN to the AIX host, are used. Their AIX hdisk information is listed:

- ▶ hdisk1 - hdisk4: IBM MPIO FC 2145 (V7000)
- ▶ hdisk5 - hdisk8: IBM FlashSystem 840 Storage

The following steps are based on AIX 7.1.

Create a file system on spinning disk

Example 5-14 shows the steps to create a file system on AIX. In this example, hdisk1 - hdisk4 are used. All commands are preceded by a comment to the next action. Always check the command parameters against your current AIX version.

Example 5-14 AIX file system creation

```
#
# # Create a file system on normal disks

# # list physical disks
# lsdev -C -c disk
hdisk0 Available Virtual SCSI Disk Drive

# # attach Disksystem to AIX server and check for new disks
# cfgmgr

# # list physical disks
# lsdev -C -c disk
hdisk0 Available Virtual SCSI Disk Drive
hdisk1 Available 00-00-02 MPI0 FC 2145
hdisk2 Available 00-00-02 MPI0 FC 2145
hdisk3 Available 00-00-02 MPI0 FC 2145
hdisk4 Available 00-00-02 MPI0 FC 2145

# # set path policy to your needs: round_robin, load_balance, or shortest_queue
# # check path for all disks, hdisk1 as an example
# lsattr -El hdisk1 | grep algorithm
algorithm load_balance

# # use chdev if needed
# chdev -l hdisk1 -a algorithm=round_robin
# chdev -l hdisk1 -a algorithm=shortest_queue
# chdev -l hdisk1 -a algorithm=load_balance

# # create a volume group with the four IBM 2145 FC Disks
# mkvg -B -y test_vg_1 -t 8 hdisk1 hdisk2 hdisk3 hdisk4
0516-1254 mkvg: Changing the PVID in the ODM.
0516-1254 mkvg: Changing the PVID in the ODM.
0516-1254 mkvg: Changing the PVID in the ODM.
0516-1254 mkvg: Changing the PVID in the ODM.
test_vg_1

# # list the information
# lsvg
rootvg
test_vg_1
# lsvg test_vg_1
VOLUME GROUP: test_vg_1 VOLUME IDENTIFIER:
00f6600100004c00000001464e967f3d
VG STATE: active PP SIZE: 64 megabyte(s)
VG PERMISSION: read/write TOTAL PPs: 3196 (204544 megabytes)
MAX LVs: 512 FREE PPs: 3196 (204544 megabytes)
LVs: 0 USED PPs: 0 (0 megabytes)
OPEN LVs: 0 QUORUM: 3 (Enabled)
TOTAL PVs: 4 VG DESCRIPTORS: 4
STALE PVs: 0 STALE PPs: 0
ACTIVE PVs: 4 AUTO ON: yes
MAX PPs per VG: 130048
MAX PPs per PV: 8128 MAX PVs: 16
```



```

LTG size (Dynamic): 256 kilobyte(s)      AUTO SYNC:      no
HOT SPARE:      no                      BB POLICY:      relocatable
PV RESTRICTION: none                    INFINITE RETRY: no
DISK BLOCK SIZE: 512

# # create a logical volume with file system type jfs2
# # name will be test_lv_1
# mklv -y test_lv_1 -t'jfs2' test_vg_1 3096 hdisk1 hdisk2 hdisk3 hdisk4
test_lv_1

# # create a file system on the logical volume
# # mount point /test/preferred_read will be created at the same time
# crfs -v jfs2 -d test_lv_1 -m /test/preferred_read
File system created successfully.
202893060 kilobytes total disk space.
New File System size is 405798912

# # mount new created file system
# mount /test/preferred_read

# # check
# df -g /test/preferred_read

```

Filesystem	GB blocks	Free	%Used	Iused	%Iused	Mounted on
/dev/test_lv_1	193.50	193.47	1%	4	1%	/test/preferred_read

Add the FlashSystem 840 as a mirrored copy to this file system

Example 5-15 shows the steps to extend an existing file system and to use this extension to create a mirror. In this example, the second disk, hdisk4, is used.

All commands are preceded by a comment to the next action.

Example 5-15 Create a mirrored file system on AIX

```

#
# # Add FlashSystem 840 as a mirrored copy to this file system

# # attach FlashSystem 840 to AIX server and check for new disks
# cfgmgr

# # check for new FlashSystem 840 disk, will be hdisk5 hdisk6 hdisk7 hdisk8
# lsdev -C -c disk
hdisk0 Available          Virtual SCSI Disk Drive
hdisk1 Available 00-00-02 MPI0 FC 2145
hdisk2 Available 00-00-02 MPI0 FC 2145
hdisk3 Available 00-00-02 MPI0 FC 2145
hdisk4 Available 00-00-02 MPI0 FC 2145
hdisk5 Available 00-00-02 MPI0 IBM FlashSystem Disk
hdisk6 Available 00-00-02 MPI0 IBM FlashSystem Disk
hdisk7 Available 00-00-02 MPI0 IBM FlashSystem Disk
hdisk8 Available 00-00-02 MPI0 IBM FlashSystem Disk

# # set path policy to your needs:round_robin or shortest_queue
# # check path for all disks, hdisk5 as an example
# lsattr -El hdisk5 | grep algorithm
algorithm  shortest_queue

# # use chdev if needed
# chdev -l hdisk5 -a algorithm=round_robin
# chdev -l hdisk5 -a algorithm=shortest_queue

```



```

# # list used Physical volume names
# lslv -m test_lv_1 | awk '{print $3, "\t", $5, "\t", $7}' | uniq

PV1      PV2      PV3
hdisk1
hdisk2
hdisk3
hdisk4

# # add FlashSystem 840 disk to volume group
# extendvg test_vg_1 hdisk5 hdisk6 hdisk7 hdisk8
0516-1254 extendvg: Changing the PVID in the ODM.
0516-1254 extendvg: Changing the PVID in the ODM.
0516-1254 extendvg: Changing the PVID in the ODM.
0516-1254 extendvg: Changing the PVID in the ODM.

# # create a mirror
# mklvcopy test_lv_1 2 hdisk5 hdisk6 hdisk7 hdisk8

# # list used Physical volume names and check mirror
# lslv -m test_lv_1 | awk '{print $3, "\t", $5, "\t", $7}' | uniq

PV1      PV2      PV3
hdisk1   hdisk5
hdisk2   hdisk6
hdisk3   hdisk7
hdisk4   hdisk8

# # check mirror state
# lsvg -l test_vg_1
test_vg_1:
LV NAME          TYPE      LPs      PPs      PVs  LV STATE    MOUNT POINT
test_lv_1        jfs2      3096     6192     8    open/stale  /test/preferred_read
loglv00          jfs2log   1        1        1    open/syncd  N/A

# # the mirror is stale, synchronize it
# # this command will take some time depending on volume size
# syncvg -P 32 -v test_vg_1

# # check mirror state
# lsvg -l test_vg_1
test_vg_1:
LV NAME          TYPE      LPs      PPs      PVs  LV STATE    MOUNT POINT
test_lv_1        jfs2      3096     6192     8    open/syncd  /test/preferred_read
loglv00          jfs2log   1        1        1    open/syncd  N/A

# # turn VG quorum off
# # always check your business needs, if VG quorum should be enabled or disabled
# # do this to ensure the VG will not go offline if a quorum of disks goes missing
# chvg -Q n test_vg_1

# # check VG state
# lsvg test_vg_1
VOLUME GROUP:    test_vg_1          VG IDENTIFIER:  00f6600100004c000000001464e967f3d
VG STATE:        active          PP SIZE:        64 megabyte(s)
.
OPEN LVs:        2              QUORUM:         1 (Disabled)
.

```

Now, the file system data is mirrored onto two different physical locations. The first copy is on spinning disk; the second copy is on the FlashSystem 840.

Set the correct read and write policy

IBM AIX LVM sets the scheduling policy for reads and writes to the storage systems. If you use mirrored logical volumes, the following scheduling policies for writing to disk can be set for a logical volume with multiple copies:

- ▶ **Sequential scheduling policy**
Performs writes to multiple copies or mirrors in order. The multiple physical partitions representing the mirrored copies of a single logical partition are designated primary, secondary, and tertiary. In sequential scheduling, the physical partitions are written to in sequence. The system waits for the write operation for one physical partition to complete before starting the write operation for the next one. When all write operations are complete for all mirrors, the write operation is complete.
- ▶ **Parallel scheduling policy**
Simultaneously starts the write operation for all the physical partitions in a logical partition. When the write operation to the physical partition that takes the longest to complete finishes, the write operation is complete. Specifying mirrored logical volumes with a parallel scheduling policy might improve I/O read-operation performance, because multiple copies allow the system to direct the read operation to the least busy disk for this logical volume.
- ▶ **Parallel write with sequential read scheduling policy**
Simultaneously starts the write operation for all the physical partitions in a logical partition. The primary copy of the read is always read first. If that read operation is unsuccessful, the next copy is read. During the read retry operation on the next copy, the failed primary copy is corrected by the LVM with a hardware relocation. This patches the bad block for future access.
- ▶ **Parallel write with round-robin read scheduling policy**
Simultaneously starts the write operation for all the physical partitions in a logical partition. Reads are switched back and forth between the mirrored copies.

You have to set the policy to parallel write with sequential read to get the preferred read performance of the IBM FlashSystem 840. With this option, you get these functions:

- ▶ Write operations are done in parallel to all copies of the mirror.
- ▶ Read operations are always performed on the primary copy of the devices in the mirror set.

Example 5-16 shows how to change the LVM scheduler to the parallel write with sequential read scheduling policy.

Important: Downtime of the file system is required when you change the LVM scheduler policy.

Example 5-16 Changing the scheduler to parallel write with the round-robin read scheduling policy

```
# # check current state of the LVM scheduler
# lslv test_lv_1
LOGICAL VOLUME:      test_lv_1          VOLUME GROUP:  test_vg_1
LV IDENTIFIER:       00f6600100004c00000001464e967f3d.1 PERMISSION:    read/write
VG STATE:            active/complete    LV STATE:       opened/syncd
TYPE:                jfs2                WRITE VERIFY:   off
MAX LPs:             3096                PP SIZE:       64 megabyte(s)
```



```

COPIES:                2                SCHED POLICY:    parallel
LPs:                   3096              PPs:           6192
STALE PPs:             0                BB POLICY:      relocatable
INTER-POLICY:          minimum           RELOCATABLE:    yes
INTRA-POLICY:          middle            UPPER BOUND:    16
MOUNT POINT:           /test/preferred_read LABEL:          /test/preferred_read
DEVICE UID:            0                DEVICE GID:      0
DEVICE PERMISSIONS: 432
MIRROR WRITE CONSISTENCY: on/ACTIVE
EACH LP COPY ON A SEPARATE PV ?: yes
Serialize IO ?:        NO
INFINITE RETRY:        no
# lslv test_lv_1 | grep SCHED POLICY
COPIES:                2                SCHED POLICY:    parallel

# # Logical volume must be closed.
# # If the logical volume contains a file system,
# # the umount command will close the LV device.
# umount /test/preferred_read

# # set scheduler to parallel write with sequential read-scheduling policy
# # (parallel/sequential)
# # Note: mklv and chlvs: The -d option cannot be used with striped logical volumes.
# chlvs -d ps test_lv_1
# # check changed state of the LVM scheduler
# lslv test_lv_1
LOGICAL VOLUME:         test_lv_1        VOLUME GROUP:    test_vg_1
LV IDENTIFIER:          00f6600100004c00000001464e967f3d.1 PERMISSION:       read/write
VG STATE:               active/complete  LV STATE:        closed/syncd
TYPE:                   jfs2              WRITE VERIFY:     off
MAX LPs:                3096              PP SIZE:         64 megabyte(s)
COPIES:                 2                SCHED POLICY:    parallel/sequential
LPs:                   3096              PPs:           6192
STALE PPs:             0                BB POLICY:      relocatable
INTER-POLICY:          minimum           RELOCATABLE:    yes
INTRA-POLICY:          middle            UPPER BOUND:    16
MOUNT POINT:           /test/preferred_read LABEL:          /test/preferred_read
DEVICE UID:            0                DEVICE GID:      0
DEVICE PERMISSIONS: 432
MIRROR WRITE CONSISTENCY: on/ACTIVE
EACH LP COPY ON A SEPARATE PV ?: yes
Serialize IO ?:        NO
INFINITE RETRY:        no
# lslv test_lv_1 | grep SCHED POLICY
COPIES:                2                SCHED POLICY:    parallel/sequential

# # mount file system
# mount /test/preferred_read
# # write some data to the filesystem /test/preferred_read
# # then read this data and check with iostat

```

Disks:	% tm_act	Kbps	tps	Kb_read	Kb_wrtn
hdisk1	0.0	147495.2	1036.4	737476	0
hdisk2	0.0	134964.8	736.6	674824	0
hdisk4	0.0	85035.2	448.6	425176	0
hdisk3	0.0	118422.4	684.2	592112	0

The setup of the logical volume is now preferred read from the first copy.

Set preferred read to the FlashSystem 840

Check which logical volumes are primary, which are secondary, and which are tertiary, if any. This list might be a long list. Example 5-17 lists the first 10 lines of the command output and a command for a quick overview.

Example 5-17 Get first, second, and third logical volume physical disk

```
#
# # Get first, second, and third logical volume's physical disk
# # Get reduced list of all involved hdisks
# # list used Physical volume names
# lslv -m test_lv_1 | awk '{print $3, "\t", $5, "\t", $7}' | uniq

PV1      PV2      PV3
hdisk1   hdisk5
hdisk2   hdisk6
hdisk3   hdisk7
hdisk4   hdisk8
```

Now, the spinning disk devices in the PV1 column are the primary devices. The reads are all supported by the PV1 devices. During boot, the PV1 devices are the primary copy of the mirror, and they are used as the sync point.

You have to remove the PV1 disks, so that the PV2 disks are primary. Then, you add the removed disk and set up a mirror again. Example 5-18 shows how to make the FlashSystem 840 the primary copy.

Example 5-18 Make the FlashSystem 840 the primary disk

```
# # remove the primary copy which is on spinning disk
# rmlvcopy test_lv_1 1 hdisk1 hdisk2 hdisk3 hdisk4

# # list used Physical volume names ; no mirror
# lslv -m test_lv_1 | awk '{print $3, "\t", $5, "\t", $7}' | uniq

PV1      PV2      PV3
hdisk5
hdisk6
hdisk7
hdisk8

# # add removed spinning disk to mirror data
# mklvcopy test_lv_1 2 hdisk1 hdisk2 hdisk3 hdisk4

# # list used Physical volume names ; first copy now on FlashSystem
# lslv -m test_lv_1 | awk '{print $3, "\t", $5, "\t", $7}' | uniq

PV1      PV2      PV3
hdisk5   hdisk1
hdisk6   hdisk2
hdisk7   hdisk3
hdisk8   hdisk4

# # check mirror state
# lsvg -l test_vg_1
test_vg_1:
LV NAME          TYPE      LPs      PPs      PVs  LV STATE      MOUNT POINT
test_lv_1        jfs2      3096     6192     8    open/stale    /test/preferred_read
```



```

loglv00          jfs2log    1      1      1    open/syncd    N/A

# # the mirror is stale, synchronize it
# # this command will take some time depending on volume size
# syncvg -P 32 -v test_vg_1

# # check mirror state
# lsvg -l test_vg_1
test_vg_1:
LV NAME          TYPE      LPs      PPs      PVs  LV STATE    MOUNT POINT
test_lv_1        jfs2      3096     6192     8    open/syncd  /test/preferred_read
loglv00          jfs2log    1        1        1    open/syncd  N/A

# # turn VG quorum off
# # always check your business need if VG quorum should be enabled or disabled
# # do this to ensure the VG will not go offline if a quorum of disks goes missing
# chvg -Q n test_vg_1

# # check VG state
# lsvg test_vg_1
VOLUME GROUP:    test_vg_1          VG IDENTIFIER:  00f6600100004c000000001464e967f3d
VG STATE:        active          PP SIZE:        64 megabyte(s)
.
OPEN LVs:        2              QUORUM:         1 (Disabled)
.

# # again read the data and check with iostat

Disks:          % tm_act      Kbps      tps      Kb_read  Kb_wrtn
hdisk0           0.0          8.8        1.6         44         0
hdisk5           0.0      281169.6    5695.8    1405848         0
hdisk6           0.0      261426.4    5057.2    1307132         0
hdisk7           0.0      228819.2    4434.0    1144096         0
hdisk8           0.0      115506.4    2221.0     577532         0

```

You can use the **iostat** command to see the effects of the parallel/sequential scheduler settings and the performance of the spinning disk or the FlashSystem 840 as the primary disk. Execute this command with the normal disk and again later with the FlashSystem 840 as the primary disk:

```
iostat -DR1TV 3
```

You notice that only the first disk is used for reading and that you get a big performance increase with the FlashSystem 840.

Example 5-19 shows the **iostat** output before changing the primary disk. It shows **hdisk1**, which is a spinning disk. Example 5-20 on page 153 shows the result after changing the primary. You see that **hdisk5**, which is the FlashSystem disk, is used only for reading.

Output is shortened for clarity.

Example 5-19 The iostat command checks for preferred read on the spinning disk

```

#
# # the volume group has to be in syncd state
# # use dd command to read from the mirrored logical volume
# dd if=/dev/test_lv_1 of=/dev/zero bs=16k count=100000

# # execute iostat in another windows

```



```
# iostat -DRITV 3
Disks:
-----
hdisk1
```

Example 5-20 The iostat command checks for preferred read on the IBM FlashSystem 840

```
#
# # the volume group has to be in syncd state
# # use dd command to read from the mirrored logical volume
# dd if=/dev/test_lv_1 of=/dev/zero bs=16k count=100000

# # execute iostat in another windows
# iostat -DRITV 3

Disks:
-----
hdisk5
```

Setting the FC topology on the FlashSystem 840 and AIX

The IBM FlashSystem 840 can be directly attached to an AIX host without a switch. In this case, the FC ports of the FlashSystem 840 are changed to arbitrated loop (AL) topology. You can use the **chportfc** command to change port settings on the IBM FlashSystem 840. On the AIX system, the ports also needed to be changed to AL. Example 5-21 shows changing two ports, fscsi0 and fscsi2, on the AIX system. This system has four FC ports. Ports 0 and 2 are directly attached to the FlashSystem 840 using AL and ports 1 and 3 are attached to a switch.

Example 5-21 Set the AIX port to arbitrated loop

```
# # before using these commands
# # you must first alter the port topology on the FlashSystem 840
# # all traffic has to be stopped before using this command
#
# # remove FC port fscsi0 and then configure it using cfgmgr
# rmdev -Rdl fscsi0
# cfgmgr -vl fcs0

# # remove FC port fscsi2 and then configure it using cfgmgr
# rmdev -Rdl fscsi2
# cfgmgr -vl fcs2

# check all 4 ports
# lsattr -El fscsi0 | grep attach
attach      al

# lsattr -El fscsi1 | grep attach
attach      switch

# lsattr -El fscsi2 | grep attach
attach      al

# lsattr -El fscsi3 | grep attach
attach      switch
```

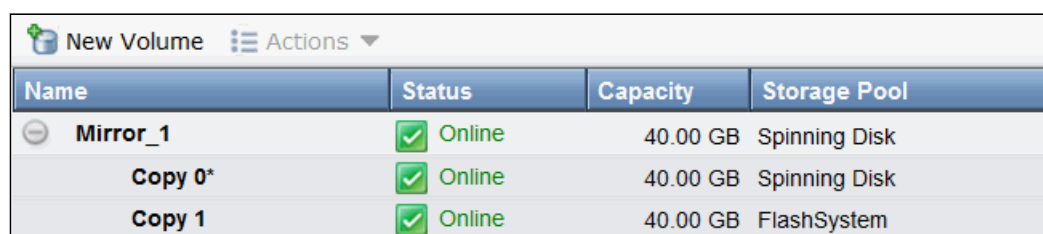
The `cfgmgr` command detects the correct topology.

Note: The topology must be set on an attached system, switch, or storage device before you run these commands.

Preferred read with the IBM SAN Volume Controller

You can set up preferred read on the IBM FlashSystem 840 with the IBM SAN Volume Controller with only one mouse click. In the IBM SAN Volume Controller GUI, go to the **Volumes** menu and right-click the FlashSystem 840 disk of the mirrored volume. Then, select **Make Primary**. You will notice the start asterisk (*) next to your primary disk, which is the preferred read disk now.

Figure 5-20 shows the mirrored IBM SAN Volume Controller volume with preferred read on spinning disk.



The screenshot shows the IBM SAN Volume Controller GUI. At the top, there are buttons for 'New Volume' and 'Actions'. Below is a table with columns: Name, Status, Capacity, and Storage Pool. The table contains three rows: 'Mirror_1' (Status: Online, Capacity: 40.00 GB, Storage Pool: Spinning Disk), 'Copy 0*' (Status: Online, Capacity: 40.00 GB, Storage Pool: Spinning Disk), and 'Copy 1' (Status: Online, Capacity: 40.00 GB, Storage Pool: FlashSystem). The 'Copy 0*' row is highlighted in light blue, indicating it is the preferred read disk.

Name	Status	Capacity	Storage Pool
Mirror_1	Online	40.00 GB	Spinning Disk
Copy 0*	Online	40.00 GB	Spinning Disk
Copy 1	Online	40.00 GB	FlashSystem

Figure 5-20 SAN Volume Controller mirrored VDisk with preferred read on spinning disk

Figure 5-21 shows the option menu to select the primary copy. The primary copy is identical to the preferred read disk.

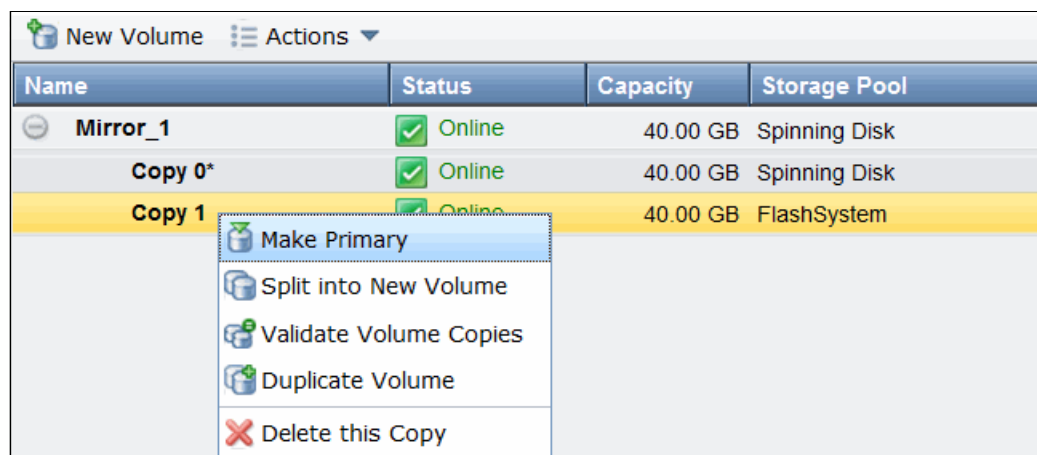


Figure 5-21 SAN Volume Controller option Make Primary

Figure 5-22 on page 155 shows the mirrored IBM SAN Volume Controller volume with preferred read using the FlashSystem 840.

New Volume Actions			
Name	Status	Capacity	Storage Pool
Mirror_1	Online	40.00 GB	FlashSystem
Copy 0	Online	40.00 GB	Spinning Disk
Copy 1*	Online	40.00 GB	FlashSystem

Figure 5-22 SAN Volume Controller mirrored VDisk with preferred read on the FlashSystem 840

Preferred read with Oracle ASM

Oracle Automatic Storage Management (ASM) in Oracle 11g includes advanced features that can use the performance of the IBM FlashSystem 840. The features of Preferred Mirror Read and Fast Mirror Resync are the two most prominent features that fit in this category.

You can set up preferred read using the following two features. You can get detailed information about these two features in the Oracle documentation:

- ▶ Preferred Mirror Read
- ▶ Fast Mirror Resync

You can read a detailed guide about Oracle ASM in the “Administering ASM Disk Groups” topic of the *Database Storage Administrator’s Guide* from the Oracle Help Center:

http://docs.oracle.com/cd/B28359_01/server.111/b31107/asmdiskgrps.htm#OSTMG137

5.5.3 Linux configuration file multipath.conf example

The name and the value of several `multipath.conf` file attributes have changed from version 5 to version 6, or from version 6 to version 6.2. The example is based on Linux 6.2 and includes commented values for other versions. Check your Linux version to set the correct `wwid` values.

Example 5-22 shows an example of a Linux `multipath.conf` file for the IBM FlashSystem 840.

Important: Check for the correct SCSI inquiry string. The actual string is "FlashSystem-9840".

Example 5-22 IBM FlashSystem 840 `multipath.conf` file for Linux 6.2

```
# multipath.conf for Linux 5.x and Linux 6.x
#
# Always check for the correct parameter names, other versions may use a different name.
# Check the correct names and values for your Linux environment.
#

defaults {
    udev_dir          /dev
    polling_interval  30
    checker_timeout   10
}

blacklist {
    wwid              "*"
}
```



```

blacklist_exceptions {
    wwid                "36005076*"
}

devices {
    device {
        vendor          "IBM"
        product          "FlashSystem-9840"
#       path_selector    "round-robin 0"          # Linux 5, Linux 6
        path_selector    "queue-length 0"         # Linux 6.2, if available
        path_grouping_policy multibus
        path_checker      tur
#       rr_min_io_rq      4                        # Linux 6.x
        rr_min_io         4                        # Linux 5.x
        rr_weight          uniform
        no_path_retry      fail
        failback            immediate
        dev_loss_tmo        300
        fast_io_fail_tmo    25
    }
}

multipaths {
# Change these example WWID's to match the FlashSystem LUN.s.
    multipath {
        wwid            360050768018e9fc15000000006000000
        alias            FlashSystem_840_6
    }
    multipath {
        wwid            360050768018e9fc15000000007000000
        alias            FlashSystem_840_7
    }
}

```

Example of Linux commands to configure the FlashSystem 840

Example 5-23 shows the commands and their results after attaching two 103 GB FlashSystem 840 volumes to the Linux 6.2 host. In this example, a 100 GB FlashSystem 820 LUN was already attached to the system, and the two FlashSystem 840 LUNs are new. The multipath.conf file of Example 5-22 on page 155 must be extended by entries for the IBM FlashSystem 840, which are shown in Example 5-25 on page 158.

Example 5-23 Commands to create the FlashSystem 840 devices

```

#
# # list current devices
# multipath -l
mpatha (1ATA      SAMSUNG HE161HJ                S209J90S) dm-4 ATA,SAMSUNG
HE161HJ
size=149G features='0' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=0 status=active
   `- 0:0:0:0 sda 8:0  active undef running
FlashSystem_820_1 (20020c24001117377) dm-1 IBM,FlashSystem
size=100G features='0' hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=0 status=active
   |-- 5:0:0:1 sdc 8:32  active undef running
   |-- 4:0:0:1 sdj 8:144 active undef running

# # check for new devices
]# multipath

```



```

create: FlashSystem_840_7 (360050768018e9fc15000000007000000) undef IBM,FlashSystem-9840
size=103G features='0' hwhandler='0' wp=undef
`-+- policy='queue-length 0' prio=1 status=undef
    |- 5:0:1:1 sde 8:64 undef ready running
    |- 6:0:0:1 sdg 8:96 undef ready running
    |- 7:0:0:1 sdi 8:128 undef ready running
    `-- 4:0:1:1 sdl 8:176 undef ready running
create: FlashSystem_840_6 (360050768018e9fc15000000006000000) undef IBM,FlashSystem-9840
size=103G features='0' hwhandler='0' wp=undef
`-+- policy='queue-length 0' prio=1 status=undef
    |- 6:0:0:0 sdf 8:80 undef ready running
    |- 5:0:1:0 sdd 8:48 undef ready running
    |- 7:0:0:0 sdh 8:112 undef ready running
    `-- 4:0:1:0 sdk 8:160 undef ready running

# # list multipath devices
# multipath -l
FlashSystem_840_6 (360050768018e9fc15000000006000000) dm-2 IBM,FlashSystem-9840
size=103G features='0' hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=0 status=active
    |- 6:0:0:0 sdf 8:80 active undef running
    |- 5:0:1:0 sdd 8:48 active undef running
    |- 7:0:0:0 sdh 8:112 active undef running
    `-- 4:0:1:0 sdk 8:160 active undef running
mpatha (1ATA SAMSUNG HE161HJ S209J90S) dm-4 ATA,SAMSUNG
HE161HJ
size=149G features='0' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=0 status=active
    `-- 0:0:0:0 sda 8:0 active undef running
FlashSystem_820_1 (20020c24001117377) dm-1 IBM,FlashSystem
size=100G features='0' hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=0 status=active
    |- 5:0:0:1 sdc 8:32 active undef running
    `-- 4:0:0:1 sdj 8:144 active undef running
FlashSystem_840_7 (360050768018e9fc15000000007000000) dm-0 IBM,FlashSystem-9840
size=103G features='0' hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=0 status=active
    |- 5:0:1:1 sde 8:64 active undef running
    |- 6:0:0:1 sdg 8:96 active undef running
    |- 7:0:0:1 sdi 8:128 active undef running
    `-- 4:0:1:1 sdl 8:176 active undef running

# # list devices with ls
# ls -l /dev/mapper/
lrwxrwxrwx. 1 root root 7 Oct 16 09:30 FlashSystem_820_1 -> ../dm-1
lrwxrwxrwx. 1 root root 7 Oct 16 09:58 FlashSystem_840_6 -> ../dm-2
lrwxrwxrwx. 1 root root 7 Oct 16 09:58 FlashSystem_840_7 -> ../dm-0

```

Example 5-24 shows the creation of an aligned partition in Linux 6.2.

Example 5-24 Creating a Linux partition

```

#
# fdisk /dev/mapper/FlashSystem_840_6

```

The device presents a logical sector size that is smaller than the physical sector size. Aligning to a physical sector (or optimal I/O) size boundary is recommended, or performance may be impacted.

WARNING: DOS-compatible mode is deprecated. It's strongly recommended to switch off the mode (command 'c') and change display units to sectors (command 'u').

Command (m for help): u
Changing display/entry units to sectors

Command (m for help): n
Command action
 e extended
 p primary partition (1-4)
p
Partition number (1-4): 1
First sector (63-216006655, default 1024): 128
Last sector, +sectors or +size{K,M,G} (128-216006655, default 216006655):
Using default value 216006655

Command (m for help): w
The partition table has been altered!

Calling ioctl() to reread partition table.
Syncing disks.

Using FlashSystem 820 and FlashSystem 840 with Linux client hosts

The FlashSystem 820 and IBM FlashSystem 840 have different product names that need to be configured in the multipath.conf file. Example 5-25 shows the configuration file for the FlashSystem 820 and the FlashSystem 840.

Example 5-25 Multipath.conf file extension for the FlashSystem 820 and the FlashSystem 840

```
# multipath.conf for Linux 5.x and Linux 6.x
#
# Always check for the correct parametername, another version may use a different name.
# Check the correct names and values for your Linux environment
#

defaults {
    udev_dir          /dev
    polling_interval  30
    checker_timeout   10
}

blacklist {
    wwid              "*"
}

blacklist_exceptions {
    wwid              "36005076*"          # FlashSystem 840
    wwid              "20020c24*"          # FlashSystem 710/810/720/820
}

devices {
    # FlashSystem 840
    device {
        vendor          "IBM"
        product          "FlashSystem-9840"
#        path_selector    "round-robin 0"      # Linux 5, Linux 6
#        path_selector    "queue-length 0"      # Linux 6.2, if available
        path_grouping_policy multibus
    }
}
```



```

        path_checker      tur
        rr_min_io_rq      4          # Linux 6.x
#       rr_min_io         4          # Linux 5.x
        rr_weight         uniform
        no_path_retry     fail
        failback          immediate
        dev_loss_tmo      300
        fast_io_fail_tmo  25
    }
    # FlashSystem 710/810/720/820
    device {
        vendor             "IBM"
        product             "FlashSystem-9840"
#       path_selector      "round-robin 0"      # Linux 5, Linux 6
        path_selector      "queue-length 0"      # Linux 6.2, if available
        path_grouping_policy multibus
        path_checker        tur
#       rr_min_io_rq       1          # 6.x, FlashSystem 710/810
#       rr_min_io         1          # 5.x, FlashSystem 710/810
        rr_min_io_rq       4          # 6.x, FlashSystem 720/820
        rr_min_io         4          # 5.x, FlashSystem 720/820
        rr_weight           uniform
        no_path_retry       fail
        failback            immediate
        dev_loss_tmo        300
        fast_io_fail_tmo    25
    }
}

multipaths {

# Change these example WWID's to match the FlashSystem LUN.s.
    multipath {
        wwid              360050768018e9fc15000000006000000
        alias              FlashSystem_840_6
    }
    multipath {
        wwid              360050768018e9fc15000000007000000
        alias              FlashSystem_840_7
    }
}

```

Linux tuning

The Linux kernel buffer file system writes data before it sends the data to the storage system. With the IBM FlashSystem 840, better performance can be achieved when the data is not buffered but directly sent to the IBM FlashSystem 840. When setting the scheduling policy to no operation (NOOP), the fewest CPU instructions possible are used for each I/O. Setting the scheduler to NOOP gives the best write performance on Linux systems. You can use the following setting in most Linux distributions as a boot parameter: `elevator=noop`.

Current Linux devices are managed by the device manager *Udev*. You can define how *Udev* manages devices by adding rules to the `/etc/udev/rules.d` directory.

Example 5-26 shows the rules for the IBM FlashSystem 840 with Linux 6.

Example 5-26 Linux device rules Linux 6.x

```

#
cat 99-IBM-FlashSystem.rules

ACTION=="add|change", SUBSYSTEM=="block",ATTRS{device/model}=="FlashSystem-9840",
ATTR{queue/scheduler}="noop",ATTR{queue/rq_affinity}="1",
ATTR{queue/add_random}="0",ATTR{device/timeout}="5"

```



```
ACTION=="add|change", KERNEL=="dm-*",  
PROGRAM="/bin/bash -c 'cat /sys/block/$name/slaves/*/device/model | grep FlashSystem-9840'",  
ATTR{queue/scheduler}="noop",ATTR{queue/rq_affinity}="1",ATTR{queue/add_random}="0"
```

This udev rules file contains two lines originally. It has been broken up to multiple lines for better readability. If you use this file, make sure that each line starts with the keyword ACTION.

Example 5-27 shows the rule for the IBM FlashSystem 840 with Linux 5.

Example 5-27 Linux device rules Linux 5.x

```
#  
cat 99-IBM-FlashSystem.rules  
  
ACTION=="add|change", SUBSYSTEM=="block",SYSFS{model}=="FlashSystem-9840",  
RUN+="/bin/sh -c 'echo noop > /sys/$DEVPATH/queue/scheduler'"
```

This udev rules file contains one line originally. It has been broken up to multiple lines for better readability. If you use this file, make sure that the line starts with the keyword ACTION.

You can apply the new rules by using the Example 5-28 commands.

Example 5-28 Restarting udev rules

```
# linux 6.2  
/sbin/udevadm control --reload-rules  
/sbin/start_udev
```

5.5.4 Example of a VMWare configuration

You can set the number of I/Os for each path on VMware with this command:

```
esxcli nmp roundrobin setconfig --device <device> --iops=10 --type "iops"
```

This example sets 10 I/Os for each path.

5.6 FlashSystem 840 and Easy Tier

You can implement the IBM FlashSystem 840 with IBM SAN Volume Controller Easy Tier. SAN Volume Controller Easy Tier moves hot, frequently used data to the FlashSystem 840 automatically and cold, less frequently or never used data to the traditional disk system. An example of implementing SAN Volume Controller Easy Tier is shown in Figure 5-23 on page 161.

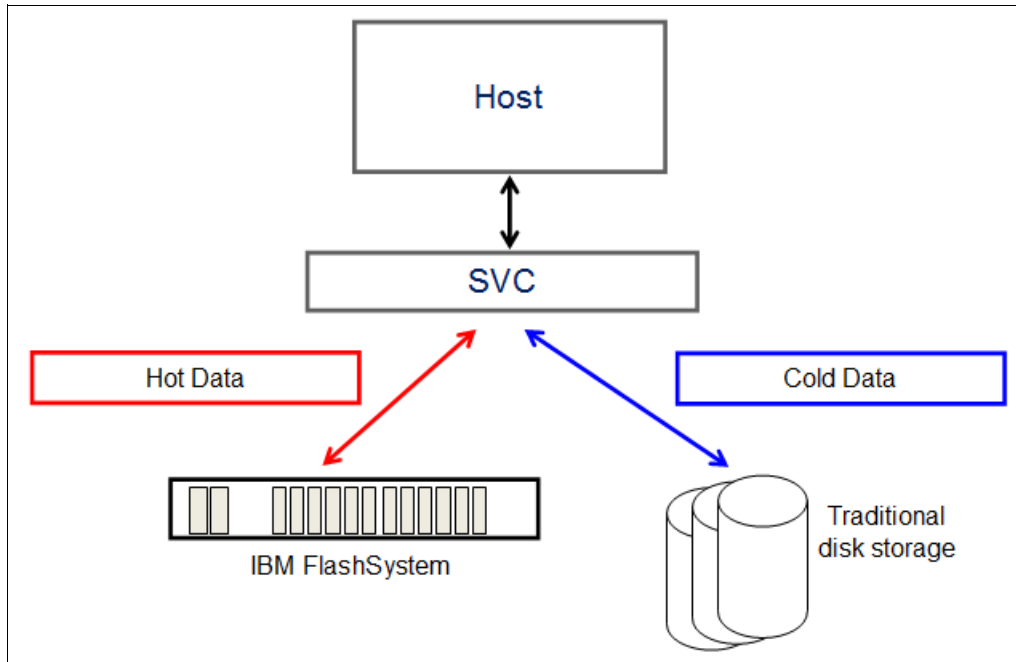


Figure 5-23 Easy Tier with SAN Volume Controller

For more details about the IBM FlashSystem 840 and SAN Volume Controller solution, see 8.1.4, “IBM Spectrum Virtualize - SAN Volume Controller advanced functionality” on page 284 and 9.2, “Tiering” on page 307.

5.7 Troubleshooting

Troubleshooting information for issues that you might encounter when configuring InfiniBand or creating file systems is described.

5.7.1 Troubleshooting Linux InfiniBand configuration issues

The following list shows potential Linux configuration issues, troubleshooting guidance, and resolutions:

- ▶ When you install OFED-X.X.X, an error occurs saying that you failed to build the ofa_kernel RPM Package Manager (originally Red Hat Package Manager).
The kernel that is used by the server might not be supported by OpenFabrics Enterprise Distribution (OFED). If the Install All option was chosen, try the Customize option in the OFED installation menu and select only the components that are needed. If this process does not work, try installing a different version of OFED.
- ▶ Loading the driver module fails.
The HCA used might not be supported by OFED, or the driver was not installed correctly. Obtain the latest drivers for the host channel adapters (HCAs) from the HCA vendor’s website.
- ▶ When you try to install OFED, an error occurs, such as “<module 1> is required to build <module 2>”.

This error means that certain dependencies that are required by OFED are not installed on the server. You must install all of the required dependencies:

- To search for the necessary RPM (if the **yum** package-management tool is available), enter this command:

```
# yum provides <dependency_name>
```

- To install the RPM, enter this command:

```
# yum install <dependency_rpm>
```

Note: If **yum** is not installed on the server, each dependency must be manually downloaded and installed.

- ▶ When you try to run the **srp_daemon** command, a message is displayed stating that an operation failed:
 - Make sure that the storage system is physically connected to the network and that all components are powered on.
 - Make sure that the correct cable is used and that OpenSM is running. To confirm whether OpenSM is running, enter this command:

```
# /etc/init.d/opensmd status
```

- ▶ Loading the **ib_srp** module fails.

Check that OFED is installed correctly and that the necessary device drivers are also installed. If a custom OFED installation was performed, make sure that **ibutils** and all packages that are related to **srp** were selected.

5.7.2 Linux fdisk error message

You might get an error message, such as “Re-reading the partitioning table failed”, when creating a partition on Linux. Also, the corresponding device is not created in the **/dev/mapper** directory. You solve this problem by issuing the **partprobe** command, as depicted in Example 5-29.

Example 5-29 Solving the partition table failed with error 22

```
[root@localhost ~]# fdisk /dev/mapper/FlashSystem_840_3
```

```
<partition creation lines left out for clarity>
```

```
...
```

```
Command (m for help): w
```

```
The partition table has been altered!
```

```
Calling ioctl() to re-read partition table.
```

```
WARNING: Re-reading the partition table failed with error 22: Invalid argument.
The kernel still uses the old table. The new table will be used at
the next reboot or after you run partprobe(8) or kpartx(8)
Syncing disks.
```

```
[root@localhost ~]# ls -l /dev/mapper/
```

```
lrwxrwxrwx. 1 root root      7 Oct  8 08:10 FlashSystem_840_2 -> ../dm-0
```

```
lrwxrwxrwx. 1 root root      7 Oct  8 09:20 FlashSystem_840_3 -> ../dm-2
```



```
[root@localhost ~]# partprobe
```

```
[root@localhost ~]# ls -l /dev/mapper/  
lrwxrwxrwx. 1 root root      7 Oct  8 08:10 FlashSystem_840_2 -> ../dm-0  
lrwxrwxrwx. 1 root root      7 Oct  8 09:20 FlashSystem_840_3 -> ../dm-2  
brw-rw----. 1 root disk 253,  7 Oct  8 09:21 FlashSystem_840_3p1
```

You can search the /dev/mapper directory for newly generated partitions. After using the **partprobe** command, the new partition is generated.

5.7.3 Changing FC port properties

The FlashSystem 840 automatically detects the SAN topology and speed. If you want to set the speed, for example, 16 Gbps, or the topology, such as arbitrated loop or fabric explicitly, you can use the **chportfc** command. The **lsportfc** command lists the current settings.

Note: The FlashSystem 840 16 Gbps FC attachment uses fabric topology only.

5.7.4 Changing iSCSI port properties

The FlashSystem 840 uses 10 Gb Ethernet for iSCSI connections. Use the **chportip** command to set the IP address, netmask, and gateway values of the iSCSI ports. The **lsportip** command lists the current settings.



Using the IBM FlashSystem 840

In this chapter, you learn how to operate the IBM FlashSystem 840 in your business environment. We use the graphical user interface (GUI) and the command-line interface (CLI) to demonstrate how to monitor the system and work with volumes, hosts, and user security.

6.1 Overview of IBM FlashSystem 840 management tools

Note: This chapter includes the release 1.3 updates. GUI screen captures from release 1.2 that have minimal changes as of release 1.3 have not been modified.

The FlashSystem 840 can be managed from either the built-in GUI, which is a web browser-based management tool, or from the CLI.

The web-based GUI is designed to simplify storage management and to provide a fast and more efficient management tool. It is based on the IBM System Storage XIV software and has a similar look and feel.

JavaScript: You might need to enable JavaScript in your browser. Additionally, if you are using Firefox, under Advanced JavaScript Settings, you need to click **Disable or replace context menus** and allow cookies.

You must use a supported web browser to be able to manage the FlashSystem 840 by using the GUI. For a list of supported web browsers, see “Supported web browsers” on page 90.

6.1.1 Access to the graphical user interface

To log on to the GUI, point your web browser to the management IP address that was set during the initial setup of the FlashSystem 840. The default credentials are as follows:

- ▶ User name: superuser
- ▶ Password: passw0rd (with a zero in place of the letter “o”)

See Figure 6-1.



Figure 6-1 Login window

After you log in successfully, the Monitoring → System window opens, showing the home window (Figure 6-2).

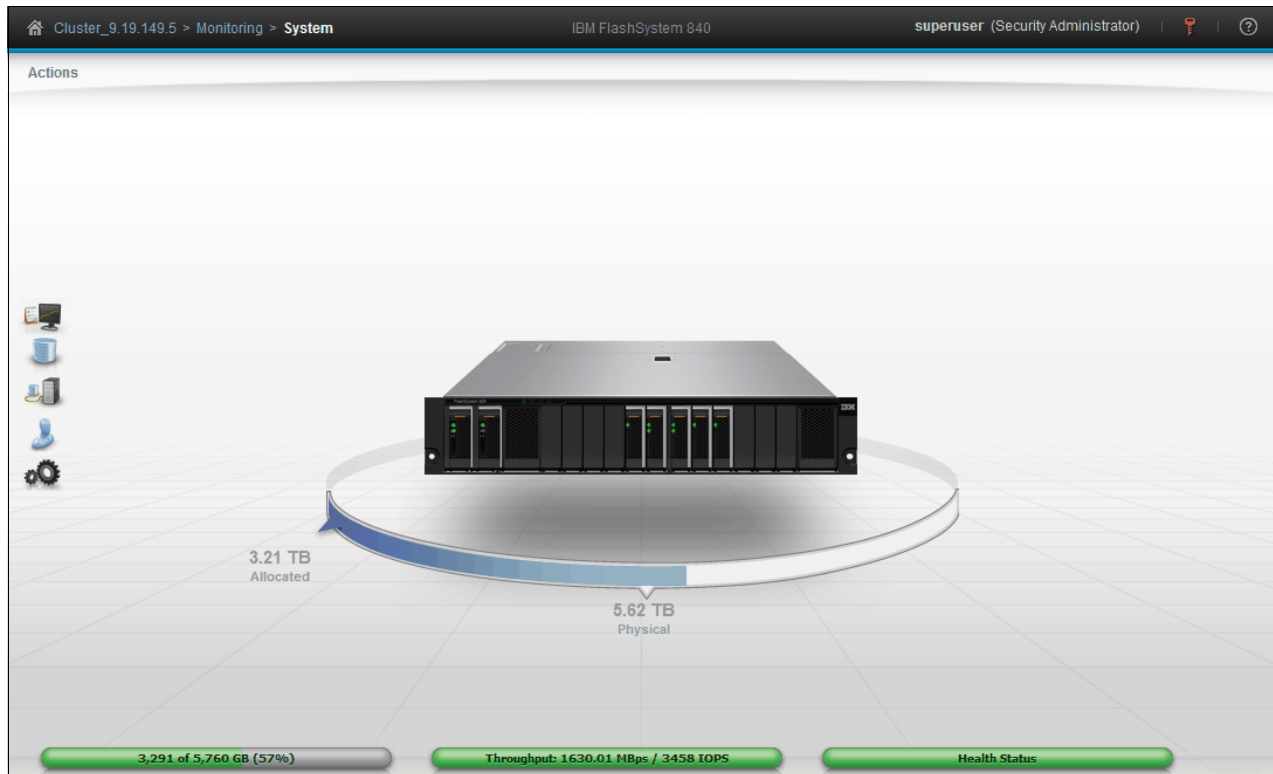


Figure 6-2 System overview window

6.1.2 Graphical user interface layout

The GUI has three main sections for navigating through the management tool (see Figure 6-3 on page 168):

- ▶ Function icons (left side)
- ▶ Status bars (bottom)
- ▶ Actions menu (upper left or right-click in the home window)

On the far left of the window are five *function icons*. The five function icons represent these areas:

- ▶ Monitoring menu
- ▶ Volumes menu
- ▶ Hosts menu
- ▶ Access menu
- ▶ Settings menu

To the upper left of the GUI is the Actions button. Actions can also be reached by right-clicking anywhere on the GUI system overview window.

In the upper-right corner of the GUI is the User security key for managing the security of the user that is logging in.

The Help (question mark (?)) icon, which provides information about licenses and gives access to the FlashSystem 840 IBM Knowledge Center, is in the upper-right corner.

At the bottom of the window are three status indicators, which provide more detailed information about the existing configuration of the FlashSystem 840 solution.

Figure 6-3 shows the main areas of the GUI system overview.

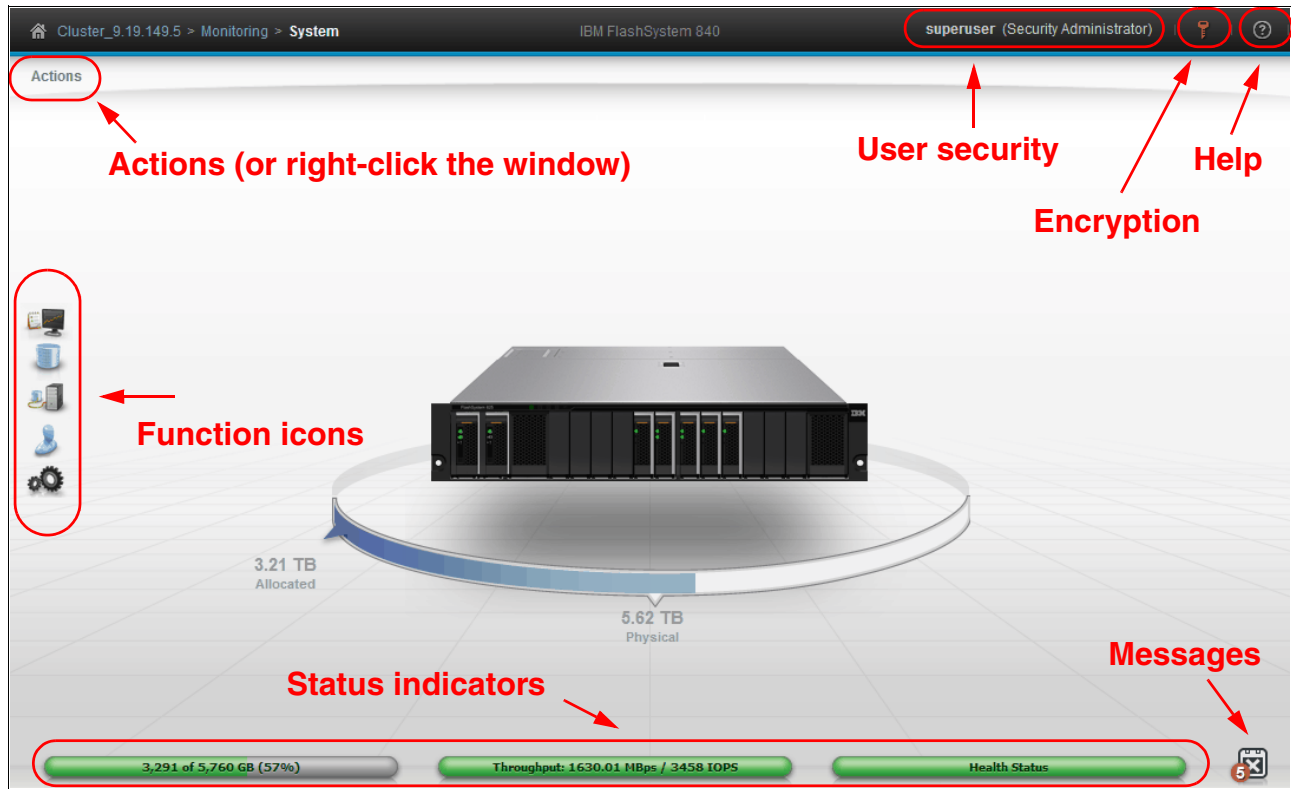


Figure 6-3 System overview window: Main areas

6.1.3 Navigation

Navigating the management tool is simple. You can hover the cursor over one of the five function icons on the left side of the window, which highlights the function icon and shows a list of options. Figure 6-4 on page 169 shows a list of the FlashSystem 840 software function icons and the associated menu options.

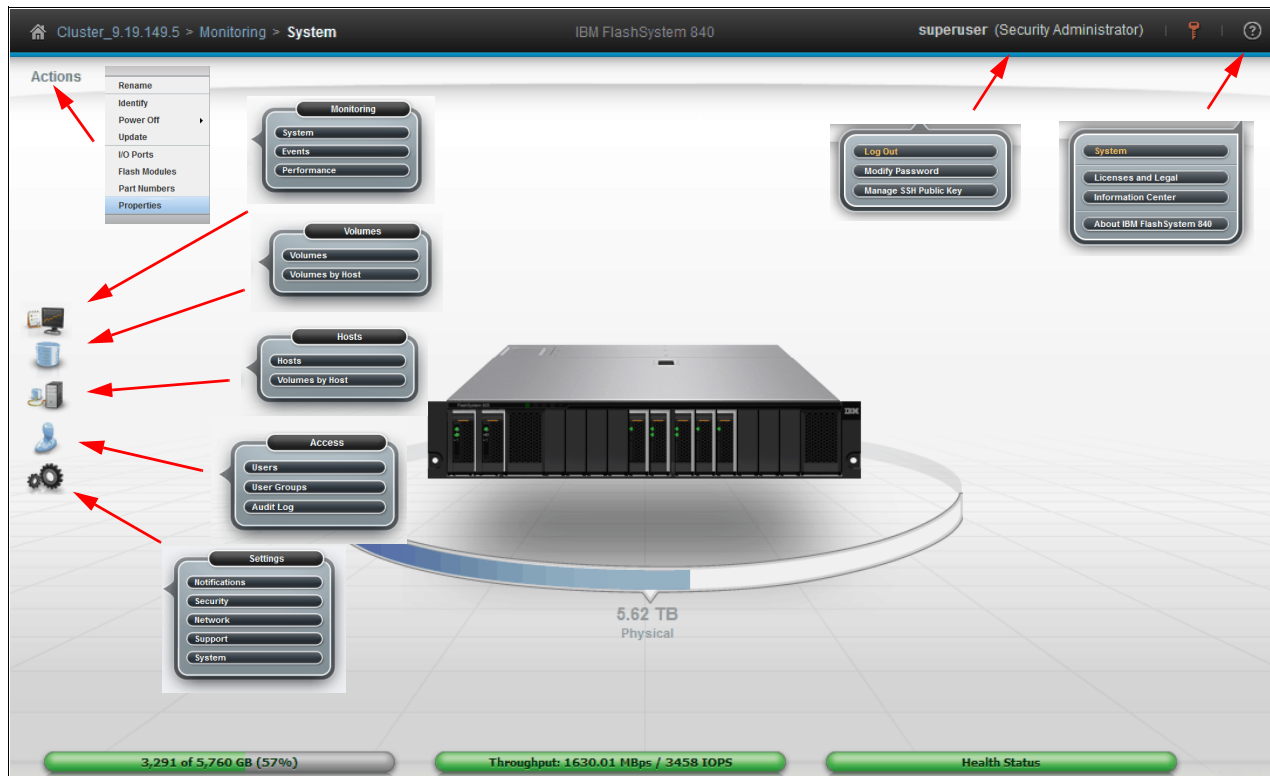
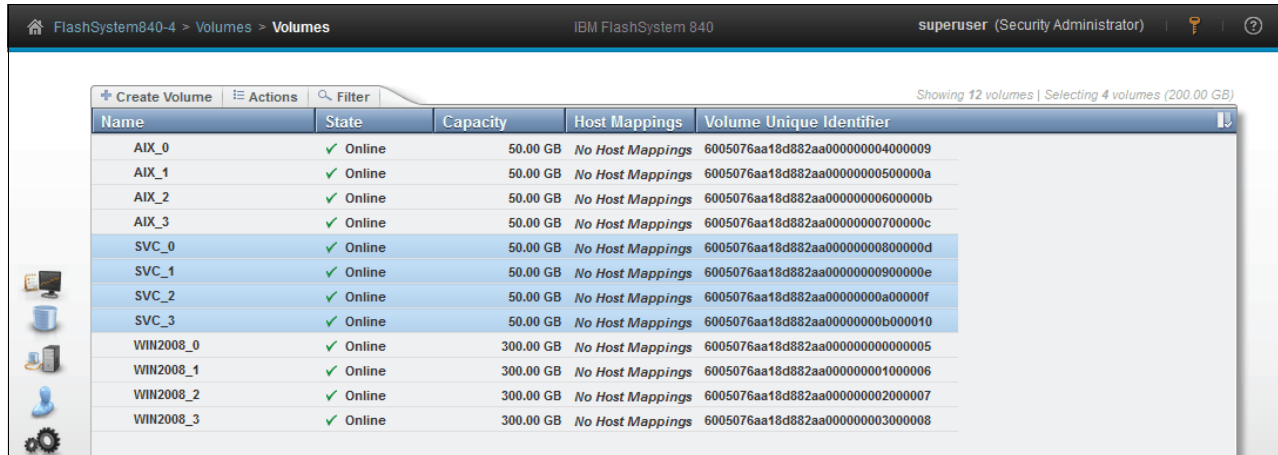


Figure 6-4 FlashSystem 840: Menu options

The following sections detail each of the five function icons and the associated menu options.

6.1.4 Multiple selections

The FlashSystem 840 management tool lets you select multiple items by using a combination of the Shift keys or Ctrl keys. To select multiple items in a display, click the first item, press and hold the Shift key, and click the last item in the list that you require. All the items in between those two items are selected. For example, Figure 6-5 from the Volumes → Volumes menu illustrates multiple selections.



Name	State	Capacity	Host Mappings	Volume Unique Identifier
AIX_0	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa000000004000009
AIX_1	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000500000a
AIX_2	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000600000b
AIX_3	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000700000c
SVC_0	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000800000d
SVC_1	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000900000e
SVC_2	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000a00000f
SVC_3	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000b000010
WIN2008_0	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa000000000000005
WIN2008_1	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa0000000001000006
WIN2008_2	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa0000000002000007
WIN2008_3	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa0000000003000008

Figure 6-5 Multiple selections by using the Shift key

This functionality is useful if the administrator wants to expand multiple volumes at the same time.

If you want to select multiple items that are not in sequential order, click the first item, press and hold the Ctrl key, and click the other items that you require (Figure 6-6).

Name	State	Capacity	Host Mappings	Volume Unique Identifier
AIX_0	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa000000004000009
AIX_1	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000500000a
AIX_2	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000600000b
AIX_3	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000700000c
SVC_0	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000800000d
SVC_1	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000900000e
SVC_2	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000a00000f
SVC_3	✓ Online	50.00 GB	No Host Mappings	6005076aa18d882aa00000000b000010
WIN2008_0	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa000000000000005
WIN2008_1	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa000000001000006
WIN2008_2	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa000000002000007
WIN2008_3	✓ Online	300.00 GB	No Host Mappings	6005076aa18d882aa000000003000008

Figure 6-6 Multiple selections using the Ctrl key

6.1.5 Status indicators

Other useful tools are the Status indicators that appear at the bottom of the window (Figure 6-7). These indicators provide information about capacity usage, throughput in megabytes per second and I/O per second (IOPS), and the health status of the system. The status indicators are visible from all windows in the FlashSystem 840 GUI.

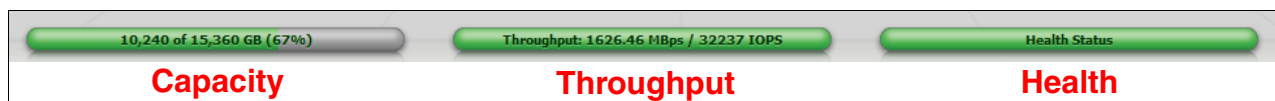


Figure 6-7 Status indicators at the bottom of the GUI window

The status indicators show the following information:

- ▶ Capacity:
 - Used gigabytes
 - Installed gigabytes
 - Percentage used
- ▶ Throughput:
 - Megabytes per second (MBps)
 - I/O per second (IOPS)
- ▶ Health one of these options:
 - Healthy (green)
 - Warning (yellow) and a link to Monitoring → Events is provided
 - Error (red) and a link to Monitoring → Events is provided

6.2 Monitoring menu

We describe the Monitoring menu and its options. The Monitoring → System menu is the default menu and home page for the FlashSystem 840 GUI. It has three options as shown in Figure 6-4 on page 169:

- ▶ System
- ▶ Events
- ▶ Performance

Part of the default window or home page is also the Actions menu where the system can be managed and information about the system can be obtained.

6.2.1 Monitoring System menu

On the home window of the FlashSystem 840 GUI, you can select Actions in the upper-left corner. Actions can also be activated by right-clicking anywhere in the GUI.

System properties

The following information can be retrieved from the Actions → Properties window:

- ▶ System name
- ▶ System state
- ▶ Hardware type
- ▶ Firmware version
- ▶ Serial number
- ▶ Model and type number
- ▶ Worldwide name
- ▶ Storage capacity
- ▶ Power on days

In our example, we right-click in the middle of the window and get the Actions menu as shown in Figure 6-8 on page 173.

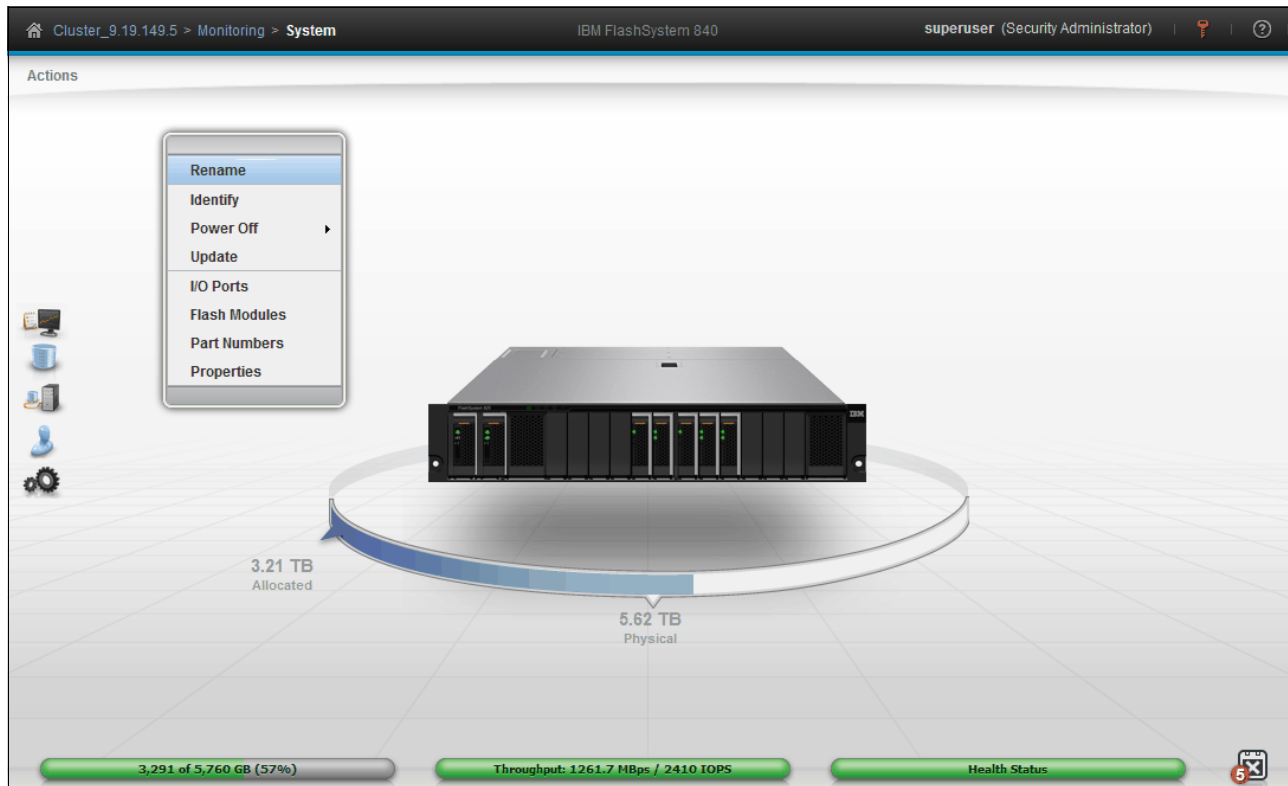


Figure 6-8 Monitoring System menu with actions displayed

We now click **Properties** where we can observe the properties of the system as shown in Figure 6-9 on page 174.

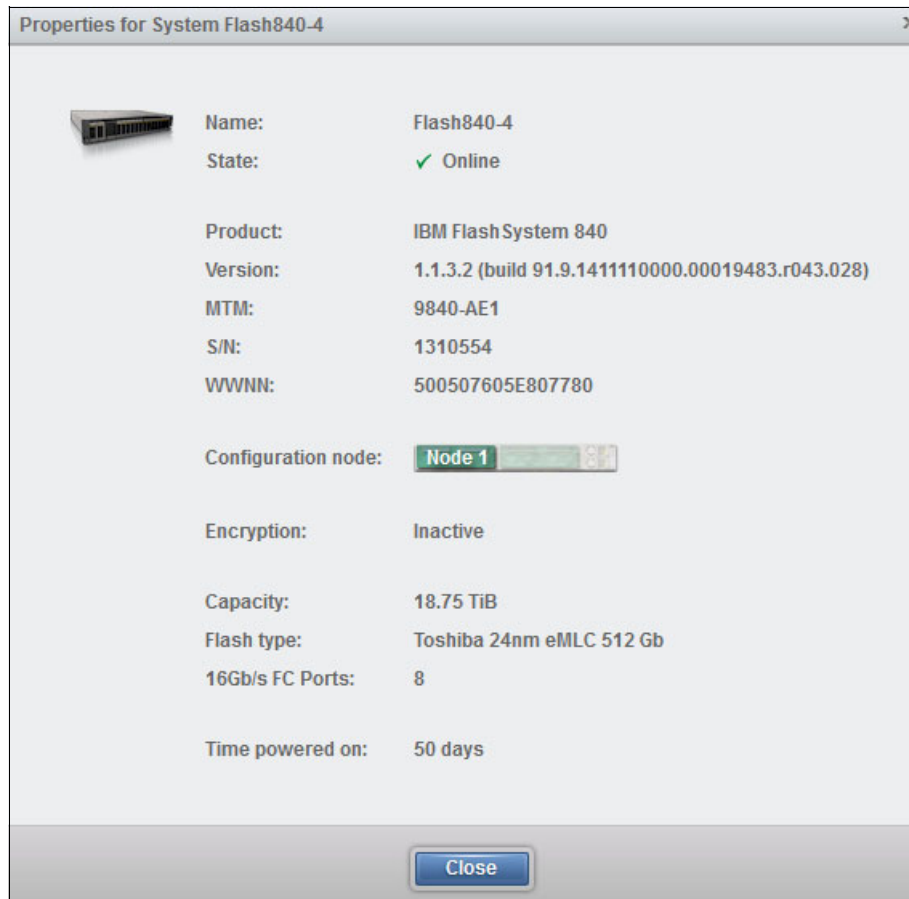


Figure 6-9 Properties for the system cluster

Rename System

The host name of the system can be changed. Click **Actions** → **Rename** and type the new name of the system as shown in Figure 6-10.

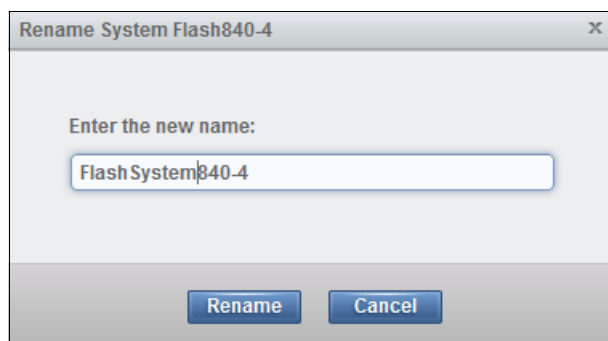


Figure 6-10 Rename the system

When the system is renamed or when any command is executed from the FlashSystem 840 GUI, the task window opens. The CLI command that the system uses to make the change can be reviewed, as shown in Figure 6-11 on page 175.

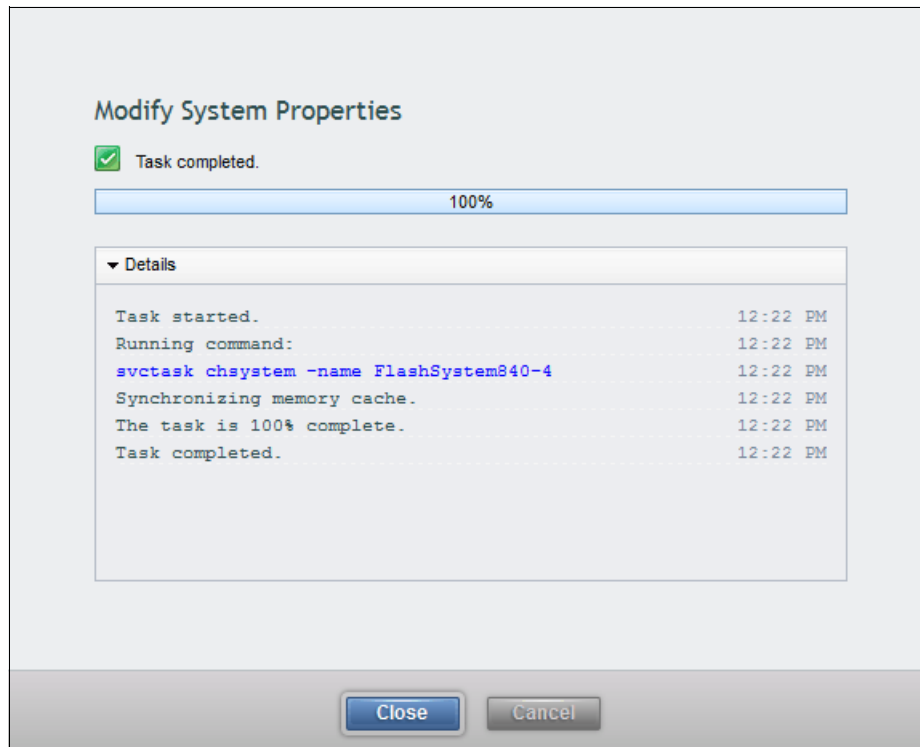


Figure 6-11 CLI command displays

The CLI commands that are displayed can also be executed by the user from within an open CLI window by using PuTTY or a similar terminal emulation tool. For more information about how to use the CLI, see 6.5.3, “Access CLI by using PuTTY” on page 229

Rename system by using CLI

When system properties and settings are changed from the FlashSystem 840 GUI, commands are executed on the system. In the preceding example, we change the system host name by using the GUI, and the Modify System Properties window opens. In this window, the CLI commands that the system uses to change system properties display.

Example 6-1 shows an example of using the CLI to change system properties. The output is shortened for clarity.

Example 6-1 Change the system name by using the CLI

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>svctask chsystem -name
FlashSystem_840
```

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>svcinfolssystem
id 000002006487F054
name FlashSystem_840
```

Identify LED

Another function of the Actions menu is the Identify function. When the Identify function is enabled, the Identify LED on the front side of the IBM FlashSystem 840 and both controller canisters turn on their blue Identify LED. The canisters are mounted from the rear side of the FlashSystem 840 and the canister Identify LEDs can be seen from the rear side of the unit.

Figure 6-12 shows the Identify LED when it is on.

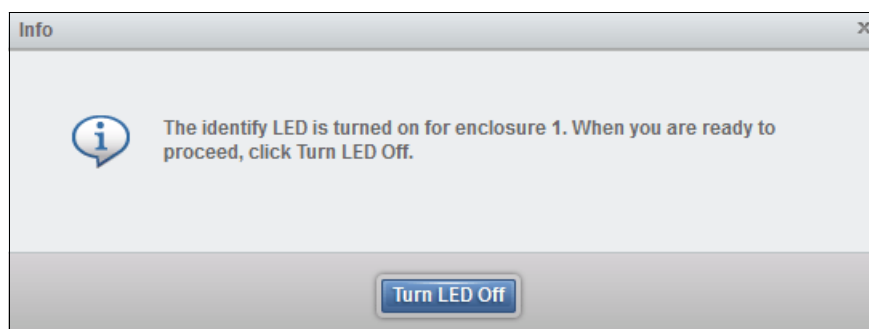


Figure 6-12 Identify LED is turned on

Also, each canister can be identified through the IBM FlashSystem 840 Service Assistant Tool. The Service Assistant Tool is described in more detail in 7.2, "Service Assistant Tool" on page 271.

Power off

The IBM FlashSystem 840 can be turned off through the Actions menu and the individual controller canisters can also be turned off. There can be many reasons to turn off the entire unit or to turn off a canister. One reason might be that you need to relocate the system to another site or that a canister must be taken out for scheduled maintenance. The power off function ensures that the system or canister is turned off securely so that data is preserved.

Figure 6-13 shows the Actions → Power Off window where the entire system or individual canisters can be turned off.

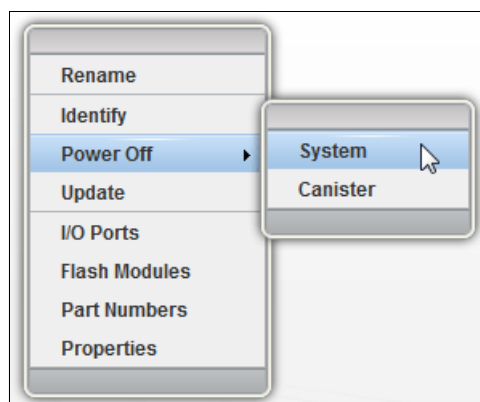


Figure 6-13 Power off entire system or a single canister

Figure 6-14 on page 177 shows the Power Off Canister window where Canister 1 is selected to be powered off.

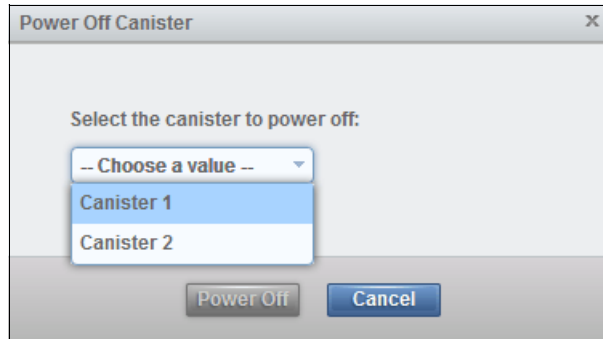


Figure 6-14 Power off a single canister

Canister 1 is the left canister when viewed from the rear side of the IBM FlashSystem 840.

Figure 6-15 shows that the Power Off window requires the administrator to type a confirmation code to prevent an accidental power-off cycle of the device.

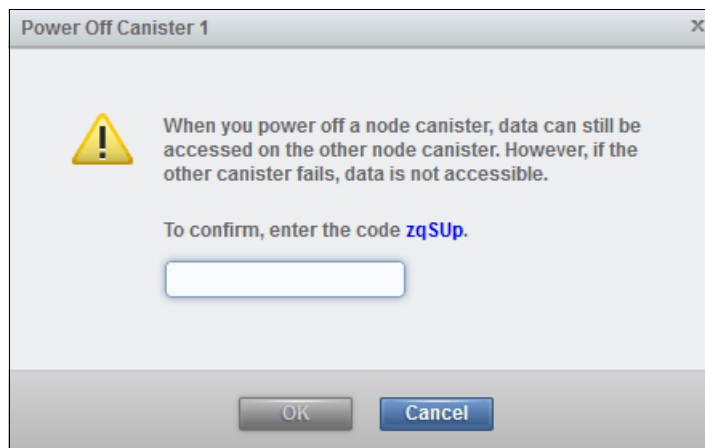


Figure 6-15 Power off Canister 1 confirmation code

The same confirmation code window appears when the entire system is to be powered off.

Note: Canister 1 is the left-side canister and Canister 2 is the right-side canister when viewed from the rear side of the IBM FlashSystem 840.

Flash module properties

The menu Actions → Flash Modules opens the Flash Module Properties information window. Flash modules within an initialized FlashSystem 840 must always be *online*, except when a flash module is in the *failed* state.

The *Use* column (parameter) shown in Figure 6-16 on page 178 can have different values depending on the RAID mode.

Note: As of FlashSystem 840 release 1.3, before enabling RAID 0, you must submit a Solution for Compliance in a Regulated Environment (SCORE)/request for price quotation (RPQ) to IBM. To submit a SCORE/RPQ, contact your IBM representative.

You have the following values when you use RAID 0:

- ▶ Candidate (ready to be a RAID 0 member).
- ▶ Member.
- ▶ RAID 0 does not support spare flash modules and does not provide redundancy for failed flash modules.

You have the following values when you use RAID 5:

- ▶ Candidate (ready to be a RAID 5 member or spare).
- ▶ Member.
- ▶ Spare.

RAID 5 provides redundancy for failed flash modules and keeps one flash module as a *spare*. The only situation in which a flash module can be a *candidate* is when there is no RAID configuration on the flash module. Including a candidate flash module into the RAID configuration requires the reinitialization of the array, which is a data destructive action. For instructions about how to reinstall the RAID configuration, see 4.4, “RAID storage modes” on page 104.

Note: The only situation in which a flash module can be a candidate is when there is no RAID configuration on the flash modules.

For more information about how to change the RAID mode, see 4.4.1, “Changing RAID modes” on page 104.

Figure 6-16 shows the Flash Module Properties information window.

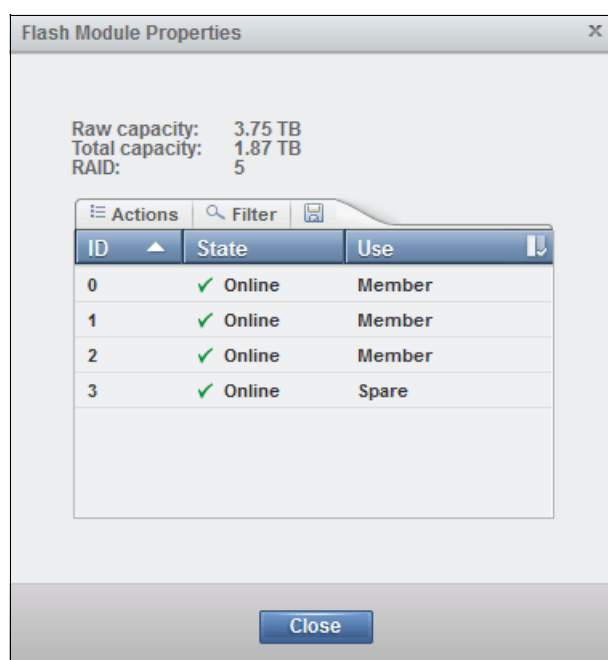


Figure 6-16 Flash Module Properties display

Part numbers

A list of part numbers for customer-replaceable units (CRUs) and field-replaceable units (FRUs) can be reviewed by selecting **Actions** → **Part Numbers**. CRUs can be replaced by IBM clients, and FRUs are replaced by either IBM Support or an IBM Service Partner.

Figure 6-17 shows the part numbers that are available for the IBM FlashSystem 840.

Part Numbers for Storage System Flash840-4			
<div> <div>Filter</div> <div></div> </div> <div>Showing 22 parts Selecting 0 parts</div>			
Item	Canister	Part Number	Part Identity
Battery 1		00DH517	11S00DH846YS11WZ43S021
Battery 2		00DH517	11S00DH846YS11WZ43S020
Canister 1 (Configuration Node)	1 (Left)	00DH520	11S00DH404YS1CWD43V544
Canister 2	2 (Right)	00DH520	11S00DH404YS1CWD43V547
Fan Module 1	2 (Right)	00DH516	11S00DJ103YS12WD42V675
Fan Module 1	1 (Left)	00DH516	11S00DJ103YS12WD43L212
Fan Module 2	2 (Right)	00DH516	11S00DJ103YS12WD43H050
Fan Module 2	1 (Left)	00DH516	11S00DJ103YS12WD43L265
Flash Module 0		00DH514	11S00DH304YS12WD3BV156
Flash Module 1		00DH514	11S00DH304YS13WD43S044
Flash Module 10		00DH514	11S00DH304YS13WD43L795
Flash Module 11		00DH514	11S00DH304YS13WD43S046
Flash Module 2		00DH514	11S00DH304YS13WD43L796
Flash Module 3		00DH514	11S00DH304YS13WD43S001

Figure 6-17 List of part numbers

I/O ports

The I/O ports of the IBM FlashSystem 840 interface cards can be managed from the Actions → I/O Ports menu. Different types of settings can be configured from the I/O Ports menu depending on the type of interface cards installed.

Port speed and topology can be configured for Fibre Channel and Fibre Channel over Ethernet (FCoE). The iSCSI IP addresses are set through this menu for Internet Small Computer System Interface (iSCSI).

Fibre Channel and FCoE

In the following example, click **Actions** → **I/O Ports** from the home window. Eight 16 Gbps Fibre Channel (FC) ports are displayed, each displaying the following information:

- ▶ Depiction of port location
- ▶ State
- ▶ Fibre Channel (FC) or Fibre Channel over Ethernet (FCoE)
- ▶ Port speed: Auto, 16 Gbps, 8 Gbps, 4 Gbps, or 2 Gbps
- ▶ Worldwide port name (WWPN)/Globally Unique Identifier (GUID)
- ▶ Topology: Fibre Channel Arbitrated Loop (FC-AL) or Fibre Channel-Peer to Peer (FC-P2P)

Figure 6-18 on page 180 shows the Fibre Channel I/O ports in the system.

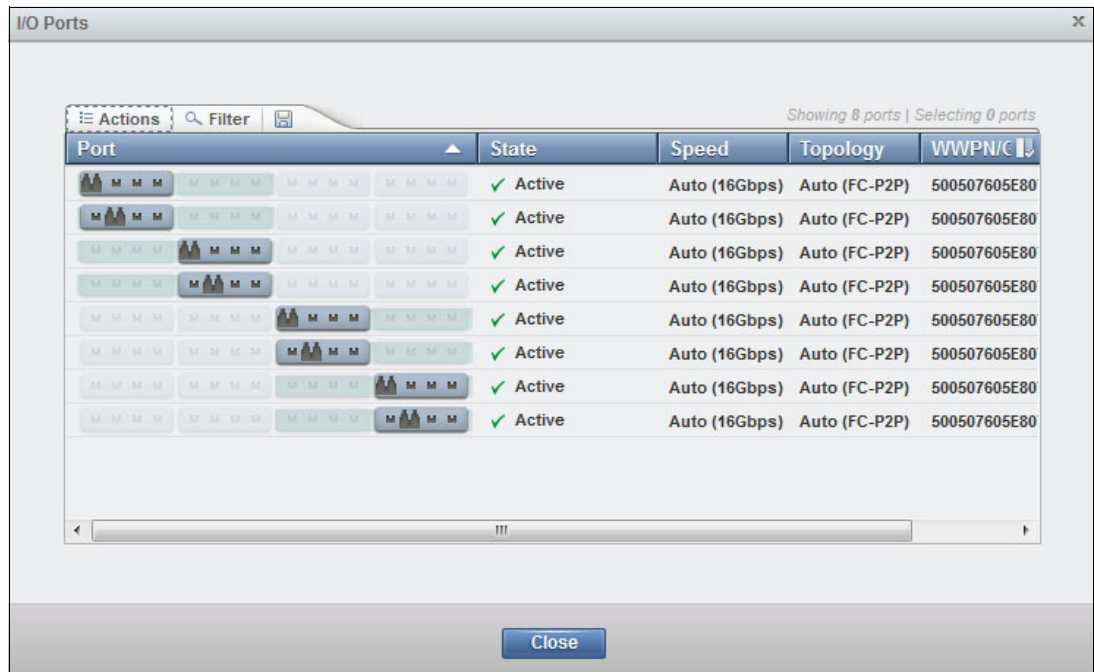


Figure 6-18 The status of the I/O ports

Any FC or FCoE port that is not connected and online has an Inactive status. In Figure 6-18, two ports on each interface card are connected and active.

FC speed and topology can be configured for each of the ports. To configure port speed, click **Actions** → **Modify Speed** as shown in Figure 6-19.

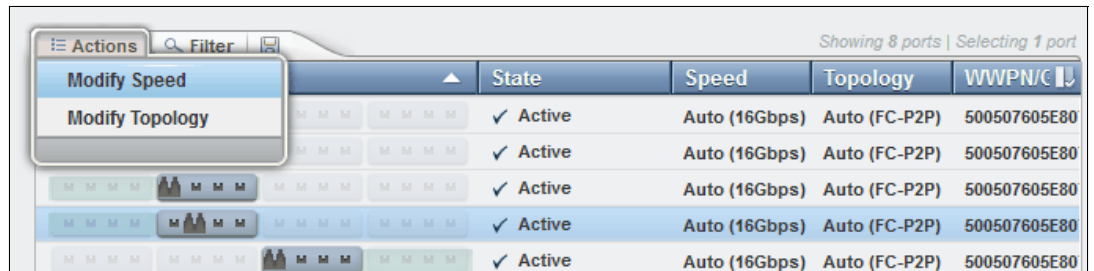


Figure 6-19 Modify speed

Next, select the speed. The default speed is *Automatic* as shown in Figure 6-20 on page 181. The Modify Speed wizard provides the feasible speed for the installed interface cards. In our example in Figure 6-20 on page 181, 16 Gbps Fibre Channel interface cards are also capable of using 8 Gbps and 4 Gbps FC speed.



Figure 6-20 Configure port speed for port 1-1-1

To configure port topology from the I/O Ports window, click **Actions** → **Modify Topology** as shown in Figure 6-21.

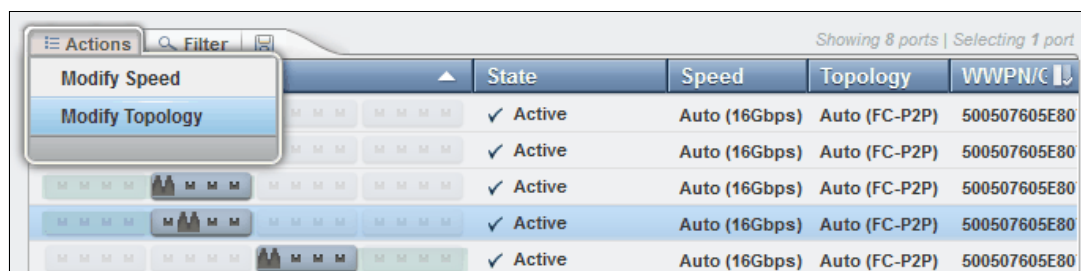


Figure 6-21 Modify topology

Next, select the topology. Select either Automatic, FC-AL (Arbitrated Loop), or FC-P2P (Point-to-Point), as shown in Figure 6-22.

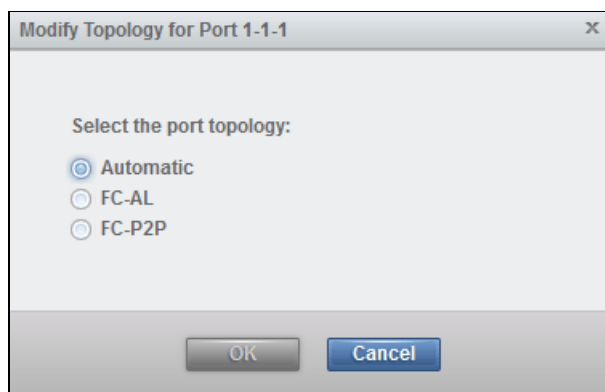


Figure 6-22 Configure topology for port 1-1-1

The FC-P2P (Point-to-Point) topology is used in situations where two FC ports connect directly to each other. FC-P2P is the default for a host that is directly connected to the FlashSystem 840, and FC-P2P is also used for a FlashSystem 840 FC port that is connected to a Fibre Channel switch.

The FC-AL (Arbitrated Loop) topology is also used to attach a host directly to the FlashSystem 840 in cases where the host only supports FC-AL, for example, when connecting a VMWare ESX server directly to the FlashSystem 840.

Note: FC-AL is not supported for ports that are connected at 16 Gbps.

Figure 6-23 shows the logical numbering of the FC ports.

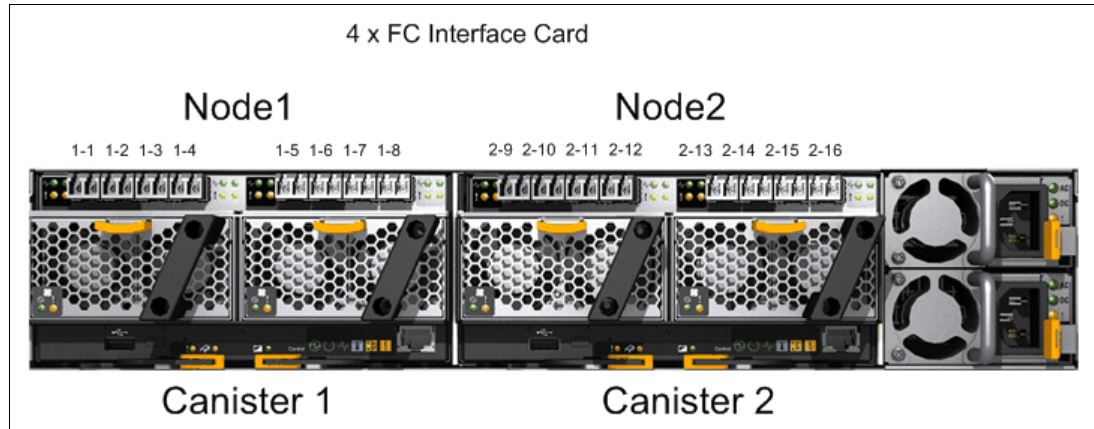


Figure 6-23 I/O port logical numbering

For a description of the physical numbering of the ports of the interface cards, see Chapter 4, “Installation and configuration” on page 65.

Note: The physical numbering and the logical numbering of the ports are not the same. The physical numbering of the ports of the interface cards starts from the left with P1 and P4 to the right, depending on the configuration. Logical numbering depends on the node name and these node names can swap, so the port names depicted in Figure 6-23 are the default names, but they might change.

iSCSI

From the FlashSystem 840 GUI, click **Actions** → **I/O Ports** from the home window. Sixteen iSCSI ports are displayed, each displaying the following information:

- ▶ Depiction of port location
- ▶ State
- ▶ IP address
- ▶ Subnet mask
- ▶ Gateway

The result is shown in Figure 6-24 on page 183.

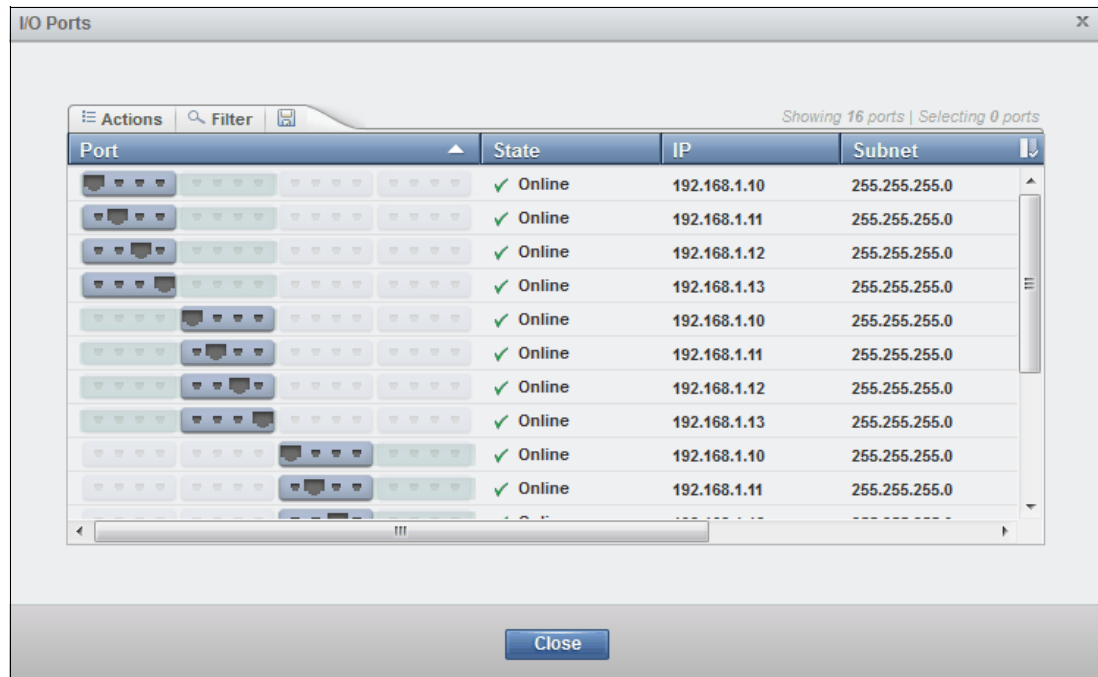


Figure 6-24 Display iSCSI I/O ports

To change the IP address of an iSCSI port, click **Actions** → **Modify** as shown in Figure 6-25.

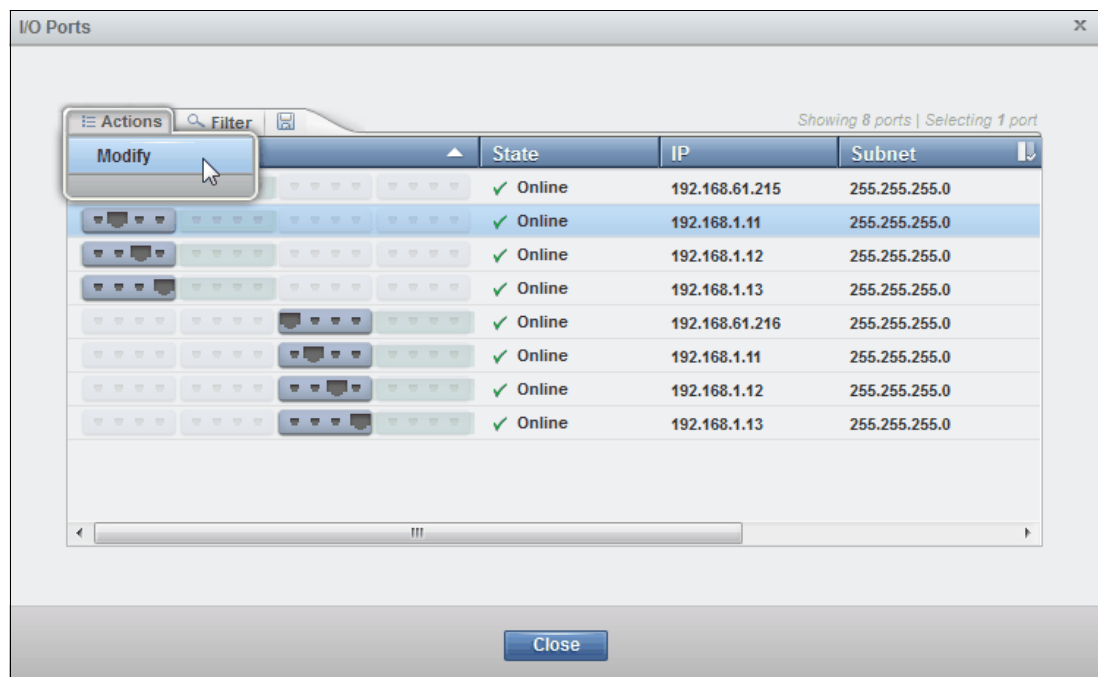


Figure 6-25 Modify iSCSI I/O ports

Next, type the new IP address, Subnet mask, and Gateway. Click **Modify** as shown in Figure 6-26 on page 184.

Modify Port 4 on Adapter 1 in Canister 1

IP Address: 192.168.61.217

Subnet mask: 255.255.255.0

Gateway: 192.168.1.1

Modify Cancel

Figure 6-26 Set the new IP address of an iSCSI I/O port

The Modify Port dialog can also be opened by selecting a port for reconfiguration, right-clicking, and clicking Modify (not shown here).

Figure 6-27 shows the final result of configuring the IP addresses of the iSCSI ports.

I/O Ports

Showing 8 ports | Selecting 0 ports

Port	State	IP	Subnet
	✓ Online	192.168.61.215	255.255.255.0
	✓ Online	192.168.61.217	255.255.255.0
	✓ Online	192.168.61.219	255.255.255.0
	✓ Online	192.168.61.221	255.255.255.0
	✓ Online	192.168.61.216	255.255.255.0
	✓ Online	192.168.61.218	255.255.255.0
	✓ Online	192.168.61.220	255.255.255.0
	✓ Online	192.168.61.222	255.255.255.0

Close

Figure 6-27 Review the final result of the iSCSI port configuration

The iSCSI ports of a FlashSystem 840 *must* be cabled and online. If not, an error indicating “Port Failure” appears in the Monitoring → Events menu. An example of a configuration where only the leftmost port on each interface card is cabled and online is shown in Figure 6-28 on page 185.

I/O Ports

Showing 16 ports | Selecting 1 port

Port	State	IP	Subnet	Gateway
10.1.0.2	Online	10.1.0.2	255.255.255.0	0.0.0.0
10.1.0.3	Online	10.1.0.3	255.255.255.0	0.0.0.0
192.168.1.10	Offline	192.168.1.10	255.255.255.0	0.0.0.0
192.168.1.10	Offline	192.168.1.10	255.255.255.0	0.0.0.0
10.1.0.4	Online	10.1.0.4	255.255.255.0	0.0.0.0
10.1.0.5	Online	10.1.0.5	255.255.255.0	0.0.0.0
192.168.1.10	Offline	192.168.1.10	255.255.255.0	0.0.0.0
192.168.1.10	Offline	192.168.1.10	255.255.255.0	0.0.0.0
10.1.0.6	Online	10.1.0.6	255.255.255.0	0.0.0.0
10.1.0.7	Online	10.1.0.7	255.255.255.0	0.0.0.0
192.168.1.10	Offline	192.168.1.10	255.255.255.0	0.0.0.0

Close

Figure 6-28 iSCSI ports not cabled

6.2.2 Monitoring events

The IBM FlashSystem 840 might show the health status indicator as green (healthy), yellow (degraded), or even red (critical). Events on the IBM FlashSystem 840 storage system are logged in the event log of the Monitoring → Events menu.

Navigating to events

To navigate to the event log, hover the cursor over the **Monitoring** icon and then click **Events** (Figure 6-29 on page 186).

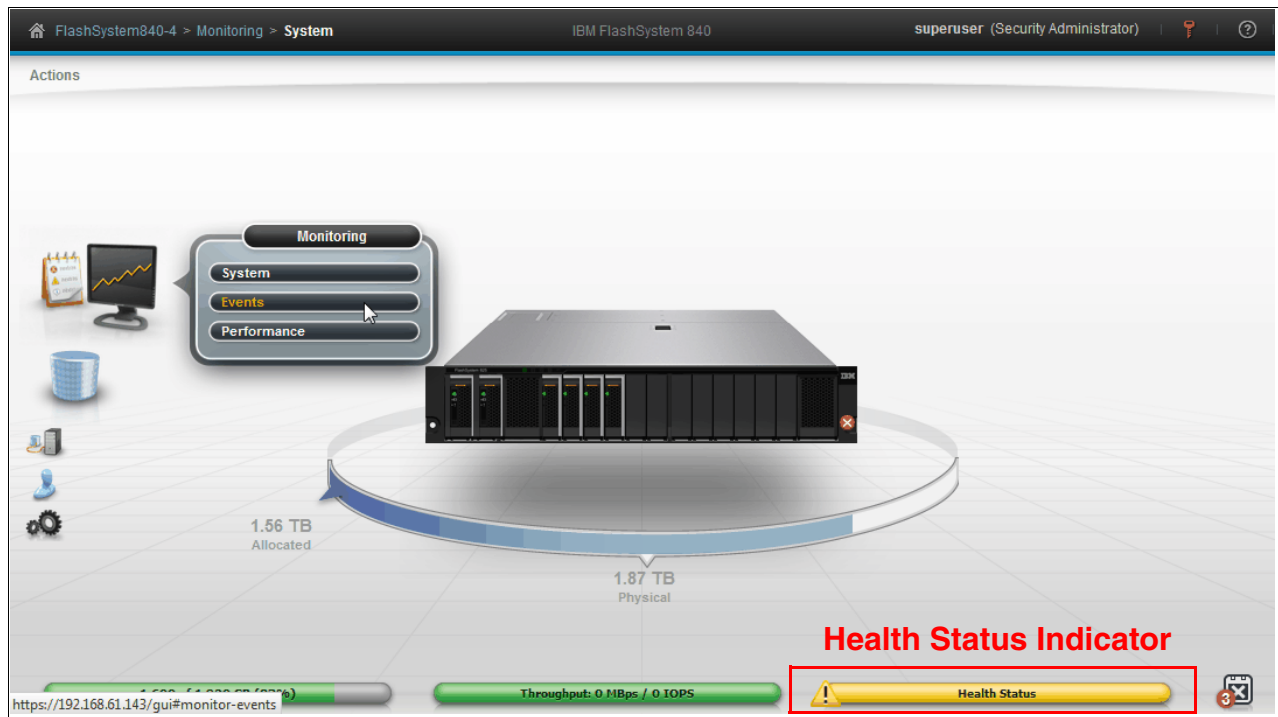


Figure 6-29 Monitoring events

You can also click the attention (!) triangle icon on the left side of the Health Status indicator to get to the Monitoring → Events menu as shown in Figure 6-29. A yellow Health Status indicator indicates a warning state or degraded.

Figure 6-30 on page 187 shows the Monitoring → Events window where the default mode is *Show All*. In Show All mode, all events, including Messages, Warnings, and Errors are displayed.

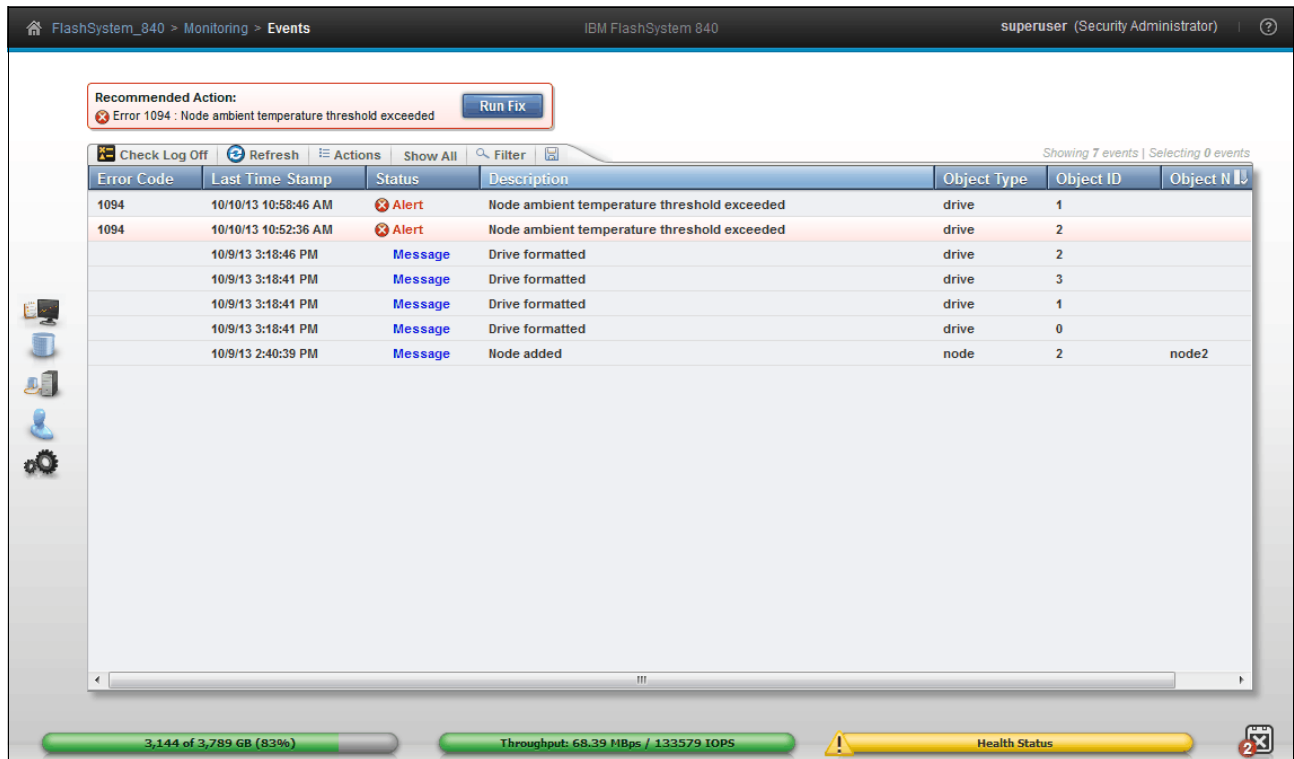


Figure 6-30 The Events window

The Monitoring → Events menu can be manipulated in several ways by using the function tabs that are displayed over the list of events.

Check Log LED off

One function of the FlashSystem 840 is its Check Log LED. This LED illuminates amber for a problem that is not isolated. An error condition results in the problem being called home. There is also a service action, and a warning condition results in a service action that the user is expected to fix. There is no correlation between the notification type of “error/warning” and the Check Log LED.

The leftmost function key of the Events menu is the Check Log Off. With Check Log Off, you turn off the Check Log LED on the front of the IBM FlashSystem 840, and only new events turn it on again.

From the Monitoring → Events menu, we click **Check Log Off**. The result is a new window as shown in Figure 6-31.

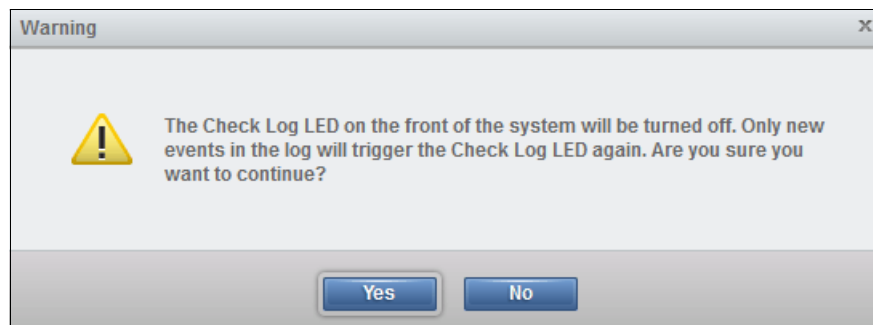


Figure 6-31 Events error LED off

We click **Yes** and the Check Log LED is off.

Change the Events view

You might want more or less information from the Monitoring → Events window. You can change the default view by right-clicking in the menu bar or by clicking the check mark icon in the upper-right corner of the Monitoring → Events window as shown in Figure 6-32.

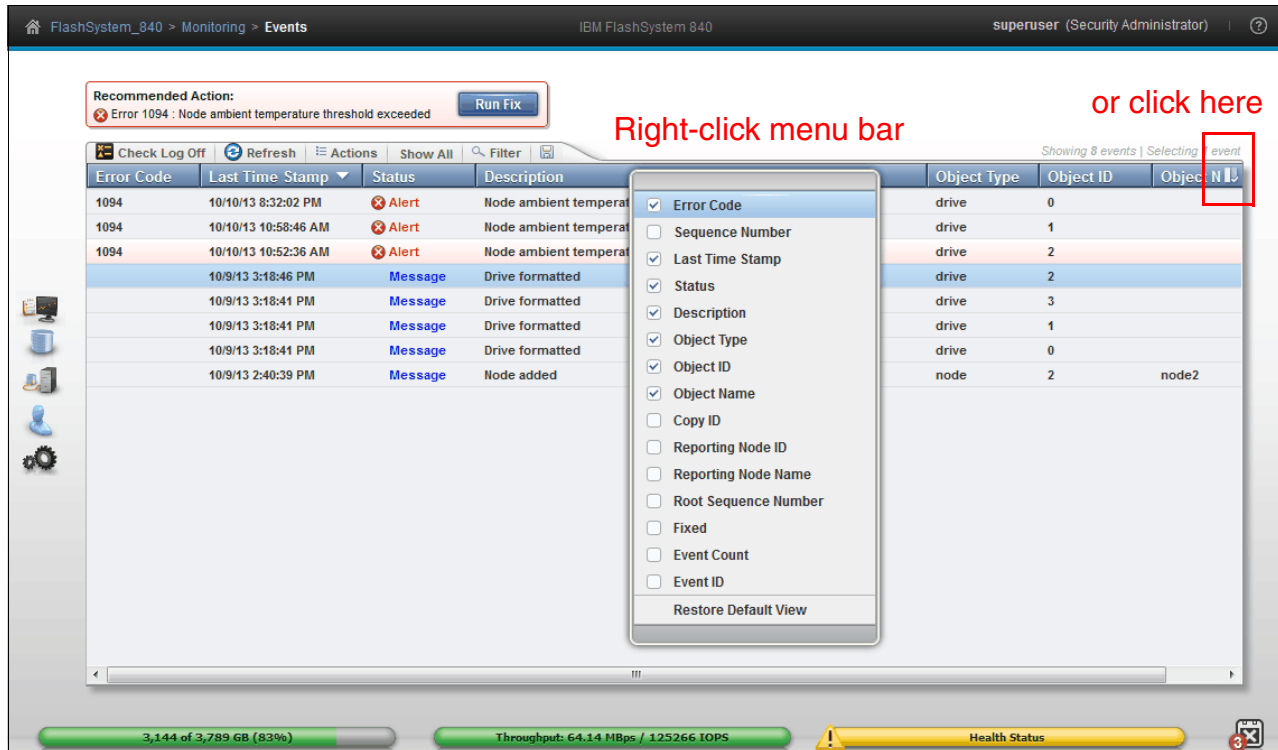


Figure 6-32 Events: Customize columns

From the Actions menu, you can perform the following actions:

- ▶ Run fix procedure on error events
- ▶ Mark informational events as fixed
- ▶ Clear the event log
- ▶ Filter displayed events on date
- ▶ Show only events from the last minutes, hours, or days
- ▶ Show the properties of an event

Figure 6-33 on page 189 shows that we select to only view events that are newer than 5 hours (show events within the last 5 hours).

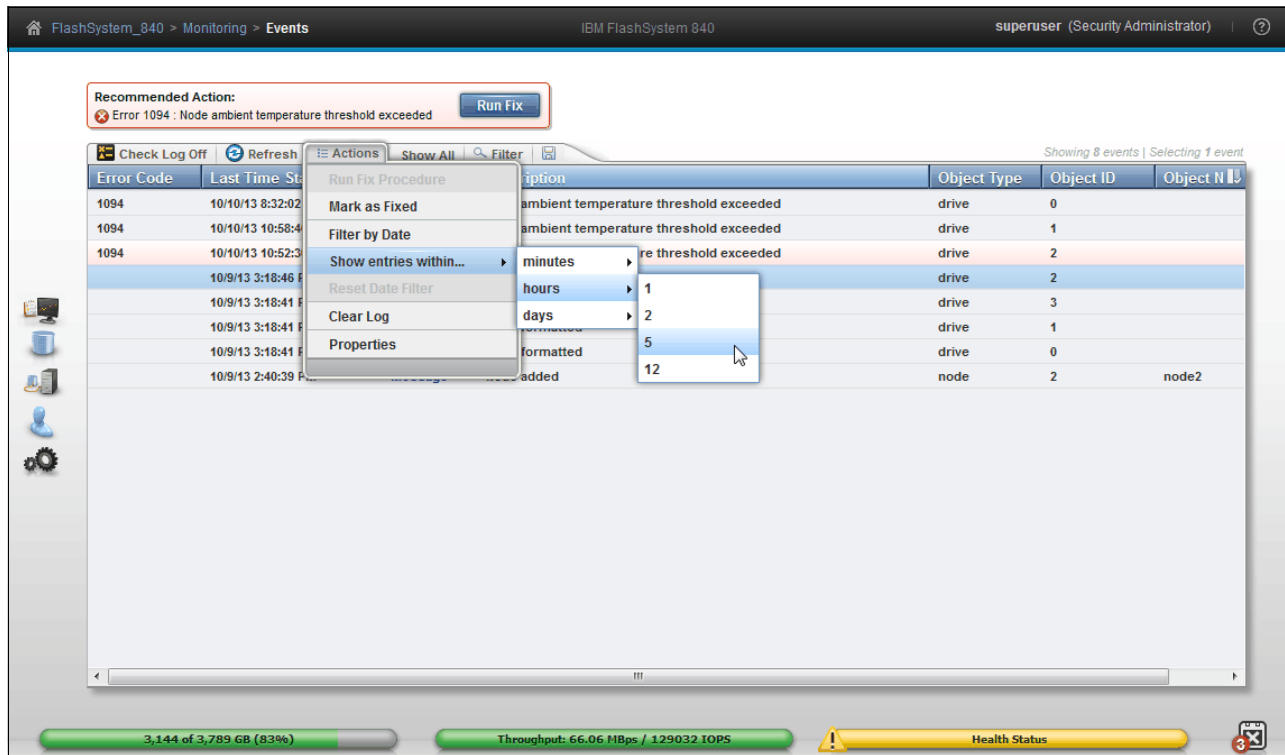


Figure 6-33 Events: Show entries within five hours

Recommended actions

In the Monitoring → Events menu, messages and alerts are displayed. If any unresolved issues exist, the Recommended Actions section displays. You can click Run Fix to initiate the Fix Procedure; the IBM FlashSystem 840 checks whether the problem still exists and fixes the issue, if possible. The fix procedure might bring the system out of a Degraded state and into a Healthy state.

In a normal situation during the daily administration of the FlashSystem 840, you are unlikely to see error events. There might however be a continuing flow of informational messages. The typical Events display is therefore to show only recommended actions.

To show only recommended actions, click **Show All** and select **Recommended Actions** as shown in Figure 6-34 on page 190.

FlashSystem_840 > Monitoring > Events IBM FlashSystem 840 superuser (Security Administrator)

Recommended Action:
 Error 1094 : Node ambient temperature threshold exceeded [Run Fix](#)

Showing 7 events | Selecting 0 events

Error Code	Last Time Stamp	Recommended Actions	Object Type	Object ID	Object Name
1094	10/10/13 10:58:46 AM	Unfixed Messages and Alerts	drive	1	
1094	10/10/13 10:52:36 AM	Show All	drive	2	
	10/9/13 3:18:46 PM	Message Drive formatted	drive	2	
	10/9/13 3:18:41 PM	Message Drive formatted	drive	3	
	10/9/13 3:18:41 PM	Message Drive formatted	drive	1	
	10/9/13 3:18:41 PM	Message Drive formatted	drive	0	
	10/9/13 2:40:39 PM	Message Node added	node	2	node2

3,144 of 3,789 GB (83%) Throughput: 67.16 MBps / 131173 IOPS Health Status

Figure 6-34 Events: Show all or only recommended actions

Figure 6-35 on page 191 shows the resulting error codes. We have two problems in our system that need attention and a fix. See Figure 6-39 on page 193 for an example of the window that displays the details that are associated with the specific event ID.

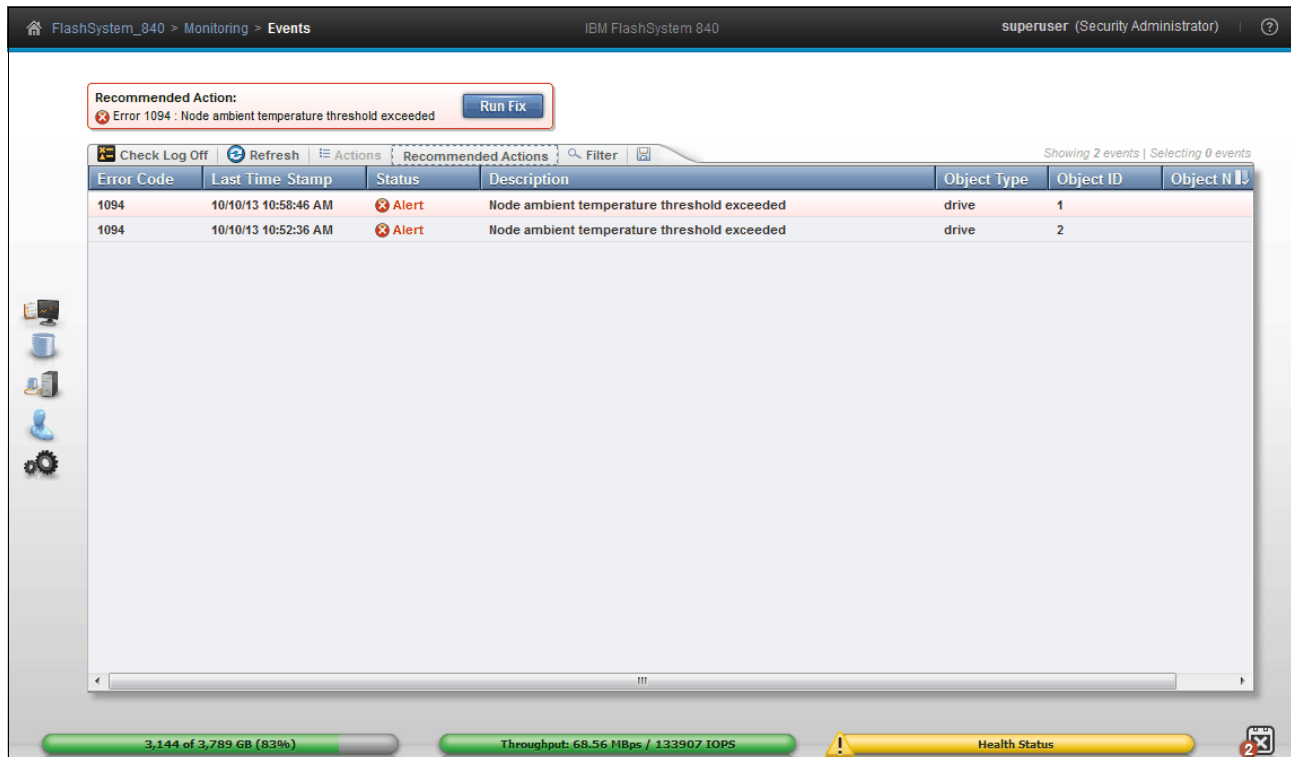


Figure 6-35 Events: Only showing recommended actions

Directed Maintenance Procedures (DMP)

There are different ways to discover that your system needs attention in a warning or error situation. If Call Home is configured on your system, which is advised, IBM Support is notified directly from the system and the system administrators are contacted by IBM for corrective actions.

The system administrator might also be in the list of email recipients and therefore is notified directly and immediately from the system as soon as an alert is sent.

For more information about how to configure Call Home, see Chapter 4, “Installation and configuration” on page 65 and “Email” on page 241

Another way of getting alert notifications is through Simple Network Management Protocol (SNMP) alerts.

For more information about how to configure SNMP alerts, see “SNMP” on page 242.

When the system administrator logs on to the GUI of the FlashSystem 840, the message “Status Alerts” shows in the lower-right corner. Hovering over the Status Alerts X icon shows the unresolved alerts as shown in Figure 6-36 on page 192.

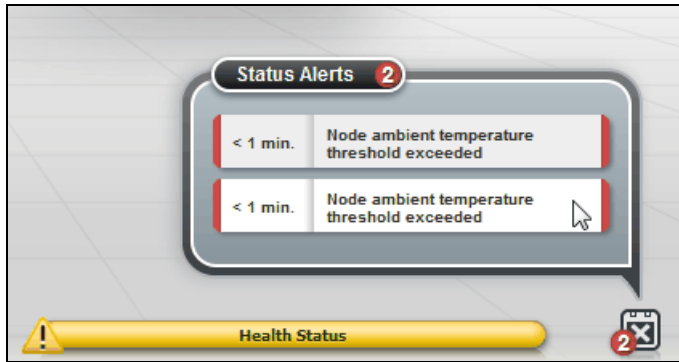


Figure 6-36 Status alerts are displayed

Status Alerts messages are also visible on the graphic of the FlashSystem 840 storage from the Home window as a red X icon on the right side of the system. Figure 6-37 shows a red X where Enclosure 1 indicates an error condition.

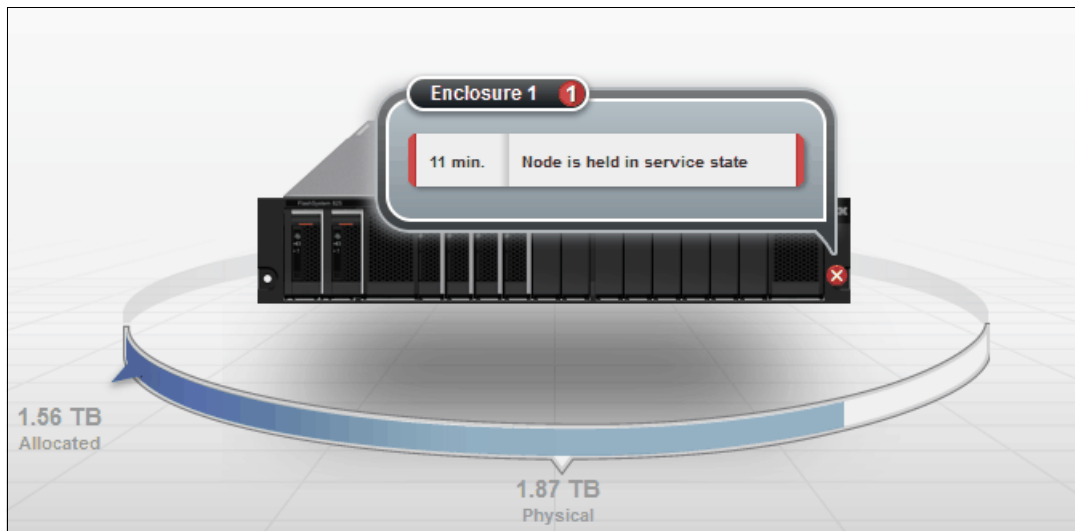


Figure 6-37 Status Alerts view displays on the FlashSystem 840 graphic

The Status Alerts in the lower-right corner of the Home window are similar to the Status Alerts X icon in the lower-right corner of the IBM FlashSystem 840 depiction. Hovering over the alerts X icon on the graphic of the system shows any unresolved alerts (Figure 6-37).

Note: These examples show different error and warning conditions.

Clicking any of the Status Alerts displayed takes you to the Monitoring → Events menu where details about the events can be reviewed. The Recommended Actions box displays and indicates that unresolved errors are in the log and need your attention. This method to fix errors is also referred to as the *Directed Maintenance Procedures (DMP)*.

To view the details of a specific event, highlight the event and click **Properties** as shown in Figure 6-38 on page 193.

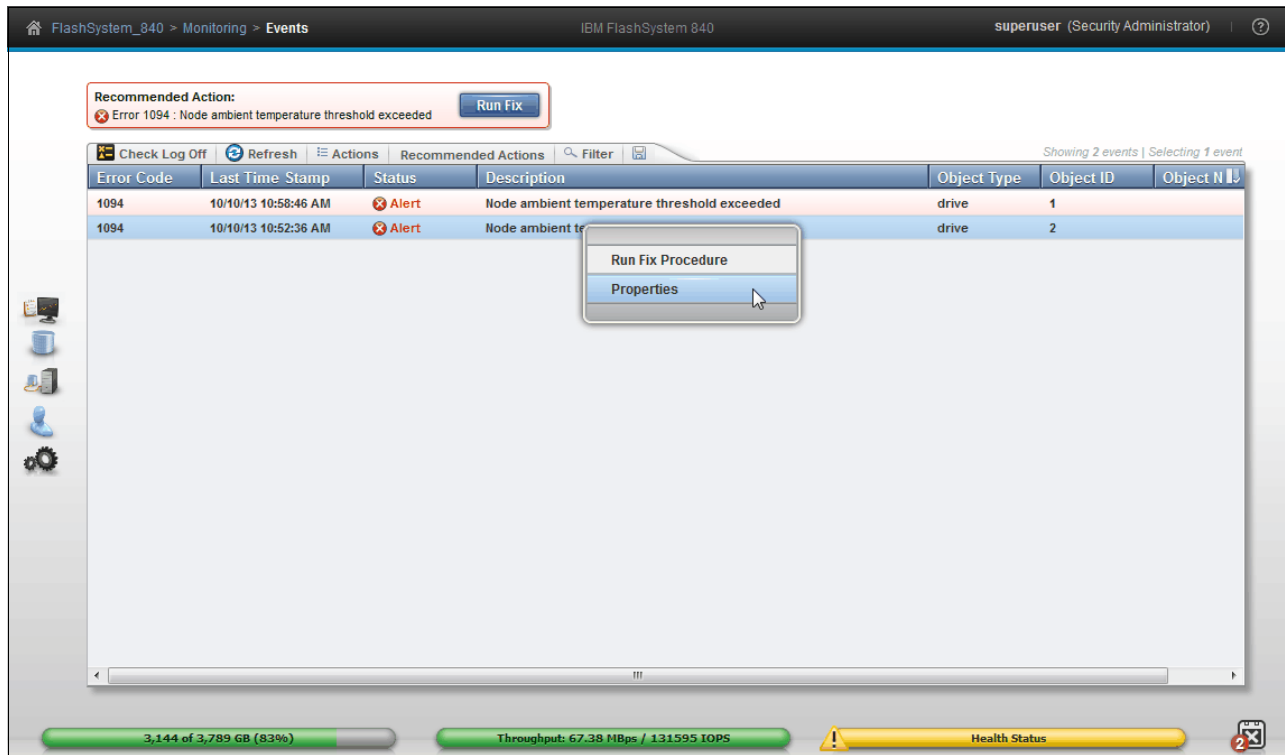


Figure 6-38 Event Properties

The Properties window now opens as shown in Figure 6-39 and you can review the details of the error.

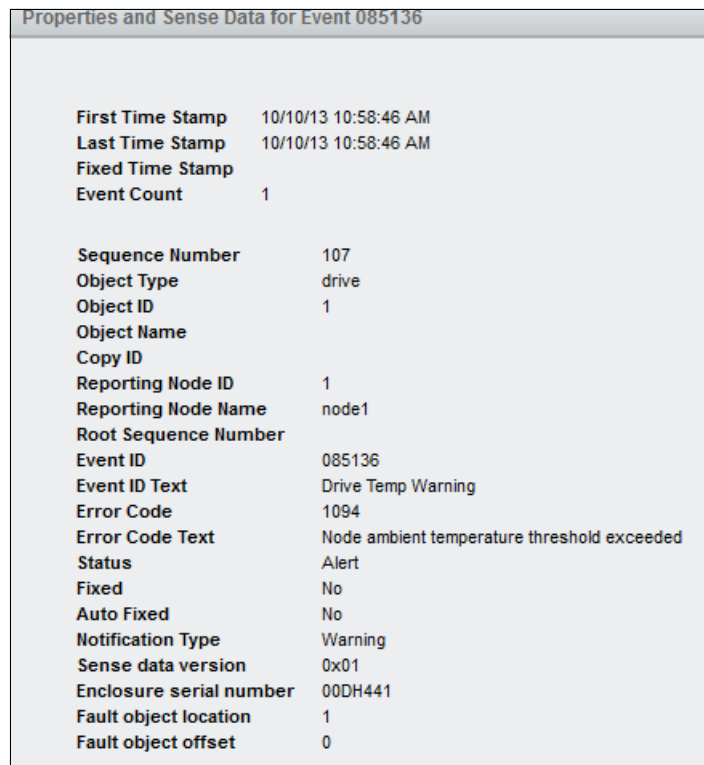


Figure 6-39 Properties for an event

In Figure 6-40, we demonstrate how an error is fixed. Before starting the fix procedure, we filter events to Recommended Actions so that only errors that require attention are displayed. Click **Run Fix** to initiate the DMP procedure.

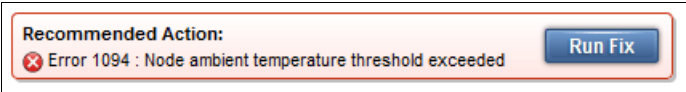


Figure 6-40 Events: Run Fix

The Run Fix procedure guides you through resolving the error event. The error message that we see in Figure 6-40 is caused by a room temperature that is too high, which might cause the system to overheat and eventually shut down if the error situation is not fixed.

Figure 6-41 shows the first step of the DMP procedure. The system reports that drive 2 (flash module 2) in slot 5 is measuring a temperature that is too high. The window also reports that all four fans in both canisters are operational and online.

Node ambient temperature threshold exceeded

Ambient temperature is greater than or equal to the warning threshold

Ambient temperature of the following component exceeds or is equal to the warning threshold.

The drive **2** that is located in slot **5** of enclosure **1** with serial number **00DH441**

Status of fans in the enclosure 1

Canister ID	Fan1 Status	Fan2 Status
1	online	online
2	online	online

Ensure that the status of all fans is online.

Run the corresponding fix procedure for any fan that is not online.

Figure 6-41 DMP procedure step 1

The next step in the DMP procedure is for the administrator to measure the room temperature and to make sure that the ambient temperature is within the specifications for the system. The instructions for this next step are shown in Figure 6-42.

Node ambient temperature threshold exceeded

Ambient Temperature is greater than or equal to the warning threshold

Measure the ambient room temperature close to the following component.

The drive **2** that is located in slot **5** of enclosure **1** with serial number **00DH441**

☐ Is the room temperature currently within the operating threshold?

Click **Next** for more information.

Figure 6-42 DMP procedure step 2

In the third step of the DMP procedure, suggestions about potential causes of overheating are provided. Overheating might be caused by blocked air vents, incorrectly mounted blank carriers in a flash module slot, or a room temperature that is too high. Instructions are displayed as shown in Figure 6-43 on page 195.

Node ambient temperature threshold exceeded
<p>Check for air flow blockages</p> <p>Ambient temperature of the following component exceeds or equal to the warning threshold.</p> <p>The drive 2 that is located in slot 5 of enclosure 1 with serial number 00DH441</p> <p>Take steps to ensure the environmental requirements of this enclosure are being met. Ensure the following:</p> <ul style="list-style-type: none"> • Environmental temperature controls are set to provide the recommended ambient operating temperature. • Enclosure vents are kept free of dust and debris. • Air flow in the vicinity of the enclosure is not impeded, for example by cables, equipment, doors, walls or other obstructions in the room. • A blank carrier is installed in each drive slot that does not contain a drive. <p>If you make any corrections, allow some time for the node to cool down before continuing this fix procedure.</p> <p>Click Next for more information.</p>

Figure 6-43 DMP procedure step 3

In this step, the DMP procedure checks whether the error condition is resolved, and all events of the same type are marked as fixed, if possible. The final step is shown in Figure 6-44.

Node ambient temperature threshold exceeded
<p>Event has been marked as fixed</p> <p>The ambient temperature events have been marked as fixed.</p> <p>Click Close to exit.</p>

Figure 6-44 DMP procedure step 4

The events indicating an error condition relating to temperature are now gone and the system is back in a Healthy state as shown in Figure 6-45 on page 196.

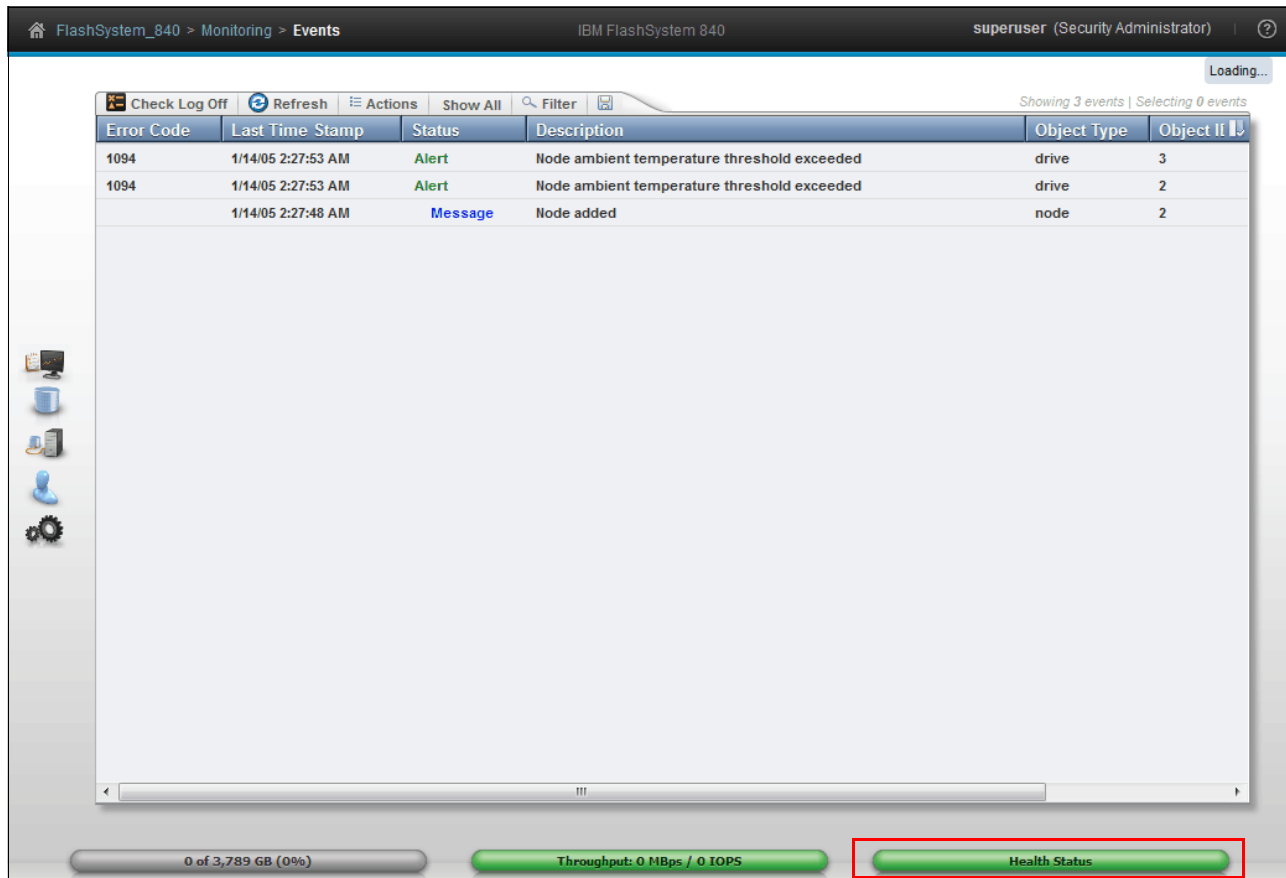


Figure 6-45 Error condition is now resolved

For more information about operational specifications for the IBM FlashSystem 840, see the following URL:

<http://www.ibm.com/systems/storage/flash>

6.2.3 Monitoring performance menu

The IBM FlashSystem 840 Performance menu gives you a good overview of how the system is performing. In the latest firmware release 1.1.3.2, the performance monitor has changed for enhanced functionality. With previous firmware releases the performance graphs represented 5 minutes of data. With the current firmware release, the default performance monitor represents a default 10 minutes of captured data and the view can be expanded for showing up to 300 days.

Performance menu overview

You enter the FlashSystem 840 performance monitor by clicking **Monitoring** → **Performance**. The first time after the browser window is opened, system latency is displayed as shown in Figure 6-46 on page 197.

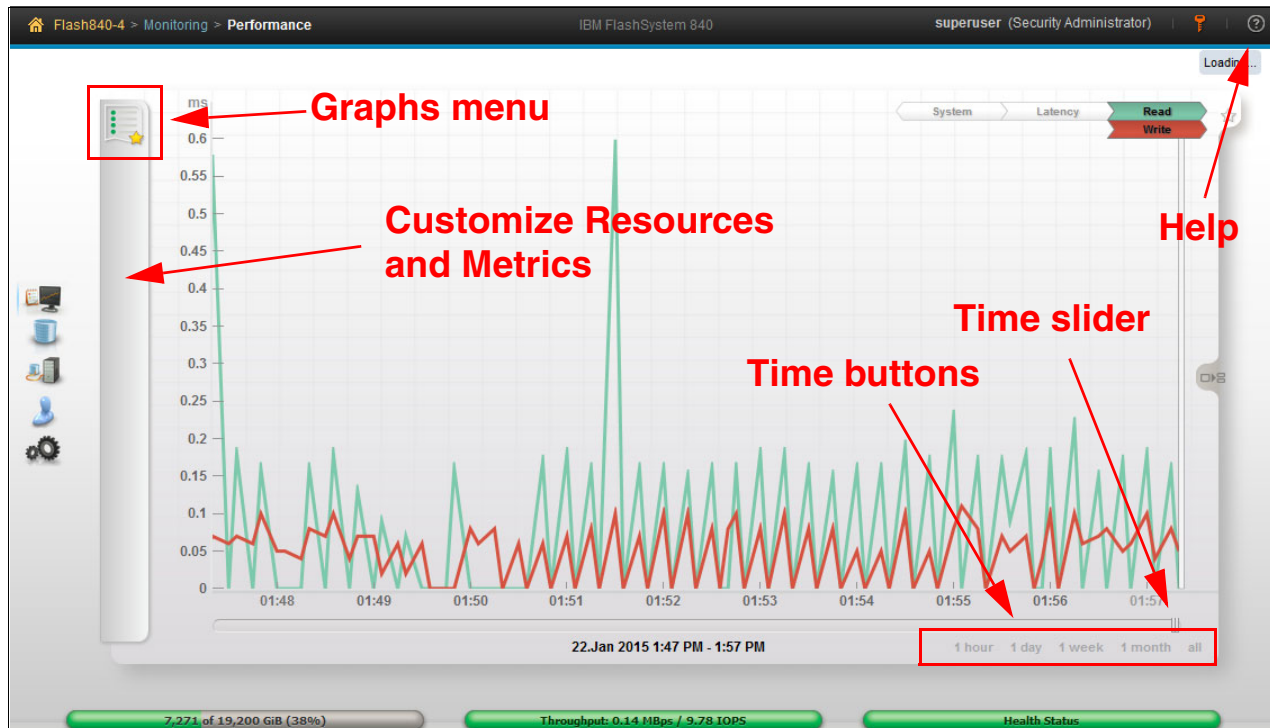


Figure 6-46 Performance monitoring default window

The horizontal part of the graph displays the timeline. You can slide the timeline to view the past. You can also adjust the granularity of the timeline by selecting one hour, one day, one week, one month, or all. All displays the year to date.

Five performance charts can be reviewed from the graphs menu:

- System I/O:

The System I/O graph displays the average number of read, write, and total I/O requests per second (IOPS) over the sample period. Each request type (read, write, and total) is represented by a different color line.

- System Latency

The System Latency graph displays the average amount of time in milliseconds (ms) each read and write I/O request takes over the given sampling period. Each request type (read and write) is represented by a different color line.

- System Bandwidth

The System Bandwidth graph displays the average number of megabytes per second (MBps) of read, write, total, and rebuild bandwidth over the sample period. Each bandwidth type (read, write, total, and rebuild) is represented by a different color line. There is one line graph for each system that is selected.

- Interface Port Total IOPS

The Port Total IOPS graph displays the average number of read, write, and total IOPS over the sample period. There is one line on the graph for each port in each host adapter in each canister. Each adapter has a different color, and all four ports on an adapter have the same color.

- Interface Port Total Queue Depth

The Port Total Queue Depth graph displays the average number of operations of that type over the sample period. There is one line on the graph for each port in each host adapter in each canister. Each adapter has a different color, and all four ports on an adapter have the same color.

Graphs menu

The Graphs menu has five default graphs defined. By clicking the icon of the graphs menu, the graphs can be selected as shown in Figure 6-47.

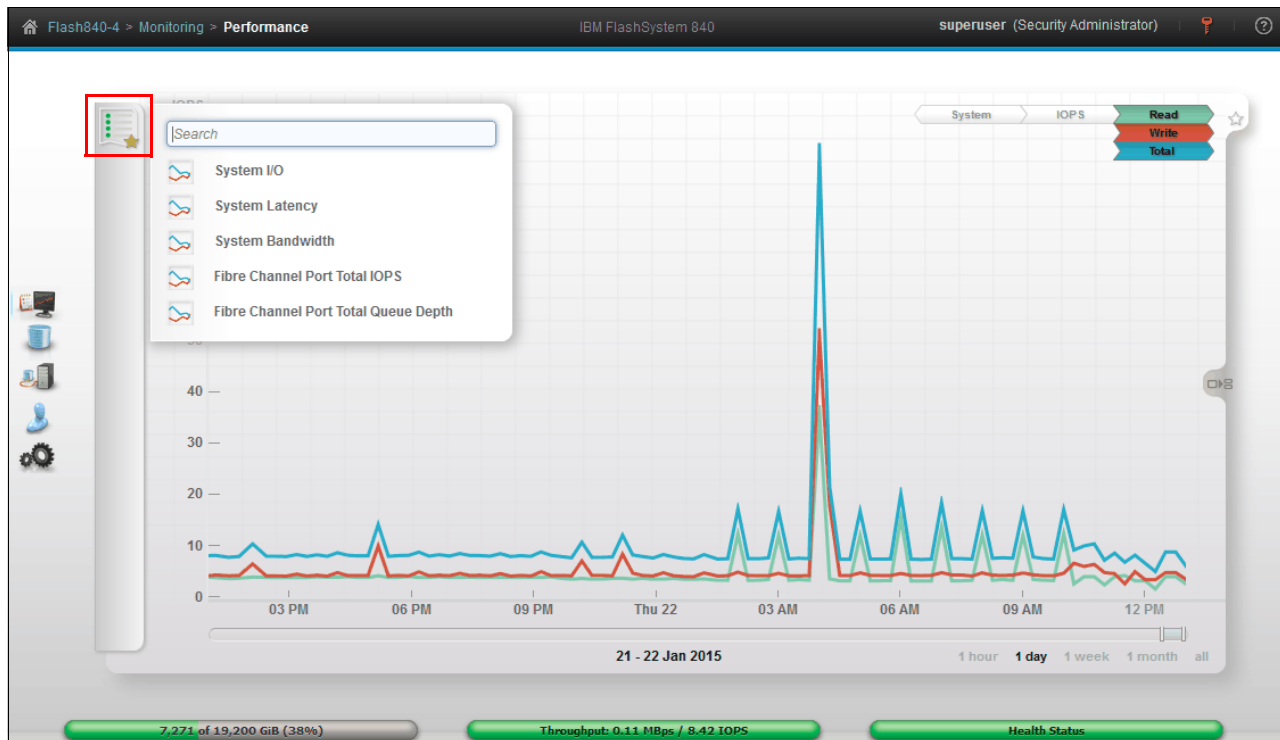


Figure 6-47 Graphs menu default window

By default the performance monitor shows System Latency. Graphs can be customized and added to the menu as we discuss in “Customize graphs menu” on page 201.

The graphs menu shows different resources for the ports depending on the FlashSystem 840 model being either InfiniBand, iSCSI, or Fibre Channel as shown in Figure 6-48 on page 199.

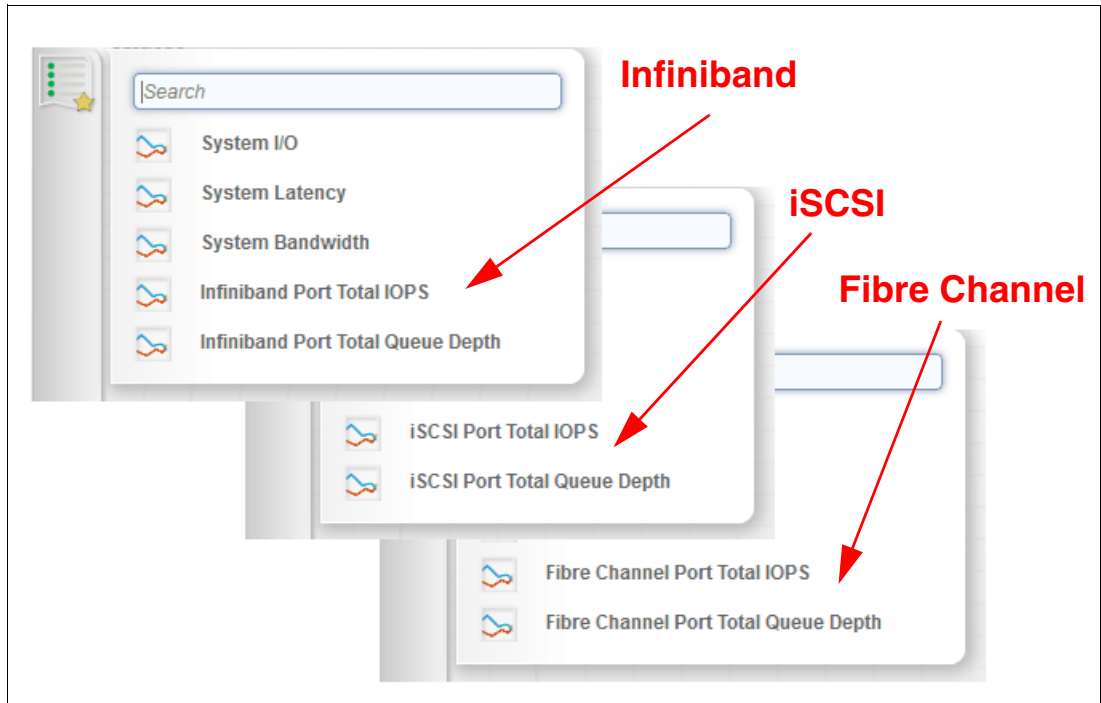


Figure 6-48 Graphs menu of three different models of FlashSystem 840

In Figure 6-49, we select **System I/O** from the graphs menu. The performance monitor now displays 10 minutes of IOPS as shown in Figure 6-49. Any time of interest can be selected by pulling the time slider.

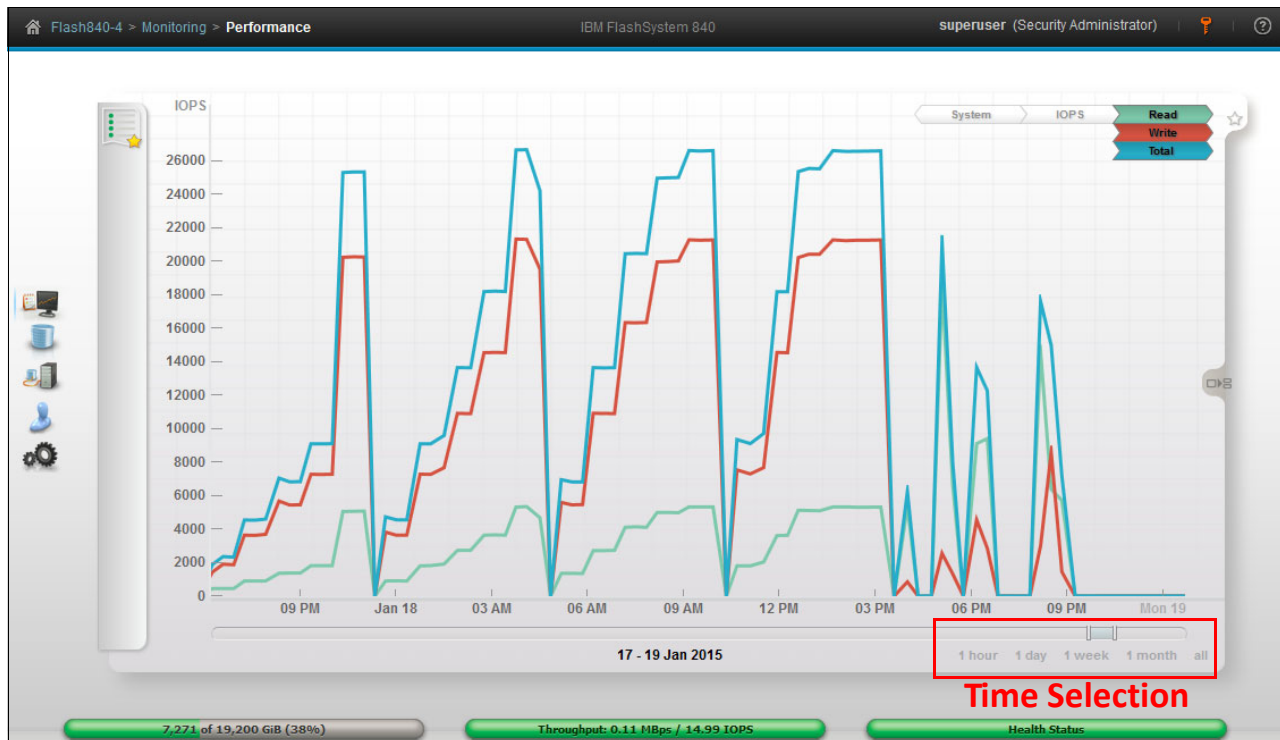


Figure 6-49 Performance graphs showing IOPS at a specific time

Storage administrators might want to know if there were long response times from storage (latency) at a given time. If the need is to compare two metrics, for example IOPS and latency, this can be done by clicking the **Click to view two charts** icon in the right side of the graphs window as shown in Figure 6-50.



Figure 6-50 Click to view two charts

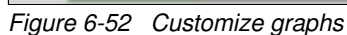
Then, select the second graph from the graphs menu. You can point to either graph (the upper one or the lower one, whichever you prefer).

Figure 6-51 shows two graphs, IOPS and latency, in the Performance window.



Figure 6-51 Two graphs are displayed

The graphs can be customized by clicking the **Customize Resources and Metrics** bar in the left side of the window. Next, resources are selected from the upper part of the window, and metrics are selected in the lower part of the window as shown in Figure 6-52.



When all resources and metrics have been selected, the blue arrow in the lower-right corner can be clicked and the resulting graphs are displayed as shown in Figure 6-53 on page 202.



Figure 6-53 Customized graphs are displayed

Figure 6-54 shows how we clicked the **Add to favorites** icon where after the two graphs shows as favorites on the graphs menu.

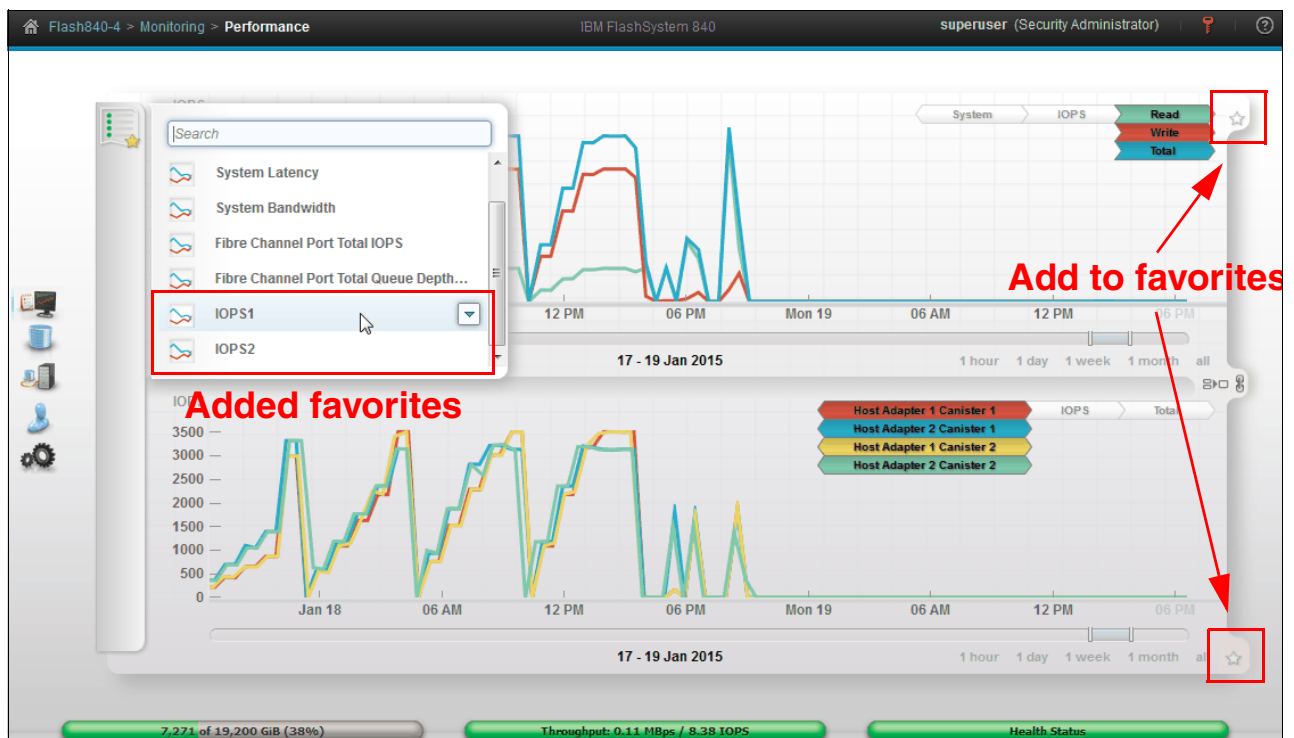


Figure 6-54 Graphs added to favorites

The graphs from the graphs menu can additionally be pinned to the toolbar where they appear as icons. The customized graphs can also be selected as the default graphs for the FlashSystem 840 performance monitor as shown in Figure 6-55.

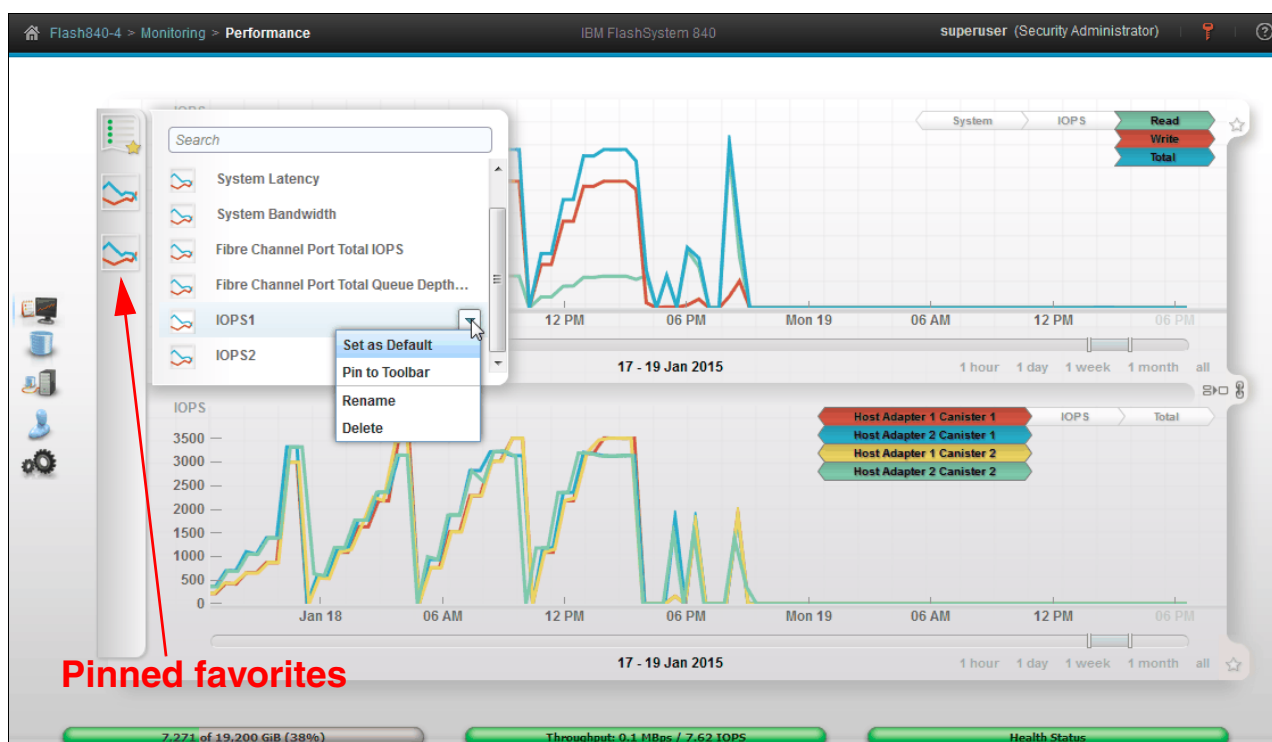


Figure 6-55 Graphs pinned to toolbar

Host port adapter numbering

Figure 6-56 shows the numbering and naming of the IBM FlashSystem 840 I/O interface ports as they correspond to the performance monitor charts for Interface Queue Depth.

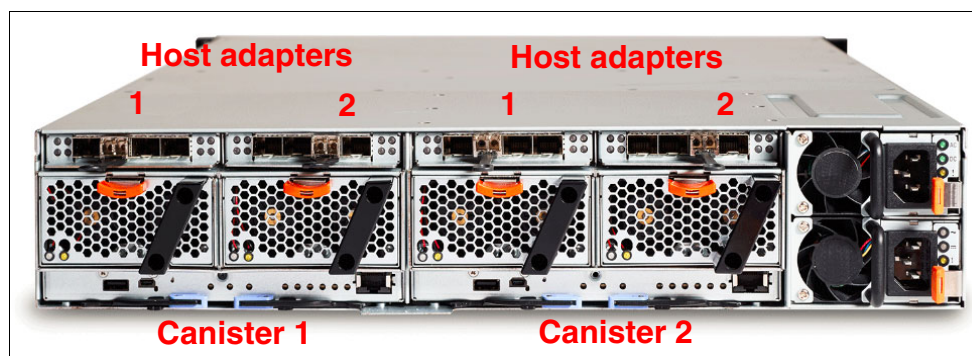


Figure 6-56 Rear-side numbering of interface cards

IBM Spectrum Control

If you need more performance monitoring, the optimal tool is IBM Spectrum Control, which delivers the functionality of IBM Tivoli Storage Productivity Center. You can manage performance and connectivity from the host file system to the physical disk, including in-depth performance monitoring and analysis of the storage area network (SAN) fabric.

IBM data management and storage management solutions deliver the functions of IBM Spectrum Control, a member of the IBM Spectrum Storage family.

For more information about IBM Spectrum Control, see this website:

<http://www.ibm.com/software/tivoli/csi/cloud-storage>

Note: For current releases of the IBM FlashSystem 840, IBM Spectrum Control does not support the product directly. The exception is if the FlashSystem 840 functions as an MDisk for the IBM SAN Volume Controller, in which case, IBM Spectrum Control supports the product through the SAN Volume Controller.

SAN Volume Controller delivers the functions of IBM Spectrum Virtualize, part of the IBM Spectrum Storage family.

6.3 Volumes

This topic provides information about managing volumes.

You can use the FlashSystem 840 GUI or CLI **svctask mkvdisk** command to create a volume. After volumes are created, they can be mapped to a host by using the **mkvdiskhostmap** command.

The volumes are built from extents in the RAID 5 or RAID 0 flash module arrays, and the volumes are presented to hosts as logical units that the host sees as external disks. We describe the Volumes menu and its options.

The Volumes menu has two options:

- ▶ Volumes
- ▶ Volumes by Host

6.3.1 Navigating to the Volumes menu

When you hover the cursor over the **Volumes** function icon, the menu shown in Figure 6-57 on page 205 opens.



Figure 6-57 Navigate to the Volumes menu

6.3.2 Volumes menu

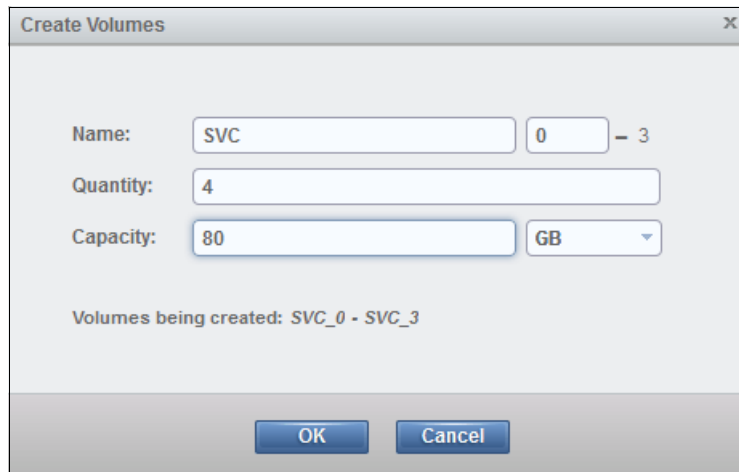
Click **Volumes** to open the window shown in Figure 6-58. You can perform tasks on the volumes, such as create, expand, rename, and delete, or you can review the properties of the volume.

Name	State	Capacity	Host Mappings	Volume Unique Identifier
AIX_0	✓ Online	50.00 GB	1	6005076aa18d882aa000000004000009
AIX_1	✓ Online	50.00 GB	1	6005076aa18d882aa00000000500000a
AIX_2	✓ Online	50.00 GB	1	6005076aa18d882aa00000000600000b
AIX_3	✓ Online	50.00 GB	1	6005076aa18d882aa00000000700000c
WIN2008_0	✓ Online	300.00 GB	1	6005076aa18d882aa0000000000000005
WIN2008_1	✓ Online	300.00 GB	1	6005076aa18d882aa000000001000006
WIN2008_2	✓ Online	300.00 GB	1	6005076aa18d882aa000000002000007
WIN2008_3	✓ Online	300.00 GB	1	6005076aa18d882aa000000003000008

Figure 6-58 Volumes window that shows all volumes

Create a volume

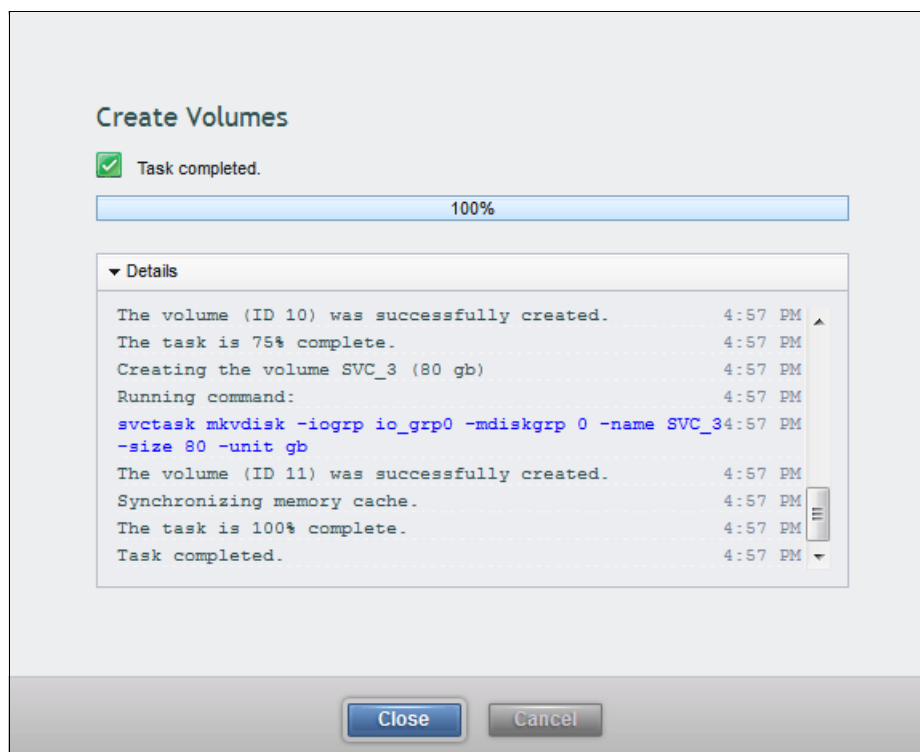
To create a volume using the GUI from the Volumes menu, click **Create Volume**. The Create Volumes window opens as shown in Figure 6-59 on page 206.



The 'Create Volumes' dialog box has a title bar with a close button. It contains three input fields: 'Name' with 'SVC' and a range '0 - 3', 'Quantity' with '4', and 'Capacity' with '80' and a unit dropdown set to 'GB'. Below these fields, it says 'Volumes being created: SVC_0 - SVC_3'. At the bottom are 'OK' and 'Cancel' buttons.

Figure 6-59 Create Volumes window

We type the volume name SVC, the requested capacity (80 GB) and a quantity of 4 volumes. We then click **OK** and the task window opens as shown in Figure 6-60.



The 'Create Volumes' task window shows a green checkmark and 'Task completed.' with a progress bar at 100%. Below is a 'Details' section with a scrollable log of events and timestamps (all 4:57 PM):

- The volume (ID 10) was successfully created.
- The task is 75% complete.
- Creating the volume SVC_3 (80 gb)
- Running command:
- `svctask mkvdisk -iogrp io_grp0 -mdiskgrp 0 -name SVC_3 -size 80 -unit gb`
- The volume (ID 11) was successfully created.
- Synchronizing memory cache.
- The task is 100% complete.
- Task completed.

At the bottom are 'Close' and 'Cancel' buttons.

Figure 6-60 Create Volumes task window

The Create Volume wizard now creates four volumes of 80 GB each. The resulting volumes can be reviewed on the Volumes menu as shown in Figure 6-61 on page 207.

Name	State	Capacity	Host Mappings	Volume Unique Identifier
AIX_0	Online	50.00 GB	1	6005076aa18d882aa000000004000009
AIX_1	Online	50.00 GB	1	6005076aa18d882aa00000000500000a
AIX_2	Online	50.00 GB	1	6005076aa18d882aa00000000600000b
AIX_3	Online	50.00 GB	1	6005076aa18d882aa00000000700000c
SVC_0	Online	80.00 GB	No Host Mappings	6005076aa18d882aa0000000080000011
SVC_1	Online	80.00 GB	No Host Mappings	6005076aa18d882aa0000000090000012
SVC_2	Online	80.00 GB	No Host Mappings	6005076aa18d882aa00000000a0000013
SVC_3	Online	80.00 GB	No Host Mappings	6005076aa18d882aa00000000b0000014
WIN2008_0	Online	300.00 GB	1	6005076aa18d882aa0000000000000005
WIN2008_1	Online	300.00 GB	1	6005076aa18d882aa0000000001000006
WIN2008_2	Online	300.00 GB	1	6005076aa18d882aa0000000002000007
WIN2008_3	Online	300.00 GB	1	6005076aa18d882aa0000000003000008

Figure 6-61 Four SAN Volume Controller volumes created

The newly created volumes have no host mappings at the time of their creation. Host mapping can be performed from the Volumes → Volumes by Host window. For the instructions to map volumes to a host, see “Mapping volumes” on page 211.

Create volumes by using the CLI

The CLI can be used for creating volumes. CLI commands execute faster than GUI commands, and administrators might prefer using the CLI.

Example 6-2 shows an example of creating a volume by using the CLI. More or fewer parameters can be applied to the **mkvdisk** command. The example shown in Example 6-2 specifies the minimum that is required.

Example 6-2 Create a volume by using the CLI

```
IBM_Flashsystem:FlashSystem_840:superuser>mkvdisk -size 15 -unit gb -name SVC_4
Virtual Disk, id [7], successfully created
```

```
IBM_Flashsystem:FlashSystem_840:superuser>lsvdisk
id name      IO_group_name status capacity vdisk_UID
0 WIN2008_1 io_grp0      online 40.00GB 0020c24000000000
1 WIN2008_2 io_grp0      online 50.00GB 0020c24001000000
2 WIN2008_3 io_grp0      online 50.00GB 0020c24002000000
3 WIN2008_4 io_grp0      online 39.99GB 0020c24003000000
4 SVC_1      io_grp0      online 15.00GB 0020c24004000000
5 SVC_2      io_grp0      online 15.00GB 0020c24005000000
6 SVC_3      io_grp0      online 15.00GB 0020c24006000000
7 SVC_4     io_grp0     online 15.00GB 0020c24007000000
```

```
IBM_Flashsystem:FlashSystem_840:superuser>
```

Performing actions on volumes

From the Volumes window, you can perform various actions on the volumes. Click **Actions** to access these operations, or you can right-click the volume name, which opens a list of operations that can be performed against the volume (Figure 6-62 on page 208).

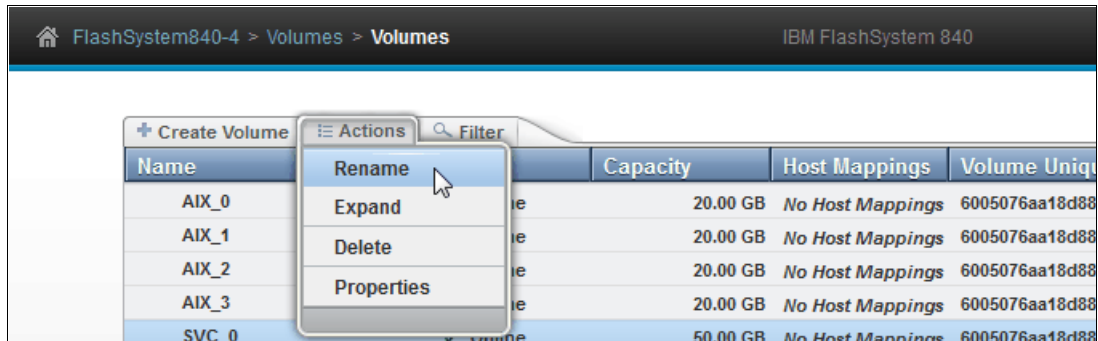


Figure 6-62 Actions of a single volume

Figure 6-63 shows the properties of a volume that indicate the volume name, its capacity, and its sector size. Each volume has a unique ID (UID), which can be discovered from the host side as a property of the logical unit. The volume is not mapped to a host.

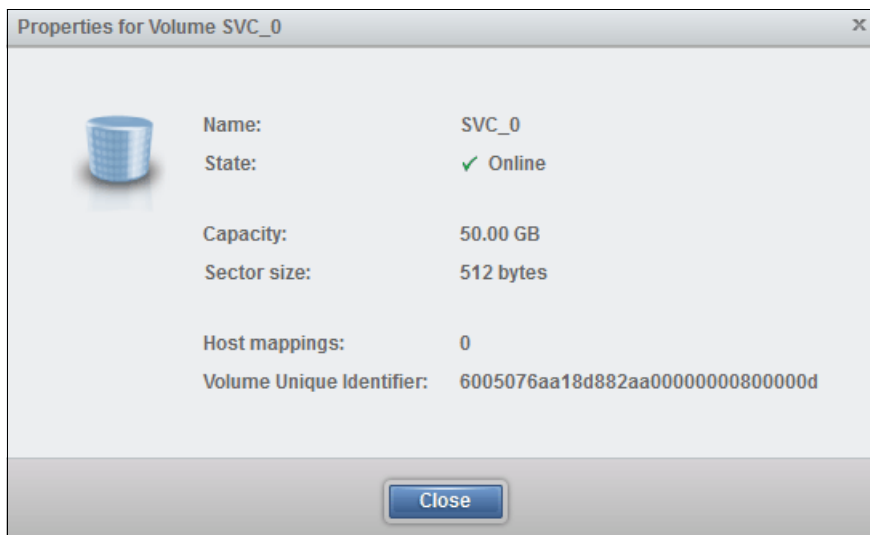


Figure 6-63 Properties of a volume

A volume can be expanded while it is online, therefore maintaining full functionality to the connected hosts. Not all operating systems, however, allow concurrent expansion of their disks so precaution must be taken that the operating system supports it. An alternative to expanding the disk is to create and map a new disk for the host.

Expanding a volume that is mapped to an AIX host

When more than one volume is selected, the actions available for the volumes are reduced to only Expand and Delete as shown in Figure 6-64 on page 209.

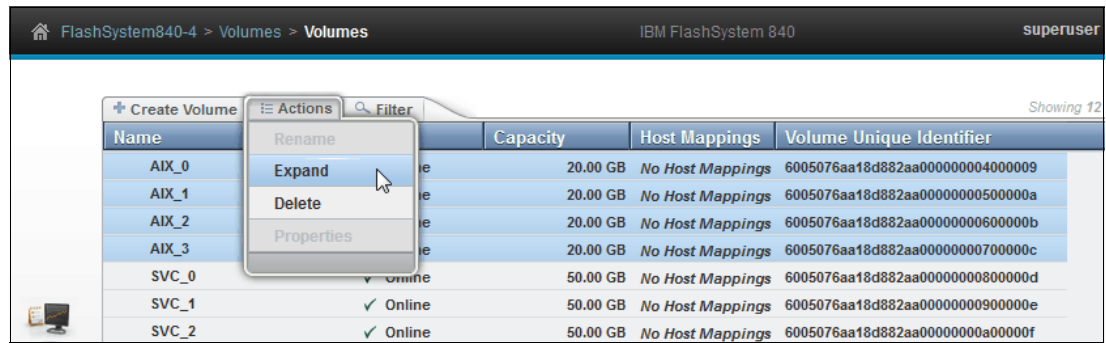


Figure 6-64 Expand four volumes

Figure 6-65 shows that we expanded each of the volumes to a size of 50 GB.

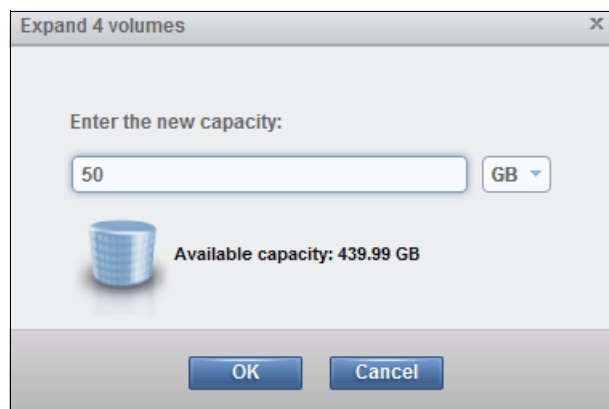


Figure 6-65 Expand to 50 GB

The resulting Volumes window displays the new capacity as shown in Figure 6-66.

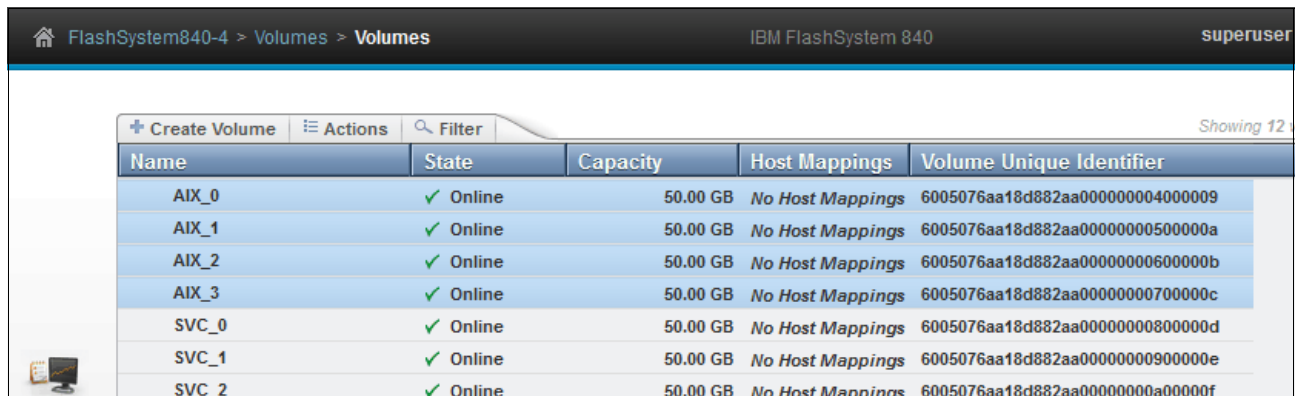


Figure 6-66 Four volumes expanded

The IBM FlashSystem 840 supports the ability to dynamically expand the size of a virtual disk (VDisk) if the AIX host is using AIX version 5.2 or later.

Use the AIX **chvg** command options to expand the size of a physical volume that the Logical Volume Manager (LVM) uses without interruptions to the use or availability of the system. For more information, see *AIX 7.1 Operating System and Device Management* at the following

web page. At this page, search for the operating system and device management:

<http://ibm.co/1cv2Ac1>

Expanding a volume that is mapped to a Microsoft Windows host

You can use the GUI and the CLI to dynamically expand the size of a volume that is mapped to a Microsoft Windows host.

After expanding the volume, using the same procedure as shown in the previous examples (Figure 6-64 on page 209 and Figure 6-65 on page 209) for AIX, start the Computer Management application and open the Disk Management window under the Storage branch.

If the Computer Management application was open before you expanded the volume, use the Computer Management application to issue a **rescan** command. You will see the volume that you expanded now has deallocated space at the right side of the disk.

If the disk is a Windows basic disk, you can create a new primary or extended partition from the deallocated space.

If the disk is a Windows dynamic disk, you can use the deallocated space to create a new volume (simple, striped, or mirrored) or add it to an existing volume.

Shrinking a volume

The shrink volume option is only provided through the CLI and cannot be performed by using the GUI. Volumes can be reduced in size, if necessary. However, if the volume contains data, do not shrink the size of the disk because shrinking a volume destroys the data.

When shrinking a volume, be aware of the following considerations:

- ▶ Shrinking a volume removes capacity from the end of the volume's address space. If the volume was used by an operating system or file system, it might be hard to predict what space was used. The file system or OS might depend on the space that is removed, even if it is reporting a high amount of free capacity.
- ▶ If the volume contains data that is used, do not attempt under any circumstances to shrink a volume without first backing up your data.

You can use the CLI **shrinkvdisksize** command to shrink the physical capacity that is allocated to the particular volume by the specified amount.

The **shrinkvdisksize** command uses this syntax:

shrinkvdisksize -size capacitytoshrinkby -unit unitsforreduction vdiskname/ID

Example 6-3 shows an example of shrinking a volume. A volume is called a *VDisk* in the CLI.

Example 6-3 Shrink a volume (VDisk)

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>lsvdisk
id name      IO_group_name status capacity vdisk_UID
0  WIN2008_1 io_grp0      online 50.00GB  0020c24000000000
1  WIN2008_2 io_grp0      online 50.00GB  0020c24001000000
2  WIN2008_3 io_grp0      online 50.00GB  0020c24002000000
3  WIN2008_4 io_grp0      online 50.00GB  0020c24003000000
```

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>shrinkvdisksize -size 10 -unit gb
WIN2008_1
```

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>lsvdisk
```


id	name	IO_group_name	status	capacity	vdisk_UID
0	WIN2008_1	io_grp0	online	40.00GB	0020c24000000000
1	WIN2008_2	io_grp0	online	50.00GB	0020c24001000000
2	WIN2008_3	io_grp0	online	50.00GB	0020c24002000000
3	WIN2008_4	io_grp0	online	50.00GB	0020c24003000000

IBM_Flashsystem:Cluster_9.19.91.242:superuser>

Important: Shrinking a volume is a data destructive action. Only shrink volumes that are not in use and that do not contain data.

6.3.3 Volumes by Host menu

Clicking the **Volumes** → **Volumes by Host** option opens the window where unmapped and mapped volumes display. This window shows which hosts are created on the system and which volumes exist. If the volumes are currently unmapped, they appear as Unmapped Volumes and a plus sign (+) is visible to the left as shown in Figure 6-67.



Figure 6-67 Volumes by Host window

Mapping volumes

When you click the plus sign (+) to expand the Unmapped Volumes window, a list of unmapped volumes is provided. We have four volumes that are named SVC_1, SVC_2, SVC_3, and SVC_4 that we want to map to our host SAN Volume Controller (SVC). We then highlight all four Volumes and click **Actions** → **Map to Host** (or right-click) as shown in Figure 6-68 on page 212.

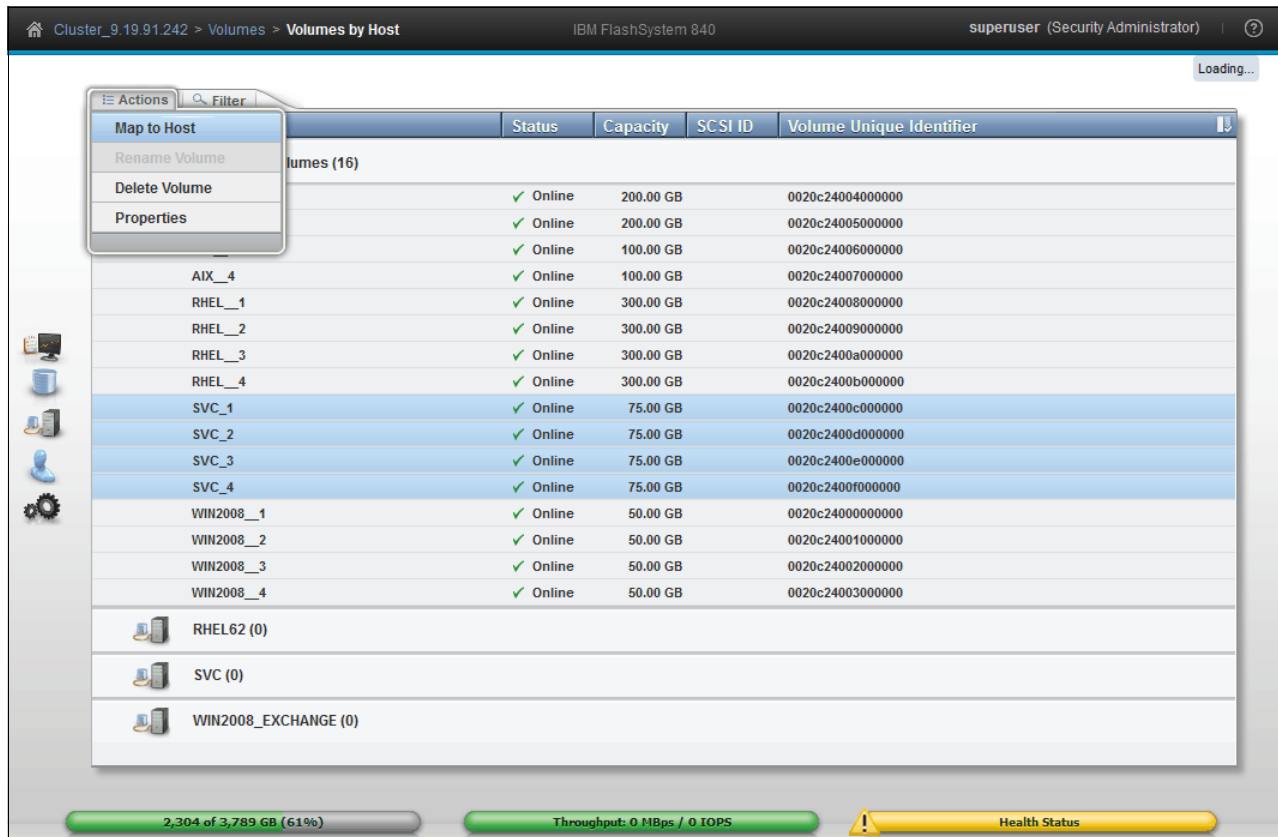


Figure 6-68 Map Volumes to Host

In the Map 4 Volumes to Host window that opens, we select the host **SVC** and click **Map** as shown in Figure 6-69.

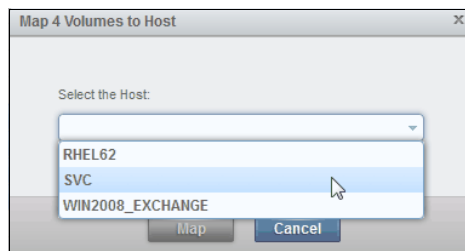


Figure 6-69 Map Volumes: Select the host SVC

Figure 6-70 on page 213 shows the Modify Mappings window where the CLI commands for mapping the volumes are executed.

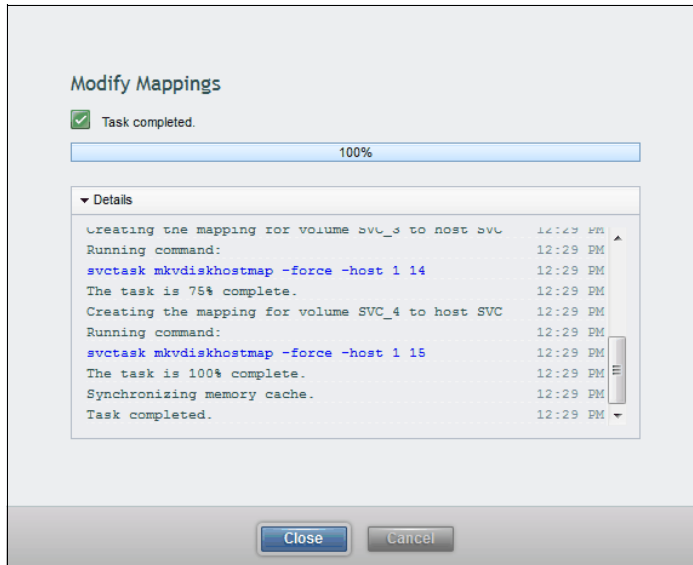


Figure 6-70 Map Volumes: CLI commands display

Figure 6-71 shows the final step in the Map 4 Volumes to Host window. The SCSI ID of the volume provided by the Map to Host action is displayed.

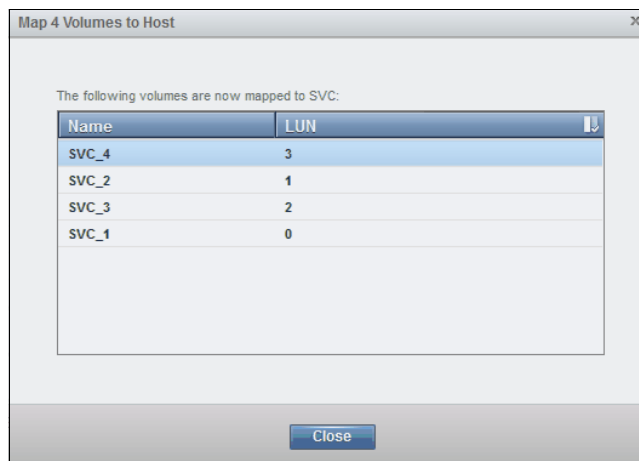


Figure 6-71 Map volumes final step

The Volumes by Host window now shows that our four SAN Volume Controller volumes are mapped and online as shown in Figure 6-72 on page 214.

The screenshot shows the 'Volumes by Host' page in the IBM FlashSystem 840 management console. The page has a breadcrumb trail: Cluster_9.19.91.242 > Volumes > Volumes by Host. The user is logged in as 'superuser (Security Administrator)'. The main content area displays a table of volumes, categorized into 'Unmapped Volumes (12)', 'RHEL62 (0)', and 'SVC (4)'. The table columns are Name, Status, Capacity, SCSI ID, and Volume Unique Identifier. All volumes are marked as 'Online'.

Name	Status	Capacity	SCSI ID	Volume Unique Identifier
Unmapped Volumes (12)				
AIX_1	✓ Online	200.00 GB		0020c24004000000
AIX_2	✓ Online	200.00 GB		0020c24005000000
AIX_3	✓ Online	100.00 GB		0020c24006000000
AIX_4	✓ Online	100.00 GB		0020c24007000000
RHEL_1	✓ Online	300.00 GB		0020c24008000000
RHEL_2	✓ Online	300.00 GB		0020c24009000000
RHEL_3	✓ Online	300.00 GB		0020c2400a000000
RHEL_4	✓ Online	300.00 GB		0020c2400b000000
WIN2008_1	✓ Online	50.00 GB		0020c24000000000
WIN2008_2	✓ Online	50.00 GB		0020c24001000000
WIN2008_3	✓ Online	50.00 GB		0020c24002000000
WIN2008_4	✓ Online	50.00 GB		0020c24003000000
RHEL62 (0)				
SVC (4)				
SVC_1	✓ Online	75.00 GB	0	0020c2400c000000
SVC_2	✓ Online	75.00 GB	1	0020c2400d000000
SVC_3	✓ Online	75.00 GB	2	0020c2400e000000
SVC_4	✓ Online	75.00 GB	3	0020c2400f000000

Figure 6-72 Volumes mapped to host SAN Volume Controller

Map a volume by using the CLI

Mapping volumes is faster using the CLI. Administrators might find it useful to perform this task by using the CLI. Volumes can be mapped by using the **svctask mkvdiskhostmap** command.

Example 6-4 shows how a volume is mapped to a host by using the CLI.

Example 6-4 Map volume by using the CLI

```
IBM_Flashsystem:FlashSystem_840:superuser>svctask mkvdiskhostmap -force -host 0 4
Virtual Disk to Host map, id [4], successfully created
```

```
IBM_Flashsystem:FlashSystem_840:superuser>lsvdiskhostmap 4
id name SCSI_id host_id host_name vdisk_UID IO_group_id IO_group_name
4 SVC_5 4 0 SVC 0020c24004000000 0 io_grp0
```

```
IBM_Flashsystem:FlashSystem_840:superuser>
```

For mapping volumes by using the CLI, we use the logical number for the host and we use the logical number for the volume. These logical numbers can be discovered by using the following commands:

- ▶ **lshost**: Shows defined hosts and their status
- ▶ **lsvdisk**: Shows defined volumes and their preferences

Unmapping volumes

When deleting a volume mapping, you are not deleting the volume itself, only the connection from the host to the volume. If you mapped a volume to a host by mistake or if you simply want to reassign the volume to another host, click **Volumes** → **Volumes by Host**. Highlight the volume or volumes that you want to unmap, right-click, and click **Unmap from Host** as shown in Figure 6-73.

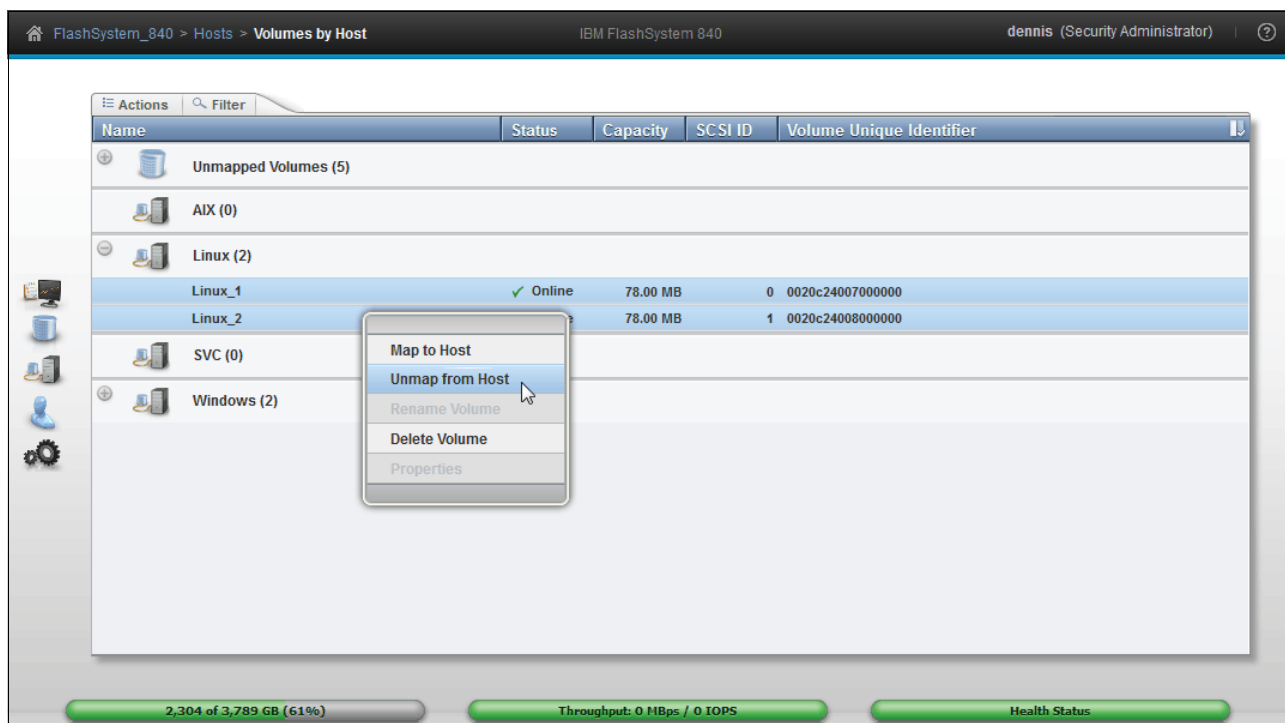


Figure 6-73 Unmap volumes from host

Next, the Unmap 2 Volumes from Host window opens to inform the administrator that the selected volumes will be unmapped as shown in Figure 6-74.

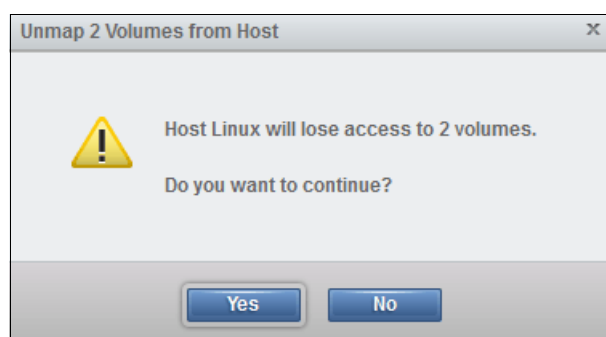


Figure 6-74 Unmapping volumes final step

By unmapping the volumes as shown in these windows, the volumes are made unavailable to the host. If data on the volumes is to be preserved, the host must unmount the disk before the volume is unmapped so that the connection to the disk is closed correctly by the host.

Note: Before unmapping a volume from a host, the host must unmount the connection to the disk or I/O errors appear.

6.4 Hosts

This topic provides information about managing hosts.

You can use the FlashSystem 840 GUI or the CLI `mkhost` command to create a logical host object. Creating a host object associates one or more host bus adapters' (HBAs') worldwide port names (WWPNs) or InfiniBand IDs with a logical host object.

You can then use the created host to map volumes (also called *virtual disks* or *VDisks*) to hosts by using the GUI or CLI `mkvdiskhostmap` command.

The Hosts menu has two options:

- ▶ Hosts
- ▶ Volumes by Host

6.4.1 Navigating to the Hosts menu

We describe the GUI Hosts menu and its options. When you hover the cursor over the Hosts function icon, a menu opens (Figure 6-75).

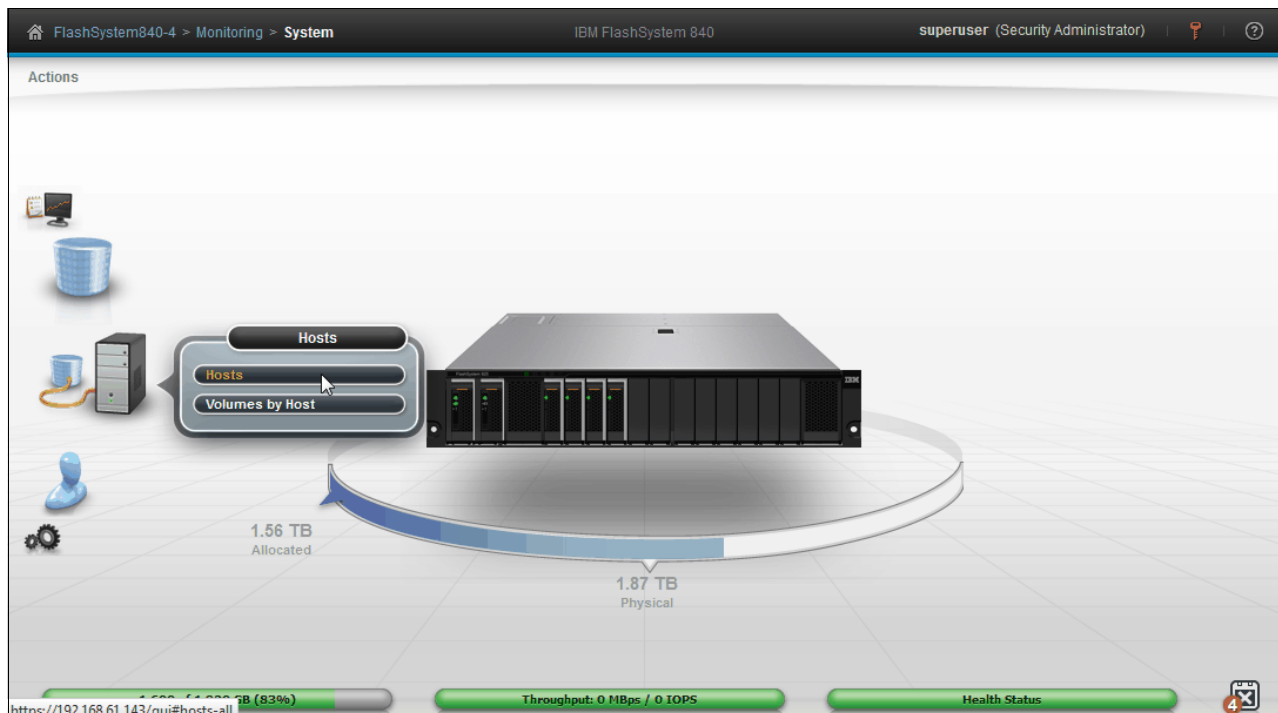


Figure 6-75 Navigate to the Hosts menu

Create host

The process of creating a host object includes specifying the host name and selecting ports for the host.

The FlashSystem 840 models are either InfiniBand, Fibre Channel (FC), Fibre Channel over Ethernet (FCoE), or iSCSI capable. However, interface cards cannot be mixed, and a single system must contain only a single type of interface card.

The FlashSystem 840 detects which type of interface cards are installed, and the Create Host wizard automatically adjusts to request the host port type for the actual model. For example, this can be the FC worldwide port name (WWPN) or the iSCSI initiator name or iSCSI qualified name (IQN).

Figure 6-76 shows the Hosts menu where the already defined hosts display.

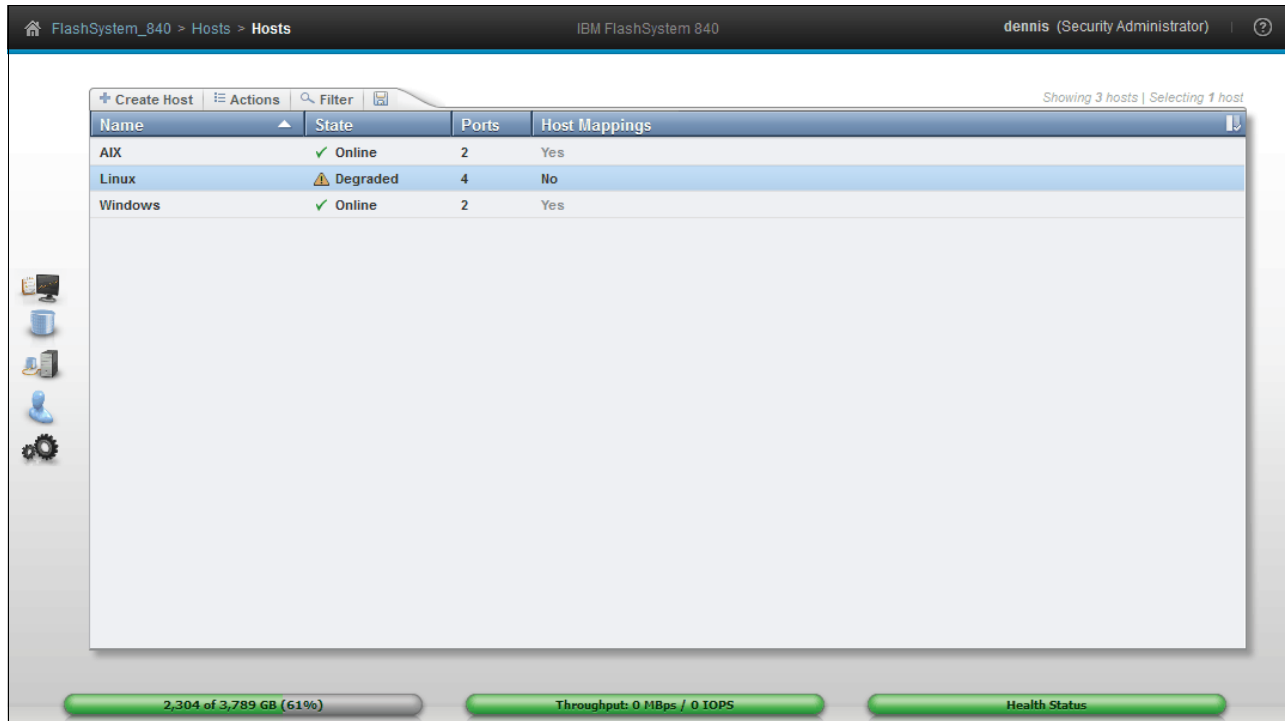


Figure 6-76 Hosts menu showing already configured hosts

In the example in Figure 6-76, our Linux host shows the Degraded state. Each connected host port *must* be zoned and connected to both canisters in the IBM FlashSystem 840. If not, the host is Degraded.

Hosts in a FlashSystem 840 configured with FC interface cards

To create a host object, click **Create Host** in the upper-left corner of the Hosts menu. The Create Host window opens as shown in Figure 6-77.

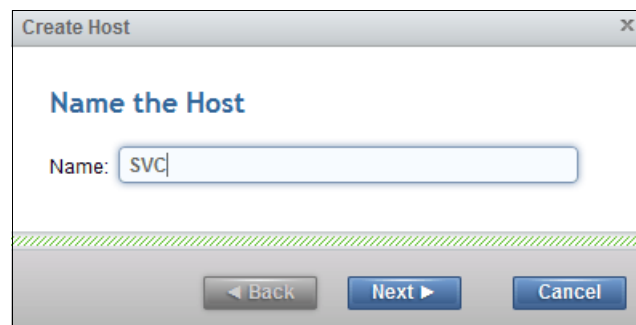


Figure 6-77 Create new host

We type the name SVC of our new host and click **Next** as shown in Figure 6-78.

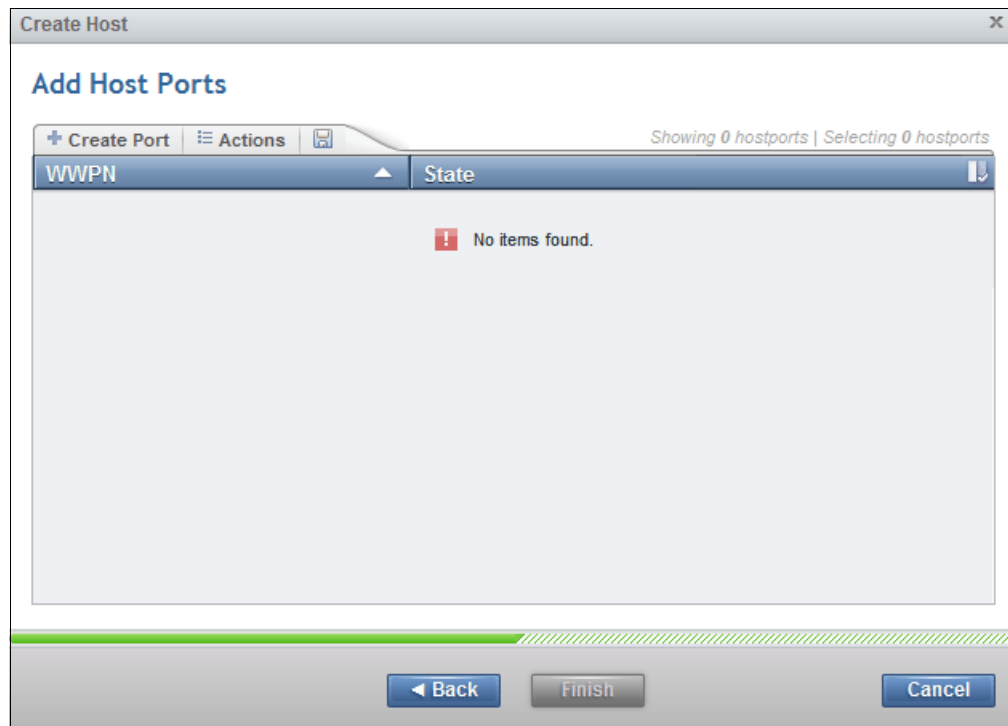


Figure 6-78 Add ports to the new host

The Add Host Ports menu displays. From this window, we add the worldwide names (WWNs) of the FC or FCoE adapters in the hosts or the InfiniBand ID if the connected host is equipped with InfiniBand adapters.

When we click **Add Port**, a new Add Host Port to Host SVC window opens as shown in Figure 6-79 on page 219.

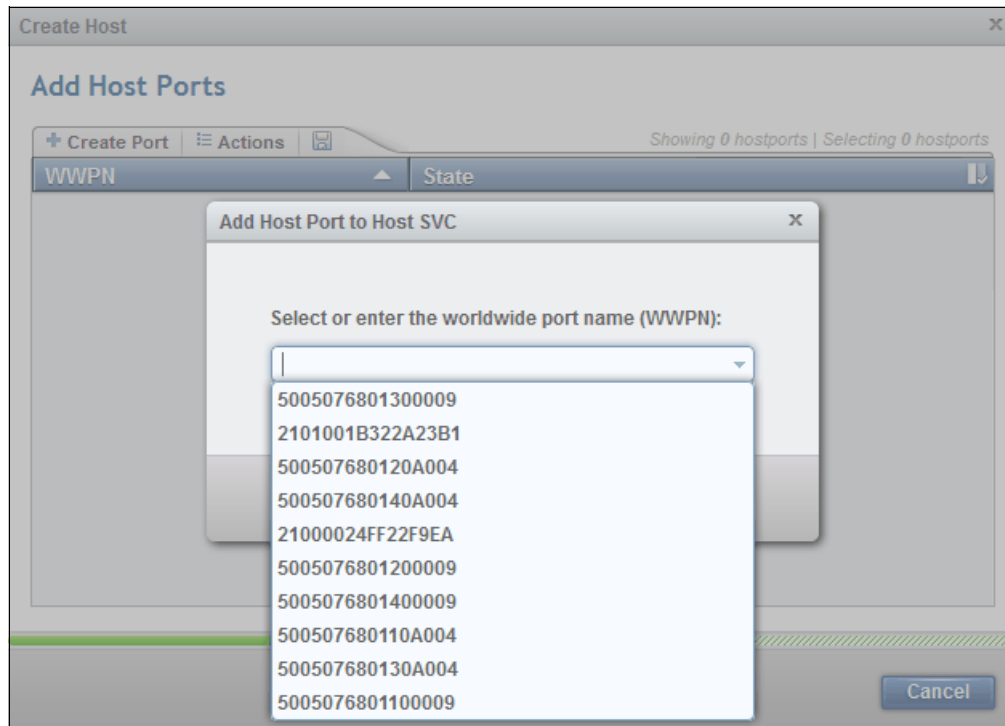


Figure 6-79 Select worldwide port names for host SAN Volume Controller

From this window, all unused WWPNs of all zoned hosts are displayed. If no WWPNs are displayed, the message “No candidate HBA ports were found” is displayed. In that case, either the WWPN is already used for another host object or the host has not been zoned correctly to the FlashSystem 840.

In our example, we are creating a SAN Volume Controller host object. Our SAN Volume Controller has eight HBA ports and each HBA port has its own WWPN. We click and highlight the ports one at a time and click **Add**. As a result, we now have a host that has eight defined ports as shown in Figure 6-80.

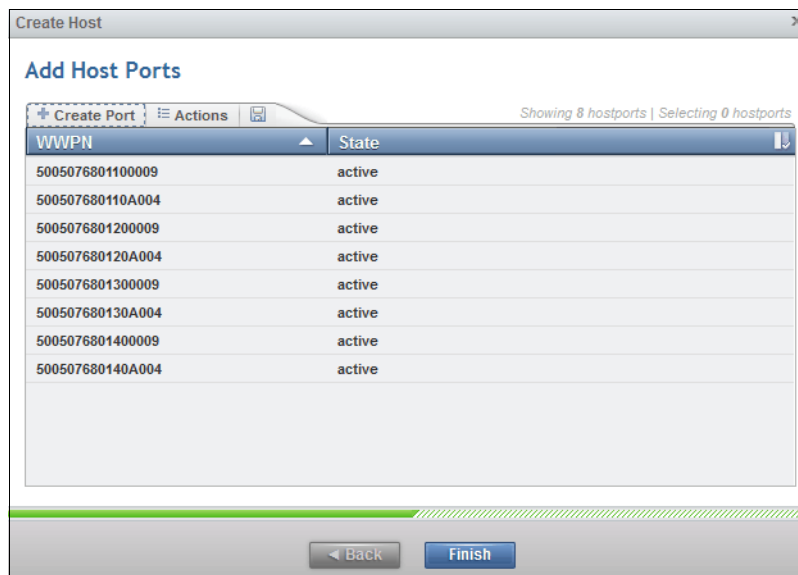


Figure 6-80 Finish creating the host SAN Volume Controller

When we click **Finish**, the host object is created. Our SAN Volume Controller host is now online and has eight ports as shown in Figure 6-81.

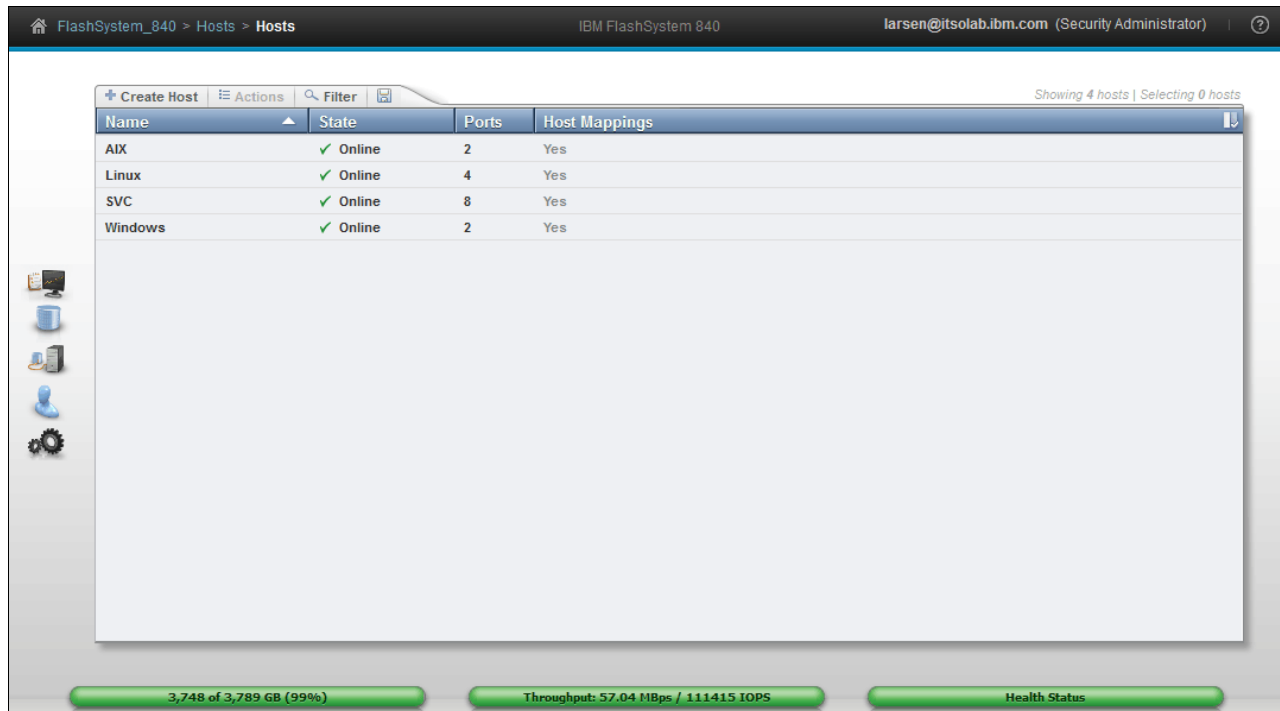


Figure 6-81 Host SAN Volume Controller is now created and online

All WWPNs on the host object are mapped to the virtual disks.

At the Hosts menu, we can now manage and examine the newly created host. We get the choice of these tasks:

- ▶ Rename
- ▶ Manage host ports
- ▶ Delete host
- ▶ View status of host ports
- ▶ View properties of the host

As shown in Figure 6-82, we get the following options when we click **Actions** of a host.

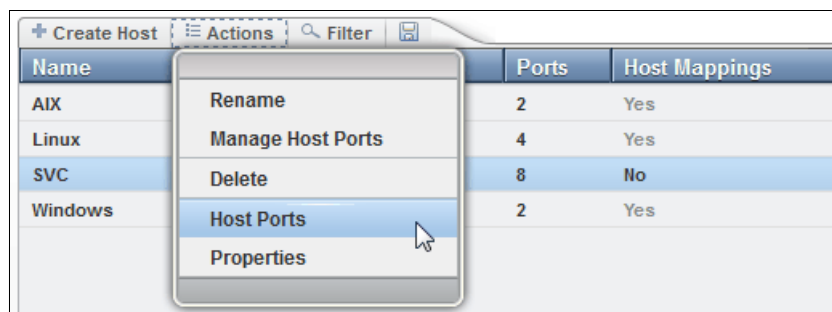
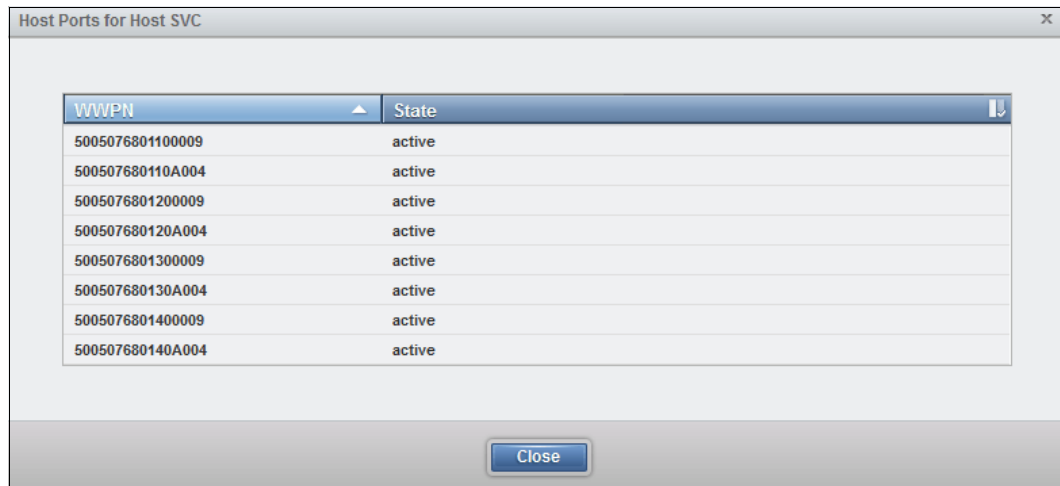


Figure 6-82 View status of host ports

When we click **Host Ports**, we get to view the properties and status of the host ports. In our example, we have eight ports for the host SAN Volume Controller. The WWPNs are displayed and the status for each port is displayed as shown in Figure 6-83 on page 221.



Host Ports for Host SVC

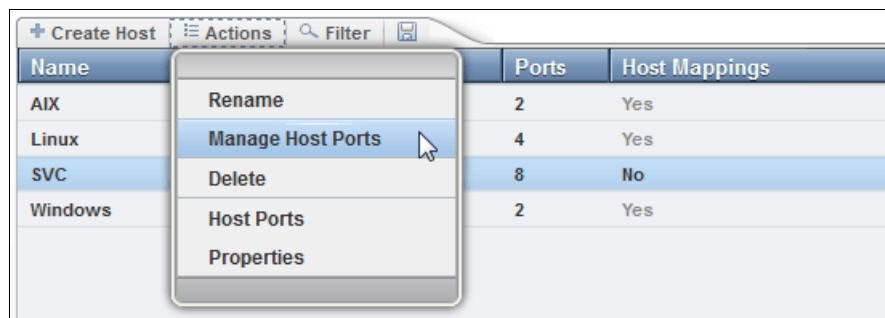
WWPN	State
5005076801100009	active
500507680110A004	active
5005076801200009	active
500507680120A004	active
5005076801300009	active
500507680130A004	active
5005076801400009	active
500507680140A004	active

Close

Figure 6-83 Host Ports for Host SVC

A correctly zoned and configured host port displays as active. By right-clicking in the blue row at the top of the listed WWPNs, we get the option to customize columns where we can select what we want to see or clear what we do not want to see.

If we want to add more host ports or if we want to remove host ports, we click **Actions** → **Manage Host Ports** as shown in Figure 6-84.



Manage Host Ports

Name	Ports	Host Mappings
AIX	2	Yes
Linux	4	Yes
SVC	8	No
Windows	2	Yes

Actions: Rename, Manage Host Ports, Delete, Host Ports, Properties

Figure 6-84 Manage Host Ports

In our example, we want to remove host ports and we highlight four of the SAN Volume Controller host ports and click **Actions** → **Remove** as shown in Figure 6-85 on page 222.

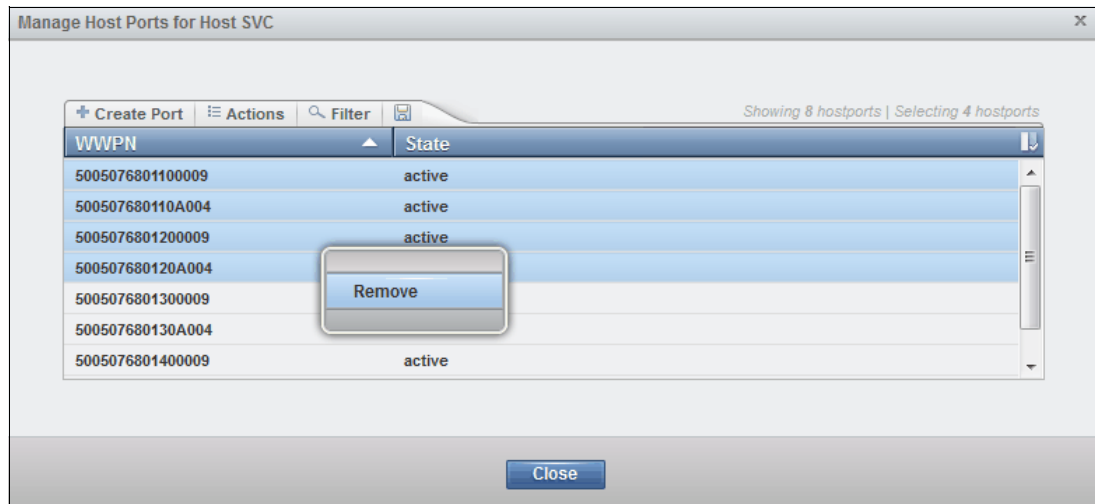


Figure 6-85 Manage host ports for host SAN Volume Controller to remove ports

The Manage Host Ports for Host SVC window now requires confirmation of how many and which host ports we want to remove as shown in Figure 6-86.

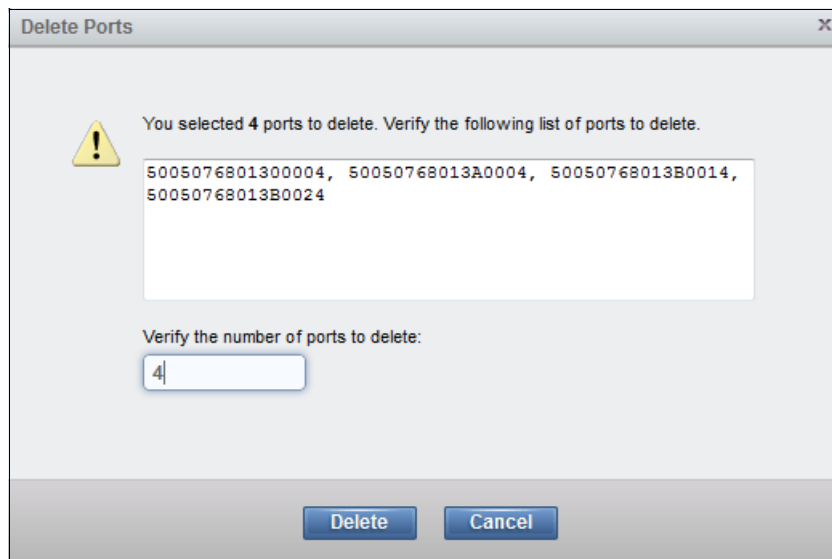


Figure 6-86 Delete four host ports

When we click **Delete**, the system removes the selected ports.

Create host through CLI

To create a host through the CLI, we use the **svctask mkhost** and **svctask addhostport** commands as demonstrated in Example 6-5.

Example 6-5 Create host and add one more port

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>svctask mkhost -fcwwpn
5005076801300004 -force -name SVC
Host, id [1], successfully created
```

```
IBM_Flashsystem:Cluster_9.19.91.242:superuser>svctask addhostport -fcwwpn
50050768013A0004 -force SVC
```


To create a new host, click **Add Host** in the upper-left corner of the Hosts window as shown in Figure 6-88. Next, type the host name and the iSCSI IQN name that is obtained from the iSCSI hosts iSCSI initiator software. Click **Add** to create the host.

The image shows a dialog box titled "Add Host". It has two input fields: "Name:" with the value "Win2008_File_SVR" and "Host port (iSCSI):" with the value "iqn.1991-05.com.microsoft:win-4j08903tko:". There are plus and minus icons to the right of the host port field. At the bottom, there are "Add" and "Cancel" buttons.

Figure 6-88 Create iSCSI host

Figure 6-89 shows the hosts, including the newly created Win2008_File_SVR host that we just created.

The image shows the "Hosts" window in the IBM FlashSystem 840 management interface. The window title is "FlashSystem840-4 > Hosts > Hosts". The user is "superuser (Security Administrator)". The window shows a table of hosts with columns: Name, State, Host Ports, and Host Mappings. The table lists five hosts: Linux_1, Linux_2, Win2008_Exchange, Win2008_File_SVR, and Win2008_SQL. The Win2008_File_SVR host is highlighted, indicating it is the newly created host.

Name	State	Host Ports	Host Mappings
Linux_1	✓ Online	1	Yes
Linux_2	✓ Online	1	Yes
Win2008_Exchange	✓ Online	1	Yes
Win2008_File_SVR	✓ Online	1	No
Win2008_SQL	✓ Online	1	Yes

Figure 6-89 New iSCSI host created

The new host does not yet have any host mappings. Host mappings for iSCSI hosts can be created in the same way as host mappings for FC and FCoE systems, which was demonstrated in “Mapping volumes” on page 211.

As discussed in “Create host” on page 216, any single host *initiator* port must be able to communicate with both FlashSystem 840 controllers called the *target* ports. If not, hosts appear as *Degraded* in the FlashSystem 840 Hosts menu. That means that an iSCSI-attached host must have its iSCSI initiator software configured so that each of its iSCSI initiators connects to at least one iSCSI target port on each FlashSystem 840 canister.

A FlashSystem 840 configured for iSCSI has four interface cards, each with four iSCSI ports. To provide enough bandwidth for failover purposes and redundancy, a more suitable configuration is that each iSCSI initiator is configured to connect to one target port on each FlashSystem 840 interface card for a total of four connections.

For more information about how to configure host connectivity, see Chapter 5, “IBM FlashSystem 840 client host attachment and implementation” on page 111.

6.4.2 Volumes by Host

The Hosts → Volumes by Host menu is functionally identical to the Volumes → Volumes by Host menu as explained in more detail in 6.3.3, “Volumes by Host menu” on page 211.

6.5 Access menu

We describe the Access menu and its options. There are a number of levels of user access to the FlashSystem 840 cluster, which are managed through the Access menu. The access levels are divided into groups. Each group has a different level of access and authority. If you want, multiple users can be defined and their access assigned to suit the tasks that they perform.

The Access menu provides the following submenus:

- ▶ Users
- ▶ User Groups
- ▶ Audit Log

6.5.1 Navigating to the Access menu

Hover the cursor over the **Access** function icon, and a menu opens (Figure 6-90).



Figure 6-90 Navigate to the Access menu

The Access menu allows user management and audit log review.

User management includes the creation of new users and the maintenance of roles and passwords for existing users. Also, part of user management is the configuration of Secure Shell (SSH) keys to provide secure access to the CLI for users.

The audit log provides a list of all commands executed on the system and also contains information about which user ran the command.

6.5.2 Users menu

Figure 6-91 shows the Users menu. This window enables you to create and delete users, change and remove passwords, and add and remove SSH keys.

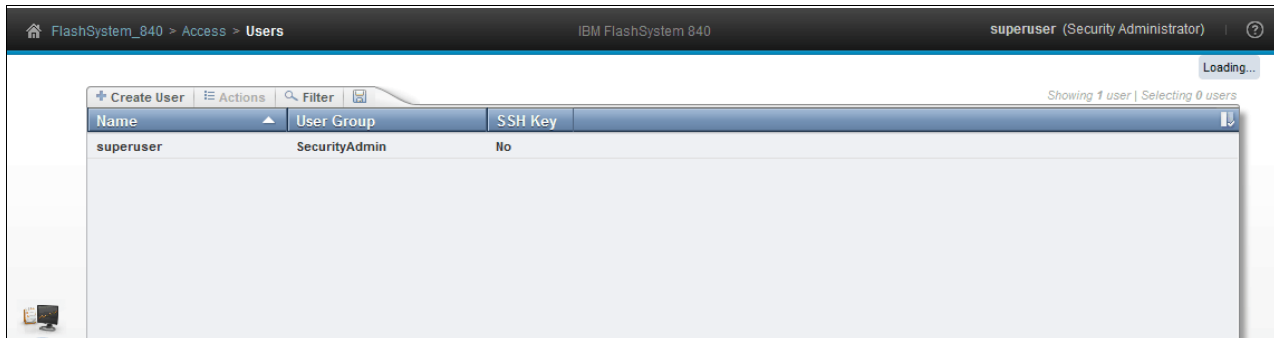


Figure 6-91 Users menu

Click **Create User** to open the window shown in Figure 6-92. You can enter the name of the user and the password, and load the SSH key (if the SSH key was generated). SSH keys are not required for CLI access. You can choose either to use SSH or a password for CLI authentication.

The 'Create User' dialog box is shown. It has a title bar with a close button. Inside, there's a user icon and a 'Name' field containing 'dennis'. Below the name is a 'Test LDAP Authentication' button. The 'Authentication Mode' section has two radio buttons: 'Local' (selected) and 'Remote (LDAP)'. The 'User Group' dropdown menu is set to 'SecurityAdmin'. The 'Local Credentials' section includes a note: 'Users must have a password, an SSH public key, or both.' It has 'Password' and 'Verify password' fields, both filled with dots. Below these is an 'SSH Public Key' field with a 'Browse...' button. At the bottom are 'Create' and 'Cancel' buttons.

Figure 6-92 Create user dennis

In the example in Figure 6-92, we create a local user dennis and configure a password that the user uses to authenticate to the system. When the user dennis opens his SSH client and points it to the IP address of the system to which he is granted access, he is prompted for the user name and password.

If the user dennis is required to authenticate using an SSH key pair, we instead enter the path for the public key in the field SSH Public Key in the Create User window. This is shown in Figure 6-93.

The screenshot shows a 'Create User' window with the following fields and options:

- Name:** dennis
- Test LDAP Authentication:** A button.
- Authentication Mode:** Radio buttons for ☒ Local and ☐ Remote (LDAP).
- User Group:** A dropdown menu showing 'SecurityAdmin'.
- Local Credentials:** A section with the instruction 'Users must have a password, an SSH public key, or both.' containing:
 - Password:** A text field with masked characters (dots).
 - Verify password:** A text field with masked characters (dots).
 - SSH Public Key:** A text field containing 'C:\Users\IBM_ADMIN\...' and a 'Browse...' button.
- Buttons:** 'Create' and 'Cancel' at the bottom.

Figure 6-93 Create user dennis as SSH key enabled

The Password and Verify Password fields are used for GUI access. If a password is not configured, the user dennis is not able to log in to the GUI.

When the SSH key is generated by using PuTTYgen, you have the choice of configuring a passphrase or not. If the SSH key pair was generated without a passphrase, the user dennis is not prompted for a password when he opens his SSH client. He is then authenticated with the private key that matches the uploaded public key. If a passphrase was configured when the SSH key pair was created, the user dennis also needs to type the passphrase password when opening the CLI to access the system.

For information about how to create SSH keys by using PuTTYgen, see “Generating an SSH key pair by using PuTTY” on page 229.

Figure 6-94 on page 228 shows that we now have a user dennis and that an SSH key is enabled for that user.

Name	User Group	SSH Key
dennis	SecurityAdmin	Yes
superuser	SecurityAdmin	No

Figure 6-94 User dennis is now created

Configuring CLI access, including how to configure SSH keys for secure access only, is described in detail in 6.5.3, “Access CLI by using PuTTY” on page 229.

Figure 6-95 shows that various actions can be performed for managing the users.

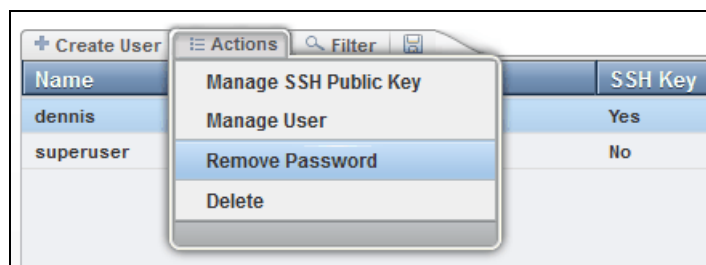


Figure 6-95 Manage user actions

To test whether the new user can log in to the GUI, we log out as user superuser and log in as user dennis. The login window is shown in Figure 6-96.



Figure 6-96 Log in as user dennis

The user area of the GUI in the upper-right corner now shows that user dennis is the current user, as shown in Figure 6-97 on page 229.



Figure 6-97 User dennis logged in

6.5.3 Access CLI by using PuTTY

Information about how to access the CLI by using PuTTY is provided.

PuTTY is a no-charge implementation of Telnet and SSH for Windows and UNIX platforms. PuTTY can be downloaded from the following URL:

<http://www.putty.org>

The CLI commands for the IBM FlashSystem 840 use the SSH connection between the SSH client software on the host system and the SSH server on the system.

You must create a system *before* you can use the CLI.

To use the CLI from a client system, follow these steps:

1. Install and set up the SSH client software on each system that you plan to use to access the CLI.
2. Authenticate to the system by using a password.
3. If you require command-line access without entering a password, use an SSH public key. Then, store the SSH public key for each SSH client on the system.

Generating an SSH key pair by using PuTTY

To use the CLI with SSH keys enabled, you must generate an SSH key pair. SSH keys can be generated from a Windows host by using the PuTTY key generator, PuTTYgen, by completing the following steps:

1. Execute **puttygen.exe**.
2. Click **SSH-2 RSA** as the type of key to generate.
Leave the number of bits in a generated key value at 1024.
3. Click **Generate** and then move the cursor around the blank area of the Key section to generate the random characters that create a unique key. When the key is completely generated, the information about the new key is displayed in the Key section.
4. Optional: Enter a passphrase in the Key passphrase and Confirm passphrase fields. The passphrase encrypts the key on the disk; therefore, it is not possible to use the key without first entering the passphrase.
5. Save the public key by performing these steps:
 - a. Click **Save public key**. You are prompted for the name and location of the public key.
 - b. Type `icat.pub` as the name of the public key and specify the location where you want to save the public key. For example, you can create a directory on your computer called `keys` to store both the public and private keys.

- c. Click **Save**.
6. Save the private key by performing these steps:
 - a. Click **Save private key**. The PuTTYgen Warning panel is displayed.
 - b. Click **Yes** to save the private key without a passphrase.
 - c. Type `icat` as the name of the private key, and specify the location where you want to save the private key. For example, you can create a directory on your computer called `keys` to store both the public and private keys. It is suggested that you save your public and private keys in the same location.
 - d. Click **Save**.
7. Close the PuTTY Key Generator window.

Figure 6-98 shows the folder where we saved the public and private keys that were generated using PuTTYgen.

Name	Date modified	Type	Size
dennis.ppk	01-10-2013 13:31	PPK File	2 KB
dennis.pub	01-10-2013 13:31	PUB File	1 KB
puttygen.exe	01-10-2013 13:30	Application	180 KB

Figure 6-98 SSH keys created using PuTTYgen

Access the CLI with SSH keys enabled

To access the FlashSystem 840 CLI by using PuTTY as the SSH client, a few configurations need to be performed with PuTTY.

First, type the IP address or name of the FlashSystem 840. The name can be used if the name resolution (Domain Name System (DNS)) is configured. Next, ensure that **SSH** is selected and that port 22 is used for the connection, as shown in Figure 6-99.

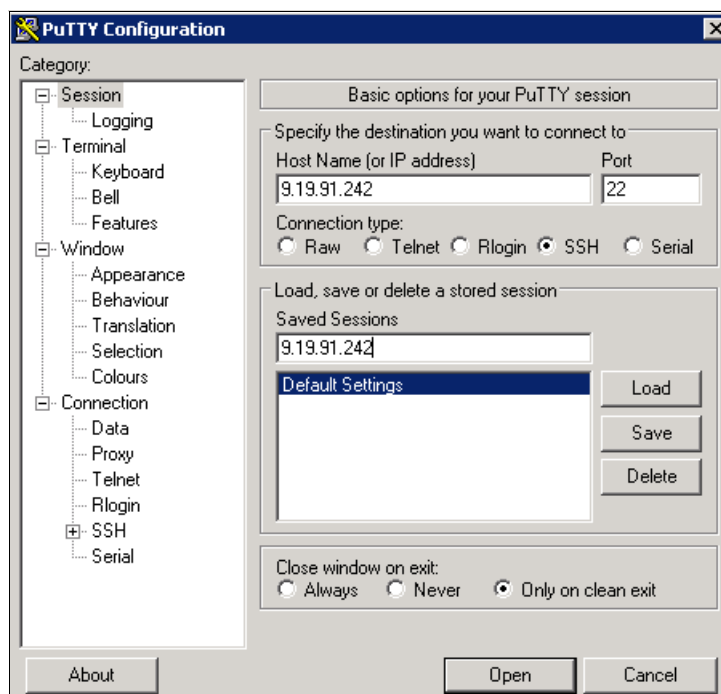


Figure 6-99 Configure PuTTY step 1 of 3

For configuring the SSH client session with the correct private key for user dennis, complete the following steps. Expand **Connection** → **SSH** and click **Auth**. In the Private key file for authentication field, type the path for the SSH private key that matches the public key loaded on the FlashSystem 840 for the user dennis.

Figure 6-100 shows that we load the private key for the user dennis.

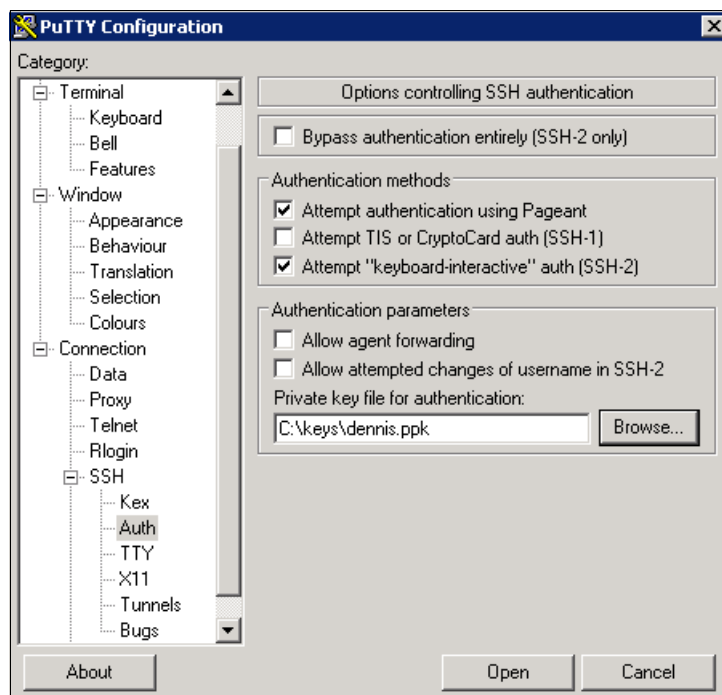


Figure 6-100 Configure PuTTY step 2 of 3

The last step for configuring the CLI access is to save the connection. We type a name and click **Save**. Now, the connection settings are saved for the next time that we use them. We click **Open** to start our CLI session.

Figure 6-101 on page 232 shows that we saved the configured session.

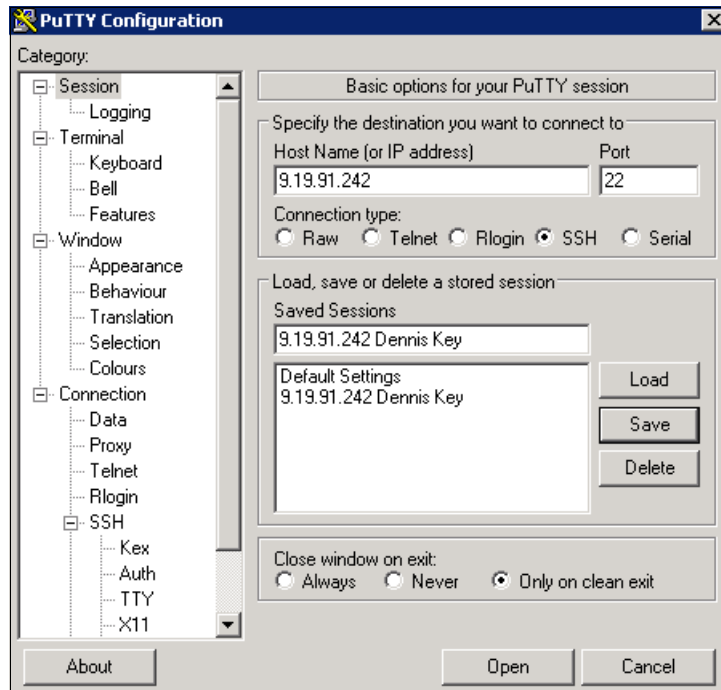


Figure 6-101 Configure PuTTY step 3 of 3

Example 6-6 shows that user dennis logs in to the CLI. He is prompted for a passphrase (password), which was configured when we generated the SSH key pair.

Example 6-6 User dennis logs in to the CLI by using PuTTY

```
login as: dennis
Authenticating with public key "rsa-key-20131022"
Passphrase for key "rsa-key-20131022":*****
Last login: Tue Oct 22 08:46:27 2013 from 9.146.197.45
```

```
IBM_Flashsystem:FlashSystem_840:dennis>
```

Setting up CLI access and generating SSH keys are described in greater detail in the FlashSystem 840 IBM Knowledge Center at the following URL:

http://www.ibm.com/support/knowledgecenter/ST2NVR_1.3.0

6.5.4 User groups

Administrators can create role-based user groups where any users that are added to the group adopt the role that is assigned to that group. Roles apply to both local and remote users on the system and are based on the user group to which the user belongs. A local user can only belong to a single group; therefore, the role of a local user is defined by the single group to which that user belongs. Users with the Security Administrator role can organize users of the system by role through user groups.

You can assign the following user roles to users of the system as shown in Table 6-1 on page 233.

Table 6-1 User groups available on the IBM FlashSystem 840

Name	Role
SecurityAdmin	Users with this role can access all functions on the system, including managing users, user groups, and user authentication.
Administrator	Users with this role can access all functions on the system, except those functions that deal with managing users, user groups, and authentication.
Monitor	Users with this role can view objects and system configuration but they cannot configure, modify, or manage the system or its resources.
CopyOperator	Users with this role have monitor-role privileges and can change and manage all Copy Services functions.
Service	Users with this role have monitor-role privileges and can view the system information, begin the disk-discovery process, and include disks that are excluded. This role is used by service personnel.

All users must be a member of a predefined user group.

Create a user group

To create a new user group, navigate to **Access** → **User Groups** and click **Create User Group**.

Figure 6-102 shows the User Groups menu, which shows the default local user groups and one remote user group.

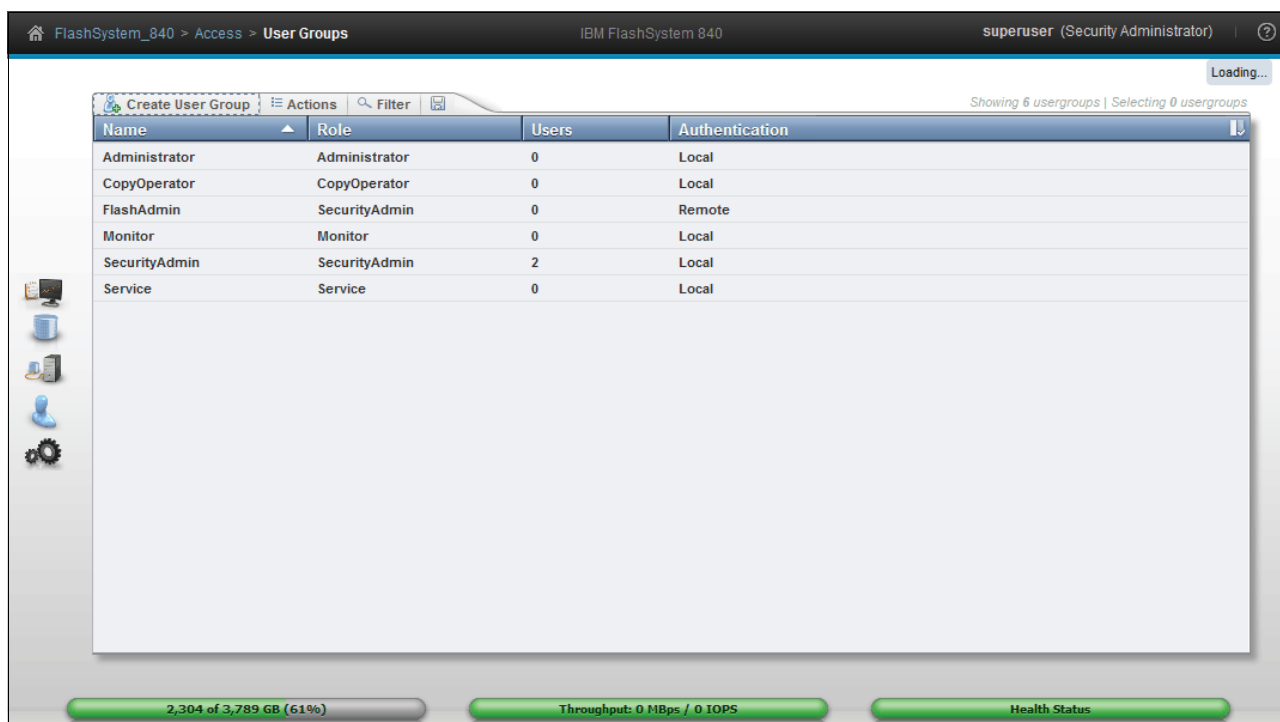
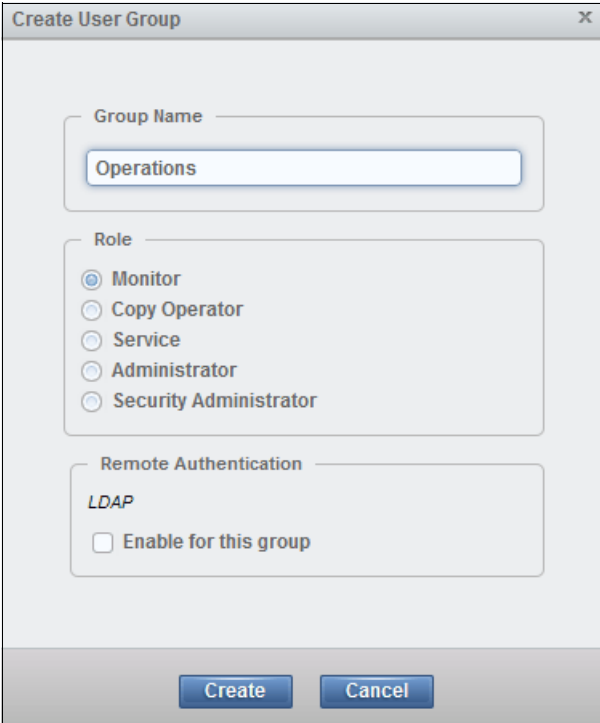


Figure 6-102 User Groups window

For more information about configuring remote authentication, see 7.1.3, “Security menu” on page 243.

The Create User Group window opens and we are prompted to provide a Group Name and select a Role for the new user. If the group that is created is for remote authentication, we check **Enable LDAP for this group** (available only for Lightweight Directory Access Protocol (LDAP) authentication; not required for the Operations group).

In our example, we want to create a new group called Operations and we want to provide the role Monitor for this new group as shown in Figure 6-103. The new group is a local group, and we do not require LDAP authentication.



The screenshot shows a window titled "Create User Group" with a close button (X) in the top right corner. The window contains three main sections: "Group Name" with a text input field containing "Operations"; "Role" with a list of radio buttons where "Monitor" is selected; and "Remote Authentication" with a checkbox labeled "Enable for this group" which is currently unchecked. At the bottom of the window are two buttons: "Create" and "Cancel".

Figure 6-103 Create a user group called Operations

When we click **Create** in the Create User Group window, the new group is created. To view the new group, click **Access** → **User Groups** as shown in Figure 6-104 on page 235.

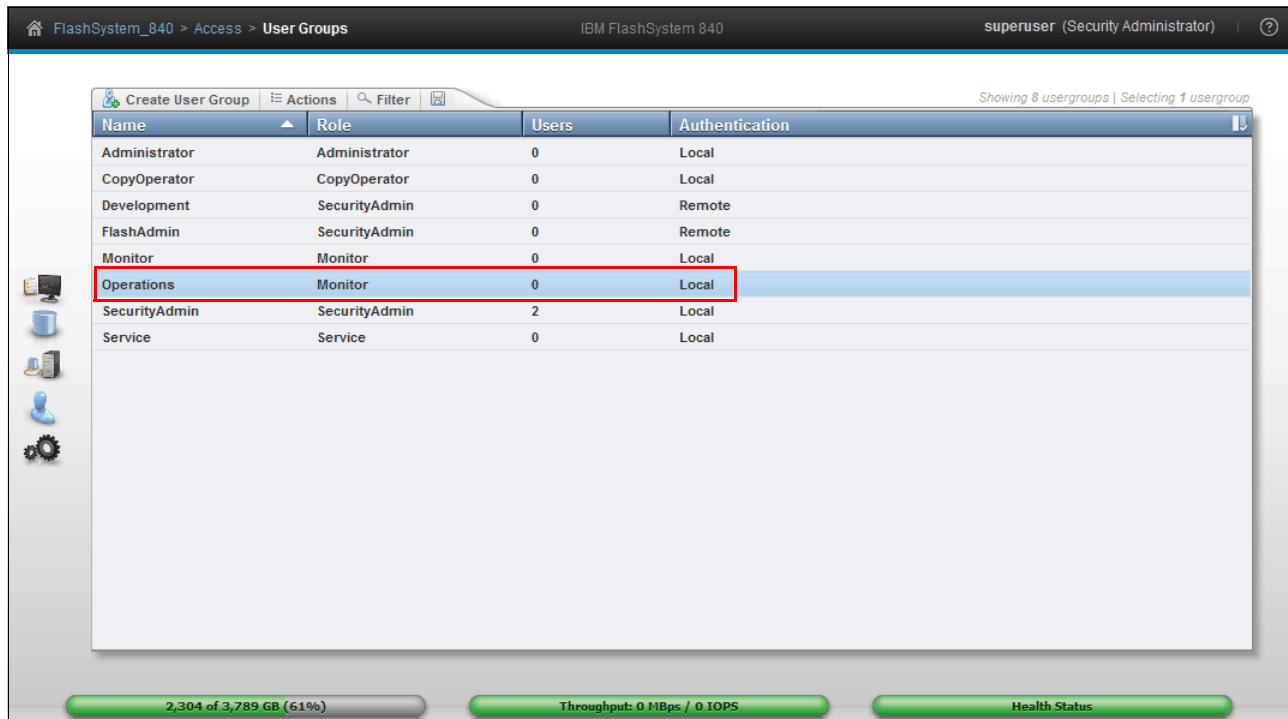


Figure 6-104 User group called Operations was created

After creating a new user group, we want to add users to that group. We now navigate to **Access** → **Users** and click **Create User**. The Create User window opens and we type the name of the new user and select the User Group **Operations** as shown in Figure 6-105 on page 236. A password must be typed and confirmed or the user is not able to log in to the GUI.

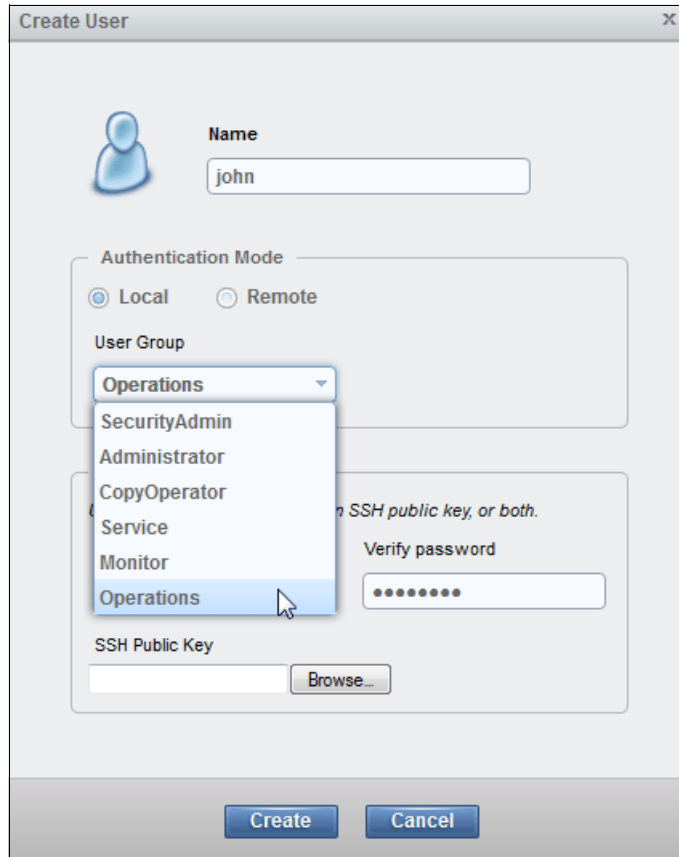


Figure 6-105 Create the user john and add it to a user group

When we click **Create**, the user john is now created and added to the Operations user group. The user john can now log in to the GUI and CLI by using the provided password. This user will have the privileges from the user group Operations, which has the role Monitor.

Note: If remote authentication is not enabled in the Settings → Security menu, the Create User Group window does not give you the option to create the group for remote authentication and LDAP (Figure 6-103 on page 234).

6.5.5 Audit log menu

An *audit log* tracks action commands that are issued through an SSH session or through the management GUI.

The audit log entries provide the following information:

- ▶ Identity of the user who issued the action command
- ▶ The name of the actionable command
- ▶ The time stamp when the actionable command was issued on the configuration node
- ▶ The parameters that were issued with the actionable command

The following commands are not documented in the audit log:

- ▶ **dumpconfig**
- ▶ **cpdumps**
- ▶ **cleardumps**

- ▶ finderr
- ▶ dumperlog
- ▶ dumpinternallog
- ▶ svcserVICetask dumperlog
- ▶ svcserVICetask finderr

The following items are also not documented in the audit log:

- ▶ Commands that fail are not logged.
- ▶ A result code of 0 (success) or 1 (success in progress) is not logged.
- ▶ The result object ID of the node type (for the **addnode** command) is not logged.
- ▶ Views are not logged.

Review the audit log

To review the audit log, navigate to **Access** → **Audit Log**. The audit log displays as shown in Figure 6-106.

Date and Time	User Name	Command	Object ID
10/15/13 4:46:55 PM	superuser	svctask mkuser -name dennis -usergrp 0 -password '###...	1
10/15/13 4:23:29 PM	superuser	svctask mkvdiskhostmap -host 2 -force 7	
10/15/13 4:23:29 PM	superuser	svctask mkvdiskhostmap -host 2 -force 8	
10/15/13 4:22:51 PM	superuser	svctask addhostport -force -fcwwpn 10008C7CFF0B0F01	2
10/15/13 4:22:32 PM	superuser	svctask addhostport -force -fcwwpn 10008C7CFF0B7881	2
10/15/13 4:22:16 PM	superuser	svctask addhostport -force -fcwwpn 10008C7CFF0B0F00	2
10/15/13 4:21:55 PM	superuser	svctask mkhost -name Linux -force -fcwwpn 10008C7CFF0...	
10/15/13 4:17:10 PM	superuser	svctask testldapserver	
10/15/13 4:16:05 PM	larsen@itsolab.ib...	svctask testldapserver -username larsen@itsolab.ibm.co...	
10/15/13 4:15:26 PM	larsen@itsolab.ib...	svctask testldapserver	
10/15/13 4:00:20 PM	superuser	svctask mkvdiskhostmap -host 0 -force 4	
10/15/13 4:00:20 PM	superuser	svctask mkvdiskhostmap -host 0 -force 5	
10/15/13 4:00:02 PM	superuser	svctask mkvdiskhostmap -host 1 -force 6	
10/15/13 3:59:30 PM	superuser	svctask addhostport -force -fcwwpn 10000000C9D49411	1
10/15/13 3:58:49 PM	superuser	svctask mkhost -name AIX -force -fcwwpn 10000000C9AA2...	
10/15/13 3:56:25 PM	superuser	svctask mkusergrp -name FlashAdmin -remote -role Secu...	5
10/15/13 3:55:15 PM	superuser	svctask rmusergrp -force 5	
10/15/13 3:54:21 PM	superuser	svctask mkldapserver -ip 9.19.91.219 -basedn dc=itsolab,dc...	0
10/15/13 3:54:21 PM	superuser	svctask chldap -username 'itsolab\administrator' -passwo...	
10/15/13 3:54:21 PM	superuser	svctask chauthservice -enable yes -type ldap	
10/15/13 3:51:20 PM	superuser	svctask addhostport -force -fcwwpn 21000024FF22F9EB	0
10/15/13 3:51:18 PM	superuser	svctask chauthservice -enable no	
10/15/13 3:51:18 PM	superuser	svctask rmlldapserver 0	

Figure 6-106 Audit Log window

The audit log can be filtered to display selected users only or within a specific time frame. The audit log can also be filtered to show from which IP address a specific command was executed.

Figure 6-107 on page 238 shows filtering options where we add the IP address of the user that is logged in.

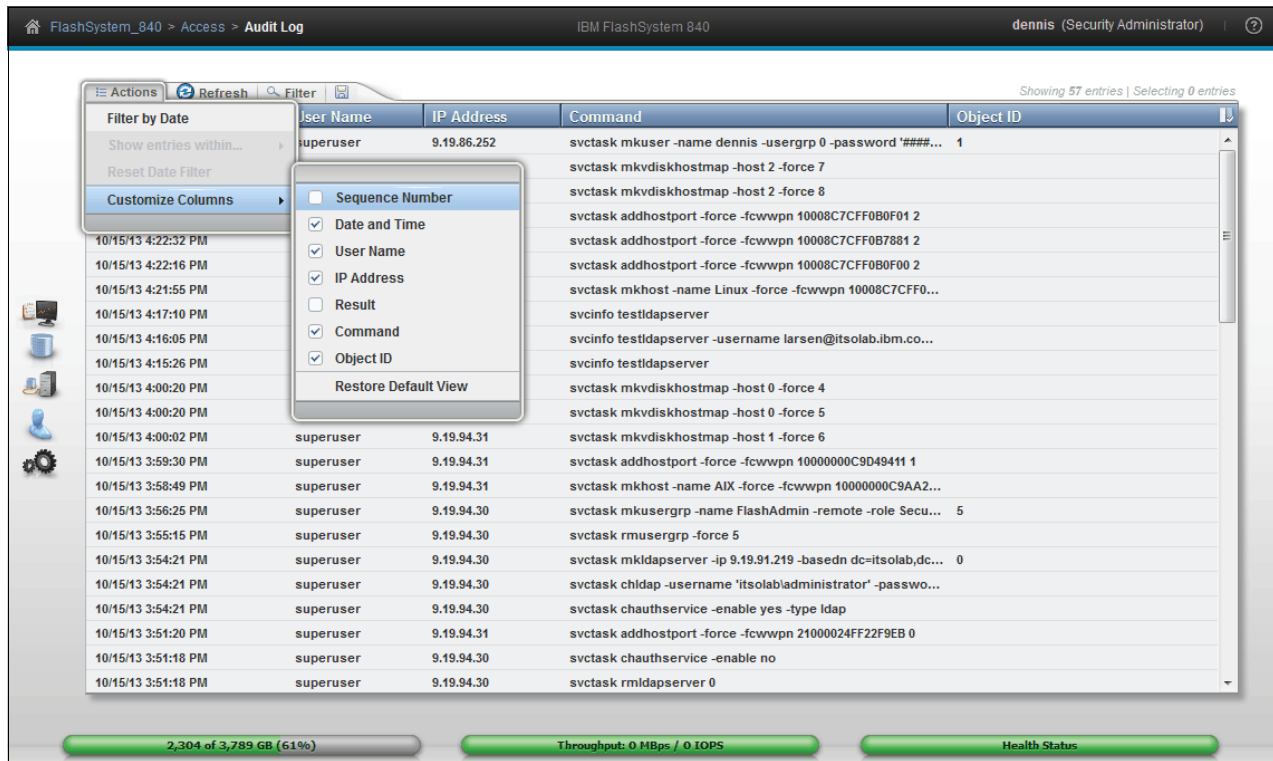


Figure 6-107 Viewing the Audit Log menu by using filtering

Also, you can export the audit log to a comma-separated file.



Configuring settings

The Settings section of the IBM FlashSystem 840 graphical user interface (GUI), as shown in Figure 7-1 on page 240, is described. The Settings function covers various options for monitoring, configuring network interfaces, and extracting support logs. It also covers remote authentication and the firmware update process. Additionally, we explain how to access IBM FlashSystem 840 Service Assistant Tool.

7.1 Settings menu

You can use the Settings panel to configure system options for event notifications, security, IP addresses, FC connectivity, and preferences related to display options in the management GUI.

The Settings menu includes five options:

- ▶ Notifications (alerting)
- ▶ Security (remote authentication with Lightweight Directory Access Protocol (LDAP))
- ▶ Network (management and service)
- ▶ Support (extract support logs)
- ▶ System (time settings, firmware update, and so on)

7.1.1 Navigating to the Settings menu

We describe the Settings menu and its options. If you hover the cursor over the Settings function icon, the Settings menu opens (Figure 7-1).



Figure 7-1 Navigate to the Settings menu

7.1.2 Notifications menu

FlashSystem 840 can use Simple Network Management Protocol (SNMP) traps, syslog messages, and Call Home email to notify you and IBM Support when significant events are detected. Any combination of these notification methods can be used simultaneously.

Notifications are normally sent immediately after an event is raised. However, there are exceptions. For example, event 085031 “Array is missing a spare flash module”, (SEC 1690) does not report until the rebuild is complete. Also, certain events might occur because of service actions that are performed. If a recommended service action is active, these events are sent only if they are still not fixed when the service action completes.

Email

The Call Home feature transmits operational and event-related data to you and IBM through a Simple Mail Transfer Protocol (SMTP) server connection in the form of an event notification email. When configured, this function alerts IBM service personnel about hardware failures and potentially serious configuration or environmental issues.

To configure email alerts, navigate to **Settings** → **Notifications** to open the window shown in Figure 7-2. From this window, you can configure the email alerts that are included in the Call Home function. During the configuration of Call Home, you configure contact information and email receivers for your own notification.

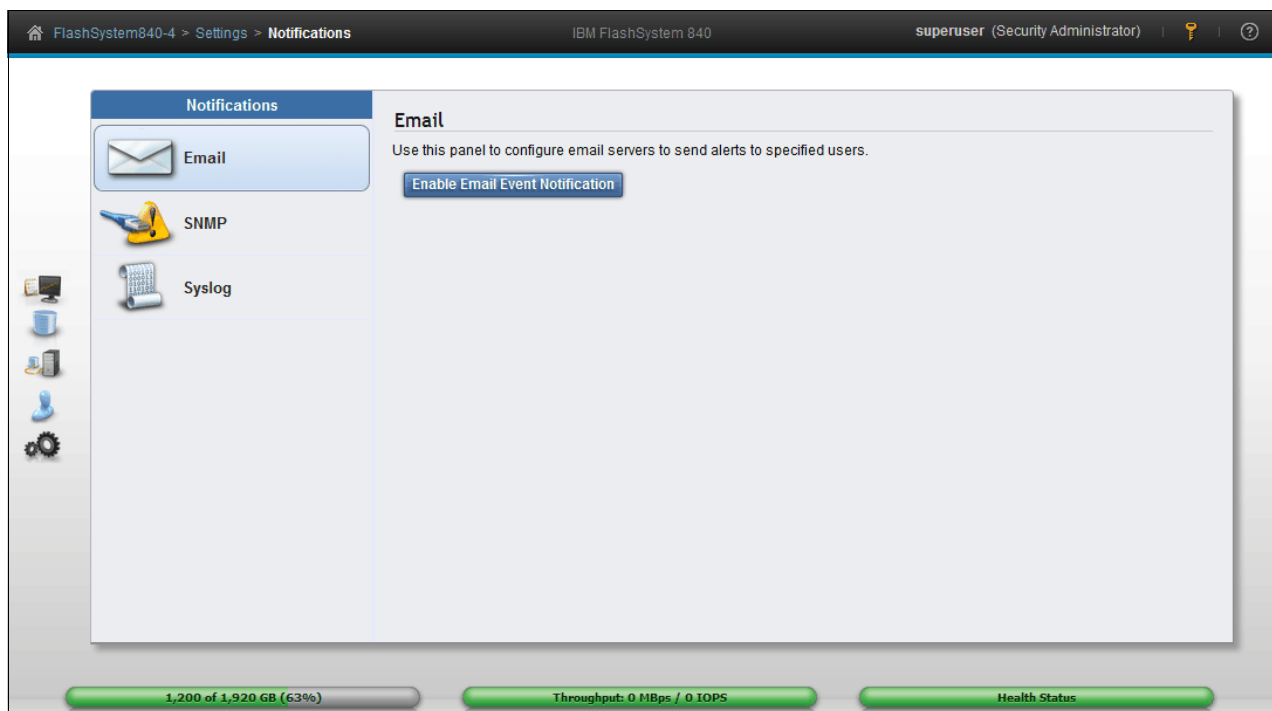


Figure 7-2 Notifications Email window

To initiate the configuration wizard, click **Enable Email Event Notification** and the setup Call Home window opens. The procedure for configuring Call Home is similar to the initialization of the IBM FlashSystem 840. The procedure for configuration is described in detail in Chapter 4, “Installation and configuration” on page 65.

SNMP

Simple Network Management Protocol (SNMP) is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that the system sends.

Figure 7-3 shows the SNMP configuration menu.

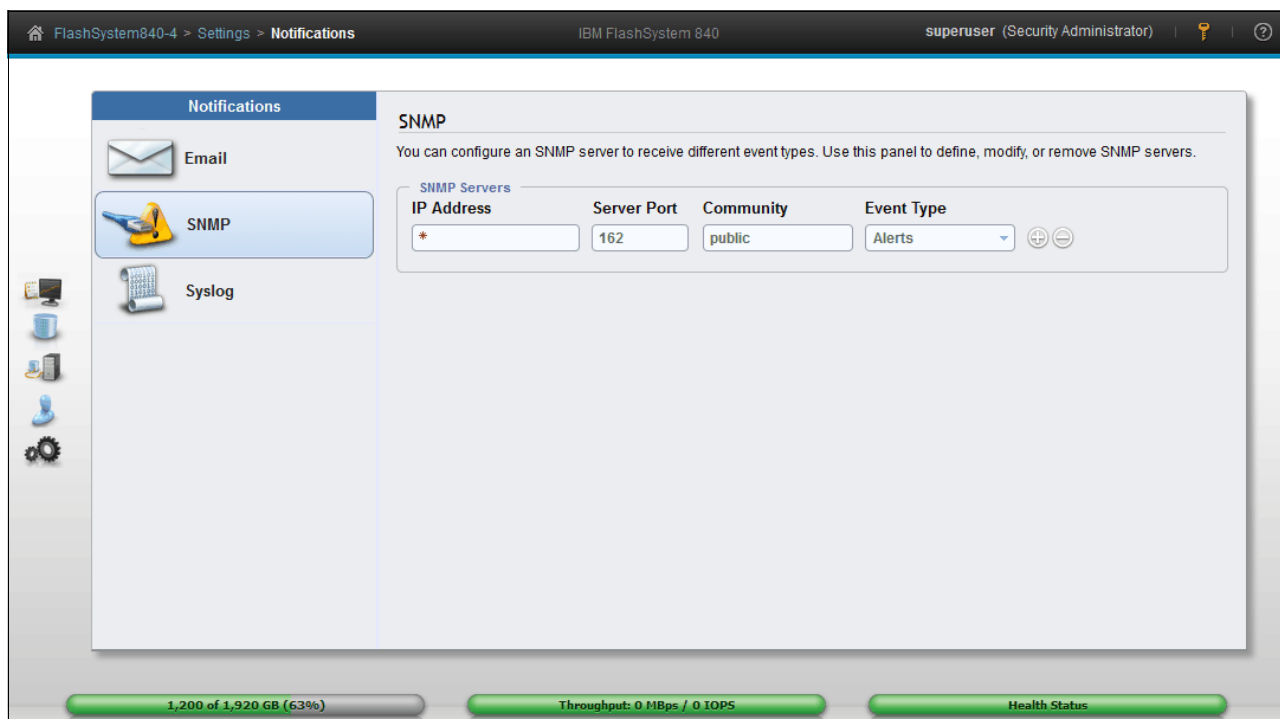


Figure 7-3 Event notifications SNMP window

In the SNMP configuration menu, you can configure one or more SNMP servers. For each of these SNMP servers, you configure the following information:

- ▶ IP address
- ▶ SNMP server port (The default is port 162.)
- ▶ SNMP community (The default is `public`.)
- ▶ Event type (The default is Alerts but it can be changed to All events.)

There are various SNMP trap receiver products on the market. These are known as *SNMP managers*. IBM Tivoli NetView® or IBM Tivoli Enterprise Console® can be used as IBM SNMP managers.

Syslog

The *syslog protocol* is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The IP network can be either IPv4 or IPv6. The system can send syslog messages that notify personnel about an event.

The IBM FlashSystem 840 can transmit syslog messages in either expanded or concise format. You can use a syslog manager to view the syslog messages that the system sends. The system uses the User Datagram Protocol (UDP) to transmit the syslog message. You can specify up to a maximum of six syslog servers (Figure 7-4 on page 243).

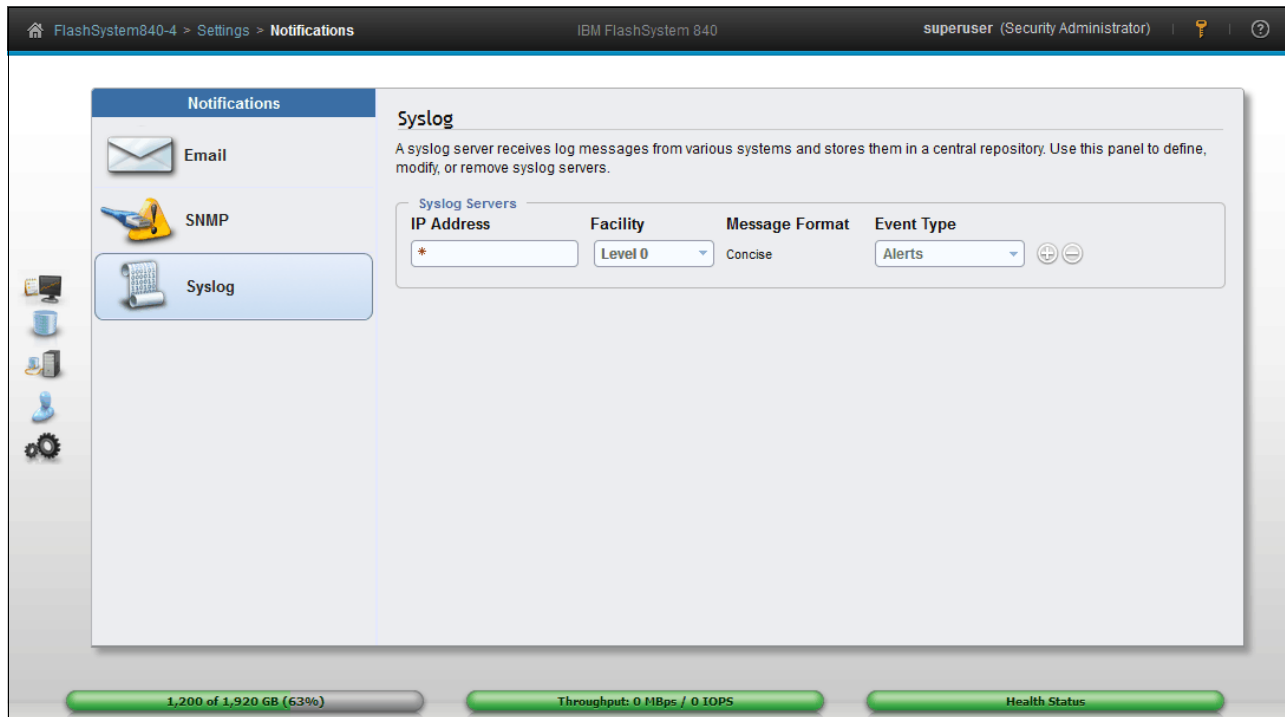


Figure 7-4 Event notifications Syslog window

In the Syslog configuration menu, you can configure one or more syslog servers. For each of these servers, you configure the following information:

- ▶ IP address
- ▶ Facility

The *facility* determines the format for the syslog messages and can be used to determine the source of the message.

- ▶ Event type (The default is Alerts but it can be changed to All events.)

There are various syslog server products on the market. Many of these products are no-charge products that can be downloaded from the Internet.

7.1.3 Security menu

When a FlashSystem 840 clustered system is created, the authentication settings default to `local`, which means that the IBM FlashSystem 840 contains a local database of users and their privileges. Users can be created on the system and can log in using the user accounts that they are given by the local `superuser` account.

You can create two types of users who can access the system. These types are based on how the users authenticate to the system. Local users are authenticated through the authentication methods that are on the IBM FlashSystem 840.

If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the command-line interface (CLI) through Secure Shell (SSH), either a password or a valid SSH key file is necessary. Local users must be part of a user group that is defined on the system. *User groups* define roles that authorize the users within that group to a specific set of privileges on the system.

For users of the FlashSystem 840 clustered system, you can configure authentication and authorization by using the CLI and the GUI as configured in the Users and User Groups menu.

A *remote user* is authenticated on a remote service with Lightweight Directory Access Protocol (LDAP) as configured in the Settings → Security section of the FlashSystem 840 GUI (Figure 7-1 on page 240). Remote users have their roles defined by the remote authentication service.

Remote authentication is disabled by default and can be enabled to authenticate users against LDAP servers.

A user who needs access to the CLI must be configured as a local user on the IBM FlashSystem 840.

Remote users do not need to be configured locally; they only need to be defined on the LDAP server.

For more information about how to configure remote authentication and authorization for users of the IBM FlashSystem 840, see the “User Authentication Configuration” section of the FlashSystem IBM Knowledge Center:

<http://ibm.co/1lRWmB0>

Reasons for using remote authentication

When remote authentication is configured, users authenticate with their domain user and password rather than a locally created user ID and password. Use remote authentication for the following reasons:

- ▶ Remote authentication saves you from having to configure a local user on every IBM storage system that exists in your storage infrastructure.
- ▶ If you have multiple LDAP-enabled storage systems, remote authentication makes it more efficient to set up authentication.
- ▶ The audit log shows the domain user name of the issuer when commands are executed. The domain user name is more informative than a local user name or just *superuser*.
- ▶ Remote authentication gives you central access control. If someone leaves the company, you only need to remove access at the domain controller, which means that there are no orphan user IDs left on the storage system.

Prepare the LDAP server

The first step in configuring LDAP is to prepare the LDAP server. We use a Microsoft Windows 2008 R2 Enterprise server, which we promoted to be a Domain Controller by using the **dcpromo** command. Next, we added the computer role *Active Directory Lightweight Directory Services*.

The privileges that the LDAP user gets on the IBM FlashSystem 840 are controlled by user groups on the storage system. There must be matching user groups on the Active Directory (AD) server and on the IBM FlashSystem 840, and the LDAP users must be added to the AD server group.

In our example, we create a group called *FlashAdmin*, which we use to manage our FlashSystem 840 storage device.

To create this group, we need to log on to the AD Domain Controller and configure Active Directory. An easy way to configure Active Directory from the AD controller is to go to **Start → Run**, type `dsa.msc`, and click **OK**. The Active Directory Users and Computers management console opens as shown in Figure 7-5.

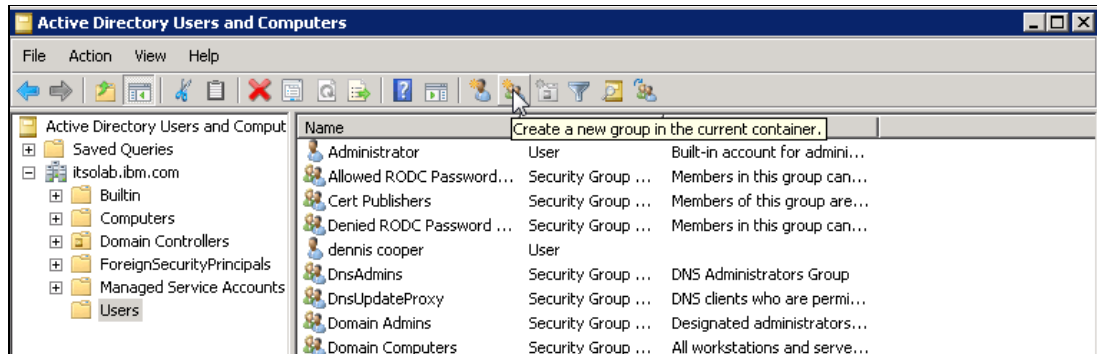


Figure 7-5 Active Directory Users and Computers window to create a new group

We click the **Create a new group in the current container** icon. The **New Object - Group** window opens as shown in Figure 7-6.

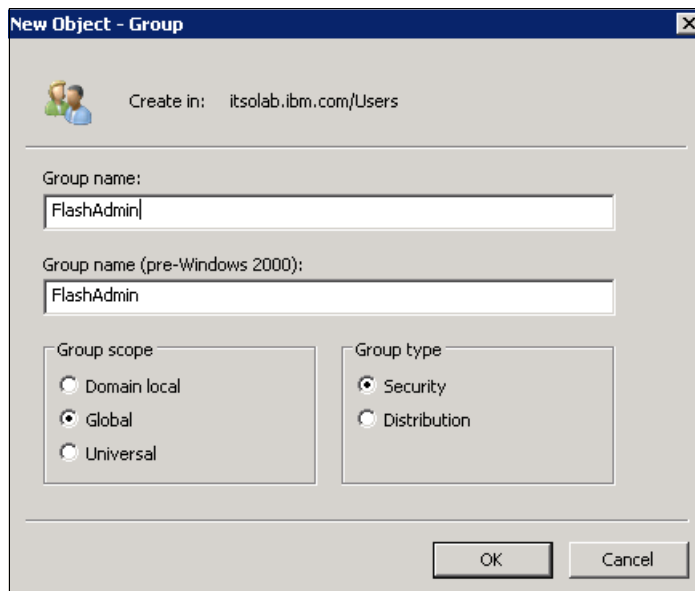


Figure 7-6 Active Directory to create a FlashAdmin group

We type `FlashAdmin` for the new group name, leave the remaining default values, and click **OK**. We now highlight the users that we want to add to the FlashSystem 840 storage administrator group and click the **Adds the selected objects to a group you specify** icon as shown in Figure 7-7 on page 246.

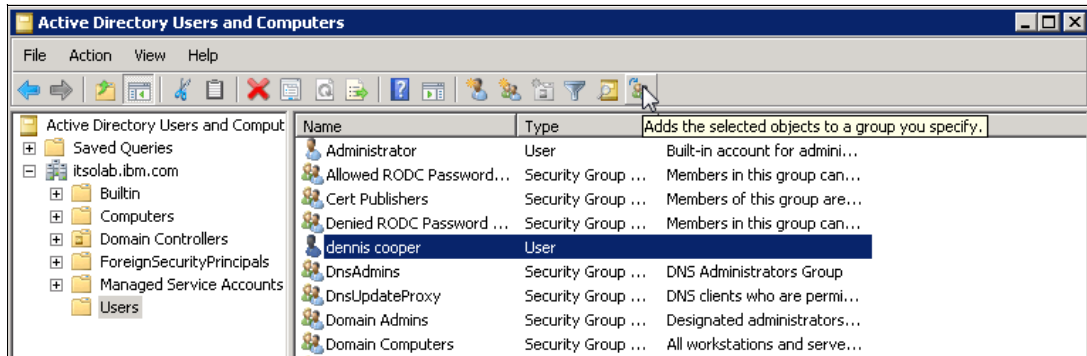


Figure 7-7 Adds the selected objects to a group you specify

In the Select Groups window, we type FlashAdmin and click **Check Names** as shown in Figure 7-8.

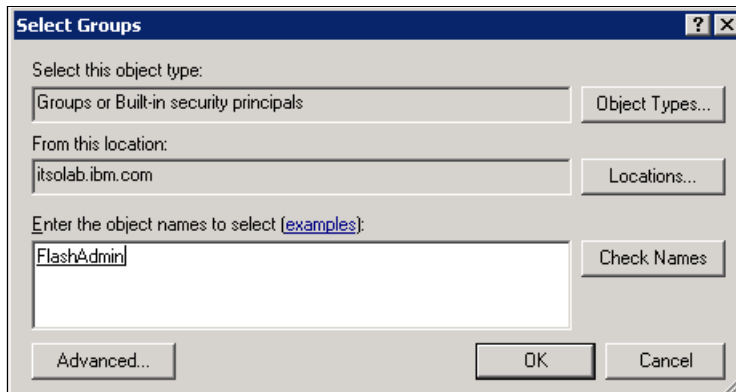


Figure 7-8 Active Directory Select Groups window to add users to the FlashAdmin group

Any other users that might be added to the FlashAdmin group get the same privileges on our FlashSystem 840.

If other users with different privileges are required, another group on the IBM FlashSystem 840 with different privileges is required. A group on the AD server with a matching name is also required.

Our LDAP server is now prepared for remote authentication.

Configure remote authentication

The next step in configuring remote authentication for the IBM FlashSystem 840 is to specify the authentication server, test connectivity, and test whether users can authenticate to the LDAP server.

From the Settings → Security menu, click **Remote Authentication**. Click **Enable Remote Authentication** as shown in Figure 7-9 on page 247.

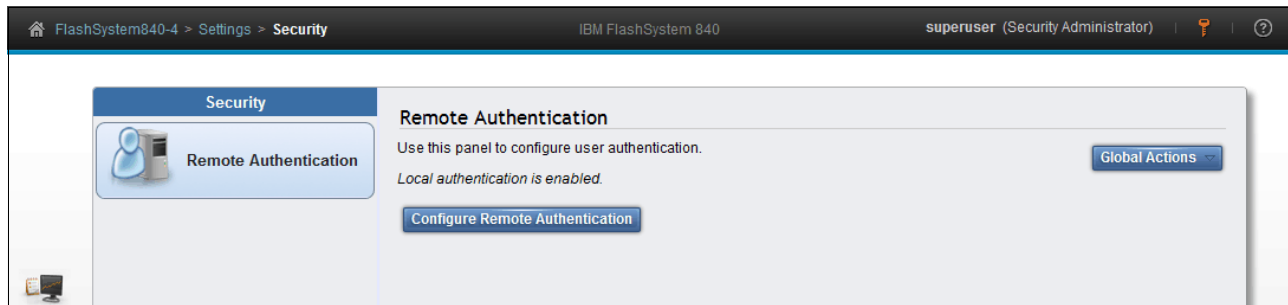


Figure 7-9 Enable remote authentication

The Configure Remote Authentication window opens. We select **LDAP** as shown in Figure 7-10.

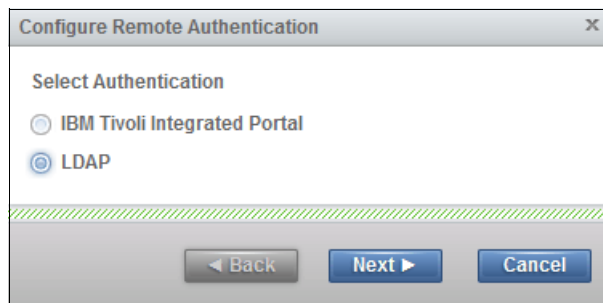


Figure 7-10 Remote Authentication wizard (step 1 of 4)

Next, we select **Microsoft Active Directory**. For Security, we select **None**, as shown in Figure 7-11.

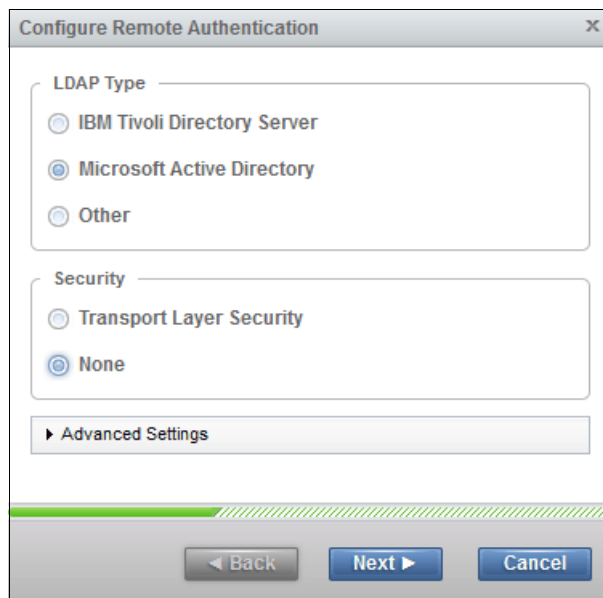


Figure 7-11 Remote Authentication wizard (step 2 of 4)

We then expand **Advanced Settings**. Any user with authority to query the LDAP directory can be used to authenticate. Our Active Directory domain is `itsolab.ibm.com`, so we use the Administrator login name on the Domain `itsolab.ibm.com` to authenticate. We then click **Next** as shown in Figure 7-12 on page 248.

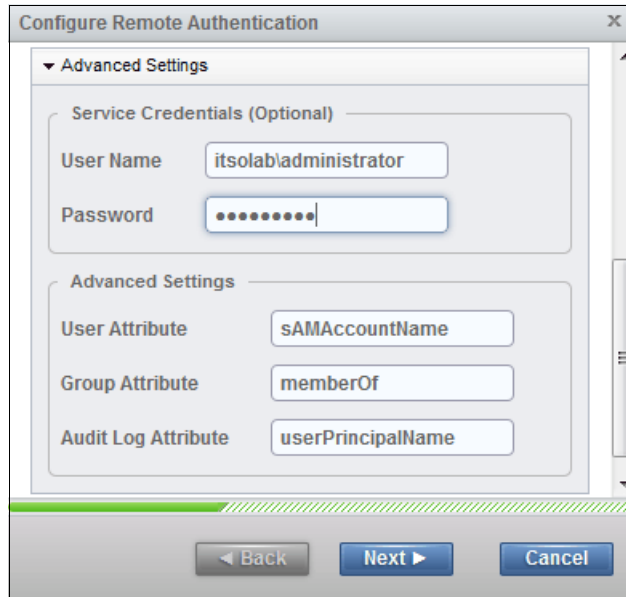


Figure 7-12 Remote Authentication wizard (step 3 of 4)

We type the IP address of our LDAP server, which is 9.19.91.219, and the LDAP Group Base Domain Name (DN) for Microsoft Active Directory.

You can obtain the LDAP User and Group Base DN for Microsoft Active Directory by using the following commands:

```
dsquery user -name <username>
dsquery group -name <group name>
```

To look up the Base DN, we log on to the LDAP server and execute the following commands (Example 7-1).

Example 7-1 Checking the LDAP server for the Base DN

```
C:\Users\Administrator>dsquery group -name FlashAdmin
"CN=FlashAdmin,CN=Users,DC=itsolab,DC=ibm,DC=com"

C:\Users\Administrator>
```

The Base DN that we need to enable LDAP authentication only requires the domain part of the output in Example 7-1. We now type **DC=itsolab,DC=ibm,DC=com** in the Base DN (Optional) field in the Configure Remote Authentication window as shown in Figure 7-13.

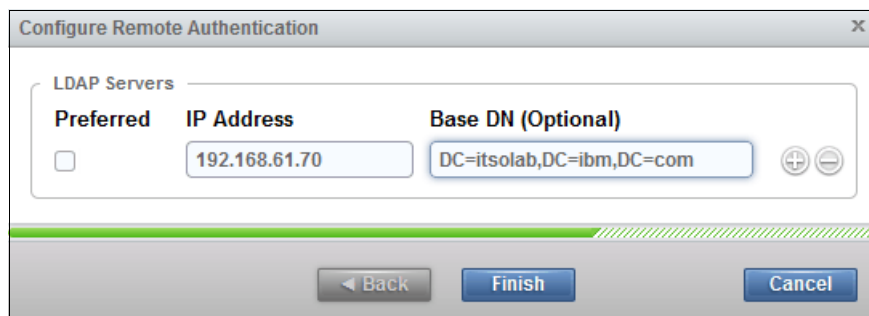


Figure 7-13 Remote Authentication wizard (step 4 of 4)

We click **Finish** and are then returned to the **Settings** → **Security** window. Figure 7-14 shows that LDAP is enabled and the window shows the preferences of the configured LDAP server.

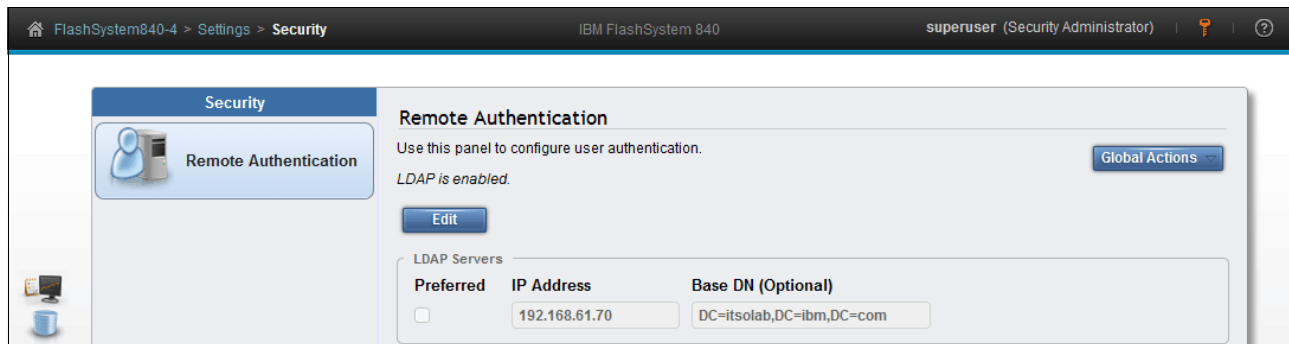


Figure 7-14 Remote Authentication enabled

Create the FlashSystem 840 LDAP-enabled user group

The first part of our LDAP configuration is complete. We need, however, to create a new user group on our FlashSystem 840 with a name that matches the name that we configured on the LDAP server. We configured the name FlashAdmin on the LDAP server.

First, click **Access** → **User Groups** as shown in Figure 7-15.

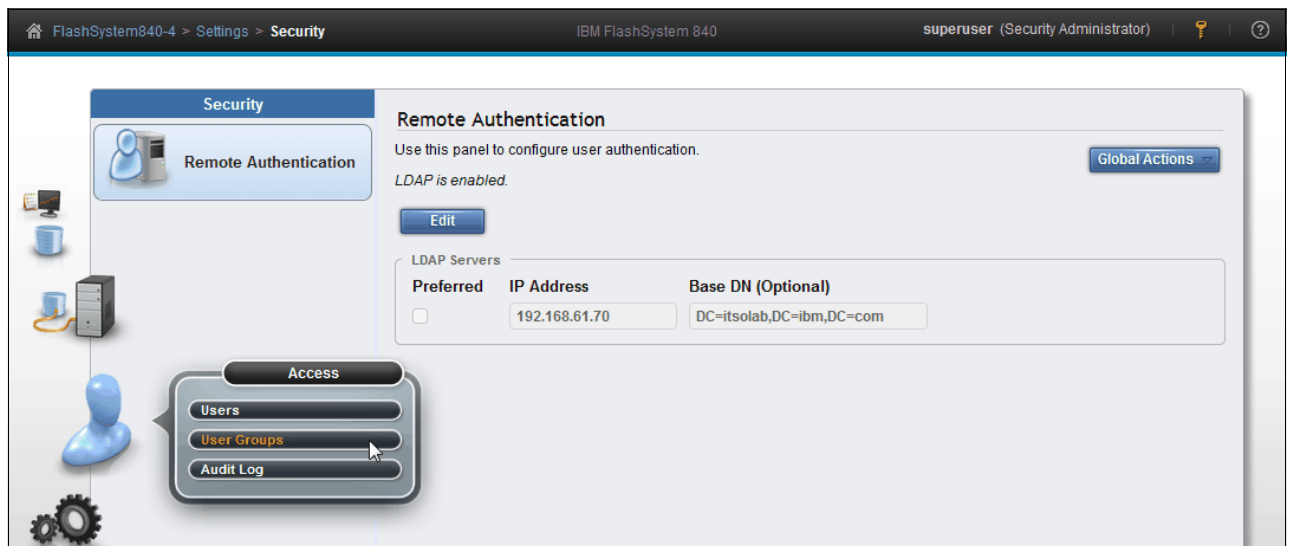


Figure 7-15 Navigate to User Groups

Figure 7-16 on page 250 shows the current configured user groups. We click **Create User Group**.

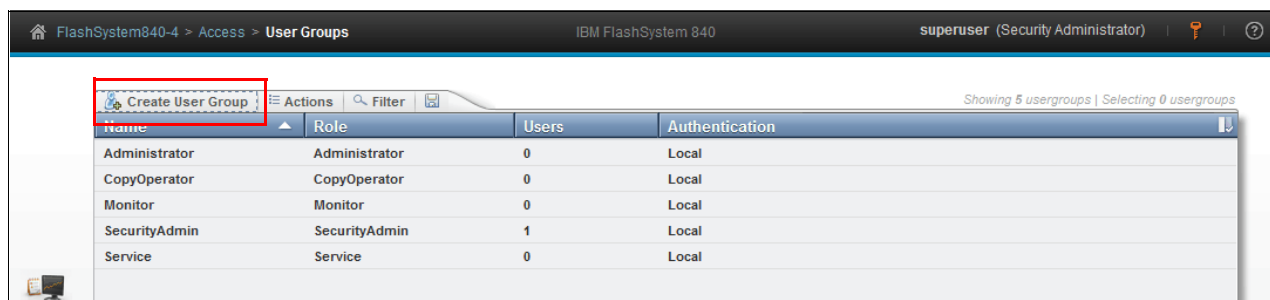


Figure 7-16 Create a user group

Figure 7-17 shows that we select **Security Administrator** and check **Enable for this group**. We type the group name FlashAdmin for the new user group.

Create User Group

Group Name:

Role:

- ☐ Monitor
- ☐ Copy Operator
- ☐ Service
- ☐ Administrator
- ☒ Security Administrator

Remote Authentication:

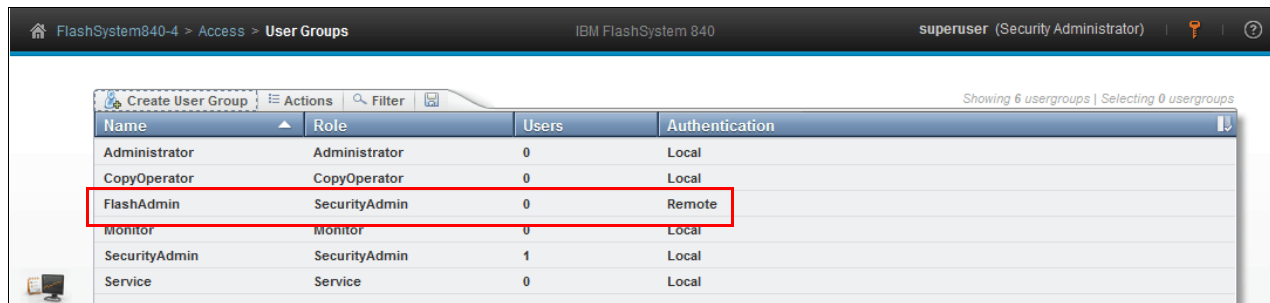
LDAP

☒ Enable for this group

Figure 7-17 Select Security Administrator

Note: If the field Remote Authentication is not visible in the Create User Groups window, remote authentication is disabled in Settings → Security.

Our new user group is created and enabled for remote authentication as shown in Figure 7-18 on page 251.



Name	Role	Users	Authentication
Administrator	Administrator	0	Local
CopyOperator	CopyOperator	0	Local
FlashAdmin	SecurityAdmin	0	Remote
Monitor	Monitor	0	Local
SecurityAdmin	SecurityAdmin	1	Local
Service	Service	0	Local

Figure 7-18 Group FlashAdmin created

Testing LDAP authentication

At this point, we can log out the user superuser and try to log in with the LDAP user. However, before we do that, the Remote Authentication window provides a capability to test LDAP.

We click **Global Actions** and select **Test LDAP Connections** as shown in Figure 7-19.

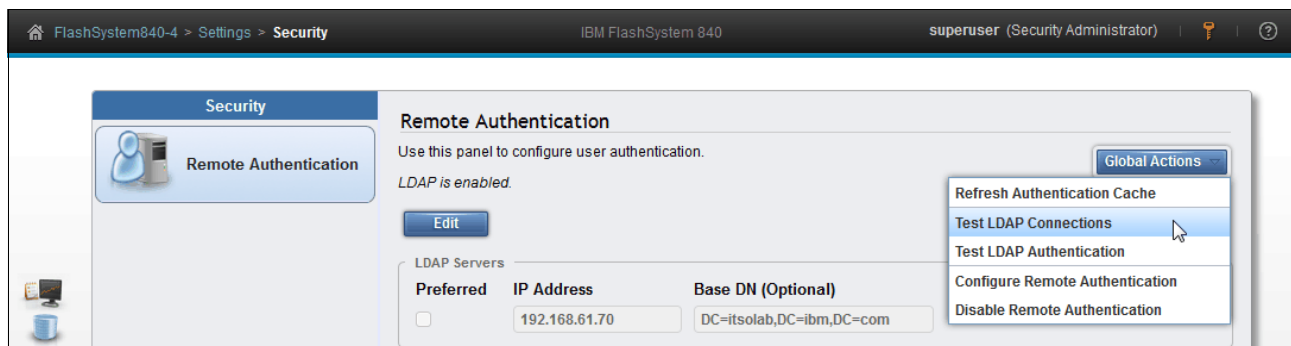


Figure 7-19 Remote Authentication: Test LDAP Connections option

The Test LDAP Connections task window opens and displays the CLI command used to test the connection. In a successful connection to the LDAP server, we get the output shown in Figure 7-20 on page 252.

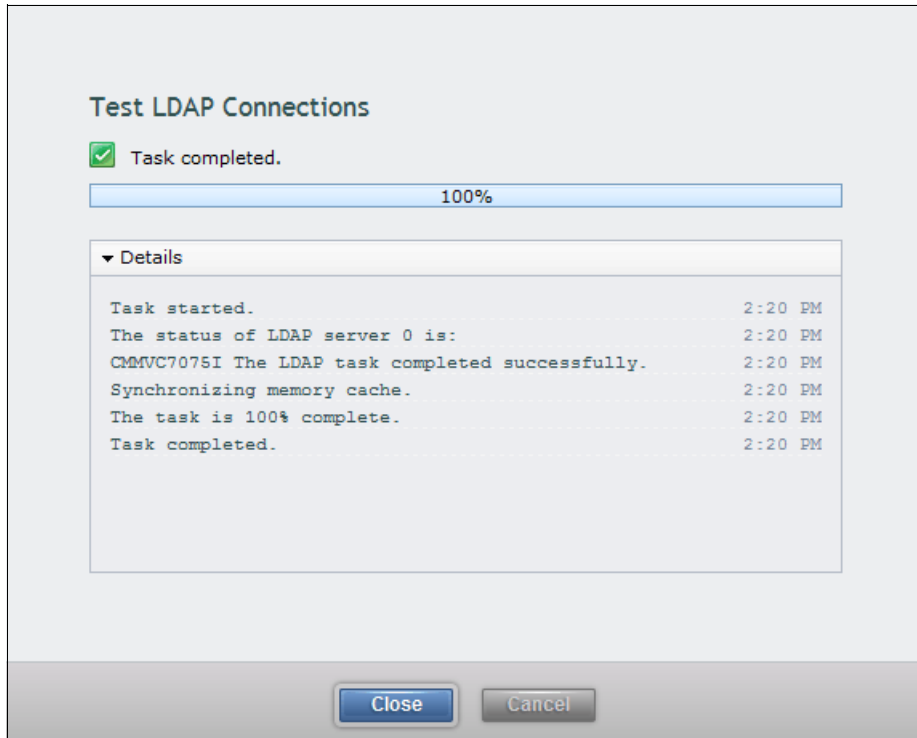


Figure 7-20 Remote Authentication: Test LDAP connections CLI result

From the Global Actions menu, we also can test whether the authentication for a specific user is functional. We click **Test LDAP Authentication** and get the window shown in Figure 7-21. We type the user credential of the LDAP user for whom we want to test authentication and click **Test**.

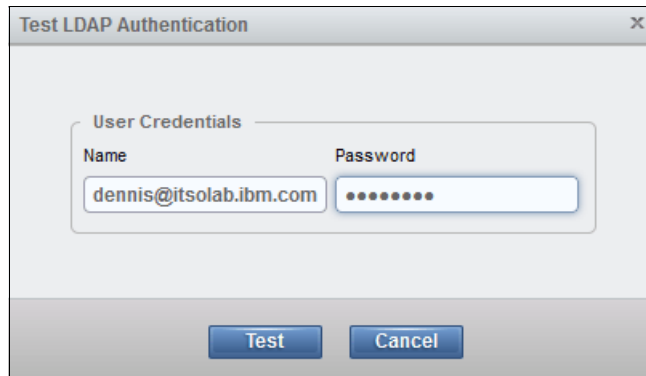


Figure 7-21 Remote Authentication: Test LDAP Authentication

When you click **Test**, the CLI command window opens. If the authentication is successful, we see the same output as shown in Figure 7-20.

If the test is unsuccessful, we see the message shown in Figure 7-22 on page 253.

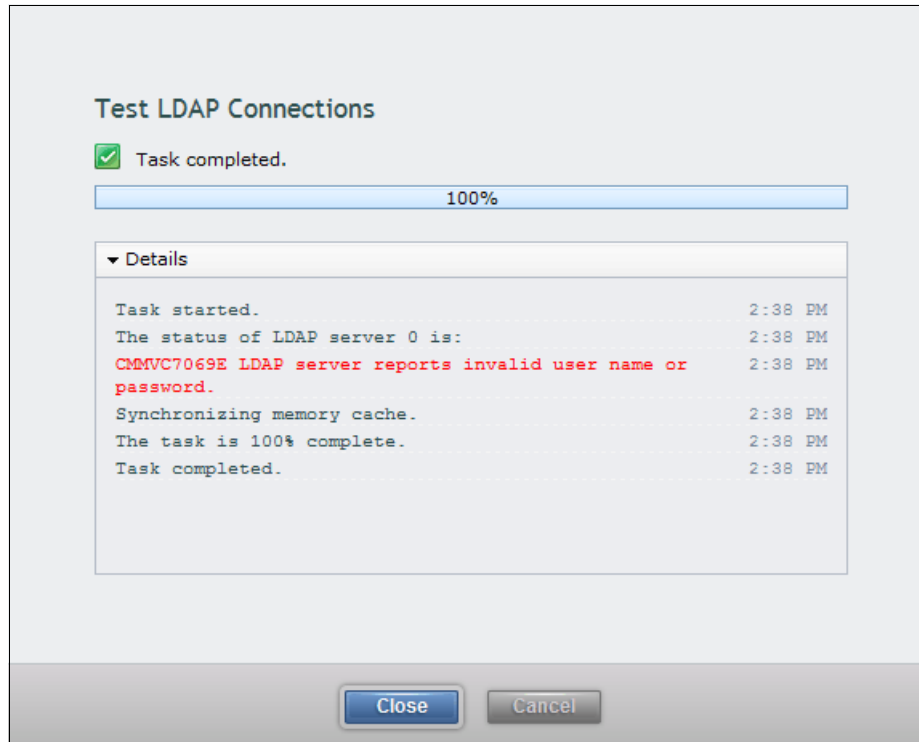


Figure 7-22 Remote Authentication: Test unsuccessful

Log in as an LDAP user

Assuming that remote authentication is successful, the superuser user can now log out and the LDAP user can log in as shown in Figure 7-23.



Figure 7-23 Login window for the LDAP user

Figure 7-24 on page 254 shows the FlashSystem 840 home window. The upper-right corner displays the user that is logged in.



Figure 7-24 Main window LDAP user logged in

Configuring remote authentication is complete.

7.1.4 Network menu

The Network menu is used for the configuration of the network setup for all the interfaces in the cluster.

Click **Settings** → **Network** to open the Network menu. You can update the network configuration, configure Service IP addresses, and view information about the Fibre Channel (FC) connections.

Management IP addresses

The *Management IP address* is the IP address of the FlashSystem 840 management interface. This interface includes the GUI and the CLI. The GUI is accessed via a web browser and the CLI is accessed via SSH using PuTTY or a similar tool. The Management IP address is a clustered IP address, which means that if any of the canisters are offline for maintenance or for any other reason, the Management IP address is still available on the surviving node.

The configured Management IP address can be reviewed or changed from the menu **Settings** → **Network** → **Management IP Address** as shown in Figure 7-25 on page 255.

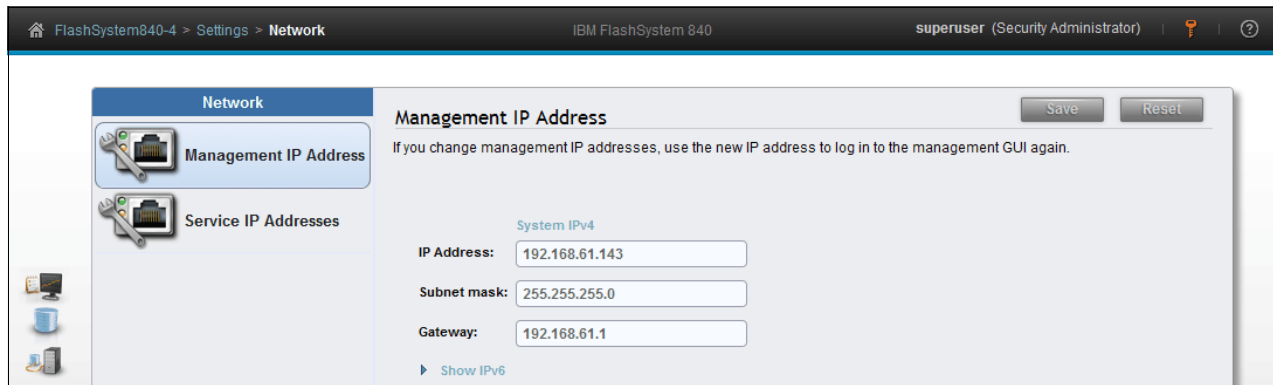


Figure 7-25 Network menu: Set management IP address

Service IP addresses

The Service IP addresses are the IP addresses of each of the two FlashSystem 840 controllers, which are called *canisters*. These canisters have their own IP addresses where several support actions can be performed, for example:

- ▶ Review installed hardware
- ▶ Place canister in the Service state
- ▶ Power cycle canister
- ▶ Identify canister
- ▶ Clear all system configuration data
- ▶ Create new cluster
- ▶ Recover a failed system (*This action is only performed by IBM Support.*)
- ▶ Update firmware manually with the controllers offline
- ▶ Extract system event logs

Note: The Service IP addresses are normally not used by the IBM FlashSystem 840 administrator. They are used *only* in troubleshooting and scheduled maintenance or when IBM Support performs certain service actions.

To configure the FlashSystem 840 Service IP addresses, click **Settings** → **Network** → **Service IP Addresses** as shown in Figure 7-26 on page 256.

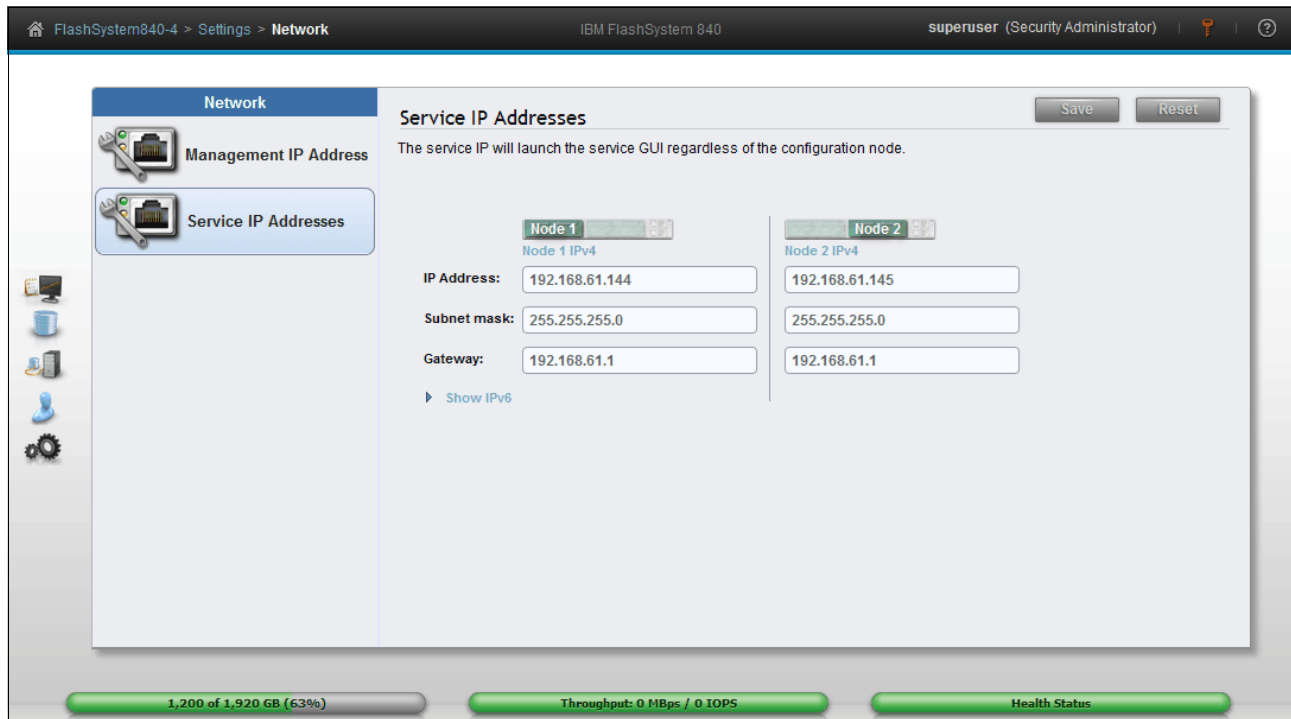


Figure 7-26 Configuring Service IP addresses

The default for setting Service IP addresses is to configure an IPv4 address for both FlashSystem 840 nodes. However, IPv6 addresses can also be set by clicking **Show IPv6** and then entering the IPv6 addresses followed by clicking **Save** (This action is not shown in Figure 7-26).

For more information about how to access and use Service Assistant Tool, see 7.2, “Service Assistant Tool” on page 271.

7.1.5 Support menu

Click **Settings** → **Support** when log files are requested by IBM Support. IBM Support often requests log files when a support case is opened by the FlashSystem 840 administrators or by the Call Home function.

The system administrator downloads the requested support package from the system and then uploads it to IBM Support. IBM Support then analyzes the data.

Download support package

To download a support package, click **Settings** → **Support** and select **Download Support Package** as shown in Figure 7-27 on page 257.

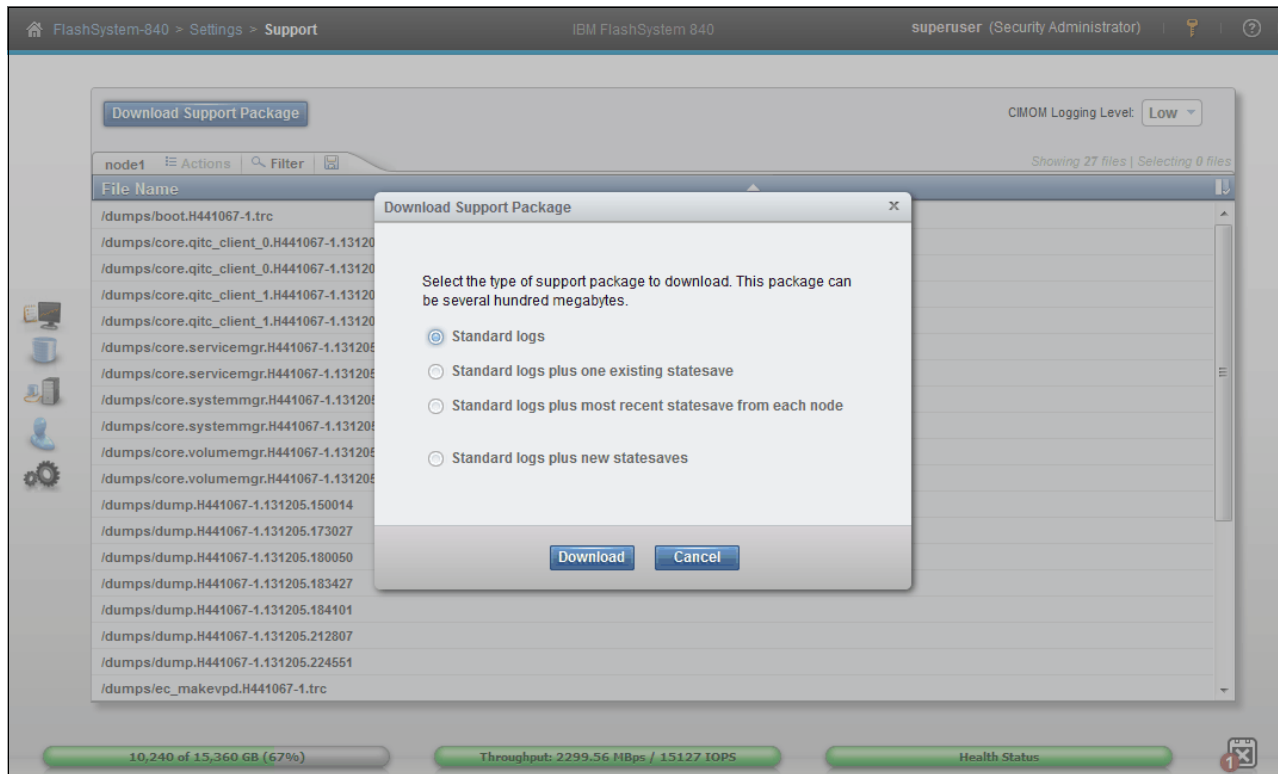


Figure 7-27 Download Support Package

IBM Support usually requests that you click **Standard logs plus new statesaves**. These logs can take from minutes to hours to download from the IBM FlashSystem 840, depending on the situation and the size of the support package that is downloaded.

The destination of the support package file is the system where the web browser was launched. Figure 7-28 shows the next step where we save the support package file on our Windows system.

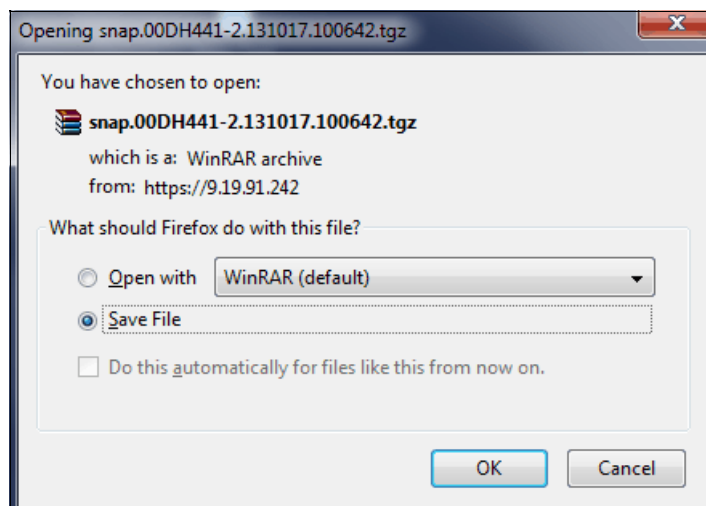


Figure 7-28 Download support package save file

IBM Support usually requests log files to be uploaded to a specific problem management record (PMR) number using EcuRep as the upload media to IBM. EcuRep is at the following URL:

<http://www.ecurep.ibm.com/app/upload>

Download individual log files

After analyzing the uploaded support package, IBM Support might request additional files. To locate these files, click **Settings** → **Support** and click **Show full log listing**. This option allows the download of specific and individual log files. An example is shown in Figure 7-29.

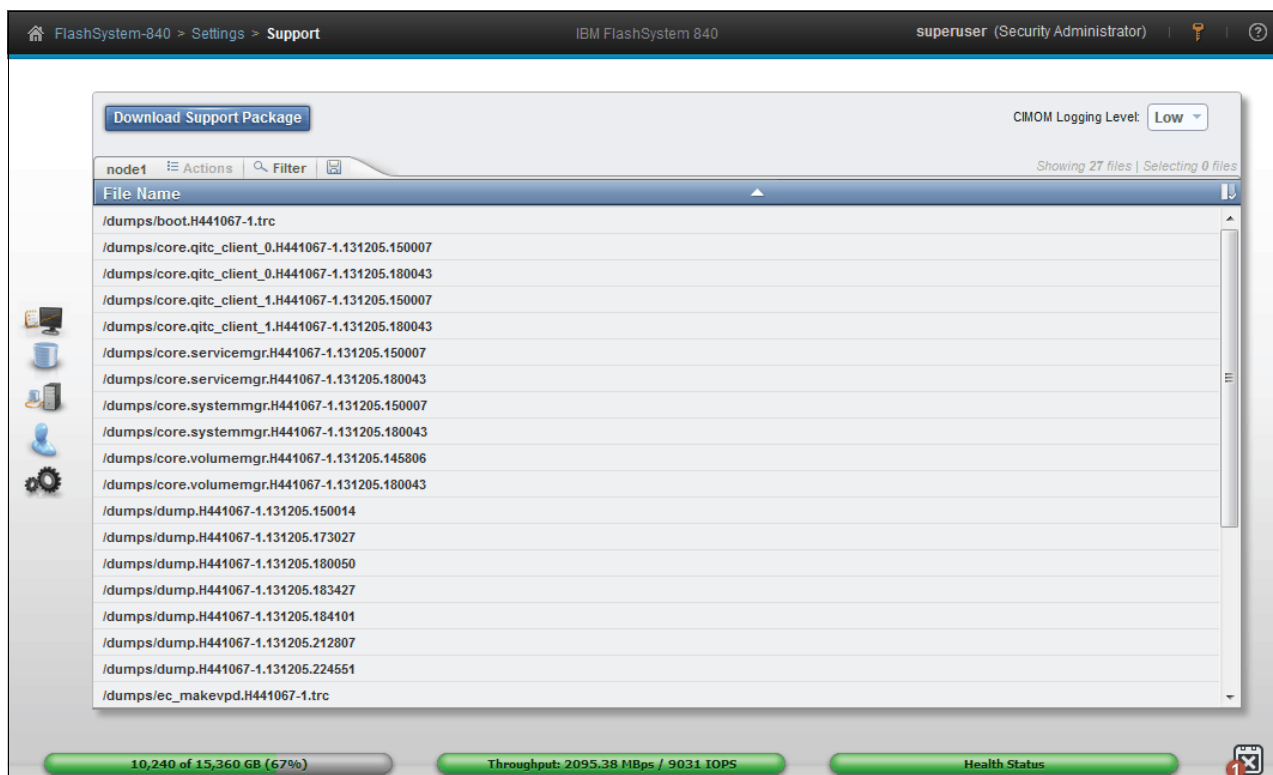


Figure 7-29 Download support: Individual files

You can download any of the various log files by selecting a single item and clicking **Actions** → **Download**. You can delete a single item in Figure 7-30 by selecting a single item and clicking **Actions** → **Delete**.

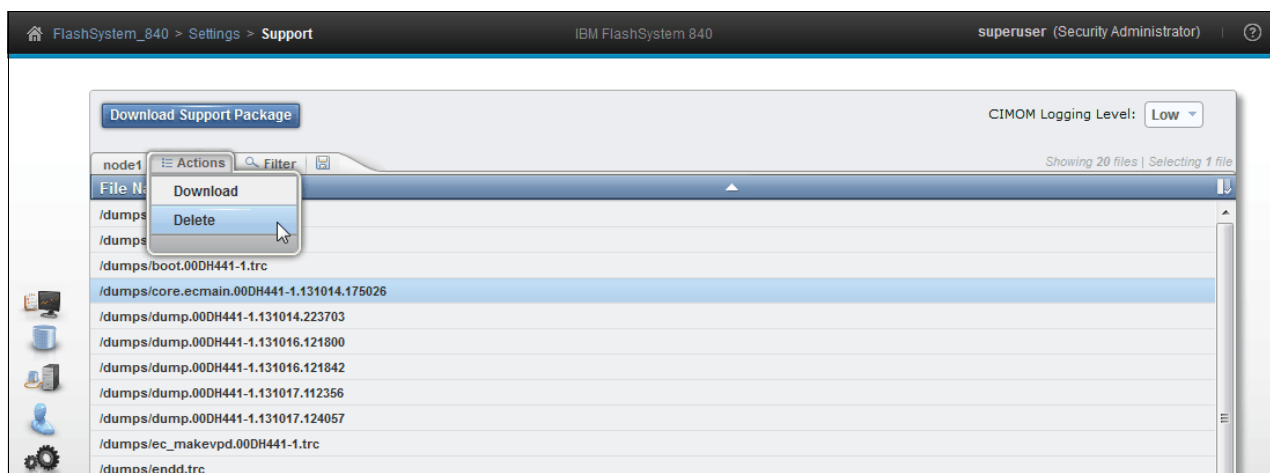


Figure 7-30 Download and Delete options of the Actions list box

Delete option: When the Delete option is not available, the file cannot be deleted because it is used by the system.

Log files are saved from each of the installed canisters, which are logically called *nodes*. In the upper left of the window, click the **node** tab to show the node1 or node2 log files (Figure 7-31).

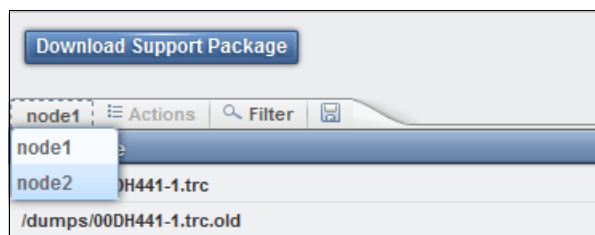


Figure 7-31 Change the log listing of the nodes canister

7.1.6 System menu

The System menu allows you to set the time and date for the cluster, enable Open Access, perform software updates for the cluster, and change the preferences for the GUI.

Date and Time option

Click the **Settings** → **System** option to open the window shown in Figure 7-32 on page 260. This window provides options to set the date and time.

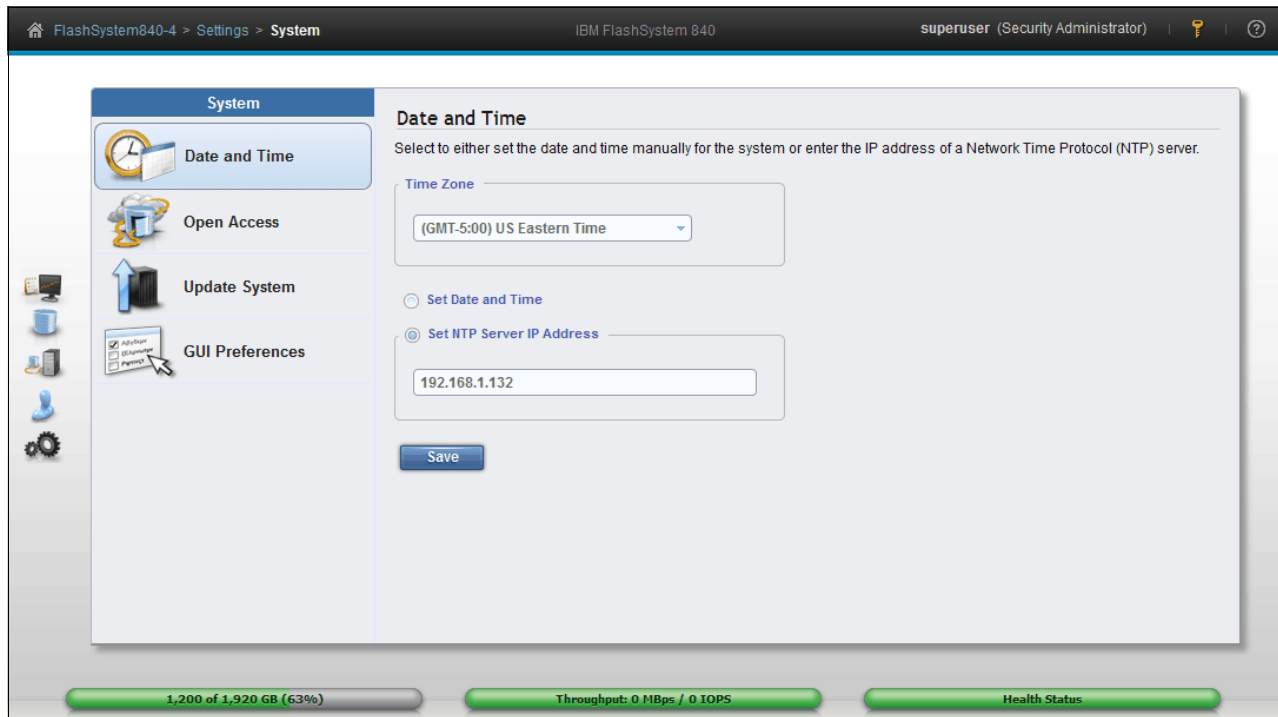


Figure 7-32 Date and time preferences

The preferred method for setting the date and time is to configure a Network Time Protocol (NTP) server. By using an NTP server, all log entries are stamped with an accurate date and time, which is important in troubleshooting. An example might be a temporarily broken FC link that caused a path failover at a connected host. To investigate the root cause of this event, logs from the host, logs from the storage area network (SAN) switches, and logs from the IBM FlashSystem 840 need to be compared. If the date and time are not accurate, events cannot be compared and matched, which makes a root cause analysis much more difficult.

Open Access

The IBM FlashSystem 840 can be configured to allow Open Access or to disallow Open Access. Open Access is feasible when the system is directly connected to a host, because then, no other hosts can connect to the system and accidentally read from or write to volumes that belong to other hosts.

Allowing Open Access can also be used in cases where the FlashSystem 840 is connected to correctly zoned SAN switches. However, disallowing Open Access and forcing the system to map its volumes only to selected hosts provides an extra layer of security.

Figure 7-33 on page 261 shows the result of clicking **Settings** → **System** → **Open Access** window.

Note: Open Access can only be enabled when no host mappings are present.

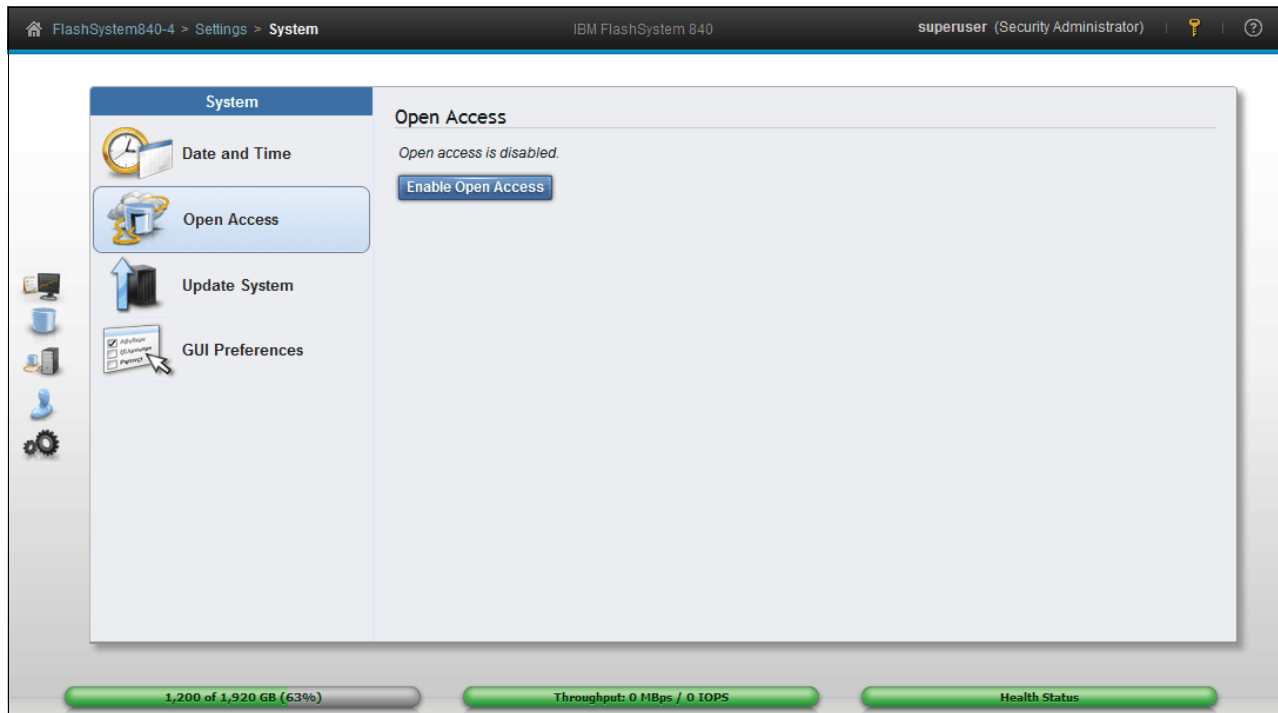


Figure 7-33 Open Access

Our system does not have any hosts defined and Open Access can be enabled. The Enable Open Access button is disabled when hosts are defined and then Open Access is not configurable.

Update software

In the following section, we demonstrate how to update firmware through the GUI of the FlashSystem 840. The firmware update that we demonstrate is from version 1.1.2.7 to version 1.1.3.2 and is therefore an update to the current and latest available firmware level. Firmware update is initiated from the **Settings** → **System menu**.

From the **Monitoring** → **System menu**, which is also called the Home window of the FlashSystem 840 GUI, right-click and click **Properties** to get information about the current firmware level as shown in Figure 7-34 on page 262.

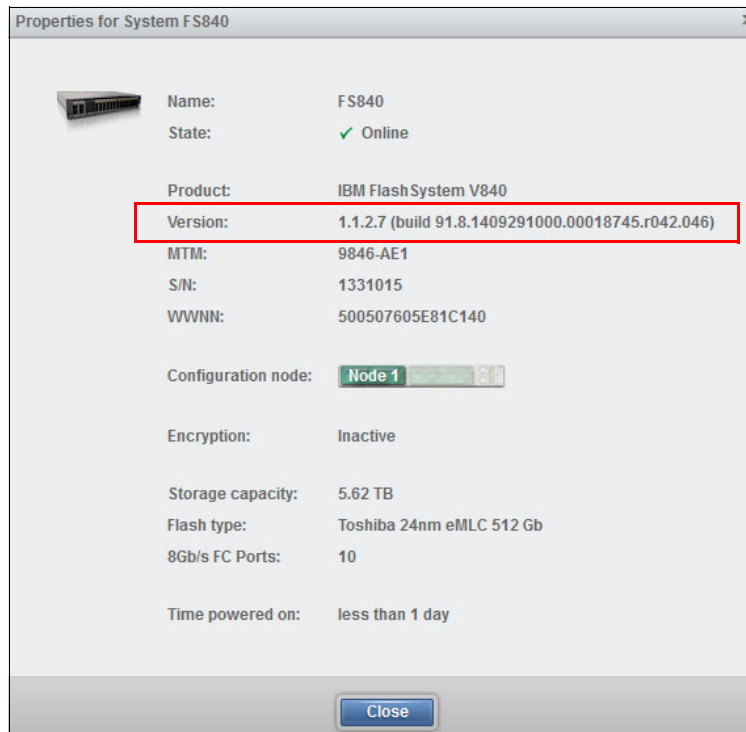


Figure 7-34 The system properties indicate the current firmware version

Note: Available for firmware update in this section was a FlashSystem V840. Regarding firmware update, the only difference is the name and product number. All steps during the firmware process are the same as FlashSystem 840.

Before starting the firmware update, the administrator must download the new firmware image file and the update test utility. A Download link is provided in the Update System window as shown in Figure 7-35 on page 263.

The latest firmware for the system can be downloaded from the Internet (if the system has access), or it can be downloaded by the administrator from the following URL:

<http://www.ibm.com/storage/support>

A firmware download requires an appropriate maintenance agreement or that the system is covered under warranty. When downloading firmware from IBM, the client must validate coverage by entering the system model number and serial number. The system model number and serial number are on the printed serial number label on the system, and they can be obtained from the GUI below the firmware version as shown in Figure 7-34.

Note: A firmware download from IBM is only possible if the system has an appropriate maintenance agreement or if the machine is under warranty.

The Settings menu allows you to update the FlashSystem 840 firmware. This update is referred to as *Concurrent Code Load* (CCL). Each node in the clustered system automatically updates in sequence while maintaining full accessibility for connected hosts.

To initiate CCL, click **Settings** → **System** → **Update System**. And then click the **Update** button as shown in Figure 7-35 on page 263.

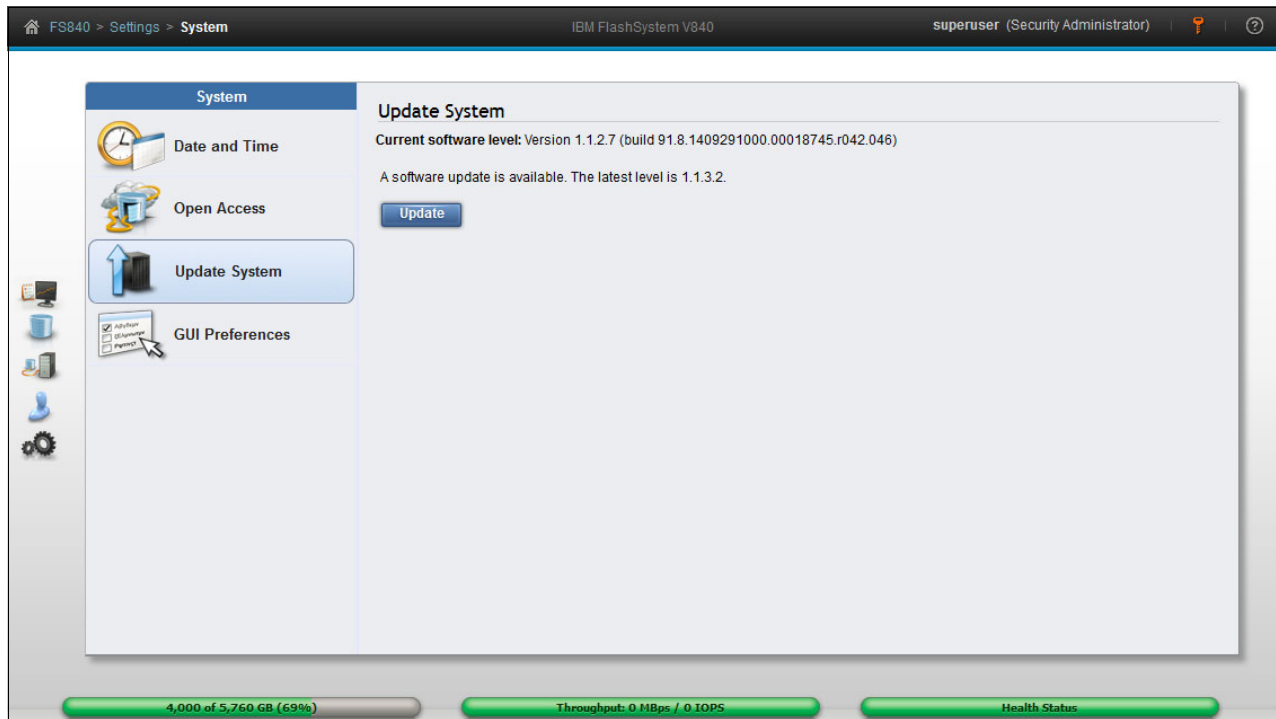


Figure 7-35 Update System

The Update System wizard begins by requesting the administrator to select the test utility and update package. The test utility checks for errors before the firmware update file can be uploaded and the update can be started. Figure 7-36 shows the Update System wizard before files are selected.

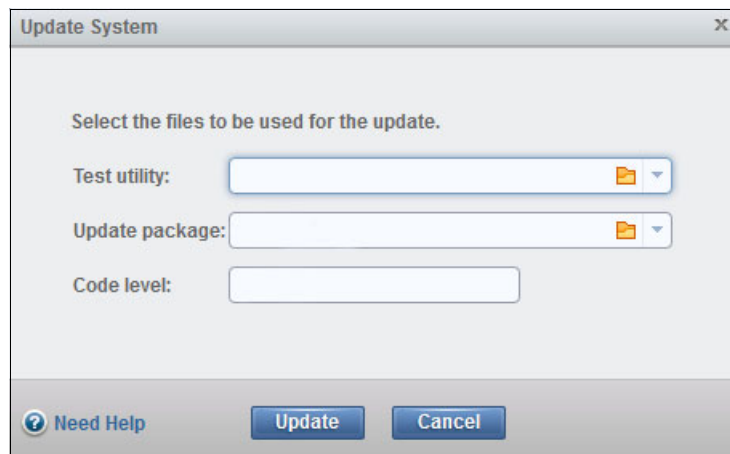


Figure 7-36 Test utility and firmware selection

Click the folder icon to locate the correct test utility and update package. The procedure for selecting the test utility file is shown in Figure 7-37 on page 264.

	Name	Date modified	Type	Size
★ Favorites				
Desktop				
Downloads				
Recent Places				
	IBM9840_INSTALL_1.1.3.2-43.28.tgz.gpg	1/20/2015 11:26 AM	GPG Encrypted File	355,187 KB
	ReleaseNotes_1.1.3.2-43.28-V840.pdf	1/20/2015 11:26 AM	Adobe Acrobat D...	255 KB
	TMS9840_INSTALL_svcupgradetest_1.4	1/20/2015 11:19 AM	4 File	19 KB
	Upgrade_Test_Utility_V1.4_Release_Notes...	1/20/2015 11:19 AM	Adobe Acrobat D...	27 KB

Figure 7-37 Update test utility file selection

The purpose of running the test utility is to verify that no errors exist and that the system is ready to update. If any issues are discovered by the test utility, the firmware update aborts with a message to the administrator of which problems to fix before the update system procedure can be repeated.

The procedure for selecting the update package is shown in Figure 7-38.

	Name	Date modified	Type	Size
★ Favorites				
Desktop				
Downloads				
Recent Places				
	IBM9840_INSTALL_1.1.3.2-43.28.tgz.gpg	1/20/2015 11:26 AM	GPG Encrypted File	355,187 KB
	ReleaseNotes_1.1.3.2-43.28-V840.pdf	1/20/2015 11:26 AM	Adobe Acrobat D...	255 KB
	TMS9840_INSTALL_svcupgradetest_1.4	1/20/2015 11:19 AM	4 File	19 KB
	Upgrade_Test_Utility_V1.4_Release_Notes...	1/20/2015 11:19 AM	Adobe Acrobat D...	27 KB

Figure 7-38 Update package file selection

As shown in Figure 7-39, we selected the appropriate files for the test utility and update package and we selected the target code level.

Update System

Select the files to be used for the update.

Test utility: TMS9840_INSTALL_svcupgradetest_1.

Update package: IBM9840_INSTALL_1.1.3.2-43.28.tgz.gp

Code level: 1.1.3.2

Need Help

Update

Cancel

Figure 7-39 Test utility and firmware selection

The system inserts the latest code level automatically, or the administrator can specify a different firmware level. In our example, we updated to 1.1.3.2.

Click **Update** to proceed. The update test utility and update package files are uploaded to the FlashSystem 840 controller nodes where after firmware update begins.

The initial part of the Update System procedure is shown in Figure 7-40 on page 265.

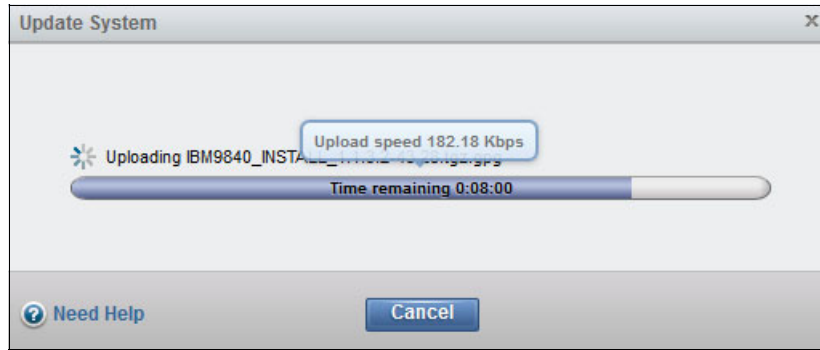


Figure 7-40 Test utility and firmware are uploading

If any errors are identified by the test utility, the administrator must resolve the errors before the firmware update can proceed. Any hardware error prevents System Update to proceed. Another example of what prevents System Update to run is if Call home is not enabled.

If an error is identified, take the correct actions to resolve the error identified by the update test utility. You can start troubleshooting in the menu **Monitoring** → **Events**. Use this menu to review and resolve any unfixed errors.

The CCL firmware update is now executing in the background. While the system updates, the progress is shown in the progress indicators as shown in Figure 7-41.

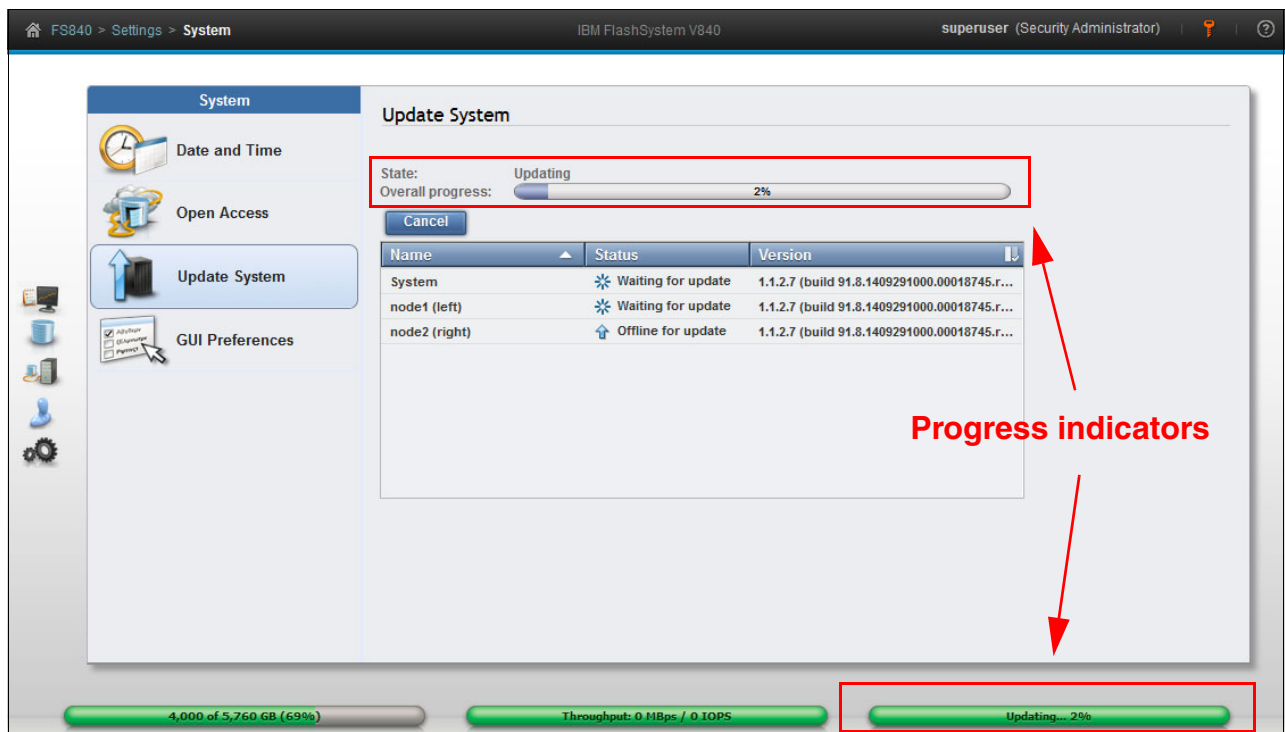


Figure 7-41 Firmware update: Node offline for update

The system can be operated normally while it is upgrading; however, no changes can be made until the firmware update completes. If trying to fix any error condition, the administrator sees a message, “Fixes cannot be applied while the system is being upgraded” as shown in Figure 7-42 on page 266.

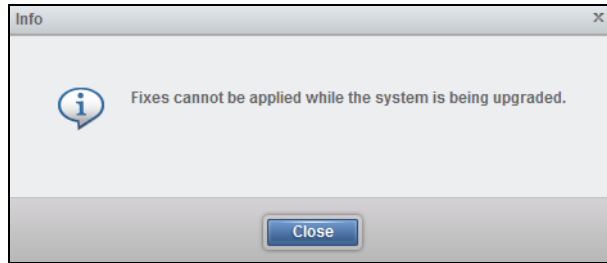


Figure 7-42 Fixes cannot be applied while upgrading

During the update, various messages show in the Update System window. In Figure 7-43, the first controller, which is node2, completes its update and the second controller is Offline for update.

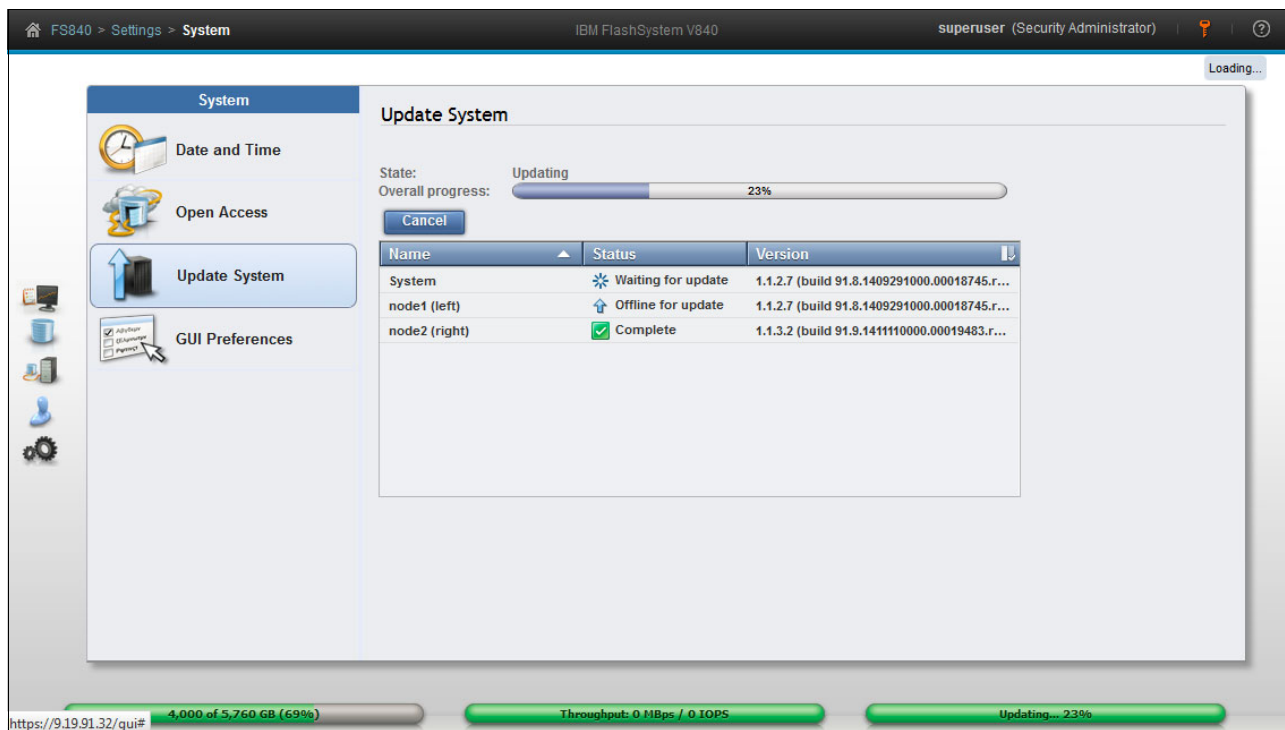


Figure 7-43 Firmware update of the first controller completes

During update, when the node that has the role of configuration node reboots, a message displays that node failover is detected. The role of the configuration node is now moved to the other controller and the browser window needs refreshing as shown in Figure 7-44 on page 267.

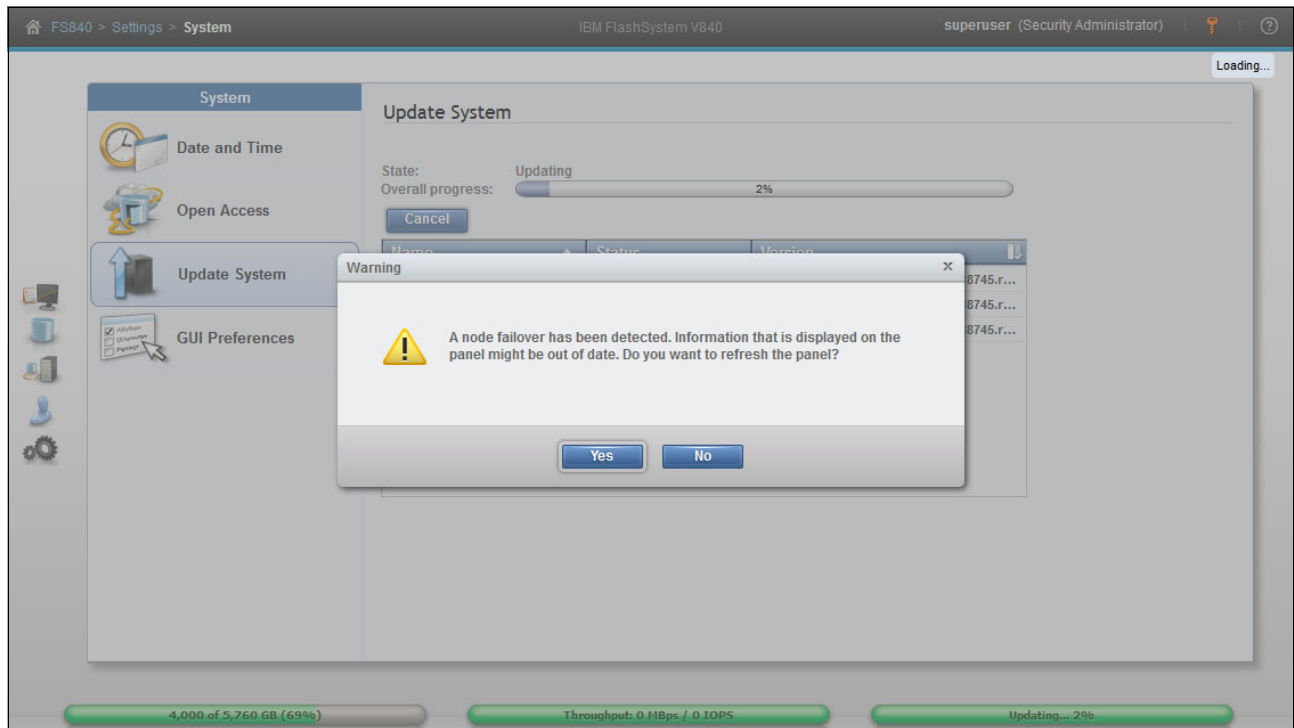


Figure 7-44 Node failover happens during update

When both controllers are updated, the FlashSystem 840 firmware update *commits* the new firmware as shown in Figure 7-45.

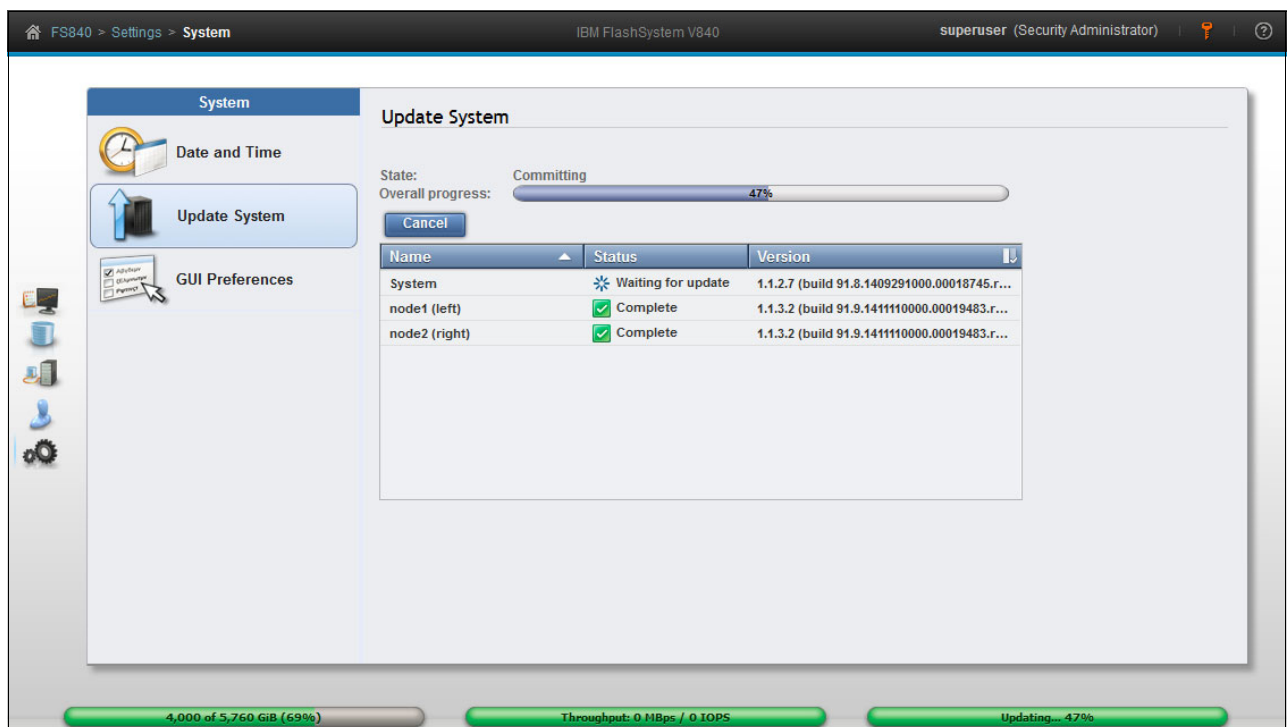


Figure 7-45 Firmware update of both controllers is complete

After committing the firmware update, the system starts the *Updating Hardware* process as shown in Figure 7-46.

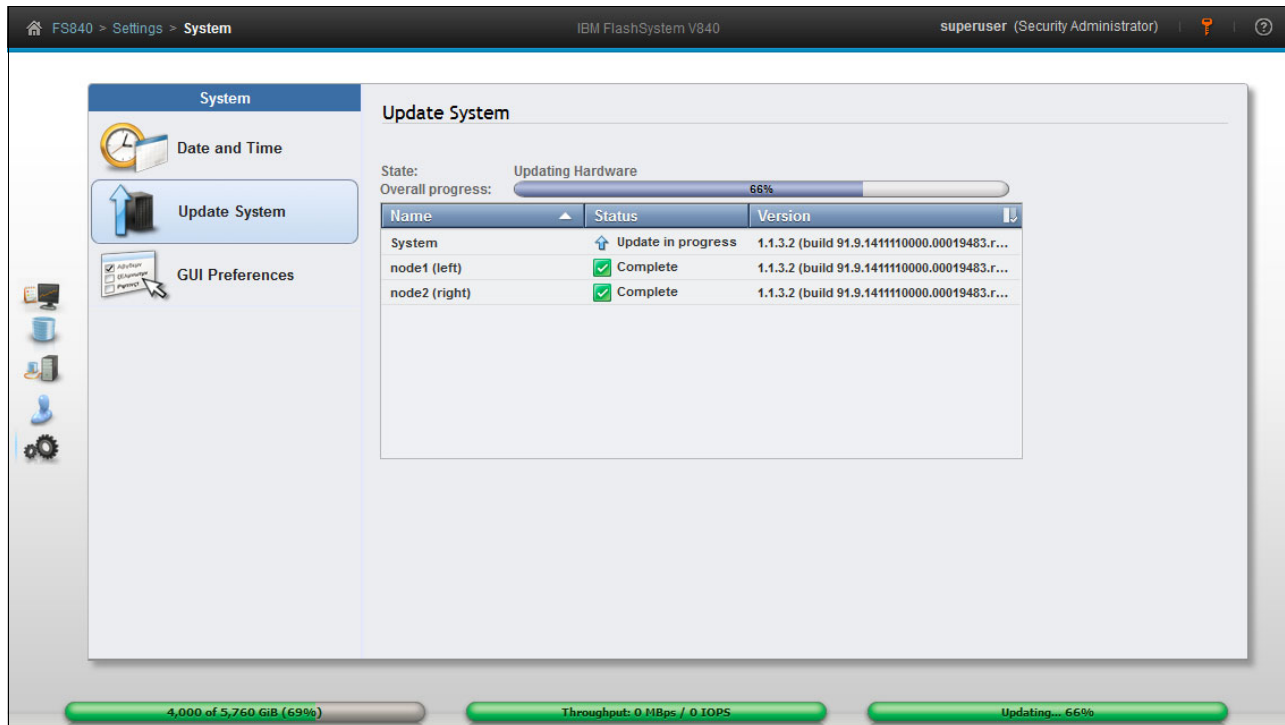


Figure 7-46 Firmware updating the hardware

During the hardware update, all individual components in the system are being firmware-updated. For example, the I/O ports are updated, during which time they are being taken offline for update one-by-one.

When the Update System wizard completes, the system returns to a healthy status. The system now has the latest firmware as shown in Figure 7-47 on page 269.

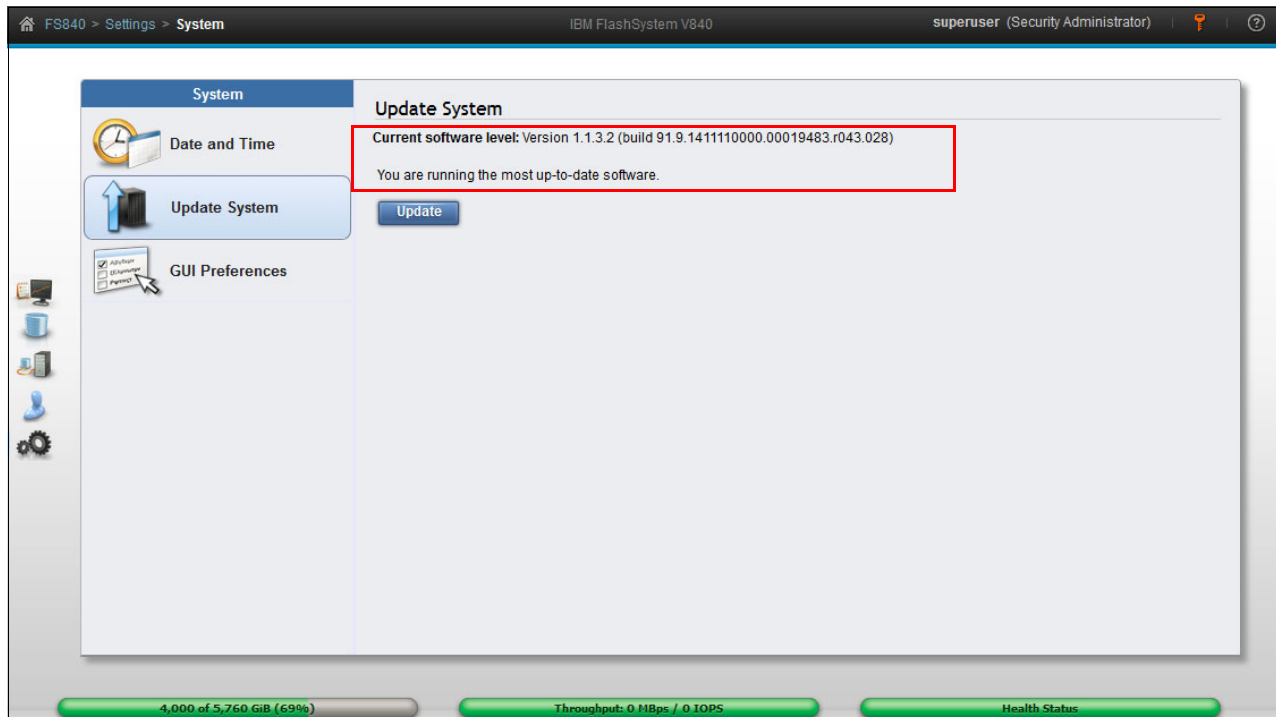


Figure 7-47 Firmware update finished

The new firmware version can be confirmed in the same way as shown in Figure 7-34 on page 262.

The system properties now display the new firmware version as shown in Figure 7-48 on page 270.

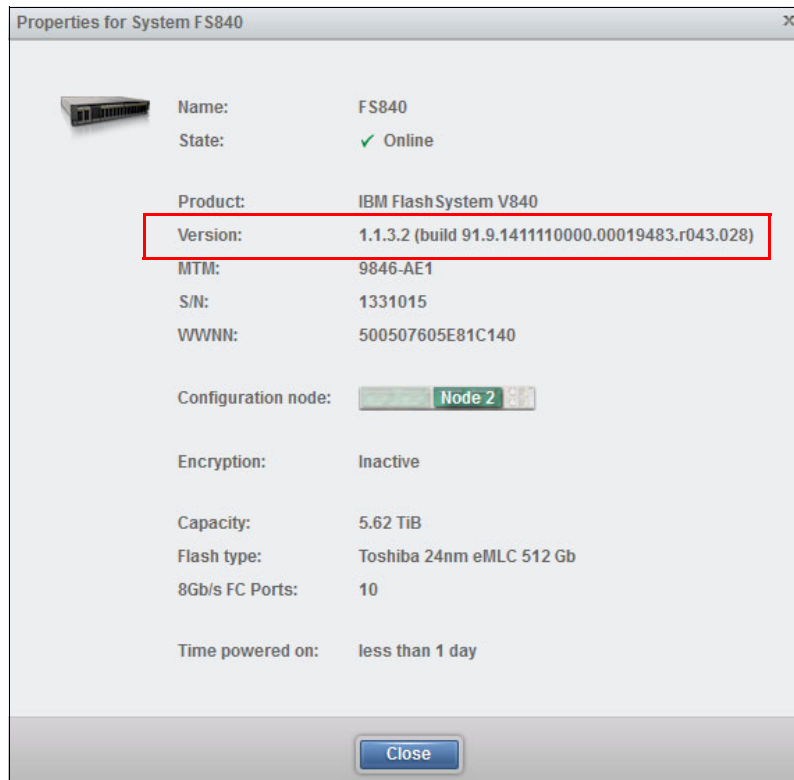


Figure 7-48 System properties with new firmware

In our example in Figure 7-48, where we updated from firmware version 1.1.2.7 to firmware version 1.1.3.2, the entire update process took approximately 2 hours to complete.

As an alternative to upgrading firmware through the GUI, you can use the FlashSystem 840 CLI instead. The process is described at the IBM Knowledge Center:

<http://ibm.co/1mGd9XP>

GUI preferences

Click **Settings** → **System** → **GUI Preferences** to change the web address of the IBM Knowledge Center, which is the help page for the IBM FlashSystem 840. This help page can be reached from any window in the management GUI by clicking the question mark (?) in the upper-right corner of the GUI as shown in Figure 7-49.

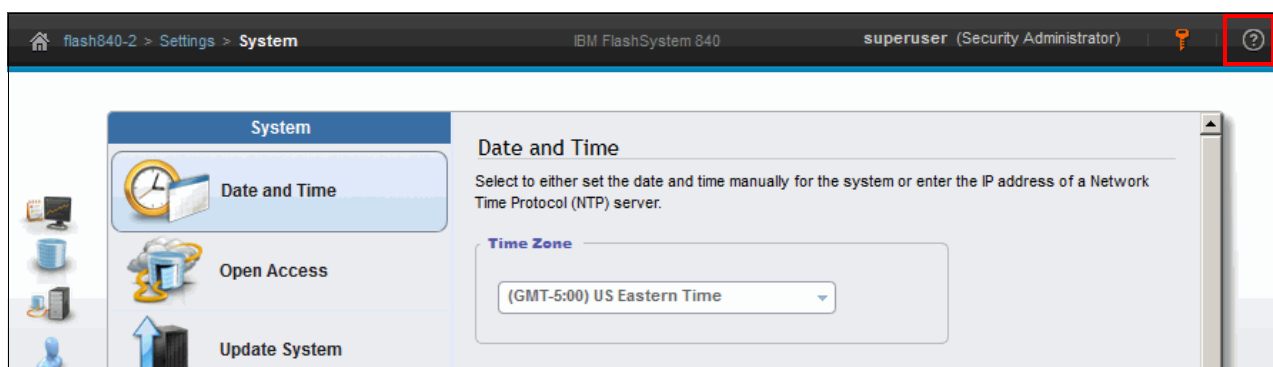


Figure 7-49 IBM Knowledge Center access

The default web address for the FlashSystem 840 IBM Knowledge Center is the Internet accessible version.

Any address can be configured in the Web Address field. If the system does not have access to the Internet, the web address can be set to access a local version of the FlashSystem 840 IBM Knowledge Center, as shown in Figure 7-50.

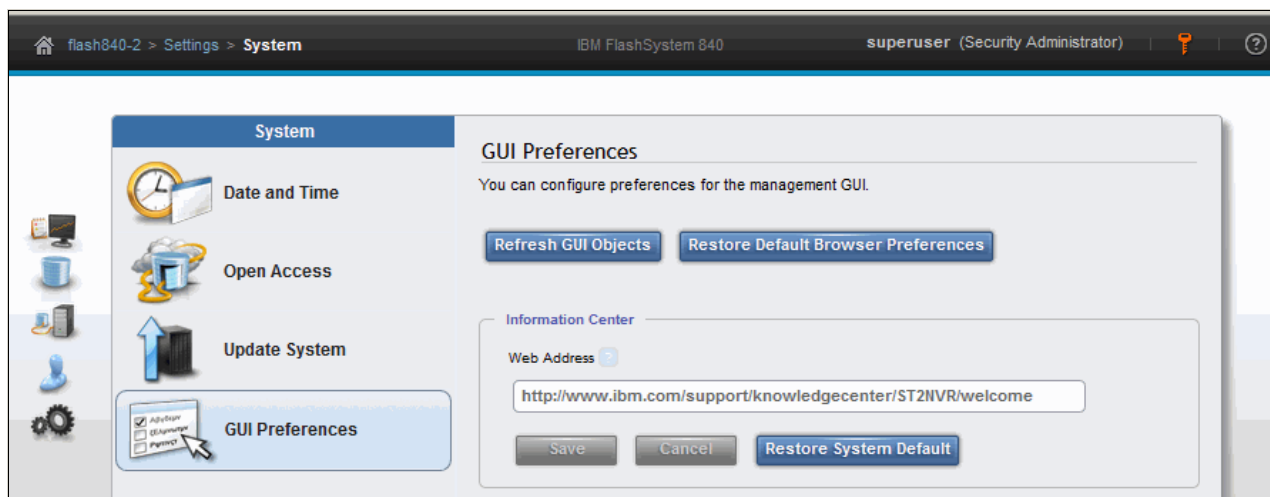


Figure 7-50 System menu: GUI Preferences

The FlashSystem 840 IBM Knowledge Center is at the following website:

http://www-01.ibm.com/support/knowledgecenter/ST2NVR_1.3.0/com.ibm.storage.flashsystem.1.3.doc/FlashSystem_840_welcome.htm

7.2 Service Assistant Tool

Service Assistant Tool is used for troubleshooting or when an IBM Support engineer directs you to use it.

7.2.1 Accessing Service Assistant Tool

Service Assistant Tool is accessed with a web browser. An example of getting access is to point your browser to the cluster management IP address followed by /service. The following URL is an example:

`https://192.168.10.10/service`

Each of the canister's service IP addresses can also be reached. There are different options for getting access to Service Assistant Tool.

The following example shows which IP addresses are configured and how they are accessed:

- ▶ 192.168.10.10: Service IP address for Canister 1:
 - `https://192.168.10.10/service` opens Service Assistant Tool for Canister 1.
 - `https://192.168.10.10` opens Service Assistant Tool for Canister 1.

- ▶ 192.168.10.11: Service IP address for Canister 2 (configuration node):
 - <https://192.168.10.11/service/> opens Service Assistant Tool for Canister 2.
 - <https://192.168.10.11> opens the cluster management GUI.
- ▶ 192.168.10.12 - Cluster IP address:
 - <https://192.168.10.12/service> opens Service Assistant Tool for the configuration node.
 - <https://192.168.10.12> opens the cluster management GUI.

Note: Canisters are named Canister 1 (view from rear left side) and Canister 2 (view from rear right side). The logical names for canisters in the Service Assistant Tool are node 1 and node 2. Which is node 1 and which is node 2 depends on the actual configuration and in which order the controllers were added to the cluster. If Canister 2 was added first, it gets the logical name node 1. There is, therefore, no direct relation between the canister number and node number.

7.2.2 Log in to Service Assistant Tool

The login window of IBM FlashSystem 840 Service Assistant Tool only allows the superuser user to log in; therefore, the user name cannot be changed. The Service Assistant Tool login window is shown in Figure 7-51.



Figure 7-51 Service Assistant Tool login

After you type the password for the superuser, you get to the Service Assistant Tool as shown in Figure 7-52 on page 273.

IBM FlashSystem 840 Service Assistant Tool

Connected to: 01 | 1 | node1

Log out

IBM

Current: 01 | 1 | node1

Status: Active

Identify

Home

Collect Logs

Manage System

Recover System

Re-install Software

Configure Enclosure

Change Service IP

Configure CLI Access

Restart Service

Home

You can view detailed status and error summary, and manage service actions for the current node. The current node is the node on which service-related actions are performed. The connected node displays the service assistant and provides the interface for working with other nodes on the system. To manage a different node, select a node from the following table.

Attention:

Only perform service actions on nodes when directed by service procedures. If used inappropriately, service actions can cause a loss of access to data, or even data loss. If the node status is active, select Monitoring-->Events in the management GUI to fix any errors that are related to the active node.

Actions:

Enter Service State

GO

Change Node

Node Name	Node Status	Error	Panel	System	Relationship
node1	Active		01-1	flash840-2	Local
node2	Active		01-2	flash840-2	Partner

Refresh

Node Errors

Node Detail

Node	Hardware	Access	Location	Ports
Node ID:	1			
Node Name:	node1			
Node Status:	Active			
Part Identity:	11S00DH404YS18WD41P156			
Node FRU:	00DH520			
Configuration Node:	Yes			
Model:	TR1			
System:	flash840-2			
System Software Build:				
Software Version:	1.1.3.2			
Software Build:	91.9.1411110000.00019483.r043.028			
Console IP:	192.168.62.16:443			
Has File Module Key:	No			

Figure 7-52 Service Assistant Tool main page

The Home window of Service Assistant Tool shows various options for examining installed hardware and revision levels and for identifying canisters or placing these canisters into the service state.

Note: Be careful when you open Service Assistant Tool. Incorrect usage might cause unattended downtime or even data loss. Only use Service Assistant Tool when IBM Support asks you to use it.

Chapter 7. Configuring settings 273



Product integration

This chapter covers the integration of the IBM FlashSystem 840 storage with the IBM SAN Volume Controller, which delivers the functionality of IBM Spectrum Virtualization, and IBM Storwize V7000. It provides an overview of the main concepts of the products involved, detailed usage considerations, port assignment, port masking, and host connectivity. Additionally, common usage scenarios are described. Throughout, suggestions and preferred practices are identified, where applicable.

These topics are covered:

- ▶ Running the FlashSystem 840 with Spectrum Virtualize - SAN Volume Controller
- ▶ Integrating FlashSystem 840 and IBM Storwize V7000 considerations

Note: Some of the following sections mention IBM SAN Volume Controller, which delivers the functions of IBM Spectrum Virtualize, part of the IBM Spectrum Storage family.

IBM Spectrum Virtualize is industry-leading storage virtualization that enhances existing storage to improve resource utilization and productivity in order to achieve a simpler, more scalable, and cost-efficient IT infrastructure.

The functionality of IBM Spectrum Virtualize is provided by IBM SAN Volume Controller.

For more information, see the following website:

<http://www.ibm.com/systems/storage/software/virtualization/svc/index.html>

8.1 Running the FlashSystem 840 with Spectrum Virtualize - SAN Volume Controller

IBM FlashSystem 840 is all about being fast and resilient to minimize latency by using IBM FlashCore™ hardware-accelerated architecture, IBM MicroLatency modules, and many other advanced flash management features and capabilities.

For clients who want advanced software features, integrating the IBM FlashSystem 840 and the IBM SAN Volume Controller, which delivers the functions of IBM Spectrum Virtualize, provide an enterprise-class solution. This solution integrates and provides functionalities and services, such as mirroring, FlashCopy, thin provisioning, Real-time Compression (RtC), and broader host support. The best way to achieve all of that function is by deploying the IBM FlashSystem 840 behind a SAN Volume Controller. For clients who need efficiency, this combination also can be used with IBM Easy Tier, with SAN Volume Controller automatically promoting hot blocks to the FlashSystem. The tiered storage solution not only efficiently uses the FlashSystem to increase performance in critical applications, but it also can reduce costs by migrating less critical data to less expensive media.

You can also order IBM FlashSystem V9000, which is a comprehensive all-flash enterprise storage solution. FlashSystem V9000 delivers the full capabilities of IBM FlashCore technology plus a rich set of storage virtualization features.

The FlashSystem V9000 improves business application availability and delivers greater resource utilization so that you can get the most from your storage resources, and achieve a simpler, more scalable, and more cost-efficient IT infrastructure. Using IBM Storwize family functions, management tools, and interoperability, this product combines the performance of the FlashSystem architecture with the advanced functions of software-defined storage to deliver performance, efficiency, and functions that meet the needs of enterprise workloads demanding IBM MicroLatency response time. For product details about the FlashSystem V9000, see the IBM Redbooks publication, *IBM FlashSystem V9000 Product Guide*, TIPS1281:

<http://www.redbooks.ibm.com/abstracts/tips1281.html?Open>

For product details about the FlashSystem V840, see the IBM Redbooks Product Guide, *IBM FlashSystem V840*, TIPS1158:

<http://www.redbooks.ibm.com/abstracts/tips1158.html?Open>

In the next topics, we show how to configure IBM FlashSystem 840 to provide storage to SAN Volume Controller and show how they are designed to operate seamlessly together, reducing management effort. At the time of the writing of this chapter, SAN Volume Controller software version 7.4 was available.

8.1.1 IBM System Storage SAN Volume Controller introduction

IBM System Storage SAN Volume Controller is a storage virtualization solution that helps to increase the utilization of existing storage capacity and to centralize the management of multiple controllers in an open system storage area network (SAN) environment.

SAN Volume Controller (machine type 2145 and accompanying software) supports attachment to both IBM and non-IBM storage systems. For the most up-to-date SAN Volume Controller supported hardware list, see the Supported Hardware List, Device Driver, and Firmware Levels & Supported Software for SAN Volume Controller at this website:

<http://www.ibm.com/support/docview.wss?uid=ssg1S1003658>

SAN Volume Controller enables storage administrators to reallocate and scale storage capacity and make changes to underlying storage systems without disruption to applications. SAN Volume Controller also provides the ability to simplify storage infrastructure, use storage resources more efficiently, improve personnel productivity, and increase application availability.

SAN Volume Controller pools storage volumes from IBM and non IBM disk arrays into a single reservoir of capacity, which can be managed from a central point. SAN Volume Controller also allows data to be migrated between heterogeneous disk arrays without disruption to applications. By moving copy services functionality into the network, SAN Volume Controller allows you to use a standardized suite of copy services tools that can be applied across the entire storage infrastructure, irrespective of storage vendor restrictions that normally apply for the individual disk controllers in use.

Additionally, SAN Volume Controller adds functions to the infrastructure that might not be present in each virtualized subsystem. Examples include thin provisioning, automated tiering, volume mirroring, and data compression.

Note: Some of SAN Volume Controller functions mentioned are included in the base virtualization license, although for other functions, an extra license might need to be purchased. Contact your IBM representative for assistance about additional licenses.

SAN Volume Controller design overview

The IBM System Storage SAN Volume Controller is designed to handle the following tasks:

- ▶ Combine storage capacity from multiple vendors into a single repository of capacity with a central management point
- ▶ Help increase storage utilization by providing host applications with more flexible access to capacity
- ▶ Help improve the productivity of storage administrators by enabling the management of combined storage volumes from a single, easy to use interface
- ▶ Support improved application availability by insulating host applications from changes to the physical storage infrastructure
- ▶ Enable a tiered storage environment, in which the cost of storage can be better matched to the value of data
- ▶ Support advanced copy services, from higher-cost devices to lower-cost devices and across subsystems from multiple vendors

SAN Volume Controller combines hardware and software into a comprehensive, modular appliance. By using Intel processor-based servers in highly reliable clustered pairs, SAN Volume Controller has no single point of failure. SAN Volume Controller software forms a highly available cluster that is optimized for performance and ease of use.

Storage utilization

SAN Volume Controller is designed to help increase the amount of storage capacity that is available to host applications. By pooling the capacity from multiple disk arrays within the SAN, it enables host applications to access capacity beyond their island of SAN storage. The Storage Networking Industry Association (SNIA) estimates that open systems disk utilization in a non-virtualized environment is only between 30 - 50%. With storage virtualization, this utilization can grow up to 80% on average.

Scalability

A SAN Volume Controller configuration can start with a single I/O group. An *I/O group* is a pair of high performance, redundant Intel processor-based servers, referred to as *nodes* or *storage engines*. Highly available I/O groups are the basic configuration of a cluster. Adding additional I/O groups (a nondisruptive process) can help increase cluster performance and bandwidth.

SAN Volume Controller can scale out to support up to four I/O groups. SAN Volume Controller supports up to 2048 host servers. For every cluster, SAN Volume Controller supports up to 8192 volumes, each one up to 256 TB, and a total virtualized capacity of up to 32 petabytes (PB).

Enhanced stretched cluster configurations provide highly available, concurrent access to a single copy of data from data centers up to 300 km apart and enable nondestructive storage and virtual machine mobility between data centers.

Note: For the most up-to-date SAN Volume Controller configuration limits, see the Configuration Limits and Restrictions website for the latest SAN Volume Controller version:

<http://www.ibm.com/support/docview.wss?uid=ssg1S1003658>

This configuration flexibility means that SAN Volume Controller configurations can start small with an attractive price to suit smaller environments or pilot projects, and then can grow with your business to manage large storage environments.

Management

SAN Volume Controller is managed at the cluster level, providing a single point of control over all the managed storage. SAN Volume Controller provides a comprehensive, easy-to-use graphical user interface (GUI) for central management. This simple interface incorporates the Storage Management Initiative Specification (SMI-S) application programming interface (API), and further demonstrates the commitment of IBM to open standards. SAN Volume Controller cluster can also be managed and monitored through a comprehensive command-line interface (CLI) via Secure Shell (SSH), enabling the use of scripts for automated repeatable operations.

The SAN Volume Controller GUI is designed for ease of use and includes many built-in IBM guidelines that simplify storage provisioning and enable new clients to get started quickly with a rapid learning curve.

Clients using IBM Spectrum Control, and IBM Systems Director can take further advantage of integration points with IBM Spectrum Virtualize, which delivers the functionality of SAN Volume Controller. Managing SAN Volume Controller under IBM Spectrum Control, enables management of the most common day-to-day activities for SAN Volume Controller without ever needing to leave the IBM Spectrum Control user interface.

For historic performance and capacity management from the perspectives of both the host and the virtualized storage devices, IBM Spectrum Control helps clients with an end-to-end view and control of the virtualized storage infrastructure. Regarding data protection, Tivoli Storage FlashCopy Manager helps integrate SAN Volume Controller FlashCopy function with major applications for consistent backups and restores.

Linking infrastructure performance to business goals

By pooling storage into a single pool, SAN Volume Controller helps insulate host applications from physical changes to the storage pool, removing disruption. SAN Volume Controller simplifies storage infrastructure by including a dynamic data-migration function, allowing for online volume migration from one device to another. By using this function, administrators can reallocate, scale storage capacity, and apply maintenance to storage subsystems without disrupting applications, increasing application availability.

With SAN Volume Controller, your business can build an infrastructure from existing assets that is simpler to manage, easier to provision, and can be changed without impact to application availability. Businesses can use their assets more efficiently and measure the improvements. They can allocate and provision storage to applications from a single view and know the effect on their overall capacity instantaneously. They can also quantify improvements in their application availability to enable better quality of service goals. These benefits help businesses manage their costs and capabilities more closely, linking the performance of their infrastructure to their individual business goals.

Tiered storage

In most IT environments, inactive data makes up the bulk of stored data. SAN Volume Controller helps administrators control storage growth more effectively by moving low-activity or inactive data into a hierarchy of lower-cost storage. Administrators can free disk space on higher-value storage for more important, active data.

Tiered storage is achieved by easily creating various groups of storage, or *storage pools*, corresponding to underlying storage with various characteristics. Examples are speed and reliability. With SAN Volume Controller, you can better match the cost of the storage that is used to the value of data placed on it.

Technology for an on-demand environment

Businesses are facing growth in critical application data that is supported by complex heterogeneous storage environments, while their staffs are overburdened. SAN Volume Controller is one of many offerings in the IBM System Storage portfolio that are essential for an on-demand storage environment. These offerings can help you to simplify your IT infrastructure, manage information throughout its lifecycle, and maintain business continuity.

8.1.2 SAN Volume Controller architecture and components

SAN-based storage is managed by SAN Volume Controller in one or more *I/O groups (pairs)* of SAN Volume Controller *nodes*, referred to as a *clustered system*. These nodes are attached to the SAN fabric, with storage controllers and host systems. The SAN fabric is zoned to allow SAN Volume Controller to “see” the storage controllers, and for the hosts to “see” SAN Volume Controller.

The hosts are not allowed to “see” or operate on the same physical storage (logical unit number (LUN)) from the storage controllers that are assigned to SAN Volume Controller, and all data transfer happens through SAN Volume Controller nodes. This design is commonly described as *symmetric virtualization*.

Storage controllers can be shared between SAN Volume Controller and direct host access if the same LUNs are not shared, and both types of access use compatible multipathing drives in the same host or operating system instance. The zoning capabilities of the SAN switch must be used to create distinct zones to ensure that this rule is enforced.

Figure 8-1 shows a conceptual diagram of a storage environment using SAN Volume Controller. It shows several hosts that are connected to a SAN fabric or LAN with SAN Volume Controller nodes and the storage subsystems that provide capacity to be virtualized. In practical implementations that have high-availability requirements (most of the target clients for SAN Volume Controller), the SAN fabric “cloud” represents a redundant SAN. A redundant SAN consists of a fault-tolerant arrangement of two or more counterpart SANs, providing alternate paths for each SAN-attached device.

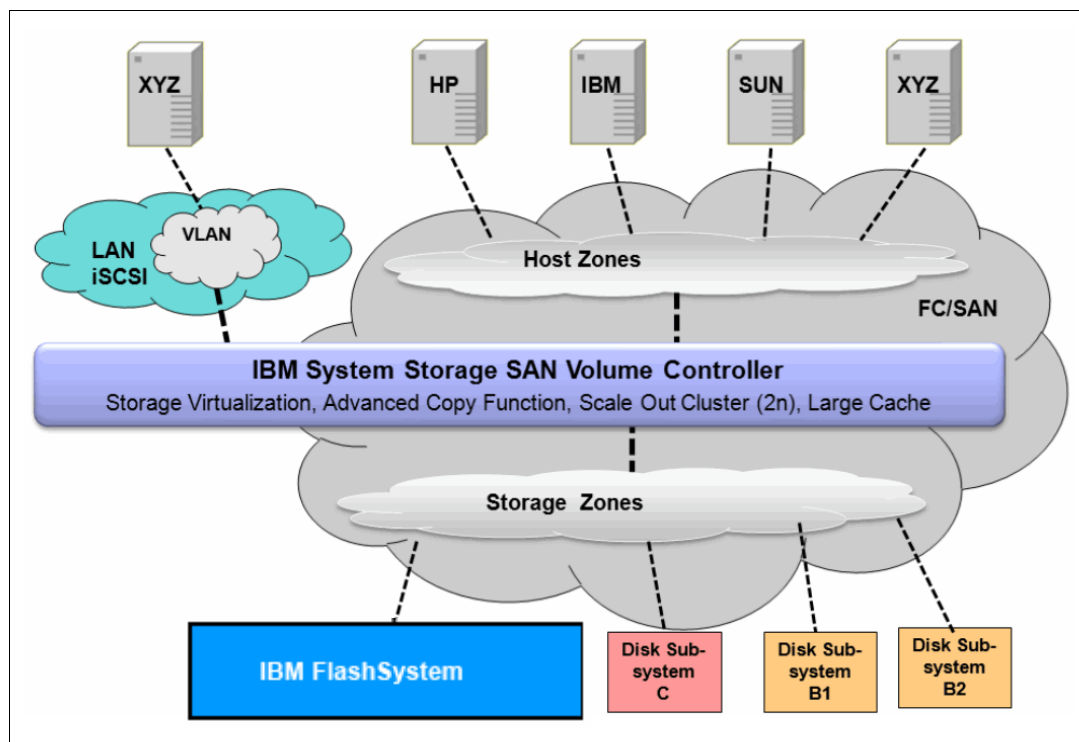


Figure 8-1 Conceptual diagram of SAN Volume Controller and the SAN infrastructure

Both scenarios (using a single network and using two physically separate networks) are supported for Internet Small Computer System Interface (iSCSI)-based and LAN-based access networks to SAN Volume Controller. Redundant paths to volumes can be provided in both scenarios. For iSCSI-based access, using two networks and separating iSCSI traffic within the networks by using a dedicated virtual local area network (VLAN) for storage traffic prevent any IP interface, switch, or target port failure from compromising the host servers' access to the volumes.

A *clustered system* of SAN Volume Controller nodes that are connected to the same fabric presents logical disks or *volumes* to the hosts. These volumes are created from managed LUNs or *managed disks* (MDisks) that are presented to SAN Volume Controller by the storage subsystems and grouped in *storage pools*. Two distinct zone sets are shown in the fabric:

- ▶ *Host zones*, in which the hosts can see and address SAN Volume Controller nodes and access volumes
- ▶ *Storage zones*, in which SAN Volume Controller nodes can see and address the MDisks/LUNs that are presented by the storage subsystems

Figure 8-2 shows the logical architecture of SAN Volume Controller, illustrating how different storage pools are built grouping MDisks, and how the volumes are created from those storage pools and presented to the hosts through I/O groups (pairs of SAN Volume Controller nodes). In this diagram, Vol12, Vol17, and Vol18 are mirrored volumes, or volumes with two copies, with each copy residing in a different storage pool. For more information about volume mirroring, see “Mirroring/Copy Services” on page 285.

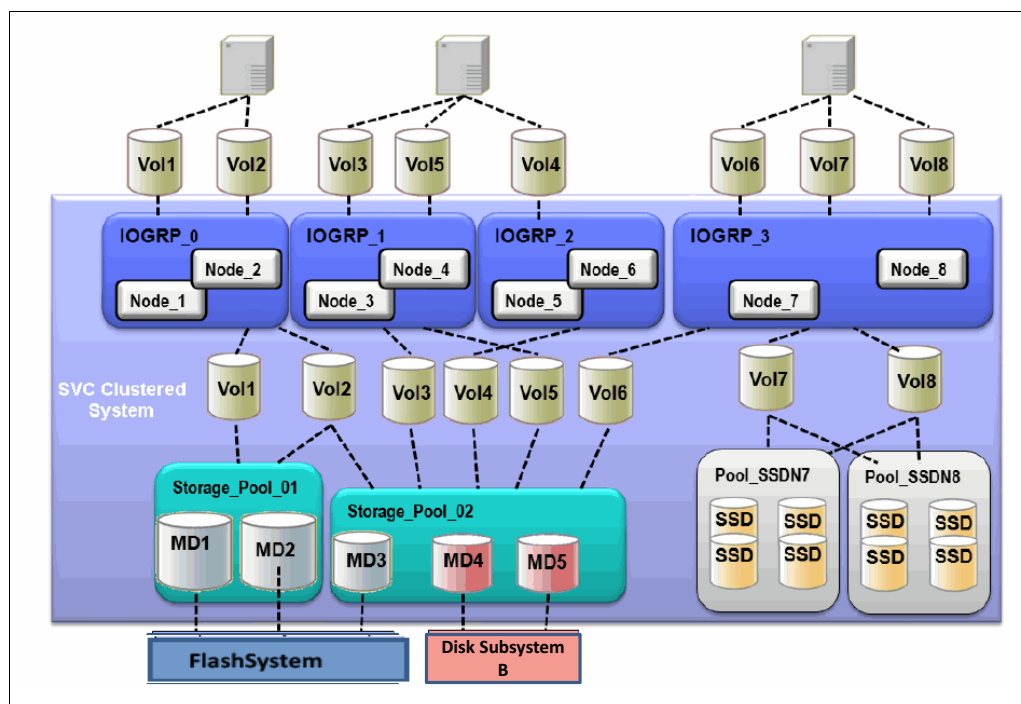


Figure 8-2 Overview of a SAN Volume Controller clustered system, hosts, and storage subsystems

Each MDisk in the storage pool is divided into a number of *extents*. The size of the extent is selected by the administrator at the creation time of the storage pool and cannot be changed later. The size of the extent ranges from 16 - 8192 MB. The volume that resides on the storage pool might be formatted in two different types: *sequential* or *striped* (default). For more details, see *Implementing the IBM System Storage SAN Volume Controller V7.4*, SG24-7933.

8.1.3 SAN Volume Controller hardware options

Throughout its lifecycle, SAN Volume Controller has used IBM System x server technology to offer a modular, flexible platform for storage virtualization that can be rapidly adapted in response to changing market demands and evolving client requirements. This “flexible hardware” design allows you to quickly incorporate differentiating features. Significant hardware options are available.

Hardware options for SAN Volume Controller

SAN Volume Controller 2145-DH8 node introduces numerous hardware enhancements. Several of these enhancements relate directly to the Real-time Compression feature and offer significant performance and scalability improvements over previous hardware versions.

Additional and enhanced CPU options

The 2145-DH8 node offers an updated primary CPU that contains 8 cores as compared to the 4 and 6 core CPUs available in previous hardware versions. Additionally, the 2145-DH8 node offers the option of a secondary 8 core CPU for use with Real-time Compression. This additional, compression-dedicated CPU allows for improved overall system performance when using compression over previous hardware models.

Note: In order to use the Real-time Compression feature on 2145-DH8 nodes, the secondary CPU is required.

Increased memory options

The 2145-DH8 node offers the option to increase the node memory from the base 32 GB to 64 GB, for use with Real-time Compression. This additional, compression-dedicated memory allows for improved overall system performance when using compression over previous hardware models.

Note: In order to use the Real-time Compression feature on 2145-DH8 nodes, the additional 32 GB memory option is required.

Quick Assist compression acceleration cards

The 2145-DH8 node offers the option to include one or two Intel Quick Assist compression acceleration cards based on the Coletto Creek chipset. The introduction of these Intel based compression acceleration cards in SAN Volume Controller 2145-DH8 node is an industry first, providing dedicated processing power and greater throughput over previous models.

Note: In order to use the Real-time Compression feature on 2145-DH8 nodes, at least one Quick Assist compression acceleration card is required. With a single card, the maximum number of compressed volumes per I/O group is 200. With the addition of a second Quick Assist card, the maximum number of compressed volumes per I/O group is 512.

For more details about the compression accelerator cards see *Implementing the IBM System Storage SAN Volume Controller V7.4*, SG24-7933.

Additional host bus adapters

The 2145-DH8 node offers the option to add one or two additional 4-port 8 Gbps or 2-port 16 Gbps FC HBA to improve connectivity options on SAN Volume Controller engine.

The following example scenarios describe where these additional ports can provide benefits:

- ▶ Isolation of node-to-node communication, potentially boosting write performance
- ▶ Isolation of node to IBM FlashSystem communication, allowing for maximum performance
- ▶ Isolation of remote copy traffic, avoiding performance problems

The additional host bus adapter (HBA) support on 2145-DH8 nodes requires SAN Volume Controller Storage Software Version 7.3 or higher, for 8 Gbps for 16 Gbps FC HBAs.

Support for a second SAN Volume Controller HBA was introduced with Software Version 7.1 and the 2145-CG8 nodes.

Note: For information about SAN Volume Controller V7.4 Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller, see this website:

<http://www.ibm.com/support/docview.wss?uid=ssg1S1004924>

Port masking

The addition of more FC HBA ports allows clients to optimize their SAN Volume Controller configuration by using dedicated ports for certain system functions. However, the addition of these ports necessitates the ability to ensure traffic isolation.

The following two examples show traffic types that you might want to isolate by using port masking:

- ▶ Local node-to-node communication
- ▶ Remote copy traffic

Port masking was introduced with SAN Volume Controller Storage Software Version 7.1. This feature enables better control of SAN Volume Controller node ports. Host port masking is supported in earlier SAN Volume Controller software versions. In those versions, host port masking provides the ability to define which SAN Volume Controller node ports were used to communicate with hosts.

The enhanced port masking in SAN Volume Controller Storage Software Version 7.1 and later provides the ability to restrict intracluster communication and replication communication to specific ports, ensuring that these traffic types only occur on the ports that you want. This capability eliminates the possibility of host or back-end port congestion due to intracluster communication or replication communication.

Notes:

- ▶ A SAN Volume Controller node attempts to communicate with other SAN Volume Controller nodes over any available path. Port masking, when enabled, ensures that this will not occur on that port.
- ▶ To use the port masking feature, use the `chsystem -localfcportmask` or `-partnerfcportmask` command.

The features in SAN zoning and the physical port assignment provide greater control and enable less congestion and better usage of SAN Volume Controller ports.

Note: When using port masking with SAN Volume Controller, follow this configuration order:

1. Configure intracluster port masking.
2. Configure replication port masking (if using replication).
3. Configure SAN zones for intracluster communication.
4. Configure SAN zones for replication communication (if using replication).
5. Configure SAN zones for all back-end storage communication.
6. Configure SAN zones for host communication.

For more information about how to configure SAN Volume Controller port masking, see:

<http://www.ibm.com/support/docview.wss?uid=ssg1S1004659>

SAN Volume Controller Stretched Cluster

The SAN Volume Controller Stretched Cluster configurations are supported by the IBM FlashSystem storage systems. When using the FlashSystem storage systems in the SAN Volume Controller Stretched Cluster environments, follow the guidance in the following Redbooks publications:

- ▶ *IBM SAN and SVC Stretched Cluster and VMware Solution Implementation*, SG24-8072
- ▶ *IBM SAN Volume Controller Stretched Cluster with PowerVM and PowerHA*, SG24-8142

8.1.4 IBM Spectrum Virtualize - SAN Volume Controller advanced functionality

The combination of the IBM FlashSystem 840 and the IBM SAN Volume Controller enables clients to take advantage of the speed of the IBM FlashSystem 840 and the robust storage management capabilities of SAN Volume Controller. IBM SAN Volume Controller offers features that enrich any storage environment with introducing minimal delay or latency in the I/O path. Next, we describe SAN Volume Controller features and benefits.

Thin provisioning

The *thin provisioning* function helps automate provisioning and improve productivity by enabling administrators to focus on overall storage deployment and utilization, and longer-term strategic requirements, without being distracted by routine everyday storage provisioning.

When creating a thin-provisioned volume, the user specifies two capacities: the real physical capacity allocated to the volume from the storage pool, and its virtual capacity available to the host. Therefore, the real capacity determines the quantity of MDisk extents that is initially allocated to the volume. The *virtual capacity* is the capacity of the volume reported to all other SAN Volume Controller components (for example, FlashCopy, cache, and Remote copy) and to the host servers.

The *real capacity* is used to store both the user data and the metadata for the thin-provisioned volume. The real capacity can be specified as an absolute value or a percentage of the virtual capacity. Thin-provisioned volumes can be used as volumes assigned to the host, by FlashCopy, and Remote Copy, to implement thin-provisioned targets, and also with the mirrored volumes feature.

FlashCopy

The *FlashCopy* function is designed to create an almost instant copy (or “snapshot”) of active data that can be used for backup purposes or for parallel processing activities. Up to 256 copies of data can be created.

FlashCopy works by creating one or two (for incremental operations) bitmaps to track changes to the data on the source volume. This bitmap is also used to present an image of the source data at the point in time that the copy was taken to target hosts while the actual data is being copied. This capability ensures that copies appear to be instantaneous.

FlashCopy permits the management operations to be coordinated, via a grouping of FlashCopy pairs, so that a common single point in time is chosen for copying target volumes from their respective source volumes. This capability is called *consistency groups* and allows a consistent copy of data for an application that spans multiple volumes.

IBM offers IBM Spectrum Control, which delivers the functionality of Tivoli Storage FlashCopy Manager that is designed to perform near-instant application-aware snapshot backups using SAN Volume Controller FlashCopy, but with minimal impact to IBM DB2, Oracle, SAP, Microsoft SQL Server, or Microsoft Exchange. FlashCopy Manager also helps reduce backup and recovery times from hours to a few minutes.

You can obtain more information about IBM Spectrum Control at this website:

<http://www.ibm.com/software/tivoli/products/storage-flashcopy-mgr>

Easy Tier

SAN Volume Controller Easy Tier function is designed to help improve performance at a lower cost through more efficient use of storage. *Easy Tier* is a performance function that

automatically migrates or moves extents off a volume to, or from, one MDisk storage tier to another MDisk storage tier. Easy Tier monitors the host I/O activity and latency on the extents of all volumes with the Easy Tier function turned on in a multitier storage pool over a 24-hour period.

Next, Easy Tier creates an extent migration plan based on this activity and then dynamically moves high activity or hot extents to a higher disk tier within the storage pool. It also moves extents whose activity has dropped off or cooled from the higher-tier MDisk back to a lower-tiered MDisk.

SAN Volume Controller Easy Tier can deliver up to a three-time performance improvement with only 5% flash storage capacity. Easy Tier can use flash storage, whether deployed in SAN Volume Controller nodes or in virtualized disk systems, to benefit all virtualized storage. This approach delivers greater benefits from flash storage than tiering systems that are limited to only a single disk system.

Because the Easy Tier function is so tightly integrated, functions, such as data movement, replication, and management, all can be used with flash in the same way as for other storage. SAN Volume Controller helps move critical data to and from flash storage as needed without application disruptions. Combining SAN Volume Controller with the FlashSystem storage devices delivers the best of both technologies: Extraordinary performance for critical applications with IBM MicroLatency, coupled with sophisticated functionality.

Mirroring/Copy Services

With many conventional SAN disk arrays, replication operations are limited to in-box or like-box-to-like-box circumstances. Functions from different vendors can operate in different ways, which make operations in mixed environments more complex and increase the cost of changing storage types. But SAN Volume Controller is designed to enable administrators to apply a single set of advanced network-based replication services that operate in a consistent manner, regardless of the type of storage being used.

SAN Volume Controller supports remote mirroring to enable organizations to create copies of data at remote locations for disaster recovery. Metro Mirror supports synchronous replication at distances up to 300 km (186.4 miles). Global Mirror supports asynchronous replication up to 8000 km (4970.9 miles). Replication can occur between any Storwize family systems, and can include any supported virtualized storage. Remote mirroring works with FC, Fibre Channel over Ethernet (FCoE), and IP (Ethernet) networking between sites.

With IP networking, the IBM Storwize family systems support both 1 GbE and 10 GbE connections and use innovative Bridgeworks SANSlide technology to optimize the use of network bandwidth. As a result, the networking infrastructure can require lower speeds (and therefore, lower costs), or users might be able to improve the accuracy of remote data through shorter replication cycles. The remote mirroring functions also support VMware vCenter Site Recovery Manager to help speed up disaster recovery.

Volume mirroring is a simple RAID 1 type function that is designed to allow a volume to remain online even when the storage pool backing it becomes inaccessible. Volume mirroring is designed to protect the volume from storage infrastructure failures by providing the ability to seamlessly mirror between storage pools.

Volume mirroring is provided by a specific volume mirroring function in the I/O stack, and it cannot be manipulated like a FlashCopy or other types of copy volumes. This feature does however provide migration functionality, which can be obtained by splitting the mirrored copy from the source or by using the “migrate to” function. Volume mirroring does not have the ability to control back-end storage mirroring or replication.

Real-time Compression

Real-time Compression is designed to enable storing up to five times¹ as much data in the same physical disk space by compressing data as much as 80%. Unlike other approaches to compression, Real-time Compression is designed to be used with active primary data, such as production databases and email systems, which dramatically expands the range of candidate data that can benefit from compression. Real-time Compression operates immediately while data is written to disk, which means that no space is wasted storing decompressed data waiting for post-processing.

The benefits of using Real-time Compression together with other efficiency technologies are significant and include reduced acquisition cost (because less hardware is required), reduced rack space, and lower power and cooling costs throughout the lifetime of the system. Real-time Compression can significantly enhance the usable capacity of your existing storage systems, extending their useful life even more.

By significantly reducing storage requirements with Real-time Compression, you can keep more information online, use the improved efficiency to reduce storage costs, or achieve a combination of greater capacity and reduced cost. Because Real-time Compression can be applied to a much wider range of data, including primary online data, the benefits of compression with SAN Volume Controller can be much greater than with alternative solutions, resulting in much greater savings. Enhancements to SAN Volume Controller nodes support up to three times the performance with Real-time Compression, enabling even larger configurations to experience compression benefits.

Starting on SAN Volume Controller version 7.1, the concurrent use of Easy Tier and Real-time Compression is supported on the same volume.

Note: In order to use the Real-time Compression feature on 2145-DH8 nodes, more CPU, memory, and compression acceleration cards are required as mentioned in 8.1.3, “SAN Volume Controller hardware options” on page 281.

8.2 SAN Volume Controller connectivity to FlashSystem 840

The IBM FlashSystem 840 is an all-flash storage array that provides extreme performance and is capable of sustaining highly demanding throughput and low latency across its FC interfaces. It includes up to 16 ports of 8 Gbps or eight ports of 16 Gbps FC. It also provides enterprise-class reliability, large capacity, and “green” data center power and cooling requirements.

To maximize the performance that you can achieve when deploying the FlashSystem 840 with SAN Volume Controller, carefully consider the assignment and usage of the FC HBA ports on SAN Volume Controller. Specifically, SAN switch zoning, coupled with port masking (a feature introduced in SAN Volume Controller Storage Software Version 7.1), can be used for traffic isolation for various SAN Volume Controller functions, reducing congestion and improving latency.

After racking, cabling, and powering on the IBM FlashSystem, you must perform several steps to configure the FlashSystem 840 optimally for use with SAN Volume Controller. The first configuration steps are described in Chapter 3, “Planning” on page 47. Follow the procedures in that chapter to set up your IBM FlashSystem 840. You do not need to create any volumes or hosts now because we describe the preferred practices of creating volumes and hosts in the following topics.

¹ Compression data based on IBM measurements. Compression rates vary by data type and content.

8.2.1 SAN Volume Controller FC cabling to SAN

When deploying a new SAN Volume Controller cluster, it is important to connect the FC ports correctly and to match the port masking and zoning configuration. Figure 8-3 shows the suggested SAN Volume Controller cabling diagram to SAN for dual redundant fabrics.

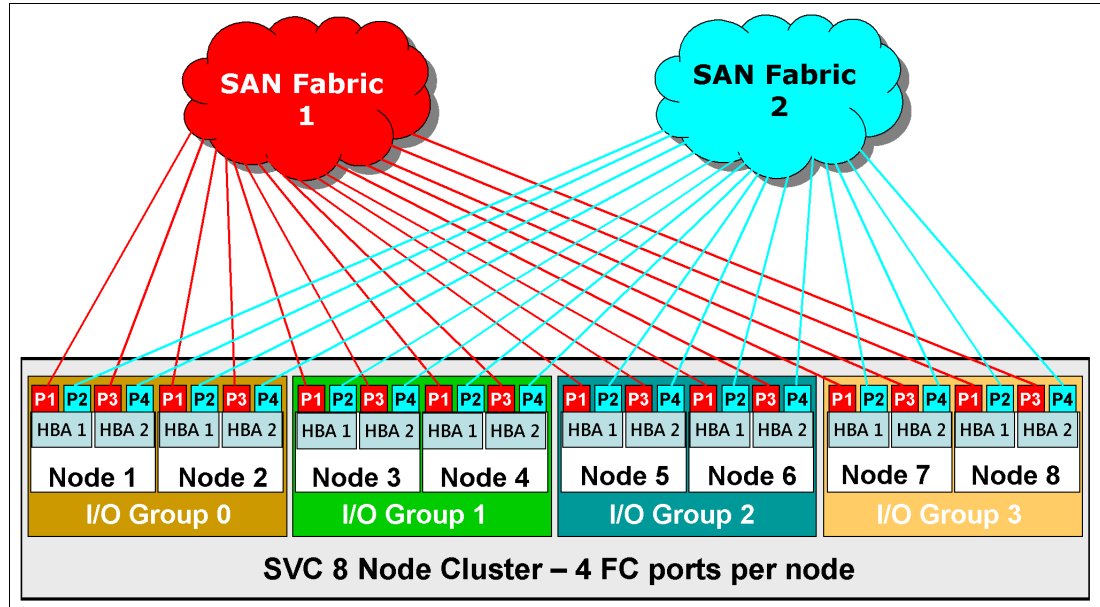


Figure 8-3 Cabling schema for a SAN Volume Controller 8-node cluster with one HBA card

For FlashSystem 840 connectivity, consider using more HBAs in SAN Volume Controller. Figure 8-4 on page 288 shows a configuration with one extra SAN Volume Controller HBA in each node, and the cabling schema for cabling to SAN fabric switches.

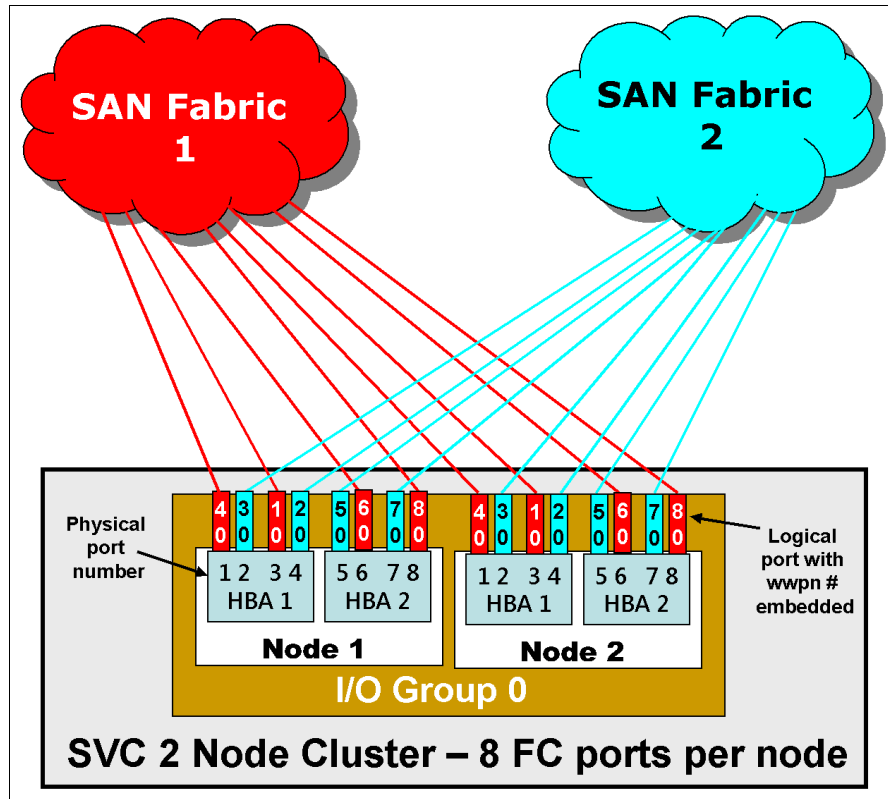


Figure 8-4 Cabling schema for SAN Volume Controller 2-node cluster with two HBA cards per node

8.2.2 SAN zoning and port designations

SAN Volume Controller can be configured with two, or up to eight SAN Volume Controller nodes, arranged in a SAN Volume Controller clustered system. These SAN Volume Controller nodes are attached to the SAN fabric, along with disk subsystems and host systems. The SAN fabric is zoned to allow the SAN Volume Controllers to “see” each other’s nodes and the disk subsystems, and for the hosts to “see” the SAN Volume Controllers. The hosts are not able to directly see or operate LUNs on the disk subsystems that are assigned to the SAN Volume Controller system. The SAN Volume Controller nodes within a SAN Volume Controller system must be able to see each other and all of the storage that is assigned to the SAN Volume Controller system.

In an environment where you have a fabric with multiple-speed switches, the preferred practice is to connect the SAN Volume Controller and the disk subsystem to the switch operating at the highest speed. SAN Volume Controller 7.4 with 16 GBps hardware supports 16 GBps, and must be connected to a supported 16 GBps capable switch.

All SAN Volume Controller nodes in the SAN Volume Controller clustered system are connected to the same SANs, and they present volumes to the hosts. These volumes are created from storage pools that are composed of MDisks presented by the disk subsystems.

The zoning capabilities of the SAN switches are used to create three distinct zones:

- SAN Volume Controller cluster system zones: Create up to two zones per fabric, and include a single port per node, which is designated for intracluster traffic. No more than four ports per node should be allocated to intracluster traffic.

- Host zones: Create a SAN Volume Controller host zone for each server accessing storage from the SAN Volume Controller system.
- Storage zones: Create one SAN Volume Controller storage zone for each storage subsystem that is virtualized by the SAN Volume Controller.

Certain limits and restrictions apply for SAN zoning and switch connectivity in SAN Volume Controller 7.4 environments. For more information, see the following website:

<http://www.ibm.com/support/docview.wss?uid=ssg1S1004924>

For information about supported SAN switches, see the IBM storage interoperability matrix at:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

Note: 4 Gbps and 8 Gbps fabrics are not supported directly connected to the 16 Gbps ports on the node hardware. 16 Gbps node ports must be connected to a supported 16 Gbps switch.

8.2.3 Port designation recommendations

The port to local node communication is used for mirroring write cache as well as metadata exchange between nodes and is critical to the stable operation of the cluster. The DH8 nodes with their 6-port, 8-port, and 12-port configurations provide an opportunity to isolate the port to local node traffic from other cluster traffic on dedicated ports, thereby providing a level of protection against misbehaving devices and workloads that could compromise the performance of the shared ports.

Additionally, there is a benefit in isolating remote replication traffic on dedicated ports as well to ensure that problems impacting the cluster-to-cluster interconnect do not adversely impact ports on the primary cluster and thereby impact the performance of workloads running on the primary cluster.

It is recommended to follow port designations for isolating both port to local and port to remote node traffic as shown in Table 8-1.

Table 8-1 SAN Volume Controller Port designations

Port	SAN	4-port nodes	8-port nodes	12-port nodes	12-port nodes, write Data Rate > 3 GBps per IO Group
C1P1	A	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Inter-node
C1P2	B	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Inter-node
C1P3	A	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Host/Storage
C1P4	B	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Host/Storage
C2P1	A		Inter-node	Inter-node	Inter-node
C2P2	B		Inter-node	Inter-node	Inter-node
C2P3	A		Replication or Host/Storage	Host/Storage	Host/Storage

Port	SAN	4-port nodes	8-port nodes	12-port nodes	12-port nodes, write Data Rate > 3 GBps per IO Group
C2P4	B		Replication or Host/Storage	Host/Storage	Host/Storage
C5P1	A			Host/Storage	Host/Storage
C5P2	B			Host/Storage	Host/Storage
C5P3	A			Replication or Host/Storage	Replication or Host/Storage
C5P4	B			Replication or Host/Storage	Replication or Host/Storage
localfcportmask		1111	110000	110000	110011

More port configurations might apply for iSCSI and FCoE connectivity.

Notes: The following notes apply to Table 8-1 on page 289:

SAN column assumes an odd/even SAN port configuration. Modifications must be made if other SAN connection schemes are used.

Care needs to be taken when zoning so that inter-node ports are not used for Host/Storage in the 8-port and 12-port configurations.

These options represent optimal configurations based on port assignment to function. Using the same port assignment but different physical locations will not have any significant performance impact in most client environments.

This recommendation provides the wanted traffic isolation while also simplifying migration from existing configurations with only 4 ports, or even later migrating from 8-port or 12-port configurations to configurations with additional ports. More complicated port mapping configurations that spread the port traffic across the adapters are supported and can be considered, but these approaches do not appreciably increase availability of the solution since the mean time between failures (MTBF) of the adapter is not significantly less than that of the non-redundant node components.

While it is true that alternate port mappings that spread traffic across HBAs may allow adapters to come back online following a failure, they will not prevent a node from going offline temporarily to reboot and attempt to isolate the failed adapter and then rejoin the cluster. Our recommendation takes all these considerations into account with a view that the greater complexity may lead to migration challenges in the future and the simpler approach is best.

Notes:

For the latest information about the DH8 node as of SAN Volume Controller 7.4, SAN zoning and SAN connections, and port designation recommendations for isolating traffic, refer to IBM Redbooks publication *IBM SAN Volume Controller 2145-DH8 Introduction and Implementation*, SG24-8229, Chapter 4 *Planning and configuration*.

When you attach your IBM FlashSystem to a SAN Volume Controller node that contains a single HBA quad port, follow the zoning and port guidelines that are suggested for any other storage back-end device. See *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

8.2.4 Verifying FlashSystem 840 connectivity in SAN Volume Controller

After you activate the zoning, you are to be able to identify the IBM FlashSystem 840 as a controller to SAN Volume Controller. You can use the SAN Volume Controller command **lscontroller** or the SAN Volume Controller GUI to navigate to **Pools** → **External Storage** to verify.

Change the controller name on the SAN Volume Controller system by one of two methods. You can issue the SAN Volume Controller command **chcontroller** with the **-name** parameter, or you can use the SAN Volume Controller GUI to navigate to **Pools** → **External Storage** and then right-click the controller that you want and select **Rename**. After you change the name, you can easily identify the IBM FlashSystem 840 as a controller to SAN Volume Controller as shown in Figure 8-5.

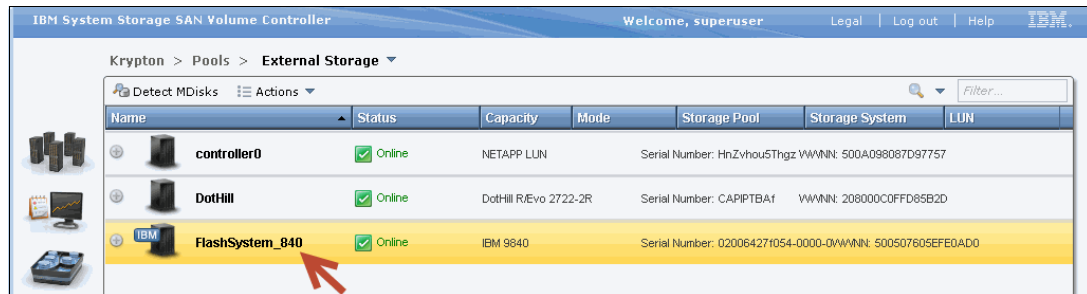


Figure 8-5 IBM FlashSystem 840 as an external storage to SAN Volume Controller

You also need to create zones between SAN Volume Controller and the host. For guidance to configure zoning for host access to SAN Volume Controller, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

IBM FlashSystem 840 host and volume creation

To provide usable storage (managed disks) to SAN Volume Controller, the IBM FlashSystem 840 must have *volumes* that are defined. These volumes must be mapped to the SAN Volume Controller *host* that also must be created. For guidance to perform these steps, see Chapter 6, “Using the IBM FlashSystem 840” on page 165.

Create one host and name it, for example, SVC. Then, add the SAN Volume Controller worldwide port names (WWPNs) that communicate with the FlashSystem (in our case, SAN Volume Controller nodes port 1 and port 5) to this newly created host. Then, create the volumes following the guidelines described in the next topic and map them to this newly created host called SVC.

SAN Volume Controller managed disk configuration

The preferred practices and considerations to design and plan the SAN Volume Controller storage pool (MDisk group) setup using the IBM FlashSystem 840 are described. There are several considerations when planning and configuring the MDisks and storage pools for the IBM FlashSystem 840 behind the SAN Volume Controller.

When using the IBM FlashSystem 840 behind the SAN Volume Controller, it is important to note the following points when planning to design and create MDisks for use in storage pools. In this case, the queue assignment and cache assignment are not as relevant as they are with traditional spindle-based disk systems due to the rapid speed with which the IBM FlashSystem 840 is able to process I/O requests.

It is advised to use MDisks in multiples of 4, which is an optimal number for CPU processing, because the MDisks and extents are equally distributed to the SAN Volume Controller CPU cores.

The second CPU with 32 GB memory feature on SAN Volume Controller Storage Engine Model DH8 provides performance benefit only when Real-time Compression is used. IBM intends to enhance IBM Storwize Family Software for SAN Volume Controller to extend support of this feature to also benefit decompressed workloads.

Storage pool extent size

When you work with an IBM FlashSystem 840 behind a SAN Volume Controller configuration, the extent size can be left at the default² of 1024 MB (1 GB). The performance of the IBM FlashSystem 840 with random I/O workload does not require the extent size to be smaller. The maximum extent size is 8192 MB, which provides for a maximum MDisk of 1024 TB or 1 Petabyte.

If there are existing systems and storage pools, you must have the same extent size for all pools to use transparent volume migration between pools. If you have different extent sizes, you can use volume mirroring to create a copy of the disk and then promote it to the master volume when the copy is completed. However, this manual process takes slightly longer to complete.

All FlashSystem versus mixed storage pools

If you use the FlashSystem 840 as the primary data storage, add all of the MDisks from the controller to a single *managed disk group* (also known as a *storage pool* in the SAN Volume Controller GUI). However, if more than one FlashSystem 840 is presented to a SAN Volume Controller cluster, a preferred practice is to create a single storage pool per controller.

If you use the FlashSystem 840 with the SAN Volume Controller Easy Tier function, you likely want to create multiple volumes for each *hybrid storage pool*. Create four or more volumes for each hybrid pool, with the combined capacity of these volumes matching the capacity that you want for the *SSD tier* in that pool. More information about SAN Volume Controller Easy Tier with the FlashSystem is in the next topic.

MDisk mapping, storage pool, and volume creation

In this section, we describe mapping MDisks from an IBM FlashSystem 840 to SAN Volume Controller. Defining a storage pool to accommodate those volumes and the process to provision a volume from the FlashSystem 840 storage pool to the hosts are also explained.

After volumes are created on the IBM FlashSystem 840 and presented to the SAN Volume Controller, they need to be recognized in the SAN Volume Controller. The first step is to click the Detect MDisks option from the SAN Volume Controller GUI to detect the newly presented

² Starting on SAN Volume Controller version 7.x

MDisks, as shown in Figure 8-6. The sequence to perform this operation is described and refers to the arrows in Figure 8-6:

1. Select the **Pools** option (arrow number 1).
2. Choose **External Storage** (arrow number 2).
3. Click the **Detect MDisks** option (arrow number 3).
4. When the task completes, click **Close** to see the newly available MDisks (arrow number 4).

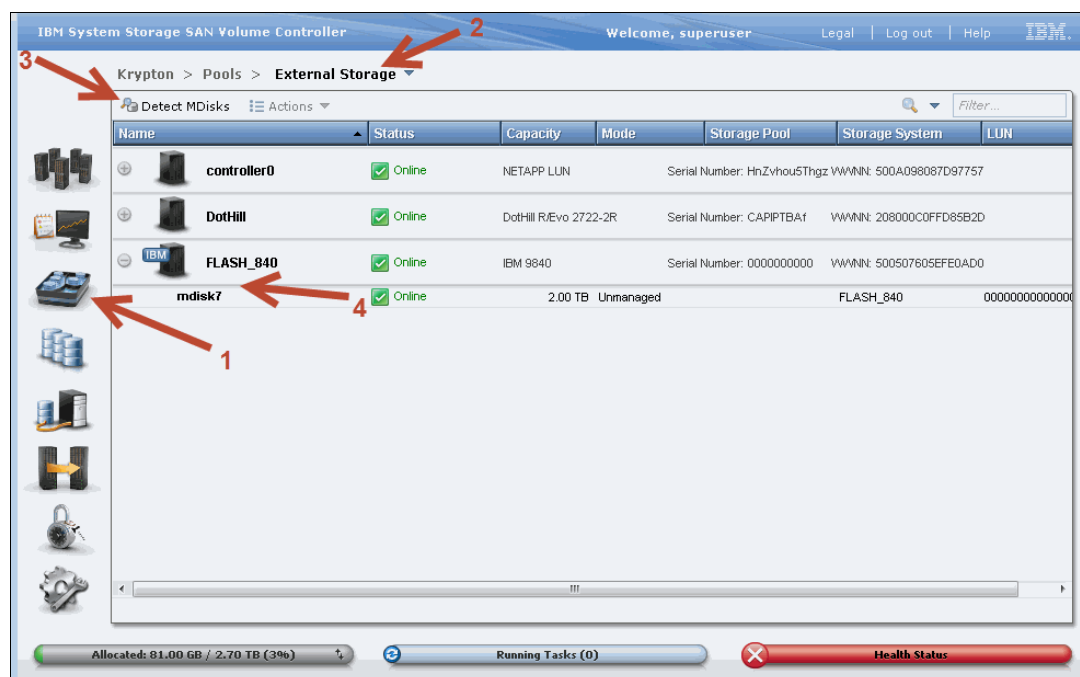


Figure 8-6 Detect MDisks option in the GUI

When this detection completes, a list of MDisks appears, as shown in Figure 8-6 (see *mdisk7*). It is important to rename the MDisks when they are added, for identification purposes. When naming an MDisk, the suggested naming convention is “%controller name_lun id on disk system%”. This defines the name of the controller from which the LUN is presented and the local identifier that is referenced on the source disk system. This information is helpful when troubleshooting.

To rename an MDisk, select the MDisk that you want, right-click, and select **Rename**. A Rename MDisks window opens. Type the new names of the MDisks and then click **Rename**.

The next step is to create a new storage pool, if wanted, as shown in Figure 8-7 on page 294. A pop-up menu is displayed that you can use to configure the storage pool name and extent size. Ensure that **Advanced Settings** is clicked to show the extent size menu. The default is 1 GB (1024 MB), which can be left as the default, as explained in “SAN Volume Controller managed disk configuration” on page 292. You enter the name of the storage pool, and when complete, click **Next**.

The sequence to perform this operation is described:

1. Select the **Pools** icon (arrow number 1).
2. Choose **MDisks by Pools** (arrow number 2).
3. Click **New Pool** (arrow number 3).
4. Enter the pool name and extent size that you want (arrow number 4).
5. Click **Next** to advance (arrow number 5).

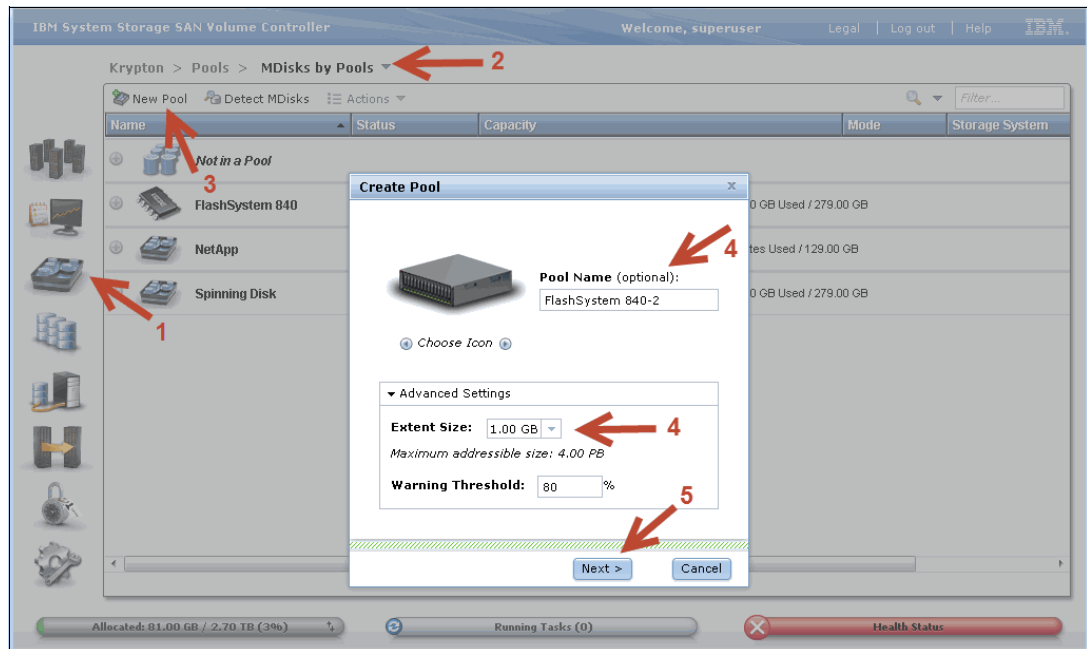


Figure 8-7 The new pool option

The MDisk that will be in this storage pool needs to be added. Select the MDisk to be a member of this group, as shown in Figure 8-8, and then click **Create**.

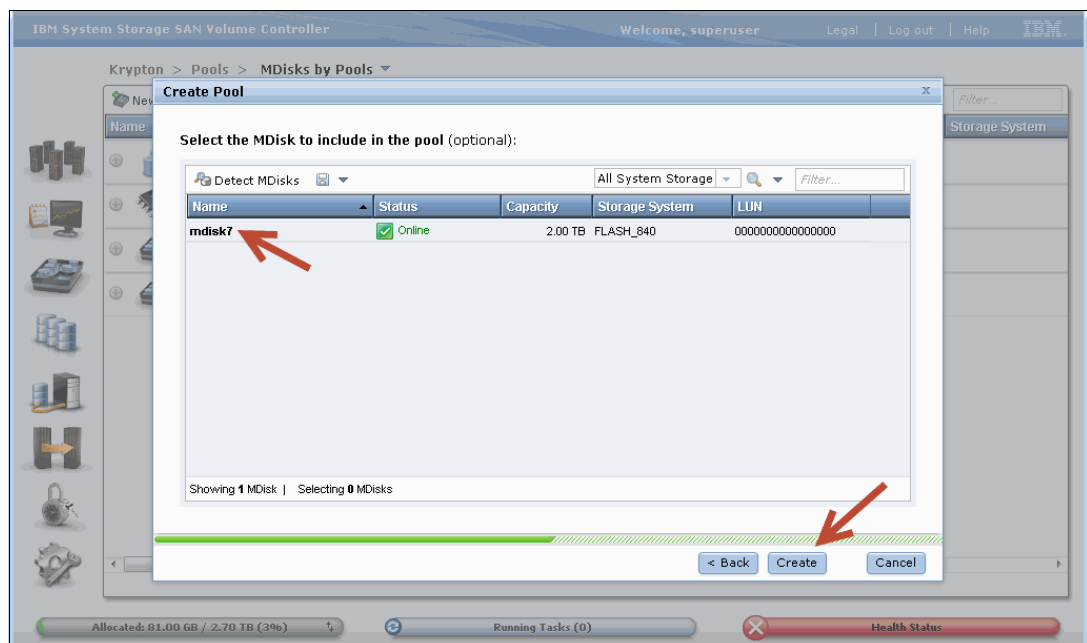


Figure 8-8 Showing the MDisk candidate to include in the storage pool being created

The process begins, as shown in Figure 8-9 on page 295. Click **Close** when completed to return to the main page.

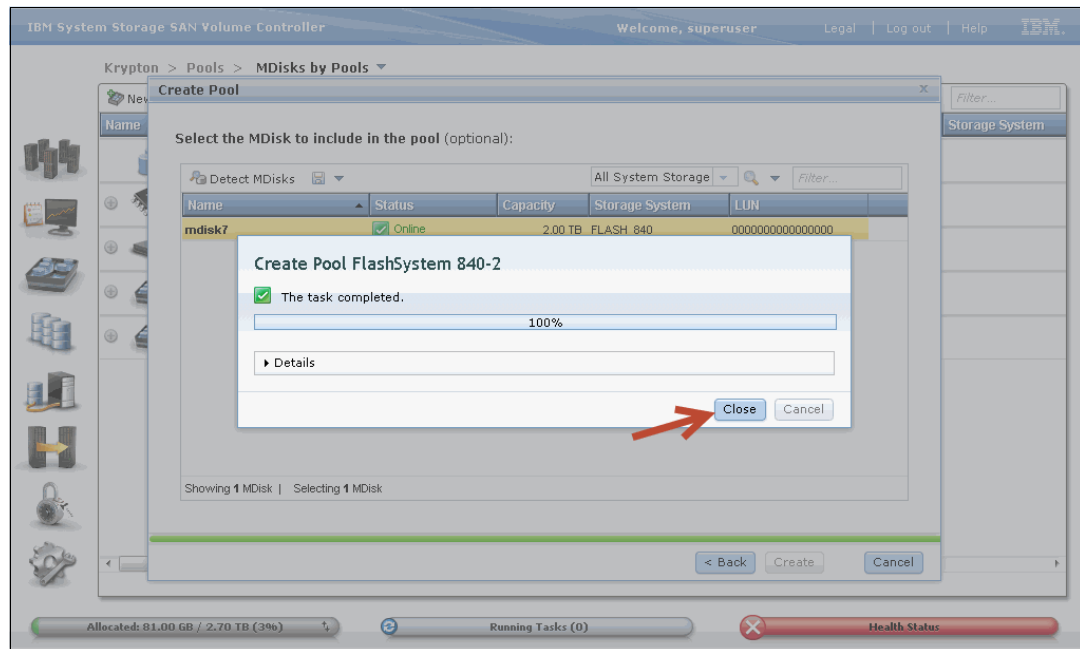


Figure 8-9 Command to create a storage pool being processed

After the command completes, the storage pool is created and the MDisk is a member of the storage pool. To confirm, check the details of the new storage pool just created as shown in Figure 8-10.

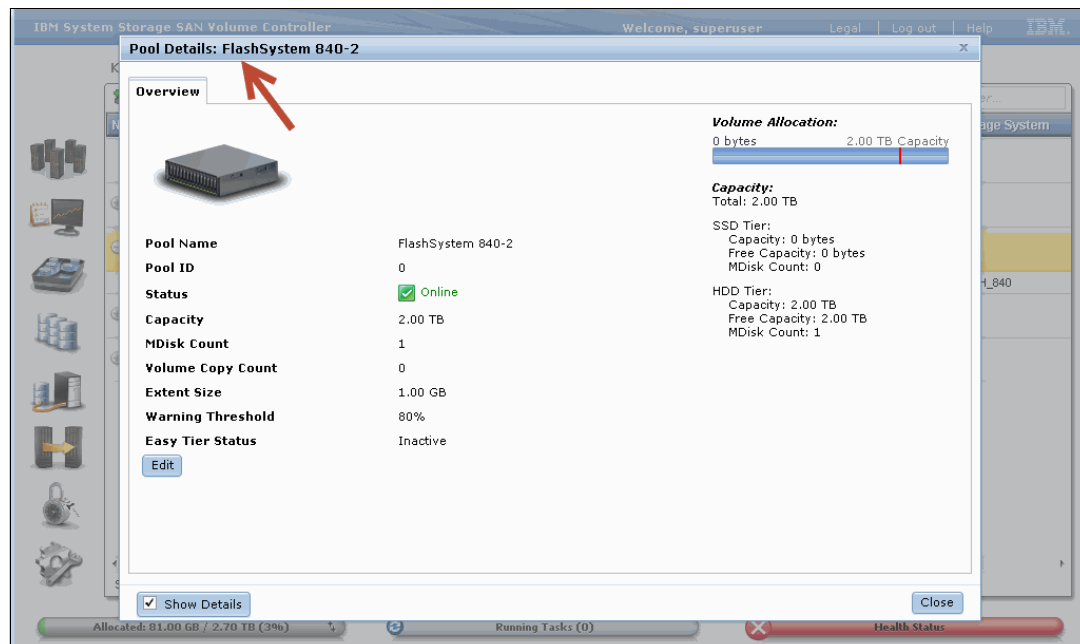


Figure 8-10 Checking storage pool details for the newly created pool

The definition of the MDisk and one storage pool is complete. If a second storage pool is necessary for volume mirroring or for more volume allocation, repeat the steps (Figure 8-7 on page 294, Figure 8-8 on page 294, Figure 8-9 on page 295, and Figure 8-10 on page 295) to create the second storage pool.

Note: It is suggested that you add all MDisks to the defined storage pool at creation time, or before starting volume allocation on that pool. You can select more than one MDisk candidate during the storage pool creation process, or you can add it manually by using the **addmdisk** command (or by using the GUI, as well).

Now that you have an available storage pool, you can create a host. Assuming that zoning is already configured between the host and SAN Volume Controller, Figure 8-11 shows the main menu to create a host on SAN Volume Controller. The sequence to perform this operation is described (the arrows refer to Figure 8-11):

1. Select the **Hosts** icon (arrow number 1).
2. Choose **Hosts** (arrow number 2).
3. Click **New Host** (arrow number 3).
4. Select the Fibre Channel port to be added to that host (arrow number 4).
5. Click **Add Port to List** (arrow number 5).
6. Check the Port Definitions to ensure that you added the port (arrow number 6).
7. Click **Create Host** to complete the process (arrow number 7).

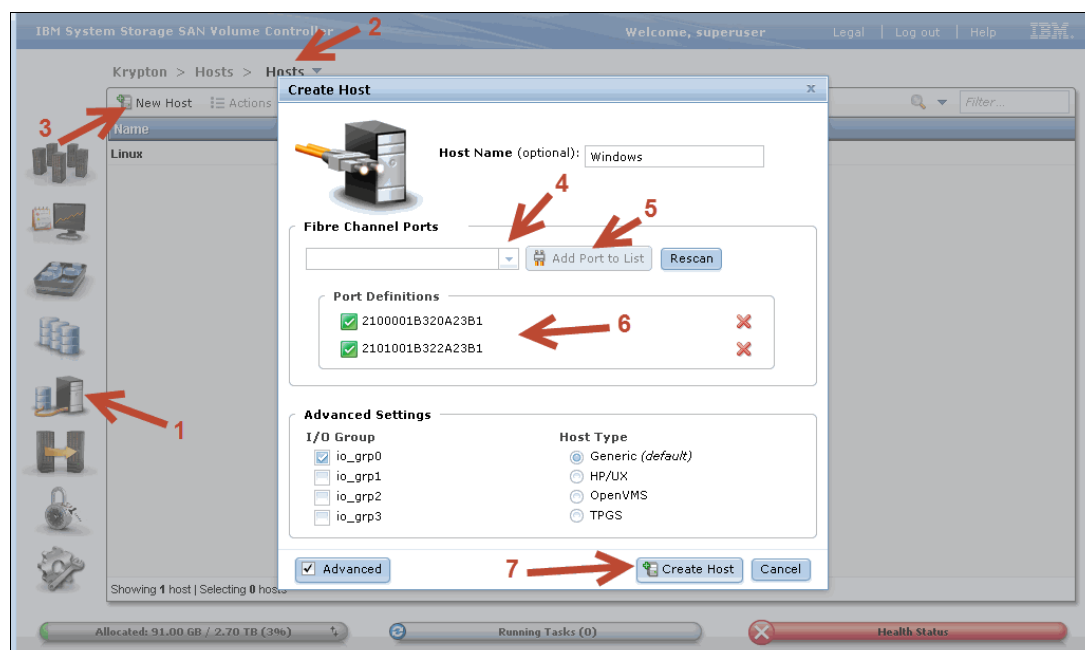


Figure 8-11 The New Host option

After you create the host, you can now create a volume as shown in Figure 8-12. The selected pool is the FlashSystem 840 pool that you created previously. The sequence to perform this operation is described (the arrows refer to Figure 8-12):

1. Select the **Volumes** icon (arrow number 1).
2. Choose **Volumes** (arrow number 2).
3. Click the **New Volume** option (arrow number 3).
4. Choose a volume preset. In this case, we used the Generic icon (arrow number 4).
5. Select a pool. In this case, we used the FlashSystem pool (arrow number 5).
6. Enter the quantity, capacity, and name in the Volume Details section.
7. Click **Create** to complete the process (arrow number 6).

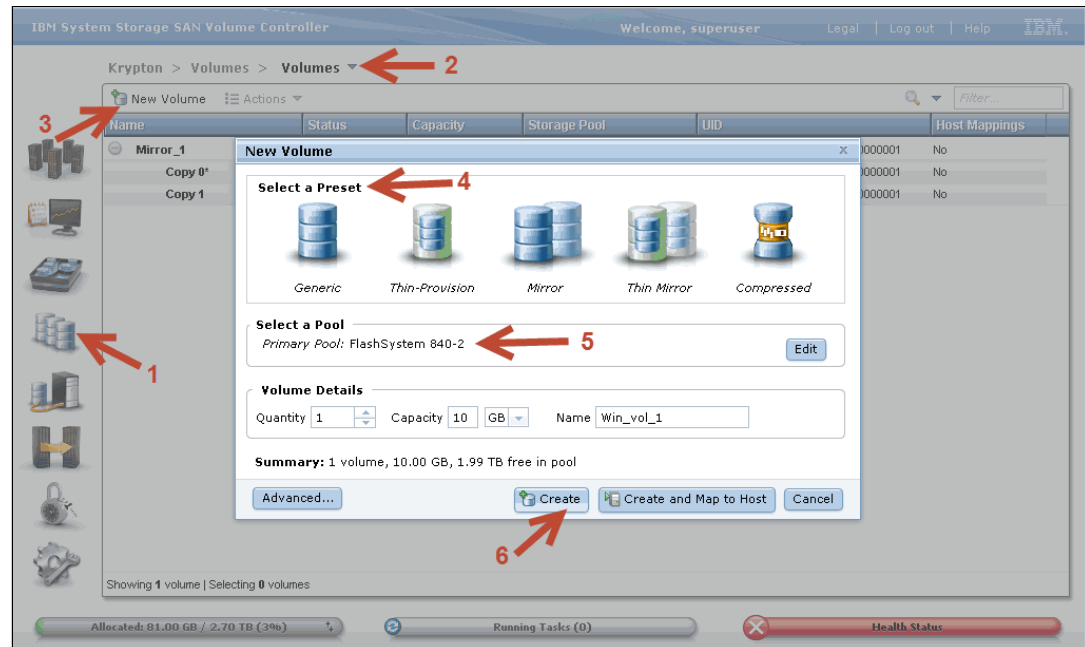


Figure 8-12 The New Volume option

The final step is to assign that recently created volume to the host that is already created. Figure 8-13 shows the process. The sequence to perform this operation is described (arrows refer to Figure 8-13):

1. Select the **Volumes** icon (arrow number 1).
2. Choose **Volumes** (arrow number 2).
3. Select the volume that you want to map (arrow number 3).
4. Click **Actions** (arrow number 4).
5. Select the **Map to Host** option (arrow number 5).

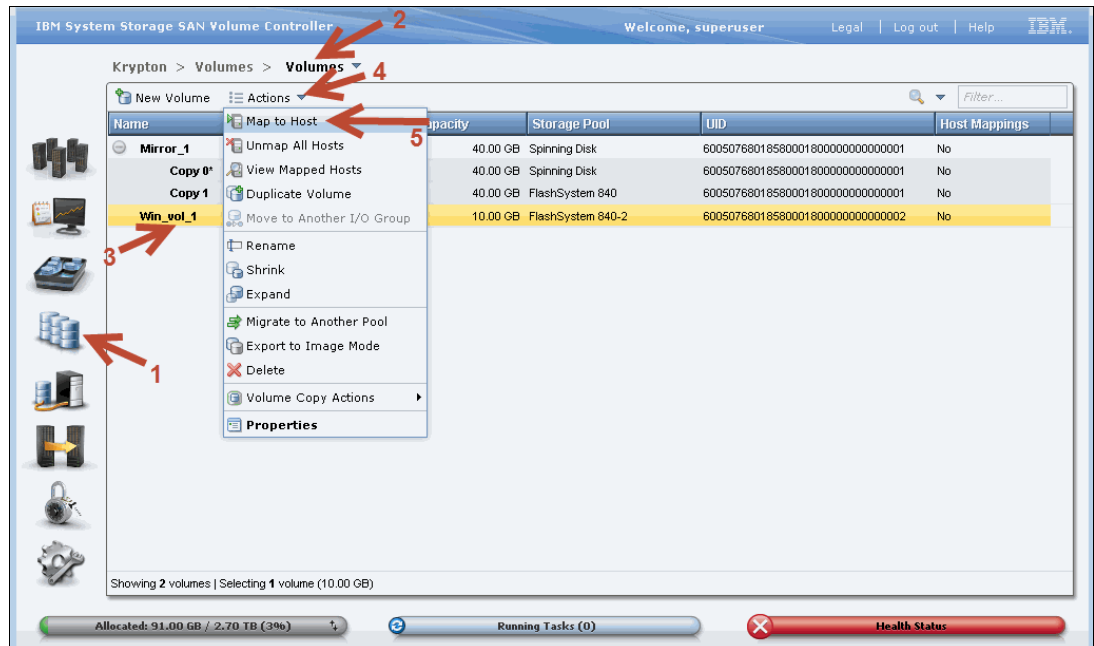


Figure 8-13 The Map to Host option

Another pop-up menu shows where you select the volumes that you want to map to the host as shown in Figure 8-14. Click **Apply** to complete the assignment.

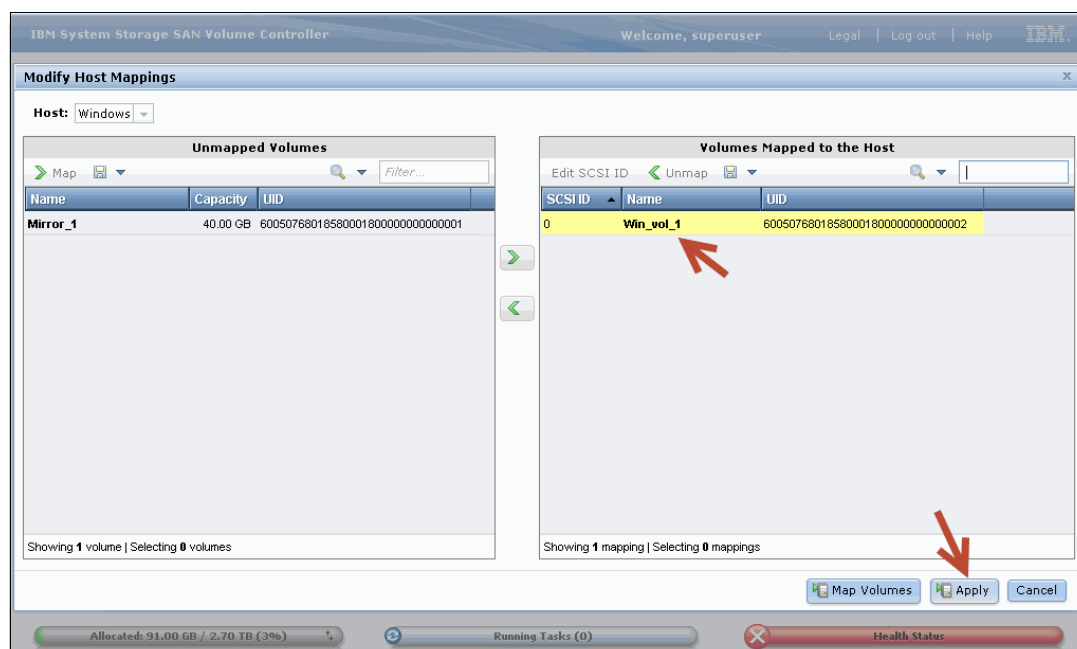


Figure 8-14 The Map to Host steps

The tasks on the SAN Volume Controller side are complete. At this point, the server is able to recognize the volume, assuming that the server is already prepared in terms of multipathing software and HBA firmware.

SAN Volume Controller volume mirroring

SAN Volume Controller *volume mirroring* is a feature that allows the creation of one volume with two copies of its extents. The two data copies, if placed in different storage pools, allow volume mirroring to eliminate the impact to volume availability if one or more MDisks, or the entire storage pool, fail.

When you design for a highly available SAN Volume Controller and FlashSystem deployment, it is suggested to have each storage pool built with MDisks coming from separate back-end storage subsystems. This allows the mirrored volumes to have each copy in a different storage pool. In this manner, volume mirroring can provide protection against planned or unplanned storage controller outages because the volume continues to be available for read/write operations from hosts with the surviving copy.

SAN Volume Controller *volume mirroring* is a simple RAID 1-type function that allows a volume to remain online even when the storage pool backing it becomes inaccessible. Volume mirroring is designed to protect the volume from storage infrastructure failures by seamless mirroring between storage pools.

Volume mirroring is provided by a specific volume mirroring function in the I/O stack, and it cannot be manipulated like a FlashCopy or other types of copy volumes. However, this feature provides migration functionality, which can be obtained by splitting the mirrored copy from the source or by using the “migrate to” function. Volume mirroring cannot control back-end storage mirroring or replication.

With volume mirroring, host I/O completes when both copies are written. Before version 6.3.0, this feature took a copy offline when it had an I/O timeout, and then resynchronized with the

online copy after it recovered. With 6.3.0, this feature is enhanced with a tunable latency tolerance. This tolerance provides an option to give preference to losing the redundancy between the two copies. This tunable timeout value is Latency or Redundancy.

The Latency tuning option, which is set with the **chvdisk -mirrowritepriority latency** command, is the default. This behavior was available in releases before 6.3.0. It prioritizes host I/O latency, which yields a preference to host I/O over availability.

However, you might have a need in your environment to give preference to redundancy when availability is more important than I/O response time. Use the **chvdisk -mirror writepriority redundancy** command.

Regardless of which option you choose, volume mirroring can provide extra protection for your environment.

Migration offers the following options:

- ▶ Volume migration by using volume mirroring and then by using Split into New Volume: By using this option, you can use the available RAID 1 functionality. You create two copies of data that initially has a set relationship (one volume with two copies: one primary and one secondary) but then break the relationship (two volumes, both primary and no relationship between them) to make them independent copies of data. You can use this to migrate data between storage pools and devices. You might use this option if you want to move volumes to multiple storage pools. Volume can have two copies at a time, which means you can add only one copy to the original volume and then you have to split those copies to create another copy of the volume.
- ▶ Volume migration by using Move to Another Pool: By using this option, you can move any volume between storage pools without any interruption to the host access. This option is a quicker version of the “Volume Mirroring and Split into New Volume” option. You might use this option if you want to move volumes in a single step or you do not have a volume mirror copy already.

Note: Volume mirroring does not create a second volume before you split copies. Volume mirroring adds a second copy of the data under the same volume so you end up having one volume presented to the host with two copies of data connected to this volume. Only splitting copies creates another volume and then both volumes have only one copy of the data.

With volume mirroring, you can move data to different MDisk within the same storage pool or move data between different storage pools. There is a benefit when using volume mirroring over volume migration because with volume mirroring, storage pools do not have to have the same extent size as is a case with volume migration.

Starting with firmware 7.3 and the introduction of the new cache architecture, mirrored volume performance has been significantly improved. Now, lower cache is beneath the volume mirroring layer, which means both copies have its own cache. This helps in cases of having copies of different types, for example generic and compressed, because now both copies use its independent cache and each copy can do its own read prefetch now. Destaging of the cache can now be done independently for each copy, so one copy does not affect performance of a second copy.

Also, because the Storwize destage algorithm is MDisk-aware it can tune or adapt the destaging process depending on MDisk type and utilization and this can be done for each copy independently.

SAN Volume Controller volume mirroring use cases with the FlashSystem

Depending on the storage system that is chosen for each of the two volume mirror copies, certain details must be considered to meet the high performance and low latency expectations of the solution. We describe usage scenarios for the SAN Volume Controller volume mirroring when virtualizing one or more FlashSystem storage systems, and the potential benefits and preferred practices for each scenario.

Use case 1: Volume mirroring between two FlashSystem storage systems

This example is the basic scenario for SAN Volume Controller volume mirroring because both copies reside on the same type of storage systems. In this scenario, there is no expected impact to performance in a failure if the two subsystems are close to the SAN Volume Controller nodes. This scenario indicates a non-stretched cluster setup. In this case, FC link latency is not likely to present an issue, and the default configuration for preferred node, primary copy, and mirroring priority is simplified.

Be careful to ensure a balanced workload between all the available resources (nodes, paths, fabrics, and storage subsystems).

However, if a SAN Volume Controller *Stretched Cluster* architecture is deployed in this scenario, plan carefully to ensure that the link latency between the SAN Volume Controller nodes, the hosts, and the two FlashSystem subsystems does not negatively affect the I/O operations from the applications. Follow these suggestions, whenever possible:

- ▶ The preferred node for each mirrored volume needs to be kept at the same site as the subsystem that contains the primary copy of that mirrored volume.
- ▶ The host that is performing the I/O operations to the mirrored volume needs to reside in the same site as the preferred node for that volume.
- ▶ The **-mirrorwritepriority** flag for the mirrored volume needs to be set to 1 latency if the access to the volume copy across the stretched cluster link represents a significant percentage of the total I/O latency. This can compromise the cache usage. Otherwise, the suggested value of redundancy still applies for every stretched cluster architecture to offer the highest level of data concurrency at both sites for protection against failure in the environment.

Note: If the primary purpose of the VDisk copy is data migration, the mirror write priority should be set to 1 latency. If, however, the primary purpose of the mirror is to provide protection against storage failure, the option should be set to redundancy.

Use case 2: Volume mirroring between a FlashSystem and a non-flash storage system

In this scenario, usually adopted for cost-reduction reasons, plan to avoid the penalty that is represented by the slowest subsystem to the overall I/O latency. Follow these suggestions:

- ▶ The primary copy of each mirrored volume needs to be set for the copy residing in the FlashSystem subsystem so that all the reads are directed to it by SAN Volume Controller. This configuration is commonly referred to as a *preferred read* configuration.
- ▶ If both subsystems are close to the SAN Volume Controller nodes, a *non-Stretched Cluster* setup, the **-mirrorwritepriority** flag for the mirrored volume needs to be set to 1 latency. Therefore, destaging to the volume copy in the slowest subsystem does not introduce a negative impact in the overall write latency and consequently, to cache usage.

Using the FlashSystem with SAN Volume Controller Easy Tier

In this scenario, the IBM FlashSystem 840 is used with SAN Volume Controller Easy Tier to improve performance on a storage pool. Due to the complexity of this scenario, there are important considerations when you design and plan to add a FlashSystem to an existing environment.

A common question is when to use the FlashSystem storage over internal solid-state disks (SSDs) in SAN Volume Controller. The suggestion is to use the FlashSystem storage when your capacity requirements for Easy Tier exceed five SSDs. At this point, consider the FlashSystem storage systems for cost efficiency and performance. For more information, see *Flash or SSD: Why and When to Use IBM FlashSystem*, REDP-5020.

When planning to use the FlashSystem 840 with SAN Volume Controller Easy Tier, first use the IBM Storage Tier Advisor Tool (STAT) to obtain a comprehensive analysis of hot extents. This allows you to estimate the amount of required FlashSystem capacity. For more information about using this tool, see *Implementing the IBM System Storage SAN Volume Controller V7.4*, SG24-7933.

Ensure that you add these MDisks as *SSD MDisks*; otherwise, SAN Volume Controller Easy Tier cannot distinguish between the spindle-based generic hard disk drives (HDDs) and the FlashSystem SSD MDisks. This task can be done by first creating the new MDisks and then change tier level by using the `chmdisk -tier ssd <mdisk>` command after adding the MDisks to a storage pool.

FlashSystem with SAN Volume Controller replication

Consider the following information when you use the IBM FlashSystem 840 with the SAN Volume Controller and replication:

- ▶ Additional latency overhead with synchronous Metro Mirror
- ▶ Distance of cross-site links and additional latency overhead
- ▶ Adequate bandwidth for cross-site links
- ▶ Amount of data to replicate and its I/O rate
- ▶ Dedicated replication ports

The IBM FlashSystem storage systems provide extremely low latency. The latency of Metro Mirror links might affect the FlashSystem storage systems to a greater degree than other traditional disk systems used with SAN Volume Controller. Metro Mirror replication distances in excess of 10 km (6.2 miles) must be carefully analyzed to ensure that they will not introduce bottlenecks or increase response times when used with the FlashSystem storage systems.

Tip: Dedicating ports for replication is strongly advised with this solution. Isolating replication ports can disperse congestion and reduce latency for replication, while protecting other ports from the impacts of increased amounts of replication traffic.

FC is the preferred connectivity method using a Dense Wavelength Division Multiplexer (DWDM), or equivalent device, between the source and target sites. Furthermore, the use of inter-switch links (ISLs) in a trunk or port channel to increase the aggregate bandwidth between the two sites is advised. Also, size the connectivity between sites according to the amount of bandwidth that you are allocating on SAN Volume Controller for replication, with additional bandwidth for future growth and peak loads. To summarize, consider the following factors when designing the replication connectivity:

- ▶ Current average and peak write activity to the volumes that you plan to replicate

- ▶ Wanted recovery point objective (RPO) and recovery time objective (RTO) values
- ▶ Existing preferred practices for SAN Volume Controller replication, described in *IBM System Storage SAN Volume Controller and Storwize V7000 Replication Family Services*, SG24-7574.

Note: When using Fibre Channel over IP (FCIP) for replication in this scenario, consider using at least 10 Gb links between sites.

8.2.5 Import/export

Import/export is useful if you use SAN Volume Controller as a migration device as described below:

- ▶ **Export to Image mode:** By using this option, you can move storage from managed mode to image mode, which is useful if you are using the SAN Volume Controller as a migration device. For example, vendor A's product cannot communicate with vendor B's product, but you must migrate existing data from vendor A to vendor B. By using Export to Image mode, you can migrate data by using Copy Services functions and then return control to the native array while maintaining access to the hosts.
- ▶ **Import to Image mode:** By using this option, you can import an existing storage MDisk or logical unit number (LUN) with its existing data from an external storage system without putting metadata on it so that the existing data remains intact. After you import it, all copy services functions can be used to migrate the storage to other locations while the data remains accessible to your hosts.

8.3 Integrating FlashSystem 840 and SAN Volume Controller considerations

We describe considerations to integrate the IBM FlashSystem 840 with the IBM SAN Volume Controller.

Figure 8-15 on page 304 shows an example of a usage scenario that combines SAN Volume Controller and the IBM FlashSystem with a tiered approach to the storage management of a storage FlashSystem. In this solution, write I/O is performed to both the FlashSystem and disk storage for both VDisk Mirror copy 1 and copy 2. Read I/O is performed from the FlashSystem to boost performance with microsecond latency.

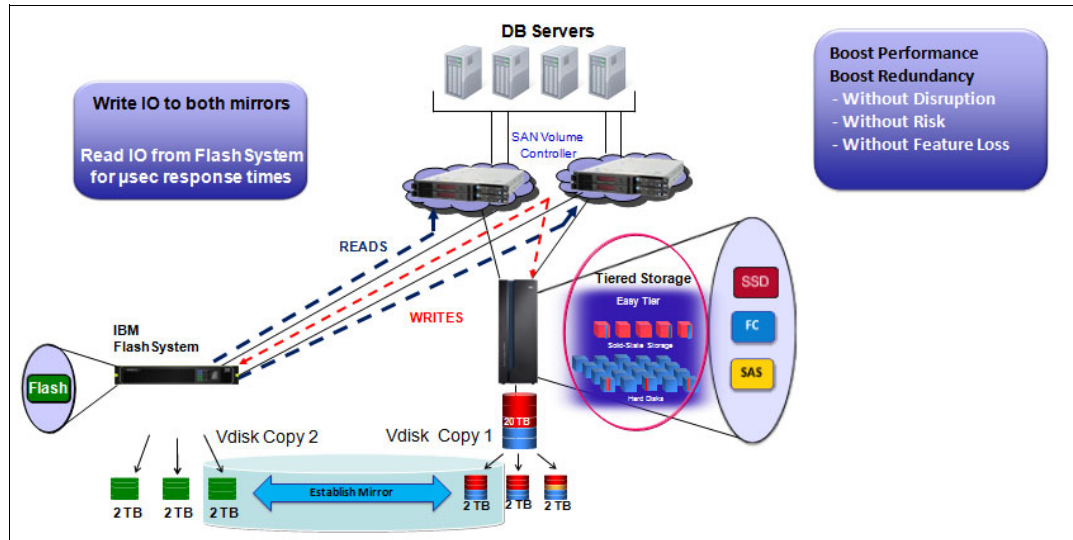


Figure 8-15 IBM FlashSystem and SAN Volume Controller tiered scenario with mirroring

8.4 Integrating FlashSystem 840 and IBM Storwize V7000 considerations

IBM Storwize V7000 is a virtualized software-defined storage system that is designed to consolidate workloads for simplicity of management, reduced cost, highly scalable capacity, performance, and high availability. This system offers improved efficiency and flexibility with built-in flash storage optimization, thin provisioning, and nondestructive migration from existing storage.

A mid-range storage system, either a racked IBM Storwize V7000 or a chassis model (IBM Flex System® V7000 Storage Node), provides rich functionality and capacity. When you examine either type of storage node, which includes the usage of Easy Tier, you want to match a certain amount of fast storage with spinning storage. This is typically a 10:1 ratio. For example, if you wanted to achieve 100 TB of capacity with Easy Tier, you can get 90 TB of spinning capacity and 10 TB of flash capacity.

In the Storwize V7000, you can achieve 10 TB of capacity with SSDs in the controller or expansion unit, but if you price the solution, you might find that 10 TBs of the FlashSystem 840 are more economical. To gain the maximum benefits in this solution, you can use the Storwize V7000 disk slots for the spinning disk capacity, and use the FlashSystem 840 as the fast tier.

Similar to SAN Volume Controller, the IBM Storwize V7000 offers the same functionalities, such as mirroring, FlashCopy, thin provisioning, Real-time Compression (RtC), and remote copy. All suggestions that are described in the SAN Volume Controller section can also be applied to the V7000 integration.

Another similarity with SAN Volume Controller is that you can integrate the FlashSystem 840 with an existing Storwize V7000 environment and a bundled solution: the FlashSystem 840 plus the Storwize V7000. For more information about ordering these bundled solutions, contact IBM services, your IBM representative, or your IBM Business Partner for assistance.

For information about how to deploy the IBM Storwize V7000, see *Implementing the IBM Storwize V7000 V7.4*, SG24-7938 and *Implementing the IBM Storwize V7000 Gen2*, SG24-8244.



Use cases and solutions

The IBM FlashSystem 840 is the perfect solution for a wide variety of use cases and business needs. The FlashSystem 840 can address the I/O issues of applications that require high performance and low latency in relation to storage access.

It is important to know how to apply the advantages that this latest technology can offer to IT market scenarios and solutions.

This chapter gives an overview of how to take advantage of the benefits of the IBM MicroLatency, macro efficiency, enterprise reliability, and extreme performance. This chapter explains where and how to use the FlashSystem as part of your business solution.

The use cases provide examples and usage of real-world implementations, design solutions, and scenarios. The use cases highlight the benefits that each scenario can offer to clients.

The basics of three solution scenarios are covered:

- ▶ Tiering
- ▶ Flash only
- ▶ Preferred read

9.1 Introduction to the usage cases

Certain applications have a natural affinity with the FlashSystem 840. These applications include applications that have a low tolerance to latency, are I/O per second (IOPS)-intense, and need to scale in size and also performance. These types of applications are key candidates for the FlashSystem because of their sometimes unique requirements for storage needs.

There is a set of applications that especially benefit from the FlashSystem solutions. The following list provides ideal candidates for the FlashSystem storage:

- ▶ Online Transaction Processing (OLTP)
- ▶ Online Analytical Processing (OLAP)
- ▶ Virtual desktop infrastructure (VDI)
- ▶ Cloud-scale infrastructure
- ▶ Computational applications

An entire book might be written about each one of these applications. The objective of this chapter is to briefly describe how the IBM FlashSystem 840 can provide value for these applications' special challenges.

Online Transaction Processing (OLTP)

OLTP applications usually rely on a database that needs to be able to serve the application as fast as possible. In OLTP, parallelism allows the application to run as many transactions as possible at the same time. From a storage perspective, too many parallel I/O requests can be challenging because the parallelism limitation of traditional storage directly related to its number of disks.

Because the FlashSystem 840 is not limited by mechanical disk, there is no longer a limitation to allow the application to reach a new level of parallelism, and to be able to run up to 10x or 20x more transactions in up to 1/3 of the time, on average.

The FlashSystem 840 allows databases to run more operations, on average up to more than 10x because of the extreme performance. More transactions can occur in less time because of the MicroLatency benefits.

Data can be collected and analyzed to see the FlashSystem benefits. For DB2, performance can be measured by extracting a db2monreport. Or, in an Oracle database, you can use an Automatic Workload Repository (AWR) report.

For more information about how to benefit from an IBM FlashSystem in a specific OLTP environment, contact your IBM representative. For more information about running the IBM FlashSystem in OLTP environments, see the following website:

<http://www.redbooks.ibm.com/abstracts/tips0973.html?Open>

Online Analytical Processing (OLAP)

OLAP applications are important for the business environment because OLAP is a key player in a wide range of Business Intelligence (BI) tools. OLAP is increasingly taking a more important role every day in industries.

Because the business environment is more complex, decisions increasingly rely on results from BI tools. BI provides the decision maker the opportunity to have a broader vision of the market. Response time makes a significant difference. In OLAP, it is even more evident. A decision made faster can help position you better in the market, but a delayed response might cause a significant negative impact to your business.

OLAP applications use a mathematics model to predict and to calculate behavior, but for that, it is necessary to access a large amount of data (*big data*), so it is here that the FlashSystem is a game changer for organizations. The FlashSystem allows clients to run their analyses and predictions at a new level of speed, helping companies to go to the market faster. It is not only a powerful processor, but the speed of access to big data plays a significant role in OLAP applications.

The FlashSystem 840 with its MicroLatency and extreme performance provides a new competitive advantage to clients.

For more information about running the IBM FlashSystem in OLAP environments, see the following website:

<http://www.redbooks.ibm.com/abstracts/tips0974.html?Open>

9.2 Tiering

A *tiering approach*, also known as *IBM Easy Tier*, is a solution that combines functionalities that can add value to other storage, in combination with the highly advanced key points that the FlashSystem 840 can offer, such as MicroLatency and maximum performance.

The great advantage of the tiering approach is the capability to automatically move the most frequently accessed data to the highest performing storage system. In this case, the FlashSystem 840 is the highest performing storage, and the less frequently accessed data can be moved to slower performing storage, which can be solid-state drive (SSD)-based storage or disk-based storage. The next topic discusses the data movement that can be done at block level or at file level.

Important: Consider that IBM Easy Tier latency with the FlashSystem is much lower than most of the high-end storage devices on the market.

9.2.1 Easy Tier or block-level tiering

Easy Tier is the IBM implementation of *block tiering*. Other vendors have different names for the block movement between different tiers of storage. This topic does not discuss the details of how tiering works. The main objective is to explain how to take the advantage of this implementation using the FlashSystem 840 and to describe which usage case best fits this solution.

Some tiering solutions do not offer the possibility of *external tier*. No external tiering means that the block movement can be done only inside the storage device. Because this is a limited approach and it does not work with the FlashSystem 840, we only discuss external tiering.

The IBM solution for external tier, also known as IBM Easy Tier, is in the IBM System Storage SAN Volume Controller and the IBM Storwize V7000 (V7000).

The SAN Volume Controller offers excellent synergy with the FlashSystem 840. For more information, see *Implementing the IBM SAN Volume Controller and FlashSystem 820*, SG24-8172.

The SAN Volume Controller and the FlashSystem address the combination of the lowest latency with the highest functionality. The V7000 and the FlashSystem address the best cost-benefit solution. Both provide the lowest latency for clients that use traditional disk array storage and need to increase the performance of their critical applications.

Usage of the IBM Easy Tier solution is indicated when there is a need to accelerate general workloads. The SAN Volume Controller will have a map of the *hot* data (more frequently accessed) and the *cold* data (less frequently accessed), and it will move the hot data to the FlashSystem and the cold data to the conventional storage. When data that was previously hot becomes cold, it will move from the FlashSystem to other storage. The inverse process occurs when cold data becomes hot (or more frequently accessed) and is moved from a traditional storage to the FlashSystem (Figure 9-1).

This solution focuses on accelerating and consolidating the storage infrastructure. It might not reach the same lowest latency that a flash-only solution offers, but it is used to improve the overall performance of the infrastructure, avoiding the acquisition of SSDs. The main requirement in this solution is to add more performance to an application or to multiple applications, especially existing applications whose performance is costly to improve by application tuning. The overall latency that the SAN Volume Controller will add to the data path will not compromise the expected latency.

The SAN Volume Controller adds other important features to the solution, such as replication at the storage layer, snapshots, and Real-time Compression.

Figure 9-1 illustrates how Easy Tier works with the FlashSystem 840. The hot data (more frequently accessed data) is moved to the FlashSystem 840 and the cold data (less frequently accessed data) is moved back to the slower disk array storage.

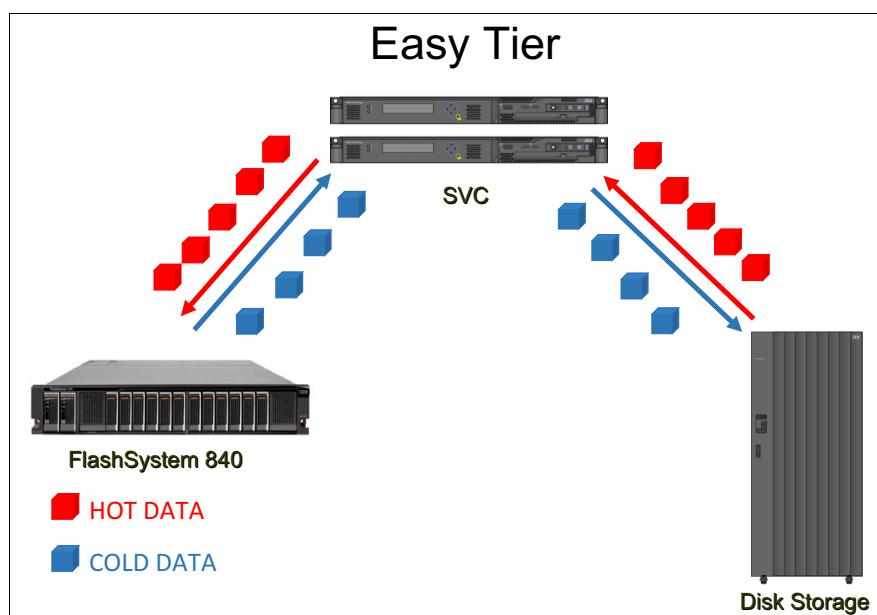


Figure 9-1 Easy Tier moving data between the tiers

Easy Tier with the FlashSystem 840 offers the following benefits to the applications:

- ▶ Capacity efficiency
- ▶ Acceleration to all applications in the same pool
- ▶ Transparent for the application
- ▶ Smart storage determining which is the hot data

9.2.2 Information Life Management or file-level tiering

There is another tiering approach that treats the information at the file level. This type is called *Information Life Management* (ILM). When using the FlashSystem 840 in an environment where the file is the crucial part of the solution, a preferred practice is to keep the most frequently accessed data in the FlashSystem storage. For more information about ILM, see the following website:

<http://www.ibm.com/services/ch/gts/pdf/br-storage-ilm-factsheet-en.pdf>

Some file systems have the capability to move the files between storage pools that contain different kinds of storage, for example, an SSD storage pool and a hard disk drive (HDD) storage pool. In IBM Spectrum Scale, which is a proven, scalable, high-performance data and file management solution (based on IBM General Parallel File System or GPFS technology, also formerly known as code name Elastic Storage), you can create policies and move the files between the pools according to a designed policy. Ideally, keep the core files on a faster disk solution, such as the FlashSystem, which provides a significant performance improvement for the application. Also, file systems typically use the concept of *metadata*, which is where the file's descriptive information is located. Considering that the most accessed information in a file system is the metadata, keeping the metadata in a FlashSystem 840 storage pool can significantly improve the overall performance of a file system. IBM Spectrum Scale is considered highly advantageous for the FlashSystem because it has the ability to operate in parallel fashion. For more information, see the following website:

<http://www.ibm.com/systems/storage/spectrum/scale>

9.3 Preferred read

It is a well-known concept for data administrators that most applications have an I/O profile that has more read events than write events. Traditionally, most databases have a 70/30 I/O profile, which means that the database spends 70% of the time reading and only 30% of the time writing.

We describe how a combination of the FlashSystem and another disk storage can add redundancy and allow the usage of software functionalities in other storage to provide outstanding performance for the applications with standard I/O behavior.

The following examples illustrate the possible combinations in a *preferred read* implementation and also can be expanded to other applications.

The preferred read solution approach allows the application to use the lowest latency for reads, or read at the speed of the FlashSystem, because the main goal of this solution is to send all read requests to the FlashSystem 840.

Preferred read presumes a *write mirroring* relationship, where the I/O is written to one or more types of storage, and all the read I/O requests are served from the FlashSystem 840.

Figure 9-2 on page 310 gives a high-level example of how preferred read works.

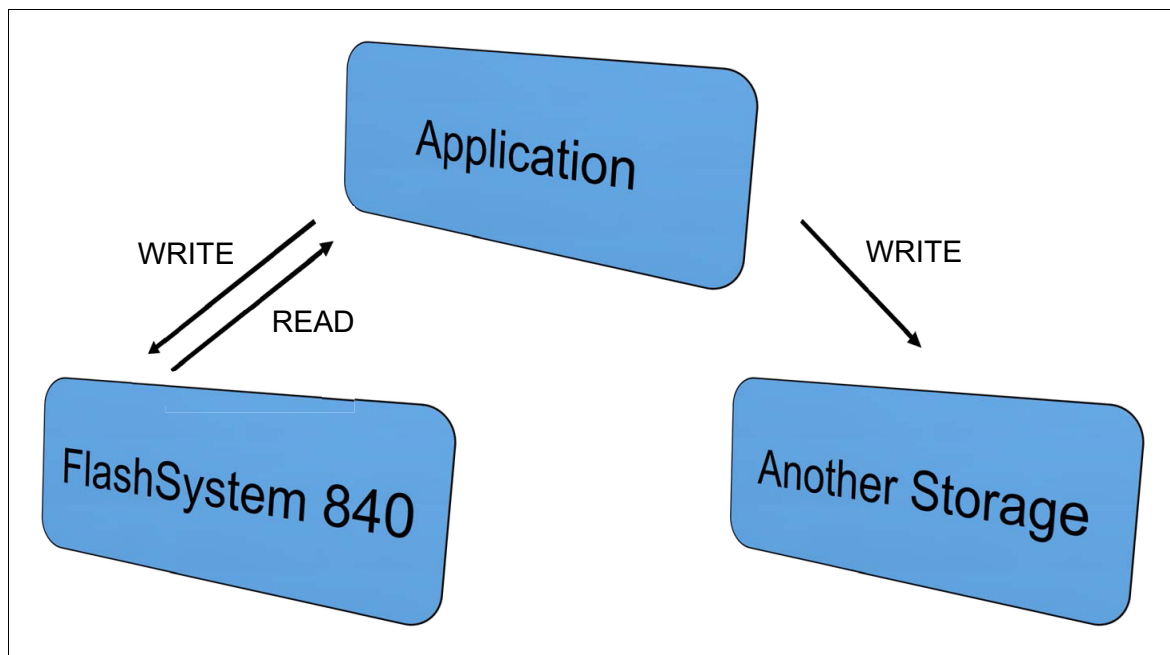


Figure 9-2 Preferred read overview

The preferred read solution adds another level of protection, as Figure 9-3 on page 311 shows. In a disaster situation, for example, if a power failure occurs on the FlashSystem 840 and it becomes unavailable, the application continues accessing the data without interruption and without data loss because in preferred read, all the data is mirrored (in both storage products).

Important: Figure 9-3 on page 311 shows a situation where the FlashSystem is the failed component. It is important to consider in a disaster recovery plan that the other storage can also fail. In this case, the replication is stopped. Consider your plan of action in this situation because preferred read continuously works with the FlashSystem. The suggestion is to set an alert in case one of the components in a preferred read relationship stops responding. Include this scenario in the disaster recovery plan.

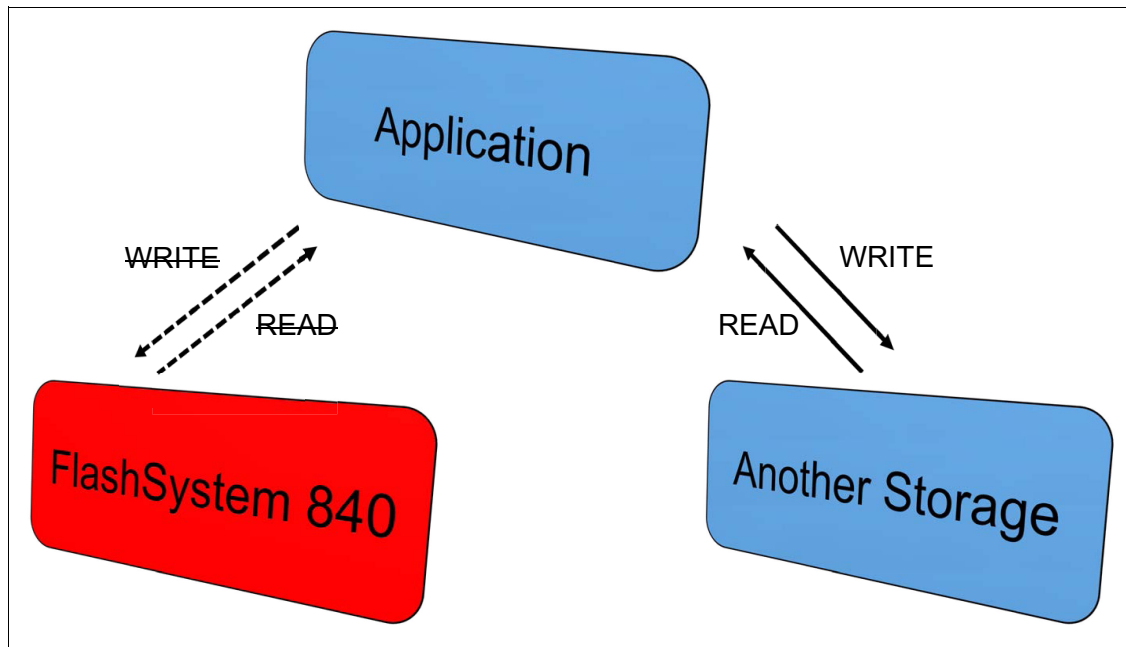


Figure 9-3 Preferred in a disaster situation on the FlashSystem 840

There is no mirroring between the FlashSystem and other storage. The writes are mirrored, or *split*, which means that every write request goes to both storage systems (the FlashSystem 840 and other storage) that are part of a preferred read relationship. The latency for write requests is always at the latency of the slower storage. You might think that this solution can seriously compromise the overall performance, but it is important to remember that, typically, in a traditional storage system, the writes are done in *cache memory*. So, most of the time, write latency is in microseconds because the cache memory has a low response time.

Therefore, the reads performed at the FlashSystem 840 speed (average 70% of the time) and the writes performed with little latency make this solution a good combination of performance and flexibility.

Preferred read can often be implemented on live systems easily and transparently and with no downtime. For more information about how to configure preferred read, see 5.5, “FlashSystem 840 preferred read and configuration examples” on page 143.

The preferred read solution allows the use of features of other storage systems. For example, if there is an existing mirroring relationship, such as Metro Mirror or Global Mirror, the preferred read does not affect this functionality because the replication continues sending the write request to the secondary site. In this scenario, we now have three copies of the data: One in the FlashSystem and the other two copies in the other storage systems.

Figure 9-4 on page 312 provides an illustration about how replication can work in a preferred read relationship.

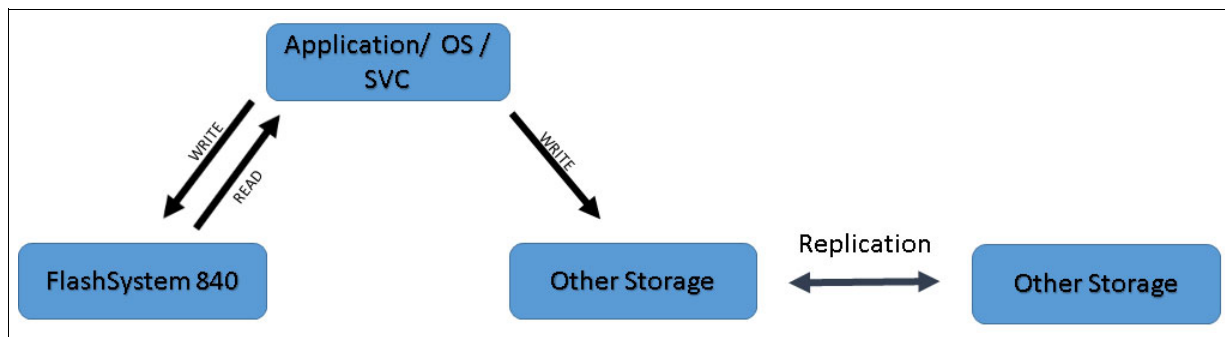


Figure 9-4 Preferred read with replication

Important: In the scenario with replication, consider that the write latency will be the latency to write the data on the secondary site and receive the acknowledgment.

Figure 9-5 gives an example of a solution using Logical Volume Manager (LVM) on site A in a preferred read design. There is also a Metro Mirror relationship between the two other disk storage systems, going from site A to site B.

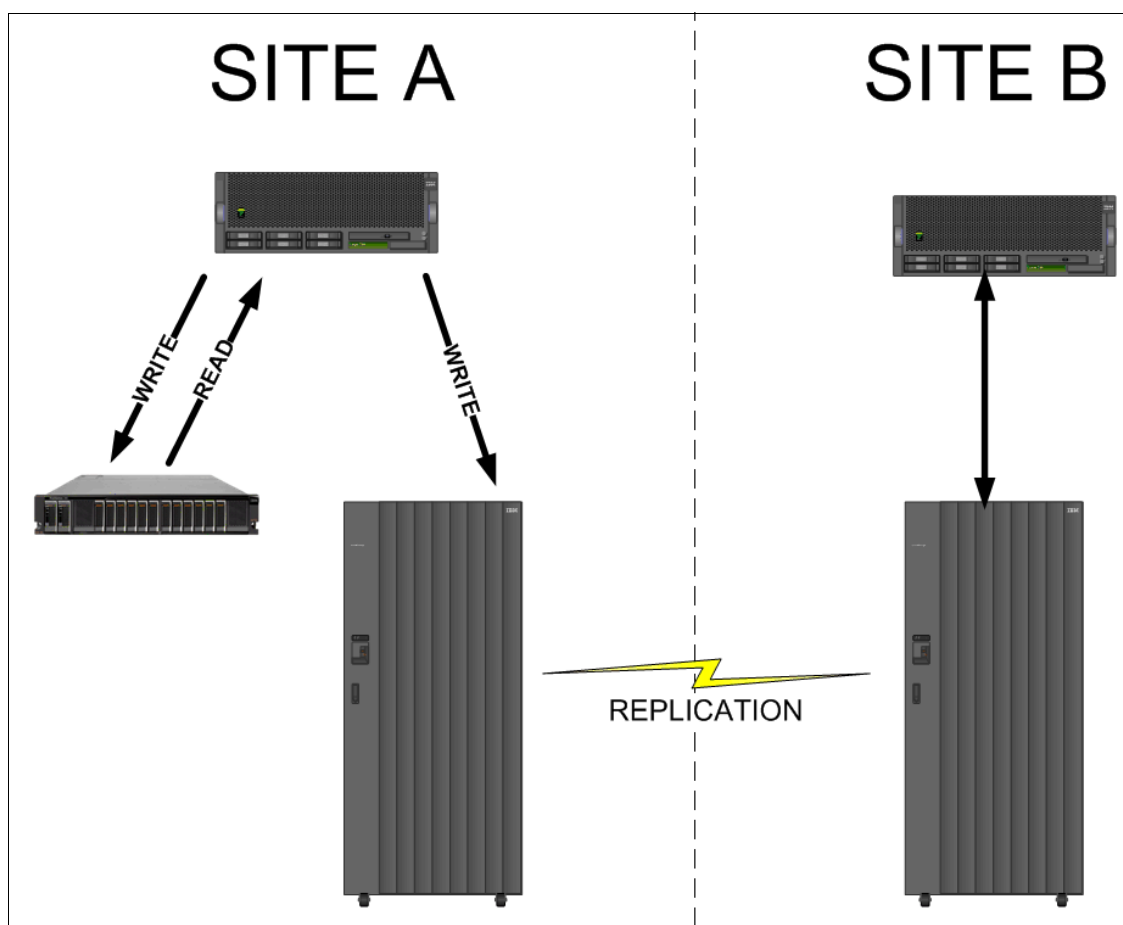


Figure 9-5 Example of preferred read with Metro Mirror in an IBM DS8000®

9.3.1 Implementing preferred read

We describe examples of preferred read implementation in some of the more frequently used scenarios.

There are three ways to implement preferred read:

- ▶ SAN Volume Controller and Storwize V7000
- ▶ Application
- ▶ Operating system

For more information about LVM, see 5.5, “FlashSystem 840 preferred read and configuration examples” on page 143.

Preferred read using the SAN Volume Controller and Storwize V7000

The IBM SAN Volume Controller and the IBM Storwize V7000 in combination with the FlashSystem 840 add functionality to the solution that is not included with only the FlashSystem 840.

Note: When referring to the SAN Volume Controller in this section, the concepts also apply to the IBM Storwize V7000 because the Storwize V7000 has the SAN Volume Controller functionality included as part of its code.

The SAN Volume Controller and the IBM Storwize V7000 can be used to implement a preferred read solution with the FlashSystem 840. Preferred read is implemented in the SAN Volume Controller by using Volume Mirroring or virtual disk (VDisk) Mirroring.

See Figure 9-6 on page 314 for an illustration of this scenario. For more information about how to implement this feature, see *IBM SAN Volume Controller and IBM FlashSystem 820: Best Practices and Performance Capabilities*, REDP-5027.

For more information about the SAN Volume Controller functionality and preferred read, see 5.5, “FlashSystem 840 preferred read and configuration examples” on page 143, and Chapter 8, “Product integration” on page 275.

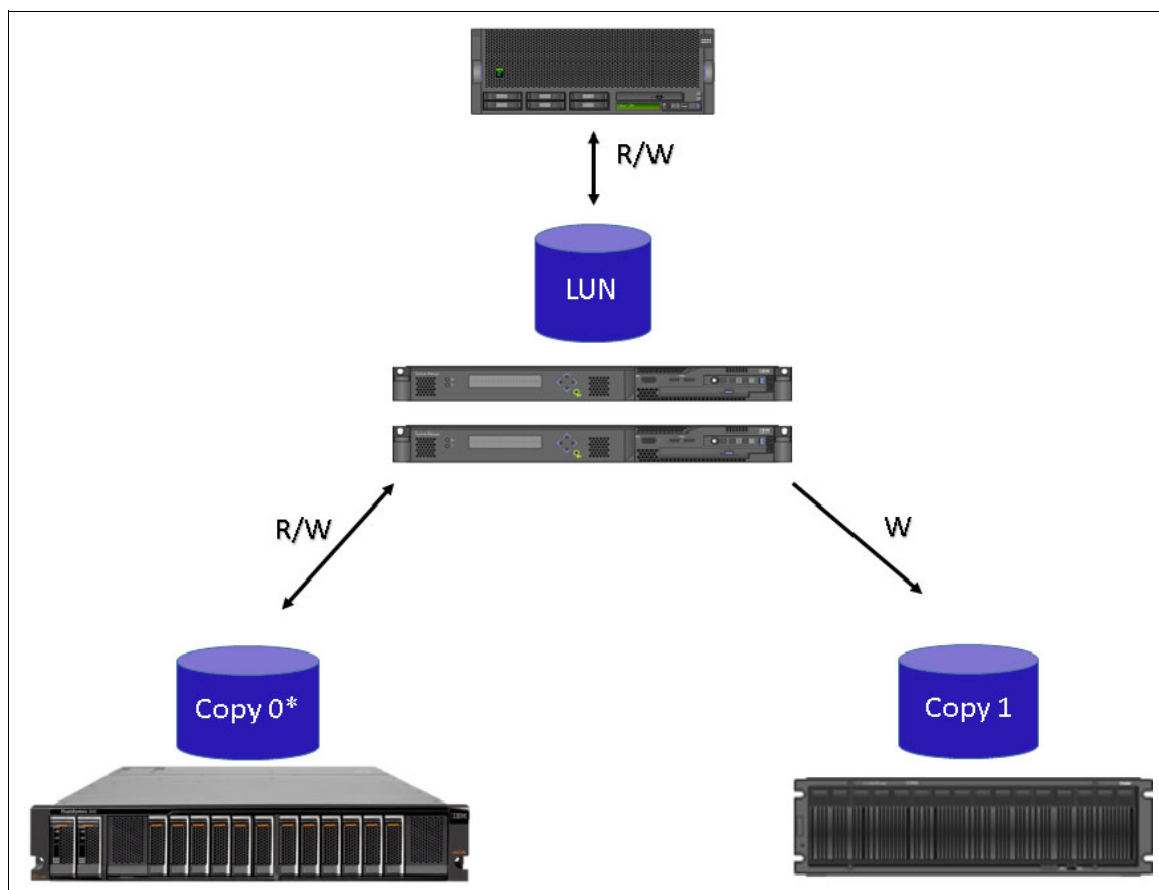


Figure 9-6 Preferred read using the SAN Volume Controller

For more information about the SAN Volume Controller, see 8.1, “Running the FlashSystem 840 with Spectrum Virtualize - SAN Volume Controller” on page 276.

Preferred read using applications

A wide range of applications can implement mirroring of write I/Os and use a preferred read design to read at the FlashSystem speed. Because the applications can be designed and implemented according to the application developer, see the documentation for your application and search by preferred read implementation, split I/O, or I/O mirroring.

For instructions about how to implement Automatic Storage Management (ASM) preferred read on Oracle environments, see 5.5, “FlashSystem 840 preferred read and configuration examples” on page 143.

Note: Always refer to the latest documentation for your application for the latest supported configurations for your environment.

Preferred read using the operating system

Many operating systems, such as Linux, IBM AIX, and Microsoft Windows, have built-in capability to split the I/O between more than one volume by performing a volume mirror at the operating system (OS) level. This kind of functionality usually allows the administrator to decide where the OS will read the data. Other operating systems perform the decision based on the speed of the volumes, using the faster path to storage, in this case, the FlashSystem.

Figure 9-7 shows LVM from AIX controlling the disk and performing a volume mirror at the OS level.

Linux LVM can also be used in this relationship. In the Solaris OS, Solaris Volume Manager (SVM) can be used. In Windows environments, you must use Veritas File System to handle the mirroring of the I/O.

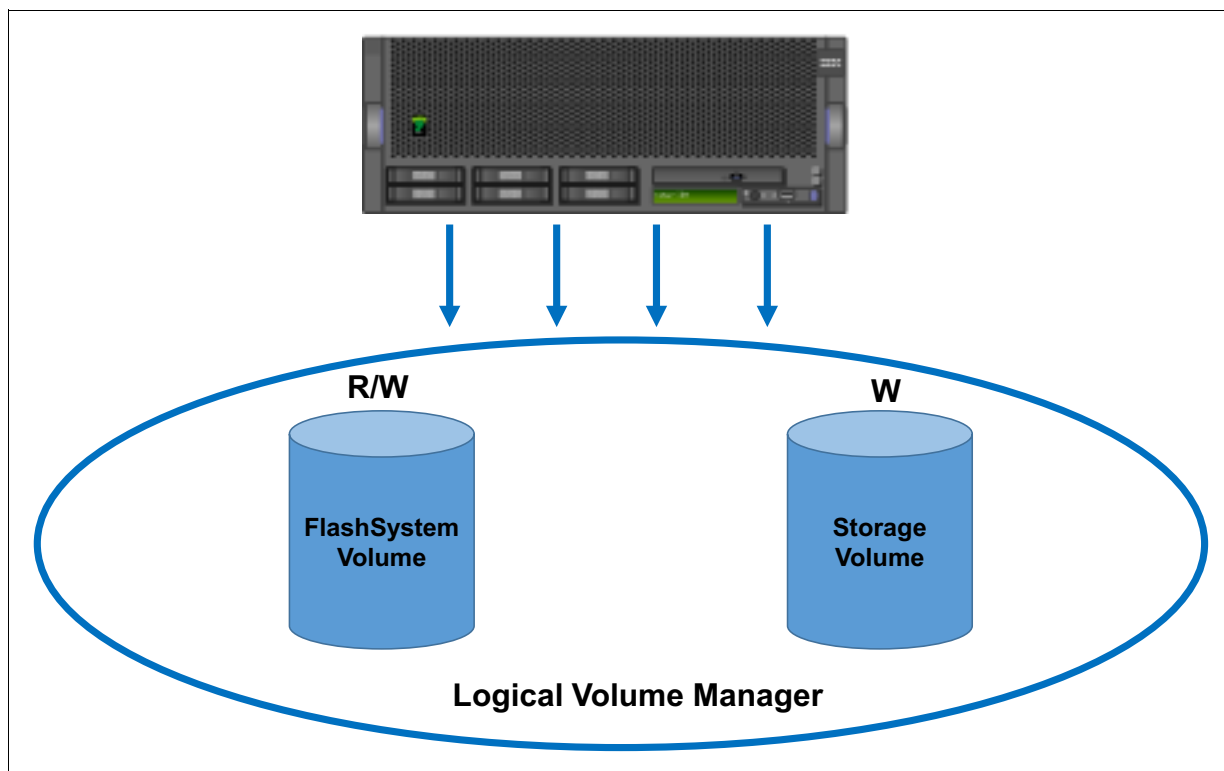


Figure 9-7 Preferred read performed by LVM in an AIX environment

9.4 Flash only

This solution design is also known as *manual placement*. Work with the FlashSystem as a dedicated approach. Either place the entire application or the most important part of the application inside the FlashSystem 840. This way, the applications benefit the most from the extreme performance and MicroLatency.

Certain databases have components that when the components are sped up, they accelerate the application without needing to accelerate the data.

Because the FlashSystem 840 has the potential to be configured with up to 48 TB (RAID 0) of capacity, it is usually more efficient to place the entire application inside the FlashSystem 840 and let the application accelerate at the maximum possible speed.

Figure 9-8 on page 316 illustrates an example where the entire database is placed inside the FlashSystem 840.

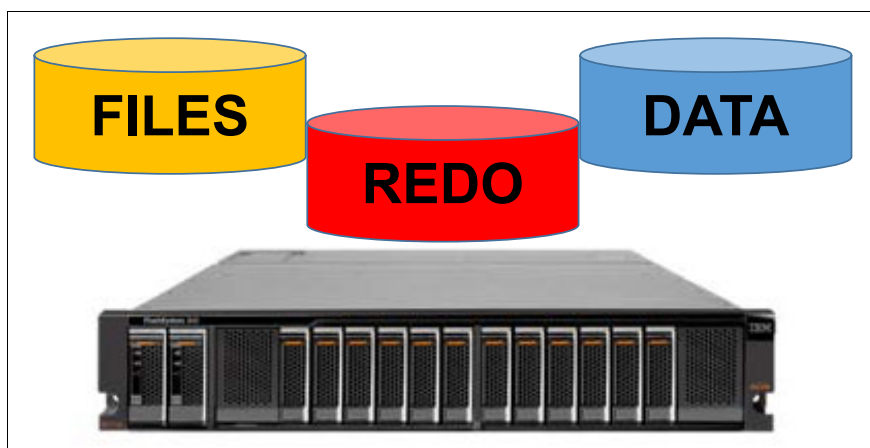


Figure 9-8 Manual data placement example

9.5 Comparison

We described various solution scenarios. We highlight the advantages of each of them to help you decide on the best solution approach to your specific application and business environment.

All applications have unique functionality, which you need to consider when you design a solution to use the FlashSystem 840. When comparing solutions, remember the following items:

- ▶ Number of I/O per second (IOPS)
- ▶ Flexibility
- ▶ Latency

The IBM FlashSystem 840 adds the following benefits to the solution:

- ▶ Extreme performance
- ▶ MicroLatency
- ▶ Macro efficiency
- ▶ Enterprise reliability

There might be situations where flexibility is a priority over latency. There might also be situations where many IOPS and latency are critical and there is less need for flexibility.

Table 9-1 summarizes and compares the benefits of each type of solution.

Table 9-1 Solution benefits comparison

Solution scenario	Number of IOPS	Flexibility	Latency
Tiering	High	Very high	Low
Flash-only or manual placement	Extremely high	Medium	Lowest possible
Preferred read	Very high	High	Very low

Figure 9-9 on page 317 shows the advantages of each solution described in Table 9-1.

EasyTier	• Best capacity efficiency
Preferred Read	• Acceleration and preserves storage investment
Flash Only	• Best performance

Figure 9-9 Highlights of the benefits that each solution can provide

When performance is important, but you need flexibility, use an IBM Easy Tier solution, because all of the SAN Volume Controller features can be used and the administrator does not need to worry about controlling which data is hot.

When every microsecond is important, a flash only solution is the best approach. Systems that need all the capacity of the extreme performance and the MicroLatency typically handle the application flexibility on another level, not on the storage level.

When a balance between the maximum possible performance and some level of flexibility is required, a preferred read solution is the best option. It helps you preserve your investment and offers good flexibility.

When you design a solution, to further understand the suggestions and solution options, you can also engage the IBM client representative to help you to find the best solution for your application.



Hints and tips

This chapter provides information about enabling or changing encryption, Fibre Channel (FC) attachment, hints, and tips for setting up a test environment for the IBM FlashSystem 840 (FlashSystem 840). Useful information that is needed to understand basic performance and troubleshooting information is provided.

10.1 Encryption hints

You can enable and disable encryption. On a system with enabled encryption, you can change the encryption access keys (Rekey):

- ▶ Enabling or disabling encryption

You can enable or disable encryption after the FlashSystem 840 is initialized. The encryption license is required before you enable encryption. Beginning with Version 1, Service Pack 3 (SP3), encryption can be enabled on an encryption-disabled FlashSystem 840 Storage Enclosure while it is running. This is a non-destructive procedure that does not impact customer data.

- ▶ Encryption Rekey

Beginning with Version 1, Service Pack 3 (SP3), encryption access keys can be changed so that the old encryption access key no longer functions, and a new key is required to unlock the system. Rekeying can be performed on an encryption-enabled FlashSystem 840 while it is running. This is a non-destructive procedure that does not impact customer data.

Note: Contact IBM Support for assistance with this procedure.

10.2 System check

Before investigating performance problems, ensure that the health status of the FlashSystem 840 environment is good. You will need to look at these components.

You can use the FlashSystem 840 graphical user interface (GUI) to access the event messages. Go to **Monitoring → Events** and check for errors. You can also use the **1seventlog** CLI command. Make sure that there are no errors.

10.2.1 Checking the Fibre Channel connections

FlashSystem 840 can be attached with up to sixteen 8 Gbps or eight 16 Gbps adapters. Example 10-1 shows the FlashSystem 840 **lsfabric840** command listing the local and remote WWPNs. In this example, the system is attached to an IBM SAN Volume Controller.

Example 10-1 FlashSystem 840 fabric840 information

```
IBM_Flashsystem:FlashSystem-840-03:superuser>lsfabric840
remote_uid      local_uid      canister_id adapter_id port_id node_id node_name type
500507680C120951 500507605E800E41 1          1          1          1      node1    fc
500507680C120951 500507605E800E51 1          2          1          1      node1    fc
500507680C120951 500507605E800E61 2          1          1          2      node2    fc
500507680C120951 500507605E800E71 2          2          1          2      node2    fc
500507680C110951 500507605E800E41 1          1          1          1      node1    fc
500507680C110951 500507605E800E51 1          2          1          1      node1    fc
500507680C110951 500507605E800E61 2          1          1          2      node2    fc
500507680C110951 500507605E800E71 2          2          1          2      node2    fc
500507680140F941 500507605E800E41 1          1          1          1      node1    fc
500507680140F941 500507605E800E51 1          2          1          1      node1    fc
500507680140F941 500507605E800E61 2          1          1          2      node2    fc
500507680140F941 500507605E800E71 2          2          1          2      node2    fc
500507680C220951 500507605E800E42 1          1          2          1      node1    fc
500507680C220951 500507605E800E52 1          2          2          1      node1    fc
```


500507680C220951	500507605E800E62	2	1	2	2	node2	fc
500507680C220951	500507605E800E72	2	2	2	2	node2	fc
500507680C210951	500507605E800E42	1	1	2	1	node1	fc
500507680C210951	500507605E800E52	1	2	2	1	node1	fc
500507680C210951	500507605E800E62	2	1	2	2	node2	fc
500507680C210951	500507605E800E72	2	2	2	2	node2	fc
500507680110F941	500507605E800E42	1	1	2	1	node1	fc
500507680110F941	500507605E800E52	1	2	2	1	node1	fc
500507680110F941	500507605E800E62	2	1	2	2	node2	fc
500507680110F941	500507605E800E72	2	2	2	2	node2	fc

Check all connections at the switch level and host level, too.

10.3 Host attachment hints

This section lists some common host attachment issues.

10.3.1 Fibre Channel link speed

It is a preferred practice to set the FlashSystem 840 port link speed and topology to fixed values.

Example 10-2 shows fixed values for a SAN-attached FlashSystem 840. The first port of every FC interface card is connected to a host, and their speed is set to 8 Gbps. The second port of every FC interface card is connected to a switch, and their speed is set to 16 Gbps.

Example 10-2 SAN-attached FlashSystem 840 with fixed values

```
#
>lsportfc
```

id	canister_id	adapter_id	port_id	type	port_speed	node_id	node_name	WWPN	nportid	status	attachment	topology
0	1	1	1	fc	8Gb	1	node1	5005076000000041	000001	active	host	pp
1	1	1	2	fc	16Gb	1	node1	5005076000000042	530500	active	switch	pp
4	1	2	1	fc	8Gb	1	node1	5005076000000051	000001	active	host	pp
5	1	2	2	fc	16Gb	1	node1	5005076000000052	570800	active	switch	pp
8	2	1	1	fc	8Gb	2	node2	5005076000000061	000001	active	host	pp
9	2	1	2	fc	16Gb	2	node2	5005076000000062	570A00	active	switch	pp
12	2	2	1	fc	8Gb	2	node2	5005076000000071	000001	active	host	pp
13	2	2	2	fc	16Gb	2	node2	5005076000000072	530400	active	switch	pp

You can set the speed and topology of the ports by using the **chportfc** command.

Example 10-3 shows an example to set a port to point-to-point topology and an 8 Gbps speed.

Example 10-3 Using the chportfc command to set a port topology and speed

```
>chportfc -topology pp 12
>chportfc -speed 8 12
>chportfc -reset 12
>lsportfc 12
id 12
canister_id 2
adapter_id 2
port_id 1
type fc
port_speed 8Gb
node_id 2
node_name node2
```



```
WWPN 5005076000000071
nportid 000001
status active
switch_WWPN 0000000000000000
fpma
vlanid
fcf_MAC
attachment host
topology pp
topology_auto no
speed_auto no
```

The use of the command-line interface (CLI) is described in 6.5.3, “Access CLI by using PuTTY” on page 229.

10.3.2 Host is in a degraded state

A host is shown in a *degraded* state if the host or the port is not connected to both FlashSystem 840 canisters. For details and examples, see 5.2.1, “Fibre Channel SAN attachment” on page 112.

10.3.3 FlashSystem port status

You can use the CLI **lspportfc** command to check the status of a FlashSystem 840 FC port. Three other values are possible for the state:

- ▶ **active**
The port is online and the link to the host/switch is established.
- ▶ **inactive_configured**
This port is online and the link to the host/switch is not working. You can check different points to get more information:
 - Cabling
 - Small form-factor pluggable (SFP)
 - Host's ports
 - Switch for disabled or incorrectly configured ports
 - Link speed and topology of all involved ports
- ▶ **inactive_unconfigured**
The port is disabled. You can verify that an SFP is installed. Ports might be disabled if an SFP was not present on the controller start. Check the logs for error port-related messages.

Use the **lspportib** or **lspportip** command to check the status of InfiniBand (IB) or iSCSI ports. The output is similar to the **lspportfc** command output. You can also check for a correct IB or iSCSI host setup.

10.3.4 AIX multipathing

You have to update AIX Object Data Manager (ODM) to recognize the new IBM FlashSystem 840 device as a multipath I/O (MPIO) device if the FlashSystem 840 is recognized as another disk instead of as an MPIO disk. You can start by using the information at the following

website and then follow the references in the line starting with “APAR is sysrouted TO” to get the correct ODM driver for your AIX operating system level:

<http://www.ibm.com/support/docview.wss?uid=isg1IV50154>

10.3.5 Direct attach hints

Always use the latest host driver for direct attachment, especially when connecting with 16 Gb FC. Check the operating system. Certain operating systems currently do not support 16 Gb direct attachment.

Note: Fibre Channel over Ethernet (FCoE) does not support direct attachment.

10.4 General guidelines for testing a specific configuration

You can evaluate the performance gain when using the FlashSystem 840 by setting up a test environment. This environment can be a dedicated testing environment, or you can use the existing environment. If you use your existing environment, you can use *preferred read* as an unobtrusive way to demonstrate the performance of the FlashSystem 840. For more information, see 10.4.2, “Test scenarios” on page 324.

You need to document the test by gathering information before the test, during the test, and at the end of the test.

Follow these steps to document the necessary information:

1. Specify expectations.

You can have a successful test when you set the correct expectations about the performance gain of the test. You can set the correct expectations by analyzing the current environment on the operating system level, or by analyzing the application. For example, the Oracle Automatic Workload Repository (AWR) contains statistics about the Oracle database I/O. To see the commands for gathering statistical data at the operating system level, see 10.4.5, “Performance data gathering basics” on page 325.

2. Create a baseline.

A *baseline* contains all the information that you will compare after the test to check the success of the test, and to verify whether you have fulfilled the expectations. You create the baseline with the current storage environment. The baseline contains values for the latency, I/O per second (IOPS), bandwidth, and run time of batch jobs. This is the first baseline created in the current environment without a FlashSystem.

3. Diagram the current configuration and the testing configuration.

A diagram helps you to visualize the current configuration and the setup of the FlashSystem 840.

4. Set up a test plan.

The test plan contains milestones for every test.

5. Have the correct data.

Ensure that you have the correct data and enough data to run the tests.

6. Create a second baseline in the testing configuration.

The first baseline was created in the configuration without the FlashSystem. This second baseline is created in the configuration with the FlashSystem. You compare these two baselines to check the success of the test.

7. Document your results.

The result section of your documentation is the most important part.

10.4.1 Save the default configuration

Before testing the FlashSystem 840 performance, it is advised to save the configuration of the system. By saving the current configuration, you create a backup of the licenses that are installed on the system. This assists you in restoring the system to the default settings at the end of the testing. You can save the configuration by using the **svcconfig backup** command.

The next two steps show how to create a backup of the configuration file and to copy the file to another system:

1. Log in to the cluster IP using a Secure Shell (SSH) client and back up the FlashSystem configuration:

```
superuser>svcconfig backup
```

```
.....
```

```
CMMVC6155I SVCCONFIG processing completed successfully
```

2. Copy the configuration backup file off the system.

Using secure copy, copy the file `/tmp/svc.config.backup.xml` from the system and store it.

For example, use **pscp.exe**, which is part of the **PuTTY** commands family. The use of the CLI is described in 6.5.3, “Access CLI by using PuTTY” on page 229:

```
pscp.exe superuser@<cluster_ip>:/tmp/svc.config.backup.xml .
```

```
superuser@ycluster_ip> password:
```

```
svc.config.backup.xml      | 163 kB | 163.1 kB/s | ETA: 00:00:00 | 100%
```

Note: Be sure to save your default configuration before you create a logical unit number (LUN) or a host.

10.4.2 Test scenarios

Test scenarios are provided that illustrate some of the simplest implementations of the FlashSystem 840, an optimal test scenario, and implementation for preferred read.

The simplest implementations of the FlashSystem 840 include the following functions:

- ▶ Easy Tier
- ▶ Preferred read (FlashSystem 840 is the faster half of mirroring with another type of storage.)
- ▶ Manual tiering (FlashSystem 840 is used as exclusive storage for critical data.)

In test situations, the most important tests are the preferred read and manual tiering variants. An optimal scenario might include these tests:

- ▶ Test the baseline by using a standard disk array

- ▶ Test by using a preferred read deployment
- ▶ Test by using a manual tiering deployment

Implementing preferred read with the FlashSystem 840 gives you an easy way to deploy the FlashSystem 840 in an existing environment. The data is secured by writing it to two different storage systems. Data is read at the FlashSystem 840 speed, because it is always read from the FlashSystem 840. This implementation does not change the existing infrastructure concepts, for example, data security, replication, backup, disaster recovery, and so on.

Examples of preferred read configurations are shown in 5.5.1, “FlashSystem 840 deployment scenario with preferred read” on page 143.

10.4.3 Data center environment

Always check the IBM System Storage Interoperation Center (SSIC) to get the latest information about supported operating systems, hosts, switches, and so on:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

For information about the prerequisites for the host operating system, see Chapter 5, “IBM FlashSystem 840 client host attachment and implementation” on page 111.

10.4.4 Secure erase of data

Some clients, especially in the financial sector, are extremely concerned about data confidentiality. The FlashSystem 840 uses encryption to secure data. If you have a license for the FlashSystem 840 encryption, you can prevent unauthorized access to the FlashSystem data.

Important: Deleting the FlashSystem 840 encryption key prevents any access to the data on the FlashSystem 840.

The flash cards can be decommissioned by using the `chdrive -task erase` CLI command. The erasure task has a quick mode (around 4 minutes) that uses a crypto-erase algorithm and a normal task that erases and overwrites flashcards with patterned data.

10.4.5 Performance data gathering basics

All technical resources (application, database, server, and storage) are described. The type of information needed to understand the performance characteristics of the environment is explained.

The focus is performance data gathering from the OS/server perspective, the database perspective, and the storage I/O-profile perspective.

The following metrics are used to determine the performance of a storage subsystem:

- ▶ Average read and write IOPS
- ▶ Average read and write bandwidth (MBps)
- ▶ Average latency/wait time/response time
- ▶ Average read/write ratio
- ▶ Average I/O queue depth
- ▶ CPU utilization

The following parameters are optional:

- ▶ Service time (response time - queue time)
- ▶ % busy
- ▶ Average request size
- ▶ Block sizes and block size ratios
- ▶ Average read and write IOPS per block size

You can check for CPU bottlenecks by gathering CPU information with the I/O information.

To evaluate the effects of the FlashSystem 840, perform a comparison before and after implementing the FlashSystem 840. For an initial performance evaluation or to estimate the performance improvement regarding storage only, three parameters are needed:

- ▶ Average read/write latency
- ▶ Average read/write IOPS
- ▶ CPU usage

More parameters help you to understand the I/O profile of an application.

Simulating the performance gain with the FlashSystem 840

You can simulate the workload by using tools, such as *IOmeter* or *vdbench*. You need a high performance server and the FlashSystem 840 to simulate the workload. You take the performance value's average read and write IOPS per block size and the block size ratio to enter these values as a workload description in *IOmeter* or *vdbench*.

When you run this workload pattern in the original environment (the environment in which you gathered the performance data), you see the same IOPS numbers as before. If you now run this workload pattern on a system with the FlashSystem, you see a better IOPS number. The ratio of the new IOPS to the original IOPS is the performance gain.

Server-side information

Different operating systems and their different methods to gather performance data are described. Operating systems have different commands and programs to collect performance data.

To estimate the performance improvement, you gather these parameters during the peak usage of an application. This is a 1 - 2-hour time frame.

AIX *iostat* command

Use the following syntax to create *iostat* information about AIX:

- ▶ ***iostat -DRt1 <INTERVAL> <COUNT>***
- ▶ ***iostat -T <INTERVAL> <COUNT>***

Example 10-4 shows the command.

Example 10-4 AIX and *iostat*

```
# iostat -DRt1 3 2400 > iostat_disk.txt
# iostat -T 3 2400 > iostat_cpu.txt
```

Set the values *INTERVAL* and *COUNT* accordingly; an interval of 3 - 5 seconds is enough. Ensure that you start both commands at the same point in time. You can add the **-V** option to the first command to get only the active disks or only the disks that are used. Example 10-5 on page 327 shows using the *iostat* command with the **-V** option.

Example 10-5 AIX and iostat with the -V option

```
# iostat -DRT1V 3 2400 > iostat_disk.txt
# iostat -T 3 2400 -V > iostat_cpu.txt
```

The main information from the output is the average response time and the CPU usage. You need to look for these values:

- ▶ CPU = % user + % system
- ▶ Await = Average Wait Time per I/O (IOPS/AvgRespTm); Average Response Time
- ▶ rps = The number of read transfers per second (read IOPS)
- ▶ avgserv (read) = The average service time per read transfer
Different suffixes are used to represent the unit of time. The default is in milliseconds (read latency).
- ▶ \$wps = The number of writes/transfers per second (write IOPS)
- ▶ avgserv (write) = The average service time per read transfer
Different suffixes are used to represent the unit of time. The default is in milliseconds (write latency).

AIX nmon

The **nmon** command displays local system statistics in interactive mode and records system statistics in recording mode. To collect performance data, use the **nmon** command in recording mode for 1 - 2 hours.

You can download the **nmon** tool from the IBM developerWorks® site:

<http://www.ibm.com/developerworks>

Use the following syntax to create **nmon** information about AIX:

```
# nmon -F <FILENAME> -T -d -A -^ -s <INTERVAL> -c <COUNT>
```

Example 10-6 shows the use of the **nmon** command.

Example 10-6 AIX and nmon

```
# nmon -F /tmp/host1_01122012.nmon -T -d -A -^ -s 3 -c 1200
```

Set the *FILENAME*, *INTERVAL*, and *COUNT* values accordingly; an interval of 3 - 5 seconds is enough.

Windows perfmon

The Windows **perfmon** program is used to monitor system activities and resources, such as the CPU, memory, network, and disk. You can use it by starting **perfmon.msc** or **perfmon.exe**.

Table 10-1 on page 328 lists the items that you select within the **perform** program to gather the performance values.

Table 10-1 Windows perfmon program

Group	Item	Description
Processor	Processor: % Processor Time	CPU time spent
Queue depth	Physical Disk: Avg. Disk Queue Length	Queue length
	Physical Disk: Avg. Disk Write Queue Length	Queue length for writes
	Physical Disk: Avg. Disk Read Queue Length	Queue length for reads
	Physical Disk: Current Disk Queue Length	Current queue length
Block sizes	Physical Disk: Avg. Disk Bytes/Read	Read block size in bytes
	Physical Disk: Avg. Disk Bytes/Write	Write block size in bytes
	Physical Disk: Avg. Disk Bytes/Transfer	R/W block size in bytes
Latency (seconds)	Physical Disk: Avg. Disk Sec/Read	Read latency
	Physical Disk: Avg. Disk Sec/Write	Write latency
	Physical Disk: Avg. Disk Sec/Transfer	Read/write latency
Bandwidth (bytes)	Physical Disk: Disk Read Bytes/sec	Bandwidth (reads)
	Physical Disk: Disk Write Bytes/sec	Bandwidth (writes)
	Physical Disk: Disk Bytes/sec	Bandwidth (total)
IOPS	Physical Disk: Disk Reads/sec	IOPS (reads)
	Physical Disk: Disk Writes/sec	IOPS (writes)
	Physical Disk: Disk Transfers/sec	IOPS (total)

Citrix iostat

Example 10-7 shows how to use the **iostat** command on Citrix.

Example 10-7 Citrix iostat

```
# iostat -x dm 3 1200 > results.txt
```

Solaris iostat

Example 10-8 shows how to use the **iostat** command on Solaris.

Example 10-8 Solaris iostat

```
# iostat -xMnz 3 1200 > results.txt
```

Linux iostat

Example 10-9 shows how to use the **iostat** command for disk and CPU information on Linux.

Example 10-9 Linux iostat

```
# iostat -xkNt 3 1200 > results.txt
# iostat -c 3 1200 > results-cpu.txt
```


Table 10-2 describes the **iostat** return values.

Table 10-2 Description of iostat return values

Item	Description
rrqm/s	The number of read requests merged per second queued to the device.
wrqm/s	The number of write requests merged per second queued to the device.
r/s	The number of read requests issued to the device per second.
w/s	The number of write requests issued to the device per second.
kB/s	The number of kilobytes written to the device per second.
rMB/s	The number of megabytes read from the device per second.
wMB/s	The number of megabytes written to the device per second.
avgrq-sz	Average size (in sectors) of the requests issued to the device.
avgqu-sz	Average queue length of the requests issued to the device.
await	Average time (ms) of I/O requests issued to the device to be served. This includes time spent in the queue and time spent servicing them.
svctm	Average service time (ms) for I/O requests issued to the device. <i>Warning:</i> Do not trust this field any longer. This field will be removed in a future sysstat version.
%util	% of CPU time during which I/O requests were issued to the device (bandwidth utilization for the device). Device saturation occurs when this value is close to 100%.

10.5 Troubleshooting

Hints and information needed to determine, report, and resolve problems are described in this section.

10.5.1 Troubleshooting prerequisites

Taking advantage of certain configuration options and ensuring that vital system access information is recorded make the process of troubleshooting easier.

Record access information

It is important that anyone who has responsibility for managing the system knows how to connect to and log on to the system. This is critical during times when system administrators are not available because of vacation or illness.

Record the following information and ensure that authorized people know how to access the information.

Management IP address

The management IP address connects to the system by using the management GUI or starts a session that runs the CLI commands. Record this address and any limitations regarding where it can be accessed from within your Ethernet network.

Follow power management procedures

Access to your volume data can be lost if you incorrectly power off all or part of a system.

Use the management GUI or the CLI commands to power off a system. Using either of these methods ensures that any volatile data is written to the flash modules and the system is shut down in an orderly manner.

Set up event notifications

Configure your system to send notifications when a new event is reported.

Correct any issues reported by your system as soon as possible. To avoid monitoring for new events by constantly monitoring the management GUI, configure your system to send notifications when a new event is reported. Select the type of event that you want to be notified about. For example, restrict notifications to only events that require immediate action. Several event notification mechanisms exist:

- Email

An event notification can be sent to one or more email addresses. This mechanism notifies individuals of problems. Individuals can receive notifications wherever they have email access, including mobile devices.

- Simple Network Management Protocol (SNMP)

An SNMP trap report can be sent to a data center management system that consolidates SNMP reports from multiple systems. Using this mechanism, you can monitor your data center from a single workstation.

- Syslog

A syslog report can be sent to a data center management system that consolidates syslog reports from multiple systems. Using this mechanism, you can monitor your data center from a single workstation.

If your system is within warranty or you have a hardware maintenance agreement, configure your system to send email events to IBM Support if an issue that requires hardware replacement is detected. This mechanism is called *Call Home*. When this event is received, IBM Support automatically opens a problem report, and if appropriate, contacts you to verify whether replacement parts are required.

If you set up Call Home to IBM Support, ensure that the contact details that you configure are correct and kept up-to-date as personnel changes.

Set up inventory reporting

Inventory reporting is an extension of the Call Home email.

Rather than reporting a problem, an email is sent to IBM Support that describes your system hardware and critical configuration information. Object names and other information, such as IP addresses, are not sent. The inventory email is sent regularly. Based on the information that is received, IBM Support can inform you whether the hardware or software that you are using requires an upgrade because of a known issue.

Back up your data

The storage system needs to back up your control enclosure configuration data to a file every day. This data is replicated on each control node canister in the system. Download this file regularly to your management workstation to protect the data. This file must be used if there is a serious failure that requires you to restore your system configuration. It is important to back up this file after modifying your system configuration.

Resolve alerts in a timely manner

Your system reports an alert when there is an issue or a potential issue that requires user attention. Perform the recommended actions as quickly as possible after the problem is reported. Your system is designed to be resilient to most single hardware failures. However, if you operate for any period with a hardware failure, the possibility increases that a second hardware failure can result in some unavailable volume data.

If there are a number of unfixed alerts, fixing any one alert might become more difficult because of the effects of the other alerts.

Keep your software up-to-date

Check for new code releases and update your code regularly. Use the management GUI or check the IBM Support website to see if new code releases are available.

The release notes provide information about new function in a release, plus any issues that are resolved. Update your code regularly if the release notes indicate an issue to which you might be exposed.

Subscribe to support notifications

Subscribe to support notifications so that you are aware of best practices and issues that might affect your system. Subscribe to support notifications by visiting the IBM Support page on the IBM website. By subscribing, you are informed of new and updated support site information, such as publications, hints and tips, technical notes, product flashes (alerts), and downloads.

Know your IBM warranty and maintenance agreement details

If you have a warranty or maintenance agreement with IBM, know the details that must be supplied when you call for support. Support personnel also ask for your customer number, machine location, contact details, and the details of the problem.

10.5.2 User interfaces for servicing your system

Your system provides a number of user interfaces to troubleshoot, recover, or maintain your system. The interfaces provide various sets of facilities to help resolve situations that you might encounter.

Management GUI interface

The management GUI is a browser-based GUI for configuring and managing all aspects of your system. It provides extensive facilities to help troubleshoot and correct problems.

You use the management GUI to manage and service your system. Click **Monitoring** → **Events** for access to problems that must be fixed and maintenance procedures that step you through the process of correcting the problem.

The information on the Events panel can be filtered three ways:

- ▶ Recommended action (default)

This method shows only the alerts that require attention. Alerts are listed in priority order and need to be fixed sequentially by using the available fix procedures. For each problem that is selected, you can perform these tasks:

- Run a fix procedure.
- View the properties.

- ▶ Unfixed messages and alerts

This method displays only the alerts and messages that are not fixed. For each entry that is selected, you can perform these tasks:

- Run a fix procedure.
- Mark an event as fixed.
- Filter the entries to show them by specific minutes, hours, or dates.
- Reset the date filter.
- View the properties.

- ▶ Show all

This method displays all event types, whether they are fixed or unfixed. For each entry that is selected, you can perform these tasks:

- Run a fix procedure.
- Mark an event as fixed.
- Filter the entries to show them by specific minutes, hours, or dates.
- Reset the date filter.
- View the properties.

Some events require a certain number of occurrences in the 25 hours before they are displayed as unfixed. If they do not reach this threshold in 25 hours, they are flagged as expired.

You can also sort events by time or error code. When you sort by error code, the most serious events, those with the lowest numbers, are displayed first. You can select any event that is listed and select **Actions** → **Properties** to view details about the event. You can view the following information:

- ▶ Recommended actions filter. For each problem that is selected, you can perform these tasks:
 - Run a fix procedure.
 - View the properties.
- ▶ Event log. For each entry that is selected, you can perform these tasks:
 - Run a fix procedure.
 - Mark an event as fixed.
 - Filter the entries to show them by specific minutes, hours, or dates.
 - Reset the date filter.
 - View the properties.

When to use the management GUI

The management GUI is the primary tool that is used to service your system.

Regularly monitor the status of the system by using the management GUI. If you suspect a problem, use the management GUI first to diagnose and resolve the problem.

Use the views that are available in the management GUI to verify the status of the system, the hardware devices, the physical storage, and the available volumes. Click **Monitoring** → **Events** for access to all problems that exist on the system. Use the Recommended Actions filter to display the most important events that need to be resolved.

If there is a service error code for the alert, you can run a fix procedure that assists you in resolving the problem. These fix procedures analyze the system and provide more information about the problem. They suggest actions to take and step you through the actions that automatically manage the system where necessary. Finally, the fix procedure checks that the problem is resolved. If there is an error that is reported, always use the fix procedures within the management GUI to resolve the problem. Always use the fix procedures for both system configuration problems and hardware failures. The fix procedures analyze the system to ensure that the required changes do not cause volumes to be inaccessible to the hosts. The fix procedures automatically perform configuration changes that are required to return the system to its optimum state.

Accessing the management GUI

This procedure describes how to access the management GUI:

1. Start a supported web browser and point the browser to the management IP address of your system.
2. When the connection is successful, you see a login panel.
3. Log on by using your user name and password.
4. When you have logged on, select **Monitoring** → **Events**.
5. Ensure that the events log is filtered by selecting **Recommended Actions**.
6. Select the recommended action and run the fix procedure.
7. Continue to work through the alerts in the order suggested, if possible.

After all the alerts are fixed, check the status of your system to ensure that it is operating as expected.

Using fix procedures

You can use fix procedures to diagnose and resolve problems with the system.

For example, to repair the system, you might perform the following tasks:

- ▶ Analyze the event log.
- ▶ Replace failed components.
- ▶ Verify the status of a repaired device.
- ▶ Restore a device to an operational state in the system.
- ▶ Mark the error as fixed in the event log.

Fix procedures help simplify these tasks by automating as many of the tasks as possible.

The example uses the management GUI to repair a system. Perform the following steps to start the fix procedure:

1. Click **Monitoring** → **Events** and ensure that you are filtering the event log to display the recommended actions.

The list might contain any number of errors that must be repaired. If multiple errors are on the list, the error at the top of the list is the highest priority and must always be fixed first. If you do not fix the higher priority errors first, you might not be able to fix the lower priority errors.

2. Select the error at the top of the list or the subsequent errors to repair.

3. Click **Run Fix Procedure**.

The panel displays the error code and provides a description of the condition.

4. Click **Next** to go forward or **Cancel** to return to the previous panel.

5. One or more panels might be displayed with instructions for you to replace parts or perform other repair activity. If you are not able to complete the actions at this time, click **Cancel** until you return to the previous panel. Click **Cancel** until you are returned to the Next Recommended Actions panel. When you return to the fix procedures, the repair can be restarted from step 1. When the actions that you are instructed to perform are complete, click **OK**. When the last repair action is completed, the procedures might attempt to restore failed devices to the system.

6. After you complete the fix, you see the statement, "Click OK to mark the error as fixed." Click **OK**. This action marks the error as fixed in the event log and prevents this instance of the error from being listed again.

7. When you see the statement, "The repair has been completed", click **Exit**. If other errors must be fixed, those errors are displayed and the fix procedures continue.

8. If no errors remain, you are shown the following statement, "There are no unfixed errors in the event log."

Command-line interface

Use the CLI to manage a system using the task commands and information commands.

10.5.3 Event reporting

Events that are detected are saved in an event log. As soon as an entry is made in this event log, the condition is analyzed. If any service activity is required, a notification is sent.

Event reporting process

The following methods are used to notify you and the IBM Support Center of a new event:

- ▶ If you enabled SNMP, an SNMP trap is sent to an SNMP manager that is configured by the client.
- ▶ If enabled, log messages can be forwarded on an IP network by using the syslog protocol.
- ▶ If enabled, event notifications can be forwarded by email by using Simple Mail Transfer Protocol (SMTP).
- ▶ *Call Home* can be enabled so that critical faults generate a problem management record (PMR) that is then sent directly to the correct IBM Support Center by using email.

Understanding events

When a significant change in status is detected, an event is logged in the event log.

Events are classified as either alerts or messages:

- ▶ An *alert* is logged when the event requires an action. Certain alerts have an associated error code that defines the service action that is required. The service actions are automated through the fix procedures. If the alert does not have an error code, the alert represents an unexpected change in the state. This situation must be investigated to see if it is expected or represents a failure. Investigate an alert and resolve it as soon as it is reported.
- ▶ A *message* is logged when a change that is expected is reported, for instance, an array build completes.

Viewing the event log

You can view the event log by using the management GUI or the CLI.

You can view the event log by clicking **Monitoring** → **Events** in the management GUI. The event log contains many entries. You can, however, select only the type of information that you need.

You can also view the event log by using the `lseventlog` CLI command.

Managing the event log

The event log has a limited size. After it is full, newer entries replace entries that are no longer required. To avoid having a repeated event that fills the event log, certain records in the event log refer to multiple occurrences of the same event. When event log entries are coalesced in this way, the time stamp of the first occurrence and the time stamp of the last occurrence of the problem are saved in the log entry. A count of the number of times that the error condition occurred is also saved in the log entry. Other data refers to the last occurrence of the event.

Describing the fields in the event log

The event log includes fields with information that you can use to diagnose problems.

Table 10-3 on page 336 describes several of the fields that are available to assist you in diagnosing problems.

Table 10-3 Description of data fields for the event log

Data field	Description
Event ID	This number precisely identifies why the event was logged.
Error code	This number describes the service action that needs to be followed to resolve an error condition. Not all events have error codes that are associated with them. Many event IDs can have the same error code because the service action is the same for all the events.
Sequence number	A number that identifies the event.
Event count	The number of events coalesced into this event log record.
Object type	The object type to which the event log relates.
Object ID	The object ID to which the event log relates.
Fixed	When an alert is shown for an error condition, it indicates whether the reason for the event was resolved. In many cases, the system automatically marks the events fixed when appropriate. Certain events must be manually marked as fixed. If the event is a message, this field indicates that you have read and performed the action. The message must be marked as read.
First time	The time when this error event was reported. If events of a similar type are coalesced together so that one event log record represents more than one event, this field is the time that the first error event was logged.
Last time	The time when the last instance of this error event was recorded in the log.
Root sequence number	If set, this number is the sequence number of an event that represents an error that probably caused this event to be reported. Resolve the root event first.
Sense data	Additional data that gives the details of the condition that caused the event to be logged.

Event notifications

Your system can use SNMP traps, syslog messages, emails, and Call Home notifications to notify you and IBM Remote Technical Support when significant events are detected. Any combination of these notification methods can be used simultaneously. Notifications are normally sent immediately after an event is raised. However, there are certain events that might occur because of service actions that are being performed. If a recommended service action is active, these events are notified only if they are still unfixed when the service action completes.

Only events recorded in the event log can be notified. Most CLI messages in response to certain CLI commands are not recorded in the event log so they do not cause an event notification.

Table 10-4 on page 337 describes the levels of event notifications.

Table 10-4 Notification levels

Notification level	Description
Error	<p>Error notification is sent to indicate a problem that must be corrected as soon as possible.</p> <p>This notification indicates a serious problem with the system. For example, the event that is being reported might indicate a loss of redundancy in the system, and it is possible that another failure might result in loss of access to data. The typical reason for sending this type of notification is a hardware failure, but certain configuration errors or fabric errors also are included in this notification level. Error notifications can be configured to be sent as a Call Home to IBM Remote Technical Support.</p>
Warning	<p>A warning notification is sent to indicate a problem or unexpected condition with the system. Always immediately investigate this type of notification to determine the effect that it might have on your operation, and make any necessary corrections. A warning notification does not require any replacement parts and therefore does not require IBM Support Center involvement. The allocation of notification type Warning does not imply that the event is less serious than an event that has notification level Error.</p>
Information	<p>An informational notification is sent to indicate that an expected event has occurred.</p> <p>No remedial action is required when these notifications are sent.</p> <p>Informational events provide information about the status of an operation. Information events are recorded in the error event log and, depending on the configuration, can be notified through email, SNMP, and syslog.</p>

Power-on self-test

A series of tests is performed to check the operation of the components and several of the options that are installed when the system is first turned on. This series of tests is called the *power-on self-test* (POST).

When the code is loaded, additional testing occurs, which ensures that all of the required hardware and code components are installed and functioning correctly.

Understanding the error codes

Error codes are generated by the event-log analysis and system configuration code.

Error codes help you to identify the cause of a problem, a failing component, and the service actions that might be needed to solve the problem.

Viewing logs and traces

The system maintains log files and trace files that can be used to manage your system and diagnose problems.

10.5.4 Resolving a problem

The management GUI provides extensive facilities to help you troubleshoot and correct problems on your system.

To run the management GUI, start a supported web browser and point it to the management IP address of your system.

After the connection is successful, you see a login panel. Log on using your user name and password. After you log on, select **Monitoring** → **Events** to view the system event log, and then, select the **Recommended Actions** filter.

An event in the log can be an informational message, or it can alert you to an error that requires fixing. Errors are prioritized by their error code, and each has a fix procedure that can be run.

A *fix procedure* is a wizard that helps you troubleshoot and correct the cause of an error. Certain fix procedures will reconfigure the system, based on your responses; ensure that actions are carried out in the correct sequence; and, prevent or mitigate the loss of data. For this reason, you must always run the fix procedure to fix an error, even if the fix might seem obvious.

To run the fix procedure for the error with the highest priority, click **Recommended Action** at the top of the Event page and click **Run This Fix Procedure**. When you fix higher priority events first, the system can often automatically mark lower priority events as fixed.

While the Recommended Actions filter is active, the event list shows only alerts for errors that have not been fixed, sorted in order of priority. The first event in this list is the same as the event displayed in the Recommended Action panel at the top of the Event page of the management GUI.

If it is necessary to fix errors in a different order, select an error alert in the event log and then click **Action** → **Run Fix Procedure**.

10.6 IBM System Storage Interoperation Center (SSIC)

IBM continuously tests and approves the interoperability of IBM products in different environments. You can search the interoperability results at the IBM System Storage Interoperation Center (SSIC).

You have to check the IBM SSIC to get the latest information about supported operating systems, hosts, switches, and so on:

<http://www.ibm.com/systems/support/storage/ssic/interoperability.wss>

If a configuration that you want is not available at the SSIC, you must submit a Solution for Compliance in a Regulated Environment (SCORE)/request for price quotation (RPQ) to IBM requesting approval. To submit a SCORE/RPQ, contact your IBM representative.



A

SAN preferred practices for 16 Gbps

This appendix provides a high-level design for your storage area network (SAN) environment and provides preferred practices guidance based on IBM 16 Gbps b-type products and features, focusing on Fibre Channel (FC) SAN design.

There are many benefits of this faster technology. Data transfers are faster, fewer links are needed to accomplish the same task, and fewer devices need to be managed. Also, less power is consumed when 16 Gbps FC is used instead of 8 Gbps FC or 4 Gbps FC. Several technological advances are increasing SAN bandwidth demands, including the IBM FlashSystem 840. Sixteen Gbps FC is keeping pace with other technological advances in the data center.

Topics in this appendix include the early planning phase, topologies, understanding possible operational challenges, monitoring, and improving a SAN infrastructure that is already implemented. The guidelines in this document do not apply to every environment but they will help guide you through the decisions that you need to make for a successful SAN design.

Sixteen Gbps Fibre Channel benefits

The latest speed developed by the T11 technical committee that defines FC interfaces is 16 Gbps FC. The FC industry is doubling the data throughput of 8 Gbps links from 800 megabytes per second (MBps) to 1600 MBps with 16 Gbps FC. Sixteen Gbps FC is the latest evolutionary step in SANs where large amounts of data are exchanged and high performance is a necessity. From host bus adapters (HBAs) to switches, 16 Gbps FC will enable higher performance with lower power consumption per bit, which is the performance required by today's leading applications.

Data transfers are faster; fewer links are needed to accomplish the same task; and fewer devices need to be managed. Also, less power is consumed when 16 Gbps FC is used instead of 8 Gbps FC or 4 Gbps FC. When high bandwidth is needed, deploy 16 Gbps FC. The first place that new, faster speeds are usually needed in SANs is in inter-switch links (ISLs) in the core of the network and between data centers. However, you can take advantage of the IBM FlashSystem 840 with a full 16 Gbps path, from the server, through the SAN and connected to the FlashSystem 840. When large blocks of data need to be transferred between arrays or sites, a faster link can accomplish the same job in less time. The 16 Gbps FC is designed to assist users in transferring large amounts of data and decreasing the number of links in the data center.

The 16 Gbps FC multimode links were designed to meet the distance requirements of most data centers. Table A-1 shows the supported link distances over multimode and single-mode fiber. The 16 Gbps FC was optimized for Optical Mode 3 (OM3) fiber and supports 100 meters (328 feet). With the standardization of Optical Mode 4 (OM4) fiber, FC has standardized the supported link distances over OM4 fiber, and 16 Gbps FC can support 125 meters (410.1 feet). If a 16 Gbps FC link needs to go farther than these distances, a single-mode link can be used that supports distances up to 10 kilometers (6.2 miles). This wide range of supported link distances enables 16 Gbps FC to work in a wide range of environments.

Table A-1 Link distance with speed and fiber type in a comparison between 8 Gbps FC versus 16 Gbps FC

Speed	OM1 link distance 62.5 um core 200 MHZ*km	OM2 link distance 50 um core 500 MHZ*km	OM3 link distance 50 um core 2000 MHZ*km	OM4 link distance 50 um core 4700 MHZ*km	OS1 link distance 9 um core ~infinite MHZ*km
8 Gbps FC	21	50	150	190	10,000
16 Gbps FC	15	35	100	125	10,000

The benefits of higher speed

The benefits of faster tools result in more work in less time. By doubling the speed, 16 Gbps FC reduces the time to transfer data between two ports. When more work can be done by a server or storage device, fewer servers, HBAs, links, and switches are needed to accomplish the same task.

Although many applications do not use the full extent of a 16 Gbps FC link yet, over the next few years, traffic and applications will grow to fill the capacity of 16 Gbps FC. The refresh cycle for networks is often longer than that of servers and storage, so 16 Gbps FC will remain in the network for years. With more virtual machines added to a physical server, performance levels can quickly escalate beyond the levels supported by 8 Gbps FC. To help position your deployments for the future, the 16 Gbps FC can be the most efficient way to transfer large amounts of data in data centers. The 16 Gbps FC will be the best performer in several applications. The 16 Gbps FC can reduce the number of ISLs in the data center or migrate a large amount of data for array migration or disaster recovery.

High performance applications, such as the virtual desktop infrastructure (VDI), solid-state drives (SSDs), and the IBM FlashSystem 840, that require high bandwidth are ideal applications for 16 Gbps FC.

As more applications demand the low-latency performance of the IBM FlashSystem, 16 Gbps FC keeps up with other advances in other components of the storage infrastructure, such as SSDs. The 16 Gbps FC combines the latest technologies in an energy-efficient manner to provide extremely high performance SANs.

Cabling considerations

Although cabling represents less than 10% of the overall data center network investment, expect it to outlive most other network components and expect it to be the most difficult and potentially costly component to replace. When purchasing the cabling infrastructure, consider not only the initial implementation costs, but subsequent costs as well. Understand the full lifecycle and study local industry trends to arrive at the best decision for your environment.

Choose the strongest foundation to support your present and future network technology needs that comply with the Telecommunications Industry Association (TIA)/International Organization for Standardization (ISO) cabling standards. The cabling itself calls for the right knowledge, the right tools, patience, a structured approach, and most of all, discipline. Without discipline, it is common to see complex cabling “masterpieces” quickly get out of control, leading to chaos.

IBM System Storage b-type Gen 5 SAN product overview

We introduce the IBM System Storage b-type Gen 5 SAN technology and the features provided by the Fabric Operating System (FOS). For the most up-to-date information, see the following website:

<http://www.ibm.com/systems/networking/switches/san/index.html>

Hardware features

The b-type Gen 5 FC directors and switches provide reliable, scalable, high-performance foundations for mission-critical storage, for example, the IBM FlashSystem 840, thanks to the new 16 Gbps FC technology. The portfolio starts with entry level 12-port fabric switches all the way up to 3456 sixteen Gbps ports (or 4608 eight Gbps ports) when connecting nine backbone chassis in a full mesh topology via UltraScale intercluster links (ICLs). These SAN platforms support 2, 4, 8, and 16 Gbps auto-sensing ports and deliver enhanced fabric resiliency and application uptime through advanced features.

The Condor3 ASIC (application-specific integrated circuit) enables the support for native 10 Gbps FC, in-flight encryption and compression, ClearLink diagnostic technology (supported only on the 16 Gbps ports), increased buffers, and Forward Error Correction (FEC). This new Gen 5 family allows a simple server deployment with dynamic fabric provisioning. *Dynamic fabric provisioning* enables organizations to eliminate fabric reconfiguration when adding or replacing servers through the virtualization of the host worldwide names (WWNs).

Fabric Vision technology

The new Brocade Fabric Vision technology is an advanced hardware and software architecture that combines capabilities from the FOS, b-type Gen 5 devices, and IBM Network Advisor. The Brocade Fabric Vision technology helps administrators address problems before they affect operations, accelerate new application deployments, and dramatically reduce operational costs.

Brocade Fabric Vision technology includes these features:

- ▶ Brocade ClearLink diagnostics: Ensures optical and signal integrity for Gen 5 FC optics and cables, simplifying deployment and support of high-performance fabrics. It uses the ClearLink Diagnostic Port (D_Port) capabilities of Gen 5 FC platforms.
- ▶ Bottleneck Detection: Identifies and alerts administrators to device or ISL congestion and abnormal levels of latency in the fabric. This feature works with Brocade Network Advisor to automatically monitor and detect network congestion and latency in the fabric, providing visualization of bottlenecks in a connectivity map and product tree, and identifying exactly which devices and hosts are affected by a bottlenecked port.
- ▶ Integration into Brocade Network Advisor: Provides customizable health and performance dashboard views to pinpoint problems faster, simplify SAN configuration and management, and reduce operational costs.
- ▶ Critical diagnostic and monitoring capabilities: Help ensure early problem detection and recovery.
- ▶ Non-intrusive and nondisruptive monitoring on every port: Provides a comprehensive end-to-end view of the entire fabric.
- ▶ Forward Error Correction (FEC): Enables recovery from bit errors in ISLs, enhancing transmission reliability and performance.
- ▶ Additional buffers: Help overcome performance degradation and congestion due to buffer credit loss.
- ▶ Real-time bandwidth consumption by hosts and applications on ISLs: Helps easily identify hot spots and potential network congestion.

Note: In the upcoming FOS 7.2, the Fabric Vision Technology will be enhanced with the following new capabilities: Brocade Monitoring and Alerting Policy Suite (MAPS) and Brocade Flow Vision.

Fabric OS features

The FOS 7 and the b-type Gen 5 platforms offer a set of advanced features. Not all of these features are available for all switch models. Some of these features are offered as optional licenses. The following list introduces the most important features with a brief explanation:

- ▶ Advanced Web Tools enable graphical user interface (GUI)-based administration, configuration, and maintenance of fabric switches and SANs.
- ▶ Advanced Zoning segments a fabric into virtual private SANs to restrict device communication and apply certain policies only to members within the same zone.

- ▶ Virtual Fabrics allow a physical switch to be partitioned into independently managed logical switches, each with its own data, control, and management paths.
- ▶ Full Fabric allows a switch to be connected to another switch. It is required to enable expansion ports (E_ports).
- ▶ Adaptive Networking Service is a set of features that provide users with tools and capabilities for incorporating network policies to ensure optimal behavior in a large SAN. FOS v7.0 supports two types of quality of service (QoS) features with the 16 Gbps fabric backbones: ingress rate limiting and source ID(SID)/destination identifier (DID)-based prioritization.
- ▶ Enhanced Group Management (EGM) enables additional device-level management functionality for IBM b-type SAN products when added to the element management. It also allows large consolidated operations, such as firmware downloads and configuration uploads and downloads for groups of devices.
- ▶ Extended Fabrics extend SAN fabrics beyond the FC standard of 10 km (6.2 miles) by optimizing internal switch buffers to maintain performance on ISLs connected at extended distances.
- ▶ Integrated Routing allows any 16 Gbps FC port to be configured as an E_port supporting FC routing.
- ▶ Integrated 10 Gbps Fibre Channel Activation enables FC ports to operate at 10 Gbps.
- ▶ Fabric Watch constantly monitors mission-critical switch operations for potential faults and automatically alerts administrators to problems before they become costly failures. Fabric Watch includes port fencing capabilities.
- ▶ Intercluster link (ICL) license with 16× (4×16 Gbps) Quad Small Form Factor Pluggable (QSFP) provides connectivity up to 100 meters (328 feet) from the switching backplane of one half of an 8-slot chassis to the other half, or to a 4-slot chassis.
- ▶ Enterprise intercluster link (ICL) license supports up to 3,840×16 Gbps universal FC ports (using 16 Gbps 48-port blades); up to 5,120×8 Gbps universal FC ports (using 8 Gbps 64-port blades); or ICL ports (32 or 16 per chassis, optical QSFP) connected to up to nine chassis in a full-mesh topology or up to 10 chassis in a core-edge topology. Connecting five or more chassis via ICLs requires an Enterprise ICL license.
- ▶ Encryption 96 Gbps disk performance upgrade activation license enables scalability of performance on the encryption blade features. The upgrade is designed to provide increased throughput for disk encryption applications up to 96 Gbps, effectively doubling encrypted throughput performance for disk-based storage with no disruption to operations.
- ▶ Advanced Performance Monitoring helps identify end-to-end bandwidth usage by host/target pairs and is designed to provide for capacity planning.
- ▶ ISL Trunking enables FC packets to be distributed efficiently across multiple ISLs between two IBM b-type SAN fabric switches and directors while preserving in-order delivery. Both b-type SAN devices must have trunking activated.
- ▶ Monitoring and Alerting Policy Suite (MAPS) is an optional SAN health monitor that allows you to enable each switch to constantly monitor its SAN fabric for potential faults and automatically alert you to problems long before they become costly failures. MAPS cannot coexist with Fabric Watch.
- ▶ Flow Vision is a comprehensive tool that enables administrators to identify, monitor, and analyze specific application data flows to maximize performance, avoid congestion, and optimize resources.

Management

IBM b-type Gen 5 FC directors and switches can be managed through several ways:

- ▶ IBM Network Advisor is a software management platform that unifies network management for SAN and converged networks. It is designed to provide users with a consistent user interface, proactive performance analysis, and troubleshooting capabilities across FC and b-type Fibre Channel over Ethernet (FCoE) installations.
- ▶ Web Tools is a built-in web-based application that provides administration and management functions on a per switch basis.
- ▶ Command-line interface (CLI) enables an administrator to monitor and manage individual switches, ports, and entire fabrics from a standard workstation. It is accessed through Telnet, Secure Shell (SSH), or serial console.
- ▶ Structure of Management Information (SMI) Agent enables integration with SMI-compliant Storage Resource Management (SRM) solutions, such as IBM Tivoli Storage Productivity Center. The SMI Agent is embedded in the IBM Network Advisor.

Monitoring

There are several monitoring tools and notification methods that allow you to monitor your entire b-type Gen 5 fabric and even integrate with external applications.

Health monitors

Fabric Watch and MAPS are monitors that allow you to enable each switch to constantly monitor its SAN fabric for potential faults and automatically alert you to problems long before they become costly failures. MAPS will only be available in FOS 7.2.0 or later.

Performance monitors

Advanced Performance Monitoring and Flow Vision are performance monitors that integrate with IBM Network Advisor. Flow Vision will only be available in FOS 7.2.0 or later.

Notification methods

There are several alert mechanisms that can be used, such as email messages, Simple Network Management Protocol (SNMP) traps, and log entries. Fabric Watch allows you to configure multiple email recipients.

An email alert sends information about a switch event to one or many specified email addresses.

The SNMP notification method is an efficient way to avoid having to log in to each switch as you do for error log notifications.

The RASLog (switch event log) can be forwarded to a central station. IBM Network Advisor can be configured as the syslog recipient for the SAN devices.

IBM Network Advisor

IBM Network Advisor is the preferred tool for managing and monitoring the IBM b-type Gen 5 SANs. It is a software management tool that provides comprehensive management for data, storage, and converged networks.

It includes an intuitive interface and provides an in-depth view of performance measures and historical data. It receives SNMP traps, syslog event messages, and customizable event alerts. It contains the Advanced Call Home feature that can automatically collect diagnostic information and send notifications to IBM Support for faster fault diagnosis and isolation.

Product description

We provide a brief description of each IBM b-type Gen 5 SAN switch or backbone.

IBM System Networking SAN24B-5

The SAN24B-5 is an entry level SAN switch that combines flexibility, simplicity, and enterprise-class functionality. It is a 1U form factor unit configurable in 12 or 24 ports and it supports 2, 4, 8, or 16 Gbps speeds. It can be deployed as a full-fabric switch or as an N-Port ID Virtualization (NPIV)-enabled access gateway, enabling the creation of dense fabrics in a relatively small space. It includes one or two power supplies based on the model (2498-24G/2498-X24 or 2498-F24 respectively).

Figure A-1 shows the IBM System Networking SAN24B-5 fabric switch.

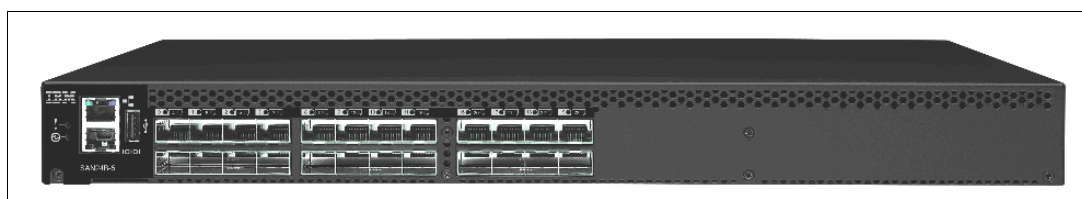


Figure A-1 SAN24B-5 switch

IBM System Networking SAN48B-5

The SAN48B-5 is a flexible, easy-to-use enterprise-class SAN switch for private cloud storage. It is a 1U form factor unit configurable in 24, 36, or 48 ports and it supports auto-sensing 2, 4, 8, or 16 Gbps, as well as 10 Gbps speeds. It can be deployed as a full-fabric switch or as an NPIV-enabled access gateway. It is also enhanced with enterprise connectivity that adds support for IBM Fibre Channel connection (FICON®). It includes dual, hot-swappable redundant power supplies with integrated system cooling fans.

Figure A-2 shows the IBM System Networking SAN48B-5 fabric switch.

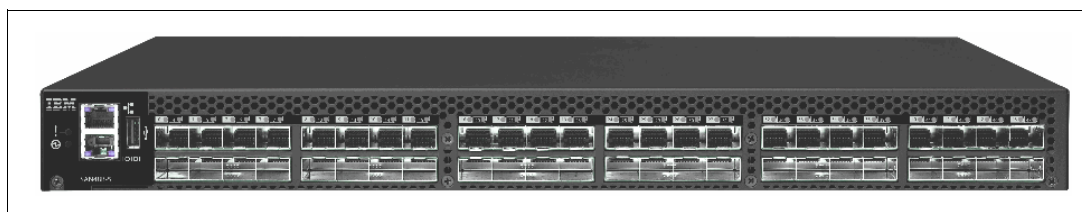


Figure A-2 SAN48B-5 switch

IBM System Networking SAN96B-5

The SAN96B-5 is a scalable enterprise-class SAN switch for highly virtualized, cloud environments. It is a 2U form factor unit configurable in 48, 72, or 96 ports and it supports auto-sensing 2, 4, 8, or 16 Gbps, as well as 10 Gbps speeds. This switch also features dual-direction airflow options to support the latest hot aisle/cold aisle configurations (2498-F96 and 2498-N96). It does not support access gateway functionality or IBM FICON connectivity.

Figure A-3 shows the IBM System Networking SAN96B-5 fabric switch.

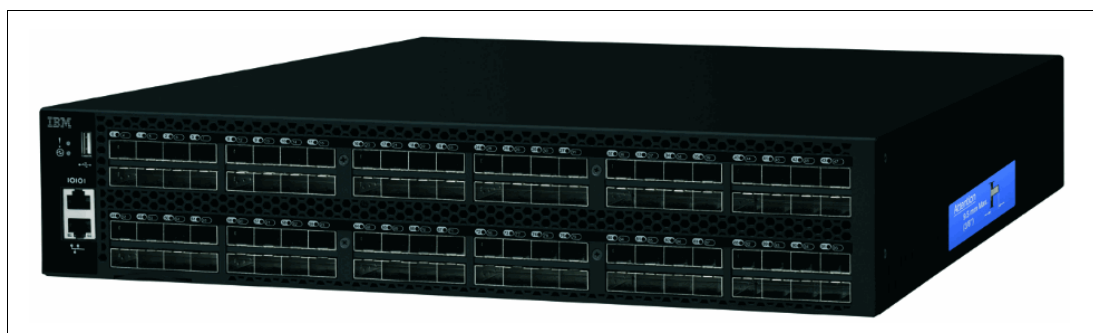


Figure A-3 SAN96B-5 switch

IBM System Networking SAN384B-2 and IBM System Networking SAN768B-2

The SAN384B-2 and SAN768B-2 backbones deliver a new level of scalability and advanced capabilities to this robust, reliable, and high-performance technology. This enables organizations to continue using their existing IT investments as they grow their businesses. In addition, they can consolidate their SAN infrastructures to simplify management and reduce operating costs. The UltraScale ICL technology available in these backbones includes new optical ports, higher port density, and support for standard optical cables up to 100 meters (328 feet). The UltraScale ICLs can connect up to 10 Brocade DCX 8510 Backbones, enabling flatter, faster, and simpler fabrics that increase consolidation while reducing network complexity and costs.

The SAN384B-2 is a 8U form factor unit designed for midsize networks. It has four horizontal blade slots to provide up to 192 sixteen Gbps FC ports.

Figure A-4 shows the IBM System Networking SAN384B-2 Backbone.



Figure A-4 SAN384B-2 Backbone

The SAN768B-2 is a 14U form factor unit designed for large enterprise networks. It has eight vertical blade slots to provide up to 384 sixteen Gbps FC ports.

Figure A-5 shows the IBM System Networking SAN768B-2 Backbone.

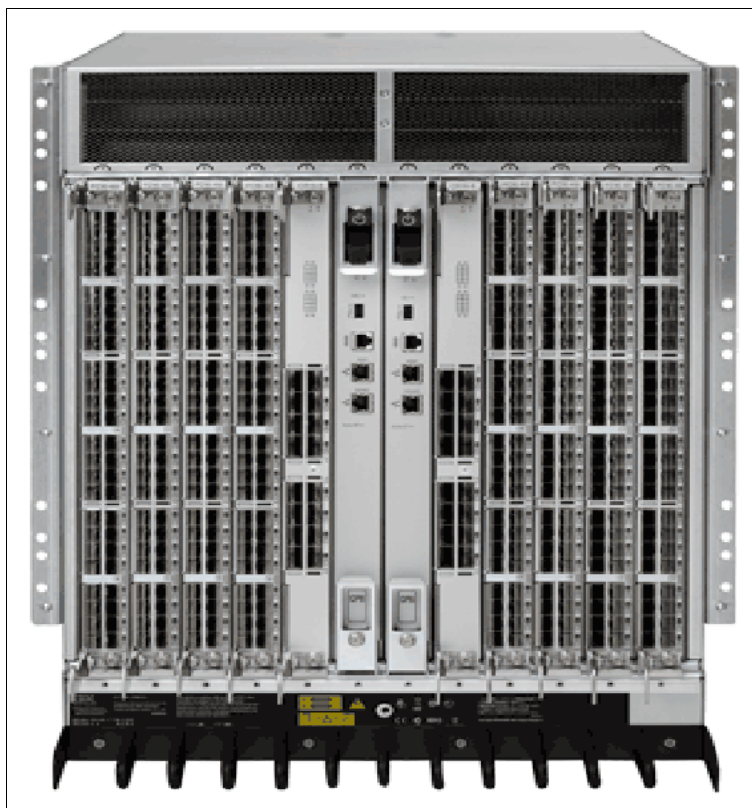


Figure A-5 SAN768B-2 Backbone

SAN design basics

The preferred practices for core-edge or edge-core-edge fabrics are described. The concepts of topology, inter-switch links, inter-chassis links, device placement, and oversubscription are described.

Topologies

A typical SAN design consists of devices on the edge of the network, switches in the core of the network, and the cabling that connects it all together. *Topology* is usually described in terms of how the switches are interconnected, such as ring, core-edge, and edge-core-edge or fully meshed. The suggested SAN topology to optimize performance, management, and scalability is a tiered, core-edge topology (sometimes called *core-edge* or *tiered core-edge*). This approach provides good performance without unnecessary interconnections. At a high level, the tiered topology has many edge switches that are used for device connectivity, and a smaller number of core switches used for routing traffic between the edge switches.

An important aspect of SAN topology is the resiliency and redundancy of the fabric. The main objective is to remove any single point of failure. Resiliency is the ability of the network to continue to function and recover from a failure. Redundancy describes the duplication of components, or even an entire fabric, to eliminate a single point of failure in the network. IBM b-type fabrics have resiliency built into Fabric OS, the software that runs on all B-Series switches, which can quickly “repair” the network to overcome most failures. For example,

when a link between switches fails, Fibre Channel shortest path first (FSPF) quickly recalculates all traffic flows. Of course, this assumes that there is a second route, which is when redundancy in the fabric becomes important.

The key to high availability and enterprise-class installation is redundancy. By eliminating a single point of failure, business continuance can be provided through most foreseeable and even unforeseeable events. At the highest level of fabric design, the complete network needs to be redundant, with two completely separate fabrics that do not share any network equipment (routers or switches).

Inter-switch link

An *inter-switch link* (ISL) is a link between two switches (referred to as *E_ports*). ISLs carry frames originating from the node ports, and those generated within the fabric. The frames generated within the fabric serve as control, management, and support for the fabric.

To maximize the performance on your ISLs, we suggest the implementation of *trunking*. This technology is ideal for optimizing performance and simplifying the management of multi-switch SAN fabrics. When two or more adjacent ISLs in a port group are used to connect two switches with trunking enabled, the switches automatically group the ISLs into a single logical ISL, or *trunk*.

ISL Trunking is designed to significantly reduce traffic congestion in storage networks. To balance workload across all of the ISLs in the trunk, each incoming frame is sent across the first available physical ISL in the trunk. As a result, transient workload peaks are much less likely to affect the performance of other parts of the SAN fabric and bandwidth is not wasted by inefficient traffic routing. ISL Trunking can also help simplify fabric design, lower provisioning time, and limit the need for additional ISLs or switches.

To further optimize network performance, b-type switches and directors support *Exchange-based routing*, also known as *Dynamic Path Selection* (DPS). Available as a standard feature in Fabric OS (starting in Fabric OS 4.4), DPS optimizes fabric-wide performance by automatically routing data to the most efficient available path in the fabric. DPS augments ISL Trunking to provide more effective load balancing in certain configurations, such as routing data between multiple trunk groups.

The preferred practice is using a minimum of two trunks, with at least two ISLs per trunk. DPS needs to be used to provide more effective load balancing, such as routing between multiple trunk groups.

Intercluster links

Intercluster links (ICLs) are high-performance ports for interconnecting multiple backbones, enabling industry-leading scalability while preserving ports for server and storage connections. Now in its second generation, the new optical UltraScale ICLs, based on Quad Small Form Factor Pluggable (QSFP) technology, connect the core routing blades of two Backbone chassis. Each QSFP-based ICL port combines four 16 Gbps links, providing up to 64 Gbps of throughput within a single cable. Available with Fabric OS (FOS) version 7.0 and later, it offers up to 32 QSFP ICL ports on the SAN768B-2 and up to 16 QSFP ICL ports on the SAN384B-2. The optical form factor of the QSFP-based ICL technology offers several advantages over the copper-based ICL design in the original platforms. First, the second generation has increased the supported ICL cable distance from 2 meters (6.5 feet) to 50 meters (164 feet) (or 100 meters (328 feet) with FOS v7.1, select QSFPs, and OM4 fiber), providing greater architectural design flexibility.

Second, the combination of four cables into a single QSFP provides incredible flexibility for deploying various different topologies, including a massive 9-chassis full-mesh design with only a single hop between any two points within the fabric. In addition to these significant advances in ICL technology, the ICL capability still provides a dramatic reduction in the number of inter-switch link (ISL) cables required, a four to one reduction compared to traditional ISLs with the same amount of interconnect bandwidth. And, because the QSFP-based ICL connections reside on the core routing blades instead of consuming traditional ports on the port blades, up to 33% more FC ports are available for server and storage connectivity.

ICL Ports on Demand are licensed in increments of 16 ICL ports. Connecting five or more chassis via ICLs requires an Enterprise ICL license.

Device placement

Device placement is a balance of traffic isolation, scalability, manageability, and serviceability. With the growth of virtualization, frame congestion can become a serious concern in the fabric if there are interoperability issues with the end devices.

Designing device connectivity greatly depends on the expected data flow between devices. For simplicity, communicating hosts and targets can be attached to the same switch.

However, this approach does not scale well. Given the high-speed, low-latency nature of FC, attaching these host-target pairs on different switches does not mean that performance is adversely affected. Although traffic congestion is possible (see Figure A-6), it can be mitigated with the correct provisioning of ISLs/ICLs. With current generation switches, locality is not required for performance or to reduce latencies. For mission-critical applications, architects might want to localize the traffic in exceptional cases, particularly if the number of ISLs available is restricted or if there is a concern for resiliency in a multi-hop environment.

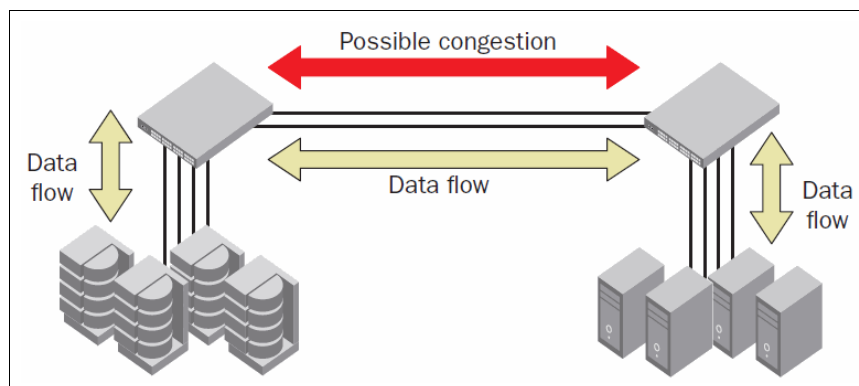


Figure A-6 Hosts and targets attached to different switches for ease of management and expansion

One common scheme for scaling a core-edge topology is dividing the edge switches into a storage tier and a host/initiator tier. This approach lends itself to ease of management and ease of expansion. In addition, host and storage devices generally have different performance requirements, cost structures, and other factors that can be readily accommodated by placing initiators and targets in different tiers.

Consider the following suggestions for device placement:

- ▶ The preferred practice fabric topology is core-edge or edge-core-edge with tiered device connectivity, or full-mesh if the port count is fewer than 1,500 ports.
- ▶ Minimize the use of localized traffic patterns and, if possible, keep servers and storage connected to separate switches.
- ▶ Select the correct optics (short wavelength (SWL)/long wavelength (LWL)/extended long wavelength (ELWL)) to support the distance between switches and devices and switches.

Disk and tape traffic isolation

A dedicated physical or virtual fabric for tape traffic is the ideal solution to get the tape and disk data flows separated. However, the main goal is designing the SAN to avoid tape and disk traffic sharing the same ISLs. Therefore, when sharing disk and tape traffic on the same fabric, tape traffic needs to be kept locally. If crossing ISLs, traffic isolation zones (TIZs) need to be implemented to ensure that disk and tape data flows do not cross the same ISLs.

Fan-in ratios and oversubscription

Another aspect of data flow is *fan-in ratio* (also called the *oversubscription ratio*, and frequently the *fan-out ratio* if viewed from the storage device perspective), both in terms of host ports to target ports and device to ISL. The *fan-in ratio* is the number of device ports that need to share a single port, whether target port or ISL/ICL.

What is the optimum number of hosts that need to connect to a storage port? This seems like a fairly simple question. However, after you consider the clustered hosts, virtual machines (VMs), and number of logical unit numbers (LUNs) storage per server, the situation can quickly become much more complex. Determining how many hosts to connect to a particular storage port can be narrowed down to three considerations: port queue depth, I/O per second (IOPS), and throughput. Of these three, throughput is the only network component. Therefore, a simple calculation is to add up the expected bandwidth usage for each host accessing the storage port.

In practice, it is highly unlikely that all hosts perform at their maximum level at any one time. With the traditional application-per-server deployment, the host bus adapter (HBA) bandwidth is over-provisioned. However, with virtual servers (kernel-based virtual machine (KVM), Xen, Hyper-V, proprietary UNIX OSs, and VMware), the topic can change radically. Network oversubscription is built into the virtual server concept. To the extent that servers use virtualization technologies, reduce network-based oversubscription proportionally. It might therefore be prudent to oversubscribe ports to ensure a balance between cost and performance. An example of 3 to 1 oversubscription is shown in Figure A-7.

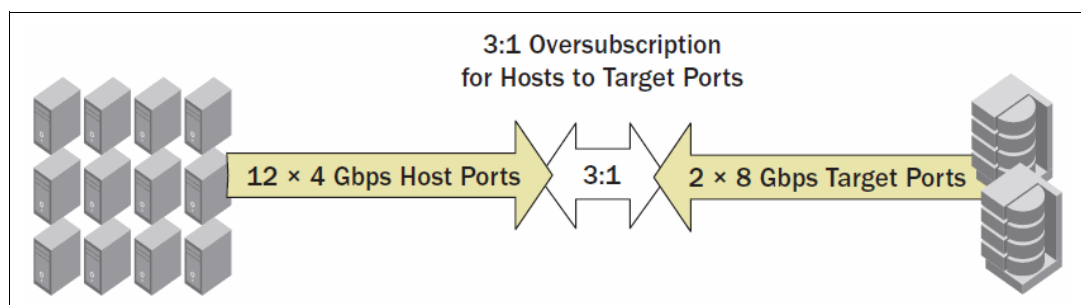


Figure A-7 Example of three-to-one oversubscription

Typical and safe oversubscription between the host and the edge is about 12:1. This is a good starting point for many deployments, but it might need to be tuned based on the requirements. If you are in a highly virtualized environment with higher speeds and feeds, you might need to go lower than 12:1.

Consider the following suggestions for avoiding frame congestion (when the number of frames is the issue rather than bandwidth utilization):

- ▶ Use more and smaller trunks.
- ▶ Bandwidth through the core (path from source/host to destination/target) needs to exceed storage requirements.
- ▶ Host-to-core subscription ratios need to be based on both the application needs and the importance of the application.
- ▶ Plan for peaks, not average usage.

For mission-critical applications, the ratio needs to exceed peak load enough so that path failures do not adversely affect the application. Have enough extra bandwidth to avoid congestion if a link fails.

Note: When the performance expectations are demanding, we suggest a 3:1 oversubscription ratio. However, 7:1 and even 16:1 are common.

FCoE as a top of rack (ToR) solution

The consolidation of server network adapters, cables, and intermediate switches provides much of the motivation for FCoE. The reduction in equipment, power, and maintenance costs is anticipated to be significant over time. With the 10 GbE widely used, FCoE technology might be a good choice for new environments with a high density of servers. It requires a Data Center Bridging (DCB) network to take advantage of new features, such as Ethernet and Priority-based flow control (PFC), Enhanced Transmission Selection (ETS), and Data Center Bridging Exchange (DCBX) Protocol. Enhanced Transmission Selection (ETS) is a transmission selection algorithm (TSA) that is specified by the IEEE 802.1Qaz draft standard. This standard is part of the framework for the IEEE 802.1 Data Center Bridging (DCB) interface.

In an FCoE edge topology, the top of rack (ToR) switch handles all the LAN and SAN traffic within a rack and forwards traffic to separate existing LAN and SAN infrastructures elsewhere in the data center. In general, vendors might allow their FC forwarders to operate in fabric mode or in NPIV mode. NPIV simplifies the implementation and avoids certain incompatibility and management issues. For IBM BladeCenter chassis, IBM provides a set of switch modules and converged network adapters that easily allow the convergence of the LAN and SAN traffic.

Innovation and market trends also play an important role when deciding to choose FCoE for new SANs. For additional information, consult *Storage and Network Convergence Using FCoE and iSCSI*, SG24-7986:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg247986.pdf>

Data flow considerations

An important consideration when designing your SAN is the understanding of data flow across the devices. Data flow might cause unwanted problems. We describe several situations and what you can do to mitigate them.

Congestion in the fabric

Congestion is a major source of poor performance in a fabric. Sufficiently impeded traffic translates directly into poor application performance.

There are two major types of congestion: traffic-based and frame-based. *Traffic-based congestion* occurs when link throughput capacity is reached or exceeded and the link is no longer able to pass more frames. *Frame-based congestion* occurs when a link has run out of buffer credits and is waiting for buffers to free up to continue transmitting frames.

Traffic versus frame congestion

After link speeds reach 4 Gbps and beyond, the emphasis on fabric and application performance shifts from traffic-level issues to frame congestion. It is difficult with current link speeds and features, such as ISL Trunking or ICLs, to consistently saturate a link. Most infrastructures today rarely see even two-member trunks reaching a sustained 100% utilization. Frame congestion can occur when the buffers available on an FC port are not sufficient to support the number of frames that the connected devices try to transmit. This situation can result in credit starvation backing up across the fabric. This condition is called *back pressure*, and it can cause severe performance problems.

One side effect of frame congestion can be large buffer credit zero counts on ISLs and fabric ports (F_ports). This is not necessarily a concern, unless counts increase rapidly in an extremely short time. There is a new feature, *Bottleneck Detection*, to more accurately assess the impact of a lack of buffer credits.

The sources and mitigation for traffic are well known and are described at length in other parts of this document. The remainder of this section focuses on the sources and mitigation of frame-based congestion.

Sources of congestion

Frame congestion is primarily caused by latencies somewhere in the SAN, usually storage devices that do not have the low-latency performance of the IBM FlashSystem, and occasionally hosts. These latencies cause frames to be held in application-specific integrated circuits (ASICs) and reduce the number of buffer credits available to all flows traversing that ASIC. The congestion backs up from the source of the latency to the other side of the connection and starts clogging up the fabric, which is called back pressure. Back pressure can be created from the original source of the latency to the other side and all the way back (through other possible paths across the fabric, to the original source again). After this situation arises, the fabric is extremely vulnerable to severe performance problems.

The following situations are potential sources of high latencies:

- ▶ Storage devices that are not optimized or where performance deteriorated over time
- ▶ Distance links where the number of allocated buffers is miscalculated or where the average frame sizes of the flows traversing the links changed over time
- ▶ Hosts where the application performance deteriorated to the point that the host can no longer respond to incoming frames in a sufficiently timely manner

Other contributors to frame congestion include behaviors where short frames are generated in large numbers:

- ▶ Clustering software that verifies the integrity of attached storage
- ▶ Clustering software that uses control techniques, such as SCSI RESERVE/RELEASE, to serialize access to shared file systems
- ▶ Host-based mirroring software that routinely sends SCSI control frames for mirror integrity checks
- ▶ Virtualizing environments, both workload and storage, that use in-band FC for other control purposes

Mitigating congestion with edge hold time

Frame congestion cannot be corrected in the fabric. Devices exhibiting high latencies, whether servers or storage arrays, must be examined and the source of poor performance eliminated. Because these are the major sources of frame congestion, eliminating them typically addresses the vast majority of cases of frame congestion in fabrics.

Edge Hold Time (EHT) is a FOS capability that allows an overriding value for Hold Time (HT). *Hold Time* is the amount of time that a Class 3 frame can remain in a queue before it is dropped while waiting for credit to be given for transmission.

The default HT is calculated from the RA_TOV, ED_TOV, and maximum hop count values configured on a switch. When using the standard 10 seconds for RA_TOV, 2 seconds for ED_TOV, and a maximum hop count of 7, a Hold Time value of 500 ms is calculated. Extensive field experience has shown that when high latencies occur even on a single initiator or device in a fabric, that not only does the F-port attached to this device see Class 3 frame discards, but the resulting back pressure due to the lack of credit can build up in the fabric and cause other flows not directly related to the high latency device to have their frames discarded at ISLs.

Edge Hold Time can be used to reduce the likelihood of this back pressure into the fabric by assigning a lower Hold Time value only for edge ports (initiators or devices). The lower EHT value will ensure that frames are dropped at the F-port where the credit is lacking, before the higher default Hold Time value used at the ISLs expires, allowing these frames to begin moving again. This localizes the impact of a high latency F-port to just the single edge where the F-port resides and prevents it from spreading into the fabric and affecting other unrelated flows.

Like Hold Time, Edge Hold Time is configured for the entire switch, and it is not configurable on individual ports or ASICs. Whether the EHT or HT values are used on a port depends on the particular platform and ASIC, as well as the type of port and also other ports that reside on the same ASIC.

Suggested settings

Edge Hold Time does not need to be set on “Core Switches” that consist of only ISLs and will therefore only use the standard Hold Time setting of 500 ms. The suggested values for platforms containing initiators and targets are based on specific deployment strategies. Users typically either separate initiators and targets on separate switches or mix initiators and targets on the same switch.

A frame drop has more significance for a target than an initiator because many initiators typically communicate with a single target port. However, target ports typically communicate with multiple initiators. Frame drops on target ports usually result in “SCSI Transport” error messages generated in server logs. Multiple frame drops from the same target port can affect multiple servers in what appears to be a random fabric or storage problem. Because the source of the error is not obvious, this can result in time wasted determining the source of the problem. Be careful therefore when applying EHT to switches where targets are deployed.

Note: The most common suggested value for EHT is 220 ms.

Only configure the lowest EHT value of 80 ms on edge switches that consist entirely of initiators. This lowest value is suggested for fabrics that are well maintained and when a more aggressive monitoring and protection strategy is deployed.

Redundancy and resiliency

An important aspect of SAN topology is the resiliency and redundancy of the fabric. The main objective is to remove any single point of failure. Resiliency is the ability of the network to continue to function and recover from a failure. Redundancy describes the duplication of components, even an entire fabric, to eliminate a single point of failure in the network. The FOS code provides resiliency built in the software, which can quickly “repair” the network to overcome most failures. For example, when a link between switches fails, routing is quickly recalculated and traffic is assigned to the new route. Of course, this assumes that there is a second route, which is when redundancy in the fabric becomes important.

The key to high availability and enterprise-class installation is redundancy. By eliminating a single point of failure, business continuance can be provided through most foreseeable and even unforeseeable events. At the highest level of fabric design, the complete network needs to be redundant, with two completely separate fabrics that do not share any network equipment (routers or switches).

Servers and storage devices need to be connected to both networks using some form of multipath I/O (MPIO) solution, so that data can flow across both networks seamlessly in either an active/active or active/passive mode. MPIO ensures that if one path fails, an alternative is readily available. Ideally, the networks are identical, but at a minimum, they need to be based on the same switch architecture. In certain cases, these networks are in the same location. However, to provide for disaster recovery (DR), two separate locations are often used, either for each complete network or for sections of each network. Regardless of the physical geography, there are two separate networks for complete redundancy.

In summary, the suggestions for the SAN design are to ensure application availability and resiliency:

- ▶ Redundancy is built into fabrics to avoid a single point of failure.
- ▶ Servers are connected to storage via redundant fabrics.
- ▶ MPIO-based failover is available from server to storage.
- ▶ Redundant fabrics are based on similar architectures.
- ▶ Separate storage and server tiers exist for independent expansion.
- ▶ At a minimum, core switches need to be of equal or higher performance compared to the edges.
- ▶ Define the highest performance switch in the fabric to be the principal switch.

In addition to redundant fabrics, redundant links need to be placed on different blades, different ASICs, or at least different port groups whenever possible, as shown in Figure A-8. Whatever method is used, it is important to be consistent across the fabric. For example, do not place ISLs on lower port numbers in one chassis and stagger them in another chassis.

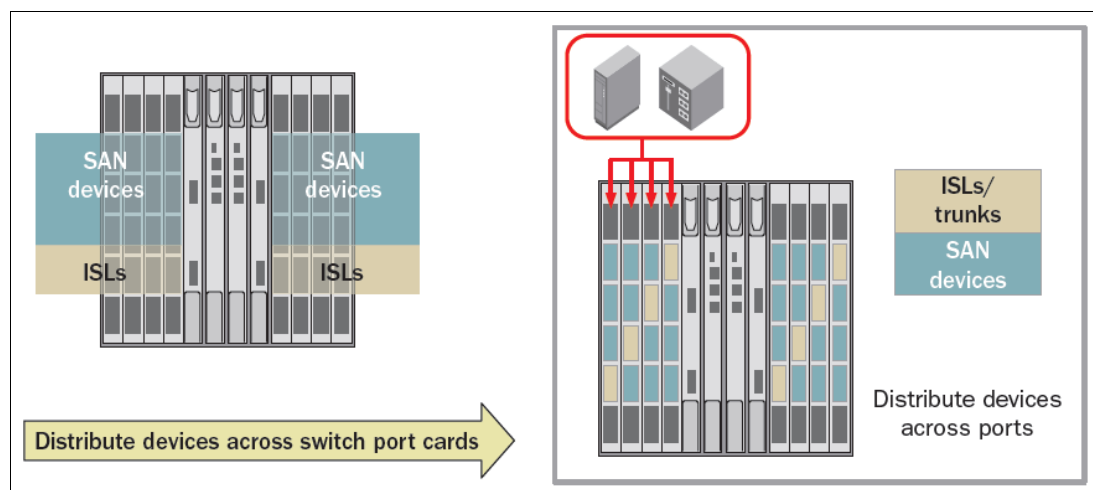


Figure A-8 Examples of distributed connections for redundancy

Note: In Figure A-8 on page 355, ISLs and SAN devices are placed on separate Application-Specific Integrated Circuits (ASICs) or port groups. Also, the Edge Hold Time (EHT) feature is ASIC-dependent, and the setting applies to all the ports on the ASIC. In environments with high-latency devices, place devices and ISLs on separate ASICs when possible.

For more details about fabric resiliency best practices, see *Fabric Resiliency Best Practices*, REDP-4722:

<http://www.redbooks.ibm.com/redpapers/pdfs/redp4722.pdf>

Single point of failure

By definition, a *single point of failure* (SPOF) is a part of a system or component that, if it fails, will stop the entire system from working. These components can be HBAs, power supplies, ISLs, switches, or even entire fabrics. Fabrics are typically deployed in pairs, mirroring one another in topology and configuration, and (unless routing is used) are completely isolated from one another.

The assessment of a potential SPOF involves identifying the critical components of a complex system that will provoke a total systems failure in a malfunction. The duplication of components (redundancy) eliminates any single point of failure of your SAN topology.

Distance and Fibre Channel over IP (FCIP) design preferred practices

For a complete DR solution, SANs are typically connected over metro or long-distance networks. In both cases, path latency is critical for mirroring and replication solutions. For native FC links, the amount of time that a frame spends on the cable between two ports is negligible, because that aspect of the connection speed is limited only by the speed of light. The speed of light in optics amounts to approximately 5 microseconds per kilometer, which is negligible compared to typical disk latency of 5 - 10 milliseconds. The Brocade Extended Fabrics feature enables full-bandwidth performance across distances spanning up to hundreds of kilometers. It extends the distance ISLs can reach over an extended fiber by providing enough buffer credits on each side of the link to compensate for latency introduced by the extended distance.

SANs spanning data centers in different physical locations can be connected through dark fiber connections using Extended Fabrics, a FOS optional licensed feature, with wave division multiplexing, such as dense wave division multiplexing (DWDM), coarse wave division multiplexing (CWDM), and time-division multiplexing (TDM). This is similar to connecting switches in the data center with one exception: additional buffers are allocated to E_ports connecting over distance. The Extended Fabrics feature extends the distance that the ISLs can reach over an extended fiber. This is accomplished by providing enough buffer credits on each side of the link to compensate for latency introduced by the extended distance. Use the buffer credit calculation above or the new CLI tools with FOS v7.1 to determine the number of buffers needed to support the required performance.

Any of the first eight ports on the 16 Gbps port blade can be set to 10 Gbps FC for connecting to a 10 Gbps line card D/CWDM without the need for a specialty line card. If connecting to DWDMs in a pass-through mode where the switch is providing all the buffering, a 16 Gbps line rate can be used for higher performance.

Consider the following suggestions:

- Connect the cores of each fabric to the DWDM.
- If using trunks, use smaller trunks and more trunks on separate port blades for redundancy and to provide more paths. Determine the optimal number of trunk groups between each set of linked switches, depending on traffic patterns and port availability.

Fibre Channel over IP (FCIP) links are most commonly used for Remote Data Replication (RDR) and remote tape applications, for Business Continuity/Disaster Recovery. Transporting data over significant distances beyond the reach of a threatening event will preserve the data so that an organization can recover from that event. A device that transports FCIP is often called a *channel extender*.

RDR is typically storage array-to-array communications. The local array at the production site sends data to the other array at the backup site. This can be done through native FC, if the backup site is within a practical distance and there is DWDM or dark fiber between the sites. However, a cost-sensitive infrastructure for IP connectivity and not native FC connectivity is more commonly available. This works out well because the current technology for FCIP is high speed and adds only a minute amount (about 35µs) of propagation delay, which is appropriate for not only asynchronous RDR and tape applications but also synchronous RDR applications.

The preferred practice deployment of FCIP channel extenders in RDR applications is to connect the FC F_ports on the channel extender directly to the FC node ports (N_ports) on the array, and not to go through the production fabric at all. On most large-scale arrays, the FC port that is assigned to RDR is dedicated to only RDR and no host traffic. Considering that the RDR port on the array can only communicate RDR traffic, there is no need to run that port into the production fabric. There are valid reasons to have to go through a production fabric, such as the IBM SAN Volume Controller. The IBM SAN Volume Controller has requirements for connectivity to the production fabric.

For RDR, the preferred practice is to use a separate and dedicated IP connection between the production data center and the backup site. Often, a dedicated IP connection between data centers is not practical. In this case, bandwidth must at least be logically dedicated. There are a few ways to do this. First, use quality of service (QoS), and give FCIP a high priority. This logically dedicates enough bandwidth to FCIP over other traffic. Second, use Committed Access Rate (CAR) to identify and rate-limit certain traffic types. Use CAR on the non-FCIP traffic to apportion and limit that traffic to a maximum amount of bandwidth, leaving the remainder of the bandwidth to FCIP. Set the aggregate FCIP rate limit on the SAN06B-R switch or Feature Code (FC) 3890 blade to use the remaining portion of the bandwidth. This results in logically dedicating bandwidth to FCIP. Last, it is possible, with massive overprovisioning of bandwidth, for various traffic types to coexist over the same IP link. Brocade FCIP uses an aggressive TCP stack called Storage Optimized TCP (SO-TCP), which dominates other TCP flows within the IP link, causing them to back off dramatically. If the other flows are User Datagram Protocol (UDP)-based, the result is considerable congestion and excessive dropped packets for all traffic.

The preferred practice is to always rate-limit the FCIP traffic on the SAN06B-R or FC3890 blade and never rate-limit FCIP traffic in the IP network, which often leads to problems that are difficult to troubleshoot. The rate-limiting technology on the SAN06B-R/FC3890 is advanced, accurate, and consistent, so there is no need to double rate limit. If a policy required you to double rate limit, the IP network needs to set its rate limiting above that of the SAN06B-R/FC3890 with plenty of extra room.

To determine the amount of network bandwidth needed, it is suggested that a month's worth of data is gathered using various tools that are host-based, fabric-based, and storage-based. It is important to understand the host-to-disk traffic because that is the amount of traffic to be replicated, or mirrored, to the remote disk.

If you are going to use synchronous RDR (RDR/S), record the peak values. If you are going to use asynchronous RDR (RDR/A), record the average value over the hour. RDR/S must have enough bandwidth to send the write I/O immediately; therefore, there must be enough bandwidth to accommodate the entire demand, which is the peak value. RDR/A needs only enough bandwidth to accommodate the high average discovered over an adequate recording period because RDR/A essentially performs traffic shaping, moving the peaks into the troughs, which works out to the average. It cannot be the average over an extremely long period because those troughs might not occur soon enough to relieve the array of the peaks. This causes excessive journaling of data, which is difficult to recover from.

Plot the values into a histogram. More than likely, you get a Gaussian curve. Most of the averages will fall within the first standard deviation of the curve, which is 68.2% of the obtained values. The second standard deviation will include 95.4% of the obtained values, which are enough samples to determine the bandwidth you will need. Outside of this, the values are corner cases, which most likely can be accommodated by the FCIP network due to their infrequency. Use a bandwidth utilization value that you are comfortable with between σ and 2σ . You can plan for a certain amount of compression, such as 2:1. However, the preferred practice is to use compression as a way to address future bandwidth needs. It is probably best not to push the limit right at the start because then you will have nowhere to go soon.

You can take advantage of FCIP Trunking to implement redundant network routes from site to site. But it is important to understand whether traffic can fail over to the alternate route transparently or whether that will affect traffic flow.

For disk extension using emulation (FastWrite), a single tunnel between sites is suggested. If multiple tunnels must be used, use Traffic Isolation (TI) zones or logical switch configuration to ensure that the same exchange always traverses by the same tunnel in both directions. Use multiple circuits instead of multiple tunnels for redundancy and failover protection.

Virtual fabrics

The IBM b-type 16 Gbps Backbones with the FC3890 Extension Blade and the SAN06B-R Extension Switch all support virtual fabrics (VFs) with no additional license. The SAN06B-R supports a maximum of four logical switches (LS) and does not support a base switch. Because there is no base switch, the SAN06B-R cannot provide support for XISL or Fibre Channel Routing (FCR) (no EX_ports and VEX_ports). VF on the SAN06B-R must be disabled if a separate RDR network is not feasible, and FCR is required to connect to production edge fabrics.

VF on the SAN06B-R/FC3890 plays a primary role in providing ways to achieve deterministic paths for protocol optimization, or for specific configuration and management requirements providing unique environments for Fibre Channel Protocol (FCP). VFs are the preferred alternative over Traffic Isolation (TI) Zones to establish the deterministic paths necessary for protocol optimization (FCIP-FW, Open Systems Tape Pipelining (OSTP), and FICON Emulation).

Protocol optimization requires that an exchange and all its sequences and frames pass through the same VE_port for both outbound and return. This means that only a single VE_port must exist within a VF LS. By putting a single VE_port in an LS, there is only one physical path between the two LSs that are connected via FCIP. A single physical path provides a deterministic path. When many devices or ports are connected for transmission across FCIP, as is the case with tape, for example, it is difficult to configure and maintain TI Zones. However, it is operationally simplistic and more stable to use VF LS.

Configuring more than one VE_port, one manually set with a higher Fibre Channel shortest path first (FSPF) cost, is referred to as a “*lay in wait*” VE_Port, and it is not supported for FCIP-FW, OSTP, or FICON Emulation. A “*lay in wait*” VE_Port can be used without protocol optimization and with RDR applications that can tolerate the topology change and some frame loss. A few FC frames might be lost when using “*lay in wait*” VE_Ports. If there are multiple VE_ports within an LS, routing across those VE_ports is performed according to the Advanced Performance Tuning (APT) policy.

VFs are significant in mixed mainframe and open system environments. Mainframe and open system environments are configured differently and only VFs can provide autonomous LSs accommodating the different configurations. Remember that RDR between storage arrays is open systems (IBM Metro/Global Mirror), even when the volume is written by FICON from the mainframe.

Understand that using a VE_port in a selected LS does not preclude that VE_Port from sharing an Ethernet interface with other VE_ports in other LSs. This is referred to as *Ethernet Interface Sharing*, which is described next.

Monitoring

Any mission-critical infrastructure must be correctly monitored. Although there are many features available in FOS to assist you with monitoring, protecting, and troubleshooting fabrics, several recent enhancements are implemented that deal exclusively with this area. An overview of the major components is provided. A complete guide to health monitoring is beyond the scope of this document. For more detailed information, see *Fabric OS Command Reference Supporting Fabric OS V7.1.0*, Part number 53-1002746-01, the *Fabric OS Troubleshooting and Diagnostics Guide, Supporting Fabric OS V7.1.0*, Part number: 53-1002751-02, and the correct SAN Health and Fabric Watch guides.

Fabric Watch

Fabric Watch is an optional health monitor that allows you to constantly monitor each director or switch for potential faults and automatically alerts you to problems long before they become costly failures.

Fabric Watch tracks various SAN fabric elements and events. Monitoring fabric-wide events, ports, and environmental parameters enables early fault detection and isolation as well as performance measurement. You can configure fabric elements and alert thresholds on an individual port basis, and you can also easily integrate Fabric Watch with enterprise system management solutions.

Fabric Watch provides customizable monitoring thresholds. You can configure Fabric Watch to provide notification before problems arise, such as reporting when network traffic through a port is approaching the bandwidth limit. This information enables you to perform preemptive network maintenance, such as trunking or zoning, and avoid potential network failures.

Fabric Watch lets you define how often to measure each switch and fabric element and specify notification thresholds. Whenever fabric elements exceed these thresholds, Fabric Watch automatically provides notification using several methods, including email messages, SNMP traps, and log entries.

Fabric Watch was significantly upgraded starting in FOS v6.4, and it continues to be a major source of early warning for fabric issues. Useful enhancements, such as port fencing to protect the fabric against misbehaving devices, are added with each new release of FOS.

Port fencing

Port fencing allows a switch to monitor specific behaviors on the port and protect a switch by fencing the port when specified thresholds are exceeded. Only implement it if the SAN management or the monitoring team have the required time and resources to monitor and quickly react to the port fencing events.

Frame Viewer

Frame Viewer was introduced in Fabric OS v7.0 to allow the fabric administrator more visibility into C3 frames dropped due to timeouts. When frame drops are observed on a switch, the user can use this feature to identify which flows the dropped frames belong to and potentially determine affected applications by identifying the endpoints of the dropped frame. Frames discarded due to timeouts are sent to the CPU for processing. Fabric OS captures and logs information about the frame, such as source ID (SID), destination ID (DID), and transmit port number. This information is maintained for a limited number of frames. The user can use the CLI (**frame log --show**) to retrieve and display this information.

Bottleneck detection

The bottleneck detection feature offers the following benefits:

- ▶ Prevent degradation of throughput in the fabric.
The bottleneck detection feature alerts you to the existence and locations of devices that are causing latency. If you receive alerts for one or more F_ports, use the CLI to check whether these F_ports have a history of bottlenecks.
- ▶ Reduce the time that it takes to troubleshoot network problems.
If you notice one or more applications slowing down, you can determine whether any latency devices are attached to the fabric and where. You can use the CLI to display a history of bottleneck conditions on a port. If the CLI shows above-threshold bottleneck severity, you can narrow the problem to device latency rather than problems in the fabric.

You can use the bottleneck detection feature with other Adaptive Networking features to optimize the performance of your fabric. Bottleneck detection requires some tuning on an environment-by-environment basis.

Credit loss

Fabric OS v7.1 and later support back-end credit loss detection for back-end ports and core blades. The credit loss detection and recovery function is enabled and disabled through the CLI by using the **bottleneckmon --cfgcredittools** command.

RAS log

The RAS log is the FOS error message log. Messages are organized by FOS component, and each message has a unique identifier, as well as severity, source, and platform information and a text message. The RAS log is available from each switch and director through the **errdump** command. The RAS log messages can be forwarded to a syslog server for centralized collection.

Audit log

The Audit log is a collection of information created when specific events are identified on an IBM b-type platform. The log can be dumped through the **auditdump** command, and audit data can also be forwarded to a syslog server for centralized collection.

Information is collected about many different events associated with zoning, security, trunking, FCIP, FICON, and others. Each release of the FOS provides more audit information.

SAN Health

SAN Health provides snapshots of fabrics showing information, such as switch and firmware levels, connected device information, snapshots of performance information, zone analysis, and ISL fan-in ratios.

Design guidelines

It is strongly advised that you implement some form of monitoring of each switch. Often, issues start out relatively benignly and gradually degrade into more serious problems. Monitoring the logs for serious and error severity messages will go a long way in avoiding many problems:

- ▶ Plan for a centralized collection of RAS logs, and perhaps Audit logs, by using syslog. You can optionally filter these messages relatively easily through some simple Perl programs.
- ▶ IBM b-type platforms are capable of generating SNMP traps for most error conditions. Consider implementing some sort of alerting mechanism by using SNMP.

Look at error logs regularly. Many users use combinations of syslog and SNMP with the Fabric Watch and the logs to maintain a close watch on the health of their fabrics. Network Advisor has many helpful features to configure and monitor your fabrics.

Scalability and supportability

IBM b-type products are designed with scalability in mind, knowing that most installations will continue to expand and that growth is supported by few restrictions. However, follow the same basic principles outlined in previous sections as the network grows. Evaluate the impact on the topology, data flow, workload, performance, and perhaps most importantly, redundancy and resiliency of the entire fabric.

If these design preferred practices are followed when the network is deployed, small incremental changes might not adversely affect the availability and performance of the network. However, if changes are ongoing and the fabric is not correctly evaluated and updated, performance and availability can be jeopardized.

In summary, although IBM SANs are designed to allow for any-to-any connectivity, and they support provision-anywhere implementations, these practices can have an adverse impact on the performance and availability of the SAN if left unchecked. As described, the network needs to be monitored for changes and routinely evaluated for how well it meets your redundancy and resiliency requirements.

Supportability

Supportability is a critical part of deploying a SAN. Follow these guidelines to ensure that the data needed to diagnose fabric behavior or problems is collected. Although not all of these items are necessary, they are all important pieces. You can never know which piece will be needed, so having all of the pieces available is best:

- ▶ Configure Fabric Watch monitoring: Use Fabric Watch to implement the proactive monitoring of errors and warnings, such as cyclic redundancy check (CRC) errors, loss of synchronization, and high-bandwidth utilization.
- ▶ Configure syslog forwarding: By keeping historical log messages and having all switch messages sent to one centralized syslog server, troubleshooting can be expedited and simplified. Forwarding switch error messages to one centralized syslog server and keeping historical log messages enable faster and more effective troubleshooting and provide simple monitoring functionality.
- ▶ Back up switch configurations: Back up switch configurations regularly so that you can restore the switch configuration if a switch has to be swapped out or to provide change monitoring functionality.
- ▶ Enable audit functionality: Provide audit functionality for the SAN and track which administrator made which changes, the usage of multiple user accounts (or RADIUS), and the configuration of change tracking or audit functionality (with the use of errorlog/syslog forwarding).
- ▶ Configure multiple user accounts (Lightweight Directory Access Protocol (LDAP)/OpenLDAP or RADIUS): Make mandatory use of the personalized user accounts part of the IT/SAN security policy, so that user actions can be tracked. Also, restrict access by assigning specific user roles to individual users.
- ▶ Establish a test “bed”: Set up a test environment to test new applications, firmware upgrades, driver functionality, and scripts to avoid missteps in a production environment. Validate functionality and stability with rigorous testing in a test environment before deploying into the production environment.
- ▶ Implement serial console server: Implement serial remote access so that switches can be managed even when there are network issues or problems during switch boot or firmware upgrades.
- ▶ Use aliases: Use “aliases” to give switch ports and devices meaningful names. Using aliases to give devices meaningful names can lead to faster troubleshooting.
- ▶ Configure **supportftp**: Configure **supportftp** for automatic file transfers. The parameters set by this command are used by supportSave and traceDump.
- ▶ Configure a Network Time Protocol (NTP) server: To keep a consistent and accurate date and time on all the switches, configure the switches to use an external time server.
- ▶ Disable the default zone: Set the default zoning mode to “No Access”.
- ▶ Enable insistent domain ID: It is a good practice to set the domain ID to be insistent to make the domain ID insistent across reboots, power cycles, and failovers.
- ▶ Set Persistent Disable for unused ports: If possible, unused ports need to be persistently disabled.
- ▶ Disable E_port capability for F_ports: Ports that are connected to storage and host devices need to have their E_port functionality persistently disabled.

Implementation

In this topic, we describe the initial setup to implement the switches. We then describe the EZSwitchSetup, a starter kit that greatly simplifies the setup and implementation of SAN switches.

Initial setup

Before you configure any IBM System Storage SAN switch, the switches need to be physically installed and correctly connected. The hardware requirements and specifications are in the specific hardware reference guide for the SAN switch model that you have acquired.

After you turn on the SAN switch, it requires you to set several initial configuration parameters. All of the b-type switches require the same initial setup. The fundamental steps have not changed from the earlier switch models.

Configuring IBM System Storage fabric backbone

The IBM System Storage fabric backbone must be configured before it is connected to the fabric, and all of the configuration commands must be entered through the active control processor (CP) blade. The SAN768B-2 that is going to be used to illustrate the configuration steps includes the following parameters:

- ▶ IP address
- ▶ Switch name
- ▶ Chassis name
- ▶ Domain ID
- ▶ Port identifier (PID) mode

Establishing a serial connection to the IBM SAN768B-2

To establish a serial connection to the console port on the IBM SAN768B-2, complete the following steps:

1. Verify that the IBM SAN768B-2 is powered on and that power-on self-test (POST) is complete by verifying that all power LED indicators on the port, control processor, and core switch blades display a steady green light.
2. Remove the shipping cap from the CONSOLE port on the active CP. Use the serial cable provided with the IBM SAN768B-2 to connect the CONSOLE port on the active CP to a computer workstation. The active CP blade is indicated by an illuminated (blue) LED.
3. Access the IBM SAN768B-2 using a terminal emulator application, such as HyperTerminal in a Windows environment or **tip** in a UNIX environment.
4. Disable any serial communication programs running on the workstation, such as synchronization programs.

5. Open a terminal emulator application, such as HyperTerminal on a personal computer, or **term**, **tip**, or **kermit** in a UNIX environment, and configure the application by using the following information:

- In a Microsoft Windows environment, use the following parameters and values:
 - Bits per second: 9600
 - Databits: 8
 - Parity: None
 - Stop bits: 1
 - Flow control: None
- In a UNIX environment, enter the following string at the prompt:
`tip /dev/ttyb -9600`
- If **ttyb** is already in use, use **ttya** instead and enter the following string at the prompt:
`tip /dev/ttya -9600`

When the terminal emulator application stops reporting information, press Enter. You receive the following login prompt:

CP0 Console Login:

6. Log in to the SAN768B-2 as admin. The default password is password. At the initial login, you are prompted to enter new administrator and user passwords. Changing the password is optional, but advised. The login appears:

```
Fabric OS (swDir)
swDir login: admin
Password:
Please change your passwords now.
Use Control-C to exit or press 'Enter' key to proceed.
swDir:admin>
```

Configuring IP addresses

The IBM SAN768B-2 requires three IP addresses, which are configured by using the **ipAddrSet** command. IP addresses are required for both CP blades (CP0 and CP1) and for the chassis management IP address (shown as switch under the **ipAddrShow** command) in the SAN768B-2.

Note: The IBM SAN768B-2 uses the following default IP addresses and host names:

- ▶ 10.77.77.75/CP0 (the CP blade in slot 6 at the time of configuration)
- ▶ 10.77.77.74/CP1 (the CP blade in slot 7 at the time of configuration)

After you log in to the active CP by using the serial cable, as described in “Establishing a serial connection to the IBM SAN768B-2” on page 363, complete the following steps:

1. Set up the Chassis IP address:

```
swDir:admin> ipAddrSet -chassis
```

Enter the information at the prompts. Specify the **-chassis** IP address.

2. Set up the CP0 IP address by entering the **ipaddrset -cp 0** command:

```
swDir:admin> ipAddrSet -cp 0
```

Enter the information at the prompts.

3. Set up the CP1 IP address by entering the **ipaddrset -cp 1** command:

```
swDir:admin> ipAddrSet -cp 1
```

Enter the information at the prompts.

After using a serial connection to configure the IP addresses for the IBM SAN768B-2, you can connect the active CP blade to the local area network (LAN) and complete the configuration by using either a serial session, Telnet, Secure Shell (SSH), or a management application, such as Web Tools or IBM Network Advisor.

Customizing the switch name

The switch name of the IBM SAN768B-2 can be up to 30 characters when using Fabric OS Release 6.3.0 or later. The switch name can include letters, numbers, hyphens, and underscore characters, and it must begin with a letter.

Type **switchName** followed by the new name in double quotation marks as shown in Example A-1.

Example A-1 Setting the switch name

```
swDir:admin> switchName "ItsoSANswitch1"
Committing configuration...
Done.
ItsoSANswitch1:admin>
```

Setting the domain ID

Each switch in a fabric must have a unique domain ID. The domain ID can be manually set through the **configure** command or it can be automatically set. The default domain ID for the IBM SAN768B-2 is 1. Use the **fabricShow** command to view the already assigned domain IDs.

Follow these steps to set the domain ID:

1. Enter **switchDisable** to disable the SAN switch:

```
ItsoSANswitch1:admin> switchDisable
```

2. Enter the **configure** command:

```
ItsoSANswitch1:admin> configure
```

3. Enter **y** at the Fabric parameters prompt:

```
Fabric parameters (yes, y, no, n): [no] y
```

4. Enter a unique domain ID:

```
Domain: (1.239) [1] 2
```

5. Complete the remaining prompts or press Ctrl+D to accept the settings and exit.

6. Enter the **switchEnable** command to reenable the switch:

```
ItsoSANswitch1:admin> switchEnable
```


Verifying the PID mode

Before connecting the IBM SAN768B-2 to the fabric, verify that the port identifier (PID) mode on the switch matches the other switches in the fabric. This parameter must be identical for all switches in the fabric. It is set by using the **configure** command.

Setting the date

Although a switch with the incorrect date and time functions correctly, it is advised that you make these values real and accurate because they are used for time stamping during the logging of events. We suggest that you set these parameters before any further operations because you might find this information helpful if you have to troubleshoot later.

Set the day and time by using the **date MMDDhhmmYY** command and these variables:

- ▶ **MM** is the month; the valid values are 01 - 12.
- ▶ **DD** is the date; the valid values are 01 - 31.
- ▶ **hh** is the hour; the valid values are 00 - 23.
- ▶ **mm** is minutes; the valid values are 00 - 59.
- ▶ **YY** is the year; the valid values are 00 - 99 (values greater than 69 are interpreted as 1970 - 1999, and values less than 70 are interpreted as 2000 - 2069).

See Example A-2 for the use of this command.

Example A-2 Setting the date and time

```
switch:admin> date
Fri Sep 28 17:01:48 UTC 2007
switch:admin> date "0913123013"
Fri Sep 13 12:30:00 UTC 2013
```

Note: Use an external Network Time Protocol (NTP) server to ensure that all switches in your environment are synchronized.

Using EZSwitchSetup

The *EZSwitchSetup* starter kit greatly simplifies the setup and implementation of supported switches. The kit ships with the switch and includes a serial cable and a CD that contains the setup software. It makes the switch setup as simple as a “click-and-go” solution. It runs only in a single switch fabric.

If you follow the standard switch configuration practice, you implement a new switch by connecting a serial cable, setting up a tool, such as Hyperterm, to communicate, and implementing the **ipaddrset** command to configure the IP address. Then, you can connect to the network using an Ethernet cable, using a web browser to access Web Tools, or alternatively using Telnet to enter CLI mode and to configure the switch further. You can set up zoning, assuming that all devices are connected and also that switch status monitoring uses Web Tools, SNMP, or an external application.

EZSwitchSetup greatly simplifies this process by walking you through all the steps automatically using an easy-to-use GUI.

EZSwitchSetup version 7.2.0 is compatible with the following GEN 4 and GEN 5 SAN switch models:

- ▶ SAN24B-5
- ▶ SAN48B-5
- ▶ SAN96B-5

- ▶ SAN24B-4
- ▶ SAN40B-4
- ▶ SAN80B-4

For full compatibility and the instructions to use EZSwitchSetup, see the *EZSwitchSetup Administrator's Guide*:

<http://www.brocade.com>

Note: Although your switch might have advanced capabilities, EZSwitchSetup is for a single-switch fabric with FC ports only. To configure and manage other features on your switch, use the CLI, Web Tools, or IBM Network Advisor.

For more information and in-depth instructions about how to deploy an IBM b-type 16 Gbps System Storage product, see *IBM b-type Gen 5 16 Gbps Switches and Network Advisor*, SG24-8186.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- ▶ *IBM FlashSystem 840 Product Guide*, TIPS1079
- ▶ *IBM FlashSystem V840*, TIPS1158
- ▶ *FlashSystem 900 Product Guide*, TIPS1261
- ▶ *IBM FlashSystem V9000 Product Guide*, TIPS1281
- ▶ *Implementing IBM FlashSystem 900*, SG24-8271
- ▶ *Introducing and Implementing IBM FlashSystem V9000*, SG24-8273
- ▶ *Implementing FlashSystem 840 with SAN Volume Controller*, TIPS1137
- ▶ *Flash or SSD: Why and When to Use IBM FlashSystem*, REDP-5020
- ▶ *Faster DB2 Performance with IBM FlashSystem*, TIPS1041
- ▶ *IBM FlashSystem in IBM PureFlex System Environments*, TIPS1042
- ▶ *Fabric Resiliency Best Practices*, REDP-4722
- ▶ *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- ▶ *Implementing the IBM Storwize V7000 V7.2*, SG24-7938
- ▶ *IBM b-type Gen 5 16 Gbps Switches and Network Advisor*, SG24-8186
- ▶ *Implementing Systems Management of IBM PureFlex System*, SG24-8060
- ▶ *Implementing the IBM SAN Volume Controller and FlashSystem 820*, SG24-8172
- ▶ *IBM SAN and SVC Stretched Cluster and VMware Solution Implementation*, SG24-8072

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- ▶ *IBM FlashSystem 840 Troubleshooting, Recovery, and Maintenance Guide*, SC27-6297
- ▶ *IBM FlashSystem 840 Installation Guide*, GI13-2871

Online resources

These websites are also relevant as further information sources:

- ▶ IBM FlashSystem family product page
<http://www.ibm.com/storage/flash>
- ▶ IBM FlashSystem 840 IBM Knowledge Center
http://www.ibm.com/support/knowledgecenter/ST2NVR_1.3.0
- ▶ IBM FlashSystem 840 support portal and product documentation
http://www.ibm.com/support/entry/portal/product/system_storage/flash_storage/flash_high_availability_systems/ibm_flashsystem_840
- ▶ IBM Redbooks Solution and Product Guides for the IBM FlashSystem family
<http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg245250.html?Open>
- ▶ IBM System Storage Interoperation Center (SSIC)
<http://www.ibm.com/systems/support/storage/ssic>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

Implementing IBM FlashSystem 840

SG24-8189-02

ISBN 0738440795



(0.5" spine)

0.475" <-> 0.873"

250 <-> 459 pages



SG24-8189-02

ISBN 0738440795

Printed in U.S.A.

Get connected

