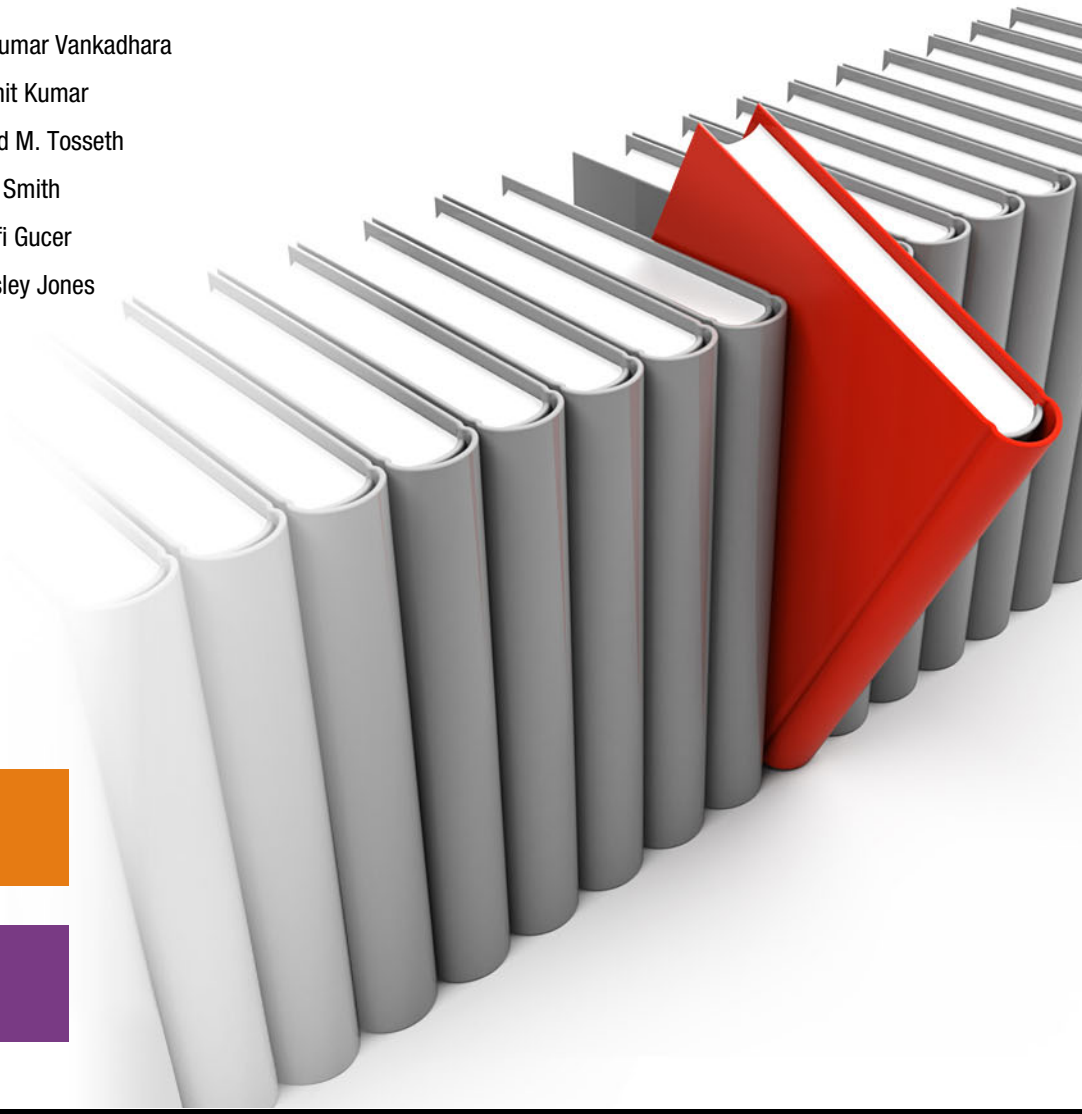
Redbooks

ibm.com/redbooks

Implementation Guide for IBM Elastic Storage System 5000

Brian Herr
Farida Yaragatti
Jay Vaddi
John Sing
Jonathan Terner
Luis Bolinches
Mary Jane Zajac
Puneet Chaudhary
Ravindra Sure
Ricardo D. Zamora Ruvalcaba
Robert Guthrie
Shradha Thakare
Stephen M. Tee
Steve Duersch
Sukumar Vankadhara
Sumit Kumar
Todd M. Tosseth
Van Smith
Vasfi Gucer
Wesley Jones



Analytics

Storage

IBM

Redbooks

IBM Redbooks

Implementation Guide for IBM Elastic Storage System 5000

December 2020

SG24-8498-00

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

First Edition (December 2020)

This edition applies to IBM Elastic Storage System 5000.

Contents

Notices v
Trademarks vi

Preface vii
Authors vii
Now you can become a published author, too xii
Comments welcome xii
Stay connected to IBM Redbooks xiii

Chapter 1. Introduction 1
1.1 IBM Spectrum Scale RAID 2
1.1.1 Product history 2
1.1.2 Distinguishing features 3
1.2 IBM Elastic Storage System 5
1.3 IBM Elastic Storage System 5000 6
1.3.1 What is new in IBM ESS 5000 6
1.3.2 Added value 6
1.4 License considerations 7

Chapter 2. IBM Elastic Storage System 5000 architecture and technical overview . . . 9
2.1 Platform 10
2.1.1 IBM ESS Management Server 10
2.1.2 I/O node 11
2.1.3 NVDIMMs 12
2.1.4 Protocol node 13
2.1.5 IBM ESS 5000 variants 14
2.1.6 IBM ESS 5000 SL series 15
2.1.7 Enclosures 16
2.2 Software enhancements 18
2.2.1 IBM ESS software solution stack overview 18
2.2.2 IBM Spectrum Scale RAID 21
2.2.3 IBM ESS solution installation and management scripts 23
2.3 Enclosure overview: Sample reliability, availability, and serviceability enhancements . 24
2.3.1 Enclosure overview 25
2.4 Reliability, availability, and serviceability features 27
2.4.1 Monitoring IBM ESS 5000 health 27
2.4.2 Monitoring the IBM ESS performance 28
2.4.3 Physical disk health 29
2.5 Software-related RAS enhancements 29
2.5.1 Integrated Call Home 29
2.5.2 Software Call Home 29
2.6 Performance 30
2.6.1 Network 31
2.6.2 Tuning 31
2.6.3 Disk topology 33
2.7 GUI enhancements 34
2.7.1 GUI users 34
2.7.2 System setup wizard 34
2.7.3 Using the GUI 38

2.7.4 Monitoring of IBM ESS 5000 hardware . . . 40
2.7.5 Storage . . . 46
2.7.6 Event notification . . . 46
2.7.7 Dashboards . . . 49
2.7.8 More information . . . 50

Chapter 3. Planning considerations . . . 51
3.1 Planning . . . 52
3.1.1 Technical and Delivery Assessment . . . 52
3.1.2 System planning worksheets . . . 52
3.1.3 Networking hardware . . . 53
3.2 Stand-alone environment . . . 59
3.2.1 Small starter environment: Stand-alone IBM ESS 5000 system with IBM Spectrum Scale client nodes . . . 60
3.2.2 Installing and deploying a remote NSD client cluster . . . 61
3.2.3 Small starter environment: IBM ESS 5000 system with IBM Spectrum Scale protocol nodes . . . 65
3.3 Mixed environment . . . 71
3.3.1 Adding an IBM ESS 5000 to an existing IBM ESS cluster . . . 71
3.3.2 Adding an IBM ESS 5000 to an IBM ESS Legacy cluster . . . 72
3.3.3 Adding an IBM ESS 5000 to an IBM ESS 3000 cluster . . . 73
3.3.4 Adding an IBM ESS 5000 system to a mixed IBM ESS Legacy and IBM ESS 3000 cluster . . . 73
3.3.5 Scenario 1: Using IBM ESS 5000 for data NSDs for the existing file system . . . 74
3.3.6 Scenario 2: Using the IBM ESS 5000 to create a file system . . . 77

Chapter 4. Use cases . . . 79
4.1 IBM ESS 5000 use cases overview . . . 80
4.2 Large capacity and data lake workloads . . . 81
4.2.1 Large capacity use cases . . . 81
4.2.2 Data lake use case . . . 83
4.3 Analytics and high-performance workloads . . . 86
4.3.1 HPC and data-intensive technical computing . . . 87
4.3.2 Data and analytics with Hadoop . . . 88
4.3.3 Storage for AI: Machine learning and deep learning . . . 90
4.4 Data optimization and resiliency . . . 93
4.4.1 Archive use case . . . 94
4.4.2 Information Lifecycle Management . . . 99
4.4.3 Resiliency . . . 99

Appendix A. Sample configuration files . . . 103
Sample 1 Gb Ethernet configuration file . . . 104
Sample IP over InfiniBand configuration files . . . 108

Related publications . . . 109
IBM Redbooks . . . 109
Online resources . . . 109
Help from IBM . . . 110

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions 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 provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises 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. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, 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 on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

Redbooks (logo)  ®
IBM®
IBM Cloud®
IBM Elastic Storage®
IBM Garage™
IBM Spectrum®
POWER®
POWER7®
POWER8®
POWER9™
Redbooks®

The following terms are trademarks of other companies:

ITIL is a Registered Trade Mark of AXELOS Limited.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Ansible, OpenShift, Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

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

Preface

This IBM Redbooks publication introduces and describes the IBM Elastic Storage Server 5000 (IBM ESS 5000) as a scalable, high-performance data and file management solution. The solution is built on IBM Spectrum Scale technology (formerly IBM General Parallel File System (IBM GPFS)).

IBM ESS is a modern implementation of software-defined storage (SDS), making it easier for you to deploy fast, highly scalable storage for artificial intelligence (AI) and big data. With the lightning-fast Non-Volatile Memory Express (NVMe) storage technology and industry-leading file management capabilities of IBM Spectrum Scale, the IBM ESS 3000 and IBM ESS 5000 nodes can grow to over yottabyte scalability and be integrated into a federated global storage system. By consolidating storage requirements from the edge to the core data center, which include Kubernetes and Red Hat OpenShift, IBM ESS can reduce inefficiency, lower acquisition costs, simplify storage management, eliminate data silos, support multiple demanding workloads, and deliver high performance throughout your organization.

This book provides a technical overview of the IBM ESS 5000 solution and helps you to plan the installation of the environment. We also explain the use cases where we believe they fit best.

Our goal is to position this book as the starting point for customers that want to use the IBM ESS 5000 as part of their IBM Spectrum Scale setups.

This book is targeted at technical professionals (consultants, technical support staff, IT Architects, and IT Specialists) who are responsible for delivering cost-effective storage solutions with IBM ESS 5000.

Authors

This book was produced by a team of specialists from around the world.

Brian Herr is a Software Engineer for IBM Spectrum Scale. He has been with IBM since 1983, working mostly in the area of high-performance computing (HPC). He has been working on the IBM ESS development team since 2008.

Farida Yaragatti is a Senior Software Engineer at IBM India. She has a BE degree in Electronics and Communication from Karnataka University, India, and has 12 years of experience in the software testing field. She has been part of manual and automation testing for IBM Spectrum Scale and IBM ESS deployment as a Senior Tester. Farida has worked at IBM for over 5 years and previously held roles within the IBM Platform Computing and IBM Smart Analytics System testing teams. She has strong engineering professional skills in software deployment testing, including automation that uses various scripting technologies, such as Python, shell scripting, the Robot framework, Ansible, and Linux.

Jay Vaddi is a Storage Performance Engineer at IBM Tucson, AZ. He has been with IBM and the performance team for over 4 years. His focus is primarily on performance analysis and evaluations of IBM Spectrum Scale and IBM Elastic Storage Server products.

John Sing is Offering Evangelist for IBM Spectrum Scale Elastic Storage Server. In his over 25 years with IBM, John has been a world-recognized IBM speaker, author, and strategist in enterprise storage, file and object storage, internet scale workloads and data center design, big data, cloud, it strategy planning, high availability (HA), business continuity, and disaster recovery (DR). He has spoken at over 40 IBM conferences worldwide, and is the author of eight IBM Redbooks publications and nine IBM Redpaper publications.

Jonathan Terner is a Software Engineer and a member of the IBM Spectrum Scale RAID team working in the United States. He recently finished his bachelor's degree in Computer Science and Mathematics from Binghamton University, where he focused on distributed systems and virtualization technology.

Luis Bolinches has been working with IBM Power Systems servers for over 15 years, and has been with IBM Spectrum Scale for over 10 years. He works 20% for IBM Systems Lab Services in the Nordic region, and the other 80% as part of the IBM Spectrum Scale development team.

Mary Jane Zajac is a Software Engineer working in the US who specializes in testing. She has been working on IBM ESS for the past 5 years, and has been the IBM ESS FVT team lead for the past 2 years. Her previous experience includes HPC Function Verification Test (FVT), integration, and system testing.

Puneet Chaudhary is a Technical Solutions Architect working with the IBM Elastic Storage Server and IBM Spectrum Scale solutions. He has worked with IBM GPFS, now IBM Spectrum Scale, for many years.

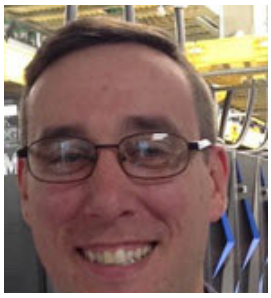
Ravindra Sure works for IBM India as a Senior System Software Engineer. He has worked on developing workload schedulers for high-performance computers, parallel file systems, computing cluster network management, and parallel programming. He has strong engineering professional skills in distributed systems, parallel computing, C, C++, Python, shell scripting, MPI, and Linux.

Ricardo D. Zamora Ruvalcaba is a Software Engineer at IBM Guadalajara, Mexico. He holds a bachelor's degree in Mechatronics. He started his career in IBM Spectrum Scale as a Test Automation Engineer where he learned about GNR/ECE. Currently, he works as developer in the IBM ESS Deployment team where he constantly implements new tools and technologies. He is also the focal point for Manufacturing Deployment problems. He has strong engineering professional skills in automation that uses various scripting technologies, such as Python, Ansible, shell scripting, Jenkins, the Robot framework, and Linux.

Robert Guthrie is a Senior Software Engineer in Austin, Texas. He works with the IBM Spectrum Scale development team on storage enclosures and NVMe. He joined IBM in 1996 working on CORBA-based enterprise management software. He is a software and systems solutions specialist with extensive expertise in networks, firmware, and middleware. He has had lead roles providing superior levels of design, development, test, and system integration functions for multiple projects serving large, international financial, insurance, and government enterprise clients. He has been working on storage products since 2008, including Information Archive, IBM Spectrum Protect, and IBM Elastic Storage Server.

Shradha Thakare is an advisory software engineer who works for IBM in Mumbai, India. She has been with IBM for 17 years. From the start of her career, she has worked in the storage domain, and her expertise is in Scale Out Network Attached Storage, IBM Spectrum Scale, and IBM ESS. Currently, she works in IBM ESS development but has worked extensively on support and customers engagements especially with Scale Out Network Attached Storage and IBM Spectrum Scale. Her key expertise is authentication and authorization, and she helps with security for IBM ESS. In the past, Shradha has co-authored four IBM Redbooks publications.

Stephen M. Tee is an IBM Elastic Storage Server and IBM GPFS Native RAID developer who is based in Austin, TX.



Steve Duersch worked on IBM Spectrum Scale for the past 19 years. He started his career with IBM as a tester before taking over management (project and people) duties. Currently, he is the program director for IBM ESS.

Sukumar Vankadhara is a Staff Software Engineer working in India who specializes in testing, IBM Spectrum Scale Native RAID, IBM Elastic Storage Server, and Erasure Code Edition Test.

Sumit Kumar is an Advisory Software Engineer at IBM India. He has a Master in Computer Application degree from IGNOU, New Delhi, India, and has 16 years of experience in the software development field. He has been part of IBM ESS deployment code development for IBM POWER8® and POWER9™ systems, including x86 servers. He also worked on IBM® Spectrum Scale as a deployment developer. Sumit has worked within the IBM for over 7 years, and previously held roles within the IBM Platform Computing and IBM Systems Director team. He has strong engineering professional skills in Software deployment and automation that use various scripting technologies, such as Python, shell scripting, Ansible, and Linux.

Todd M. Tosseth works as an IBM Spectrum Scale Develop and Test Engineer at IBM. His job responsibilities include testing the scalability and customer-like environments for IBM Spectrum Scale and IBM GPFS Data Protection Software development.

Van Smith works in the Storage Development organization focusing on reliability, availability, and serviceability (RAS) across various platforms. Previously, he was the Content Manager for Technical Training for IBM Systems Storage in the IBM Garage™ for Systems organization. He is a Certified Reliability Engineer. He has over 20 years with IBM, and has served in the subject matter expert (SME), program management, and managerial roles.

Vasfi Gucer is an IBM Technical Content Services Project Leader with IBM Garage for Systems. He has more than 20 years of experience in the areas of systems management, networking hardware, and software. He writes extensively and teaches IBM classes worldwide about IBM products. His focus has been primarily on cloud computing, including cloud storage technologies for the last 6 years. Vasfi is also an IBM Certified Senior IT Specialist, Project Management Professional (PMP), IT Infrastructure Library (ITIL) V2 Manager, and ITIL V3 Expert.

Wesley Jones serves as the test team lead for IBM Spectrum® Scale Native RAID. He also serves as one of the principle deployment architects for IBM Elastic Storage® Server. His focus areas are IBM Power Systems servers, IBM Spectrum Scale (IBM GPFS), cluster software (eXtreme Cluster Administration Toolkit (xCAT)), Red Hat Linux, networking (especially InfiniBand and Gigabit Ethernet), storage solutions, automation, and Python.

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

Doug Petteway, Erica Wazewski, Mamdouh Khamis, Rezaul Islam, Rodolfo Lopez, Wade Wallace, Amy Purdy Hirst
IBM US

Stefan Roth
IBM Germany

Ratan Swami
IBM India

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 IBM Redbooks® 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:

`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:

  `ibm.com/redbooks`

- Send your comments in an email to:

  `redbooks@us.ibm.com`

- Mail your comments to:

  IBM Corporation, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- 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

1

Introduction

This book describes the IBM Elastic Storage System 5000 (IBM ESS 5000) server in technical implementation detail. This chapter helps you evaluate, design, optimize, and deploy IBM ESS 5000 solutions.

Any IBM ESS is a storage building block in an IBM Spectrum Scale software-defined storage (SDS) cluster. If this book is your first encounter with IBM ESS, you might find a useful, high-level overview of IBM Spectrum Scale and the various models of IBM ESS in *Introduction Guide to the IBM Elastic Storage System*, REDP-5253.

This chapter introduces the technical aspects of the IBM ESS 5000 solution. It has the following sections:

- 1.1, “IBM Spectrum Scale RAID ” on page 2
- 1.2, “IBM Elastic Storage System” on page 5
- 1.3, “IBM Elastic Storage System 5000” on page 6
- 1.4, “License considerations” on page 7

1.1 IBM Spectrum Scale RAID

The IBM Elastic Storage Server family (of which the IBM ESS 5000 system is a member) is the IBM integrated solution for deploying IBM Spectrum Scale storage. Like all Elastic Storage Server systems, IBM ESS 5000 is the IBM deployment of IBM Spectrum Scale RAID erasure coding, which has proven over the past decade to provide high performance and high reliability at petabyte scale.

IBM Spectrum Scale RAID is more than erasure coding. IBM Spectrum Scale RAID is a complete SDS controller that is tightly integrated within the IBM Spectrum Scale file system. IBM Spectrum Scale RAID handles all aspects of controlling the IBM ESS 5000 physical storage media and physically storing and retrieving IBM Spectrum Scale data, including end-to-end data integrity and Disk Hospital media management. IBM Spectrum Scale RAID manages and mitigates physical media failures to provide consistent high performance at petabyte scale, even during physical media or hard disk drive (HDD) failures.

The IBM Spectrum Scale RAID software in the IBM ESS 5000 system uses Just a Bunch of Disks (JBOD) HDD drives. Because RAID functions are handled by the software, the IBM ESS 5000 system does not require an external RAID controller or acceleration hardware.

IBM Spectrum Scale RAID in the IBM ESS 5000 system supports two and three fault-tolerant RAID codes. The two-fault tolerant codes include eight data plus two parity, four data plus two parity, and 3-way replication. The three-fault tolerant codes include eight data plus three parity, four data plus three parity, and 4-way replication.

Figure 1-1 shows example RAID tracks consisting of data and parity strips.

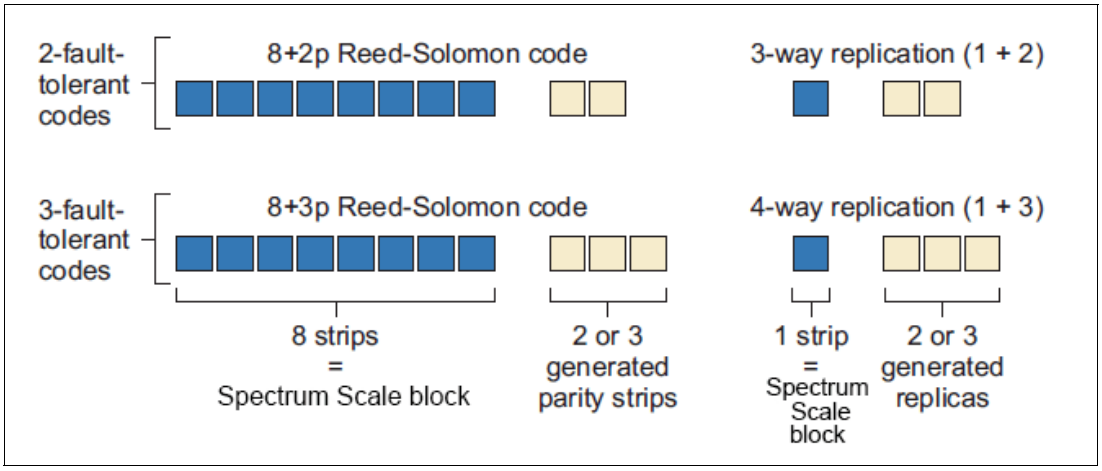


Figure 1-1 RAID tracks

1.1.1 Product history

In 2003, the Defense Advanced Research Project Agency (DARPA) started their High-Productivity Computing Systems (HPCS) program that is known as Productive, Easy-to-use, Reliable Computing System (PERCS). The IBM proposal for the DARPA HPCS project was what has become IBM Spectrum Scale RAID.

In 2007, IBM released the first market product that is based on IBM Spectrum Scale RAID: the P7IH. The system was based on the IBM POWER7® processor-based system and serial-attached SCSI (SAS) disks, and it delivered tens of gigabytes per second (GBps) of storage throughput.

Although the P7IH was, and still is, a fantastic engineering machine, in 2012 IBM released the IBM GPFS Storage Server (GSS) platform that was running what is known today as IBM Spectrum Scale RAID, but on commodity x86 hardware.

In 2014, IBM superseded the GSS with the first Elastic Storage System, which is based on the IBM POWER8 system, by using commercially available servers and disk enclosures while still being based on the same IBM Spectrum Scale RAID that was designed in 2003.

In 2019, IBM introduced the x86 based IBM ESS 3000. It uses Non-Volatile Memory Express (NVMe) drives to deliver superior performance.

IBM developed the IBM Spectrum Scale RAID technology on which IBM ESS 5000 is based from its beginning. IBM has a deep and unique understanding of this technology because it has been developing it for 17 years.

1.1.2 Distinguishing features

IBM Spectrum Scale RAID distributes data and parity information across node failure domains to tolerate unavailability or the failure of a server node. It also distributes spare capacity across nodes to maximize parallelism in rebuild operations.

IBM Spectrum Scale RAID implements end-to-end checksums and data versions to detect and correct the data integrity problems of traditional RAID. Data is checked from the pdisk blocks on the IBM ESS 5000 to the memory on the clients that connect over the network. It is the same checksum, not layers or serialized checksums that terminate in between the chain, so it really is an end-to-end checksum.

Figure 1-2 shows a simple example of a declustered RAID. The left side shows a traditional RAID layout that consists of three 2-way mirrored RAID volumes and a dedicated spare disk that uses seven drives. The right side shows the equivalent declustered layout, which still uses seven drives. Here, the blocks of the three RAID volumes and the spare capacity are scattered over the seven disks.

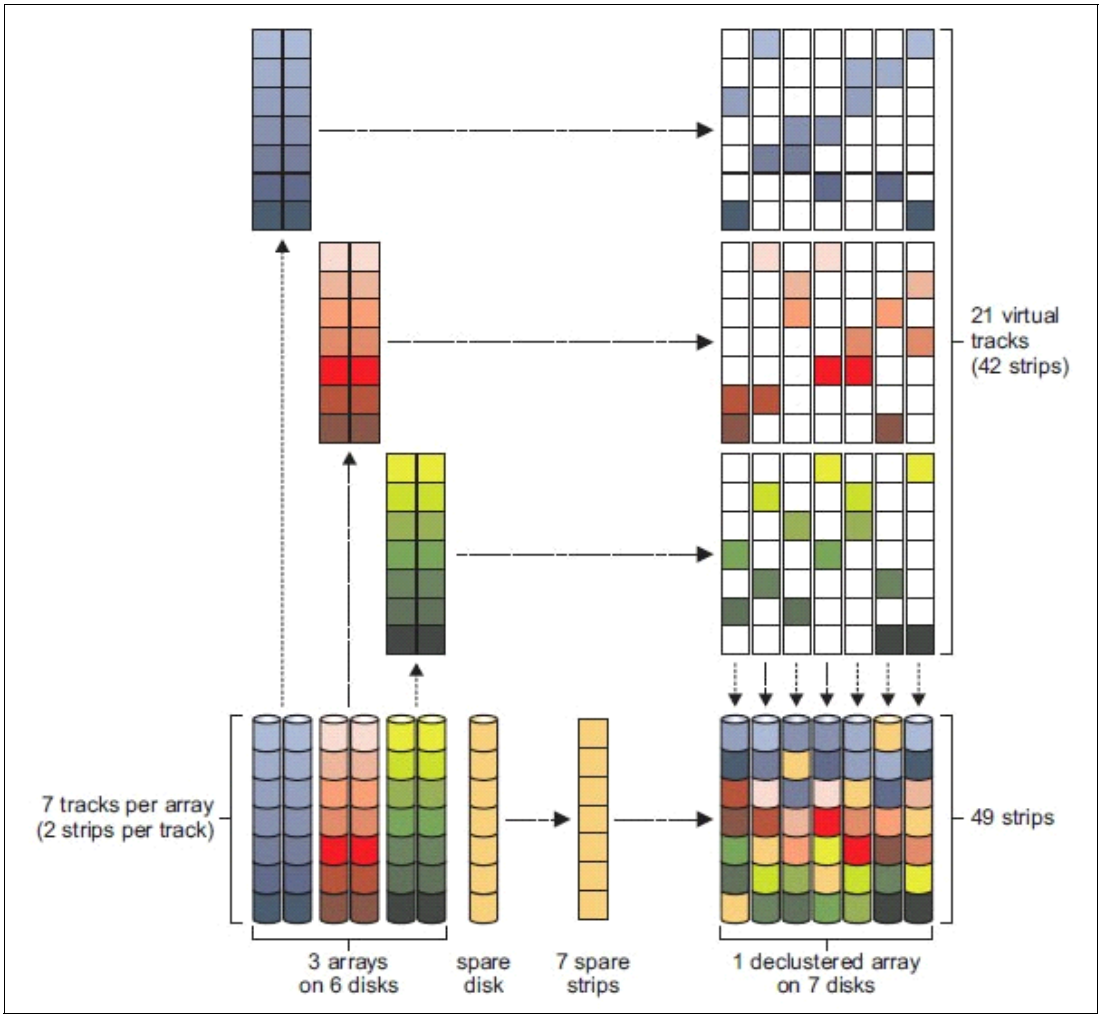


Figure 1-2 Declustered array versus 1+1 array

Figure 1-3 on page 5 shows a significant advantage of a declustered RAID layout over a traditional RAID layout after a drive failure. With the traditional RAID layout at the left of Figure 1-3 on page 5, the system must copy the surviving replica of the failed drive to the spare drive, reading only from one drive and writing only to one drive.

However, with the declustered layout that is shown at the right of Figure 1-3 on page 5, the affected replicas and the spares are distributed across all six surviving disks. This configuration rebuilds reads from all surviving disks and writes to all surviving disks, which greatly increase rebuild parallelism.

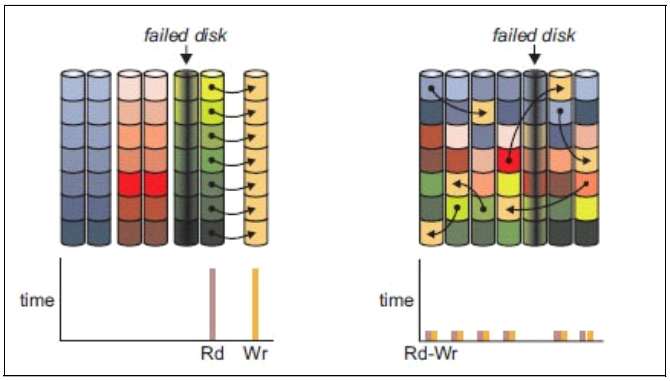


Figure 1-3 Array rebuild operation

Another advantage of the declustered RAID technology that is used by IBM ESS 5000 (and other IBM Elastic Storage Server systems) is that it minimizes the worst-case number of critical RAID tracks in the presence of multiple disk failures. IBM ESS 5000 can then deal with restoring protection to critical RAID tracks as a high priority while giving lower priority to RAID tracks that are not considered critical.

For example, consider an 8+3p RAID code on an array of 100 pdisks. In the traditional layout and declustered layout, the probability that a specific RAID track is critical is 11/100 × 10/99 × 9/98 (0.1%). However, when a track is critical in the traditional RAID array, all tracks in the volume are critical, but with a declustered RAID, only 0.1% of the tracks are critical. By prioritizing the rebuild of more critical tracks over less critical tracks, IBM ESS 5000 quickly gets out of critical rebuild and then can tolerate another failure.

IBM ESS 5000 adapts these priorities dynamically: If a *non-critical* RAID track is used and more drives fail, this RAID track's rebuild priority can be escalated to `critical`.

A third advantage of a declustered RAID is that it makes it possible to support any number of drives in the array and dynamically add and remove drives from the array. Adding a drive in a traditional RAID layout (except in the case of adding a spare) requires significant data reorganization and restriping. However, only targeted data movement is needed to rebalance the array to include the added drive in a declustered array.

1.2 IBM Elastic Storage System

IBM ESS is based on IBM Spectrum Scale RAID to provide the physical disk protection, and tightly integrated with IBM Spectrum Scale to provide the file system access over the network to all IBM Spectrum Scale clients.

A high-level overview of IBM ESS and IBM Spectrum Scale is presented in *Introduction Guide to the IBM Elastic Storage System*, REDP-5253.

Note: *Introduction Guide to the IBM Elastic Storage System*, REDP-5253 reviews how IBM ESS is an integrated part of an IBM Spectrum Scale storage cluster, and it is recommended reading to enhance the value that you receive from this book.

There are other protocols that can be used to access the IBM Spectrum Scale file system. For more information about how to access an IBM Spectrum Scale file system, see IBM Knowledge Center.

1.3 IBM Elastic Storage System 5000

The IBM ESS 5000 is the newest HDD-based IBM Spectrum Scale storage platform. This storage platform provides high-capacity and high-performance IBM Spectrum Scale storage by using HDD storage drives. IBM ESS 5000 combines high capacity with high performance, which are improvements compared to SAS-attached flash drives.

An IBM ESS 5000 system can contain 6 TB, 10 TB, 14 TB, or 16 TB HDD drives in various models.

To see the full collection of manuals and documentation for IBM ESS 5000, see IBM Knowledge Center.

Also, see *Introduction Guide to the IBM Elastic Storage System*, REDP-5253.

1.3.1 What is new in IBM ESS 5000

IBM ESS 5000 is based on IBM Spectrum Scale and IBM Spectrum Scale RAID, which is not unique because there are other IBM products that are also based on those features (all the other IBM ESS models and IBM Spectrum Scale Erasure Code Edition). However, there are some features that are new on the IBM ESS 5000, including the following prominent features:

- IBM POWER9 processor-based data servers, which provide the latest advances in IBM POWER® high memory and memory bandwidth, PCI Gen4-based internal bus transfer speeds, and fast network interface cards (NICs).
- Containerized Ansible playbooks that provide orchestration of complex tasks, such as cluster configuration, file system creation, and code update.
- Higher density and better HDD performance per rack space than any other IBM ESS that is available.

1.3.2 Added value

IBM ESS 5000 is designed to meet and beat the challenge of managing data for analytics. Packaged by using compact 4U106 or 5U92 storage HDD enclosures, IBM ESS 5000 is a proven data management solution that speeds time to value for artificial intelligence (AI), deep learning (DL), and high-performance computing (HPC) workloads because of its quick HDD storage and simple, fast containerized software installation and upgrade processes.

Its hardware and software design provides the industry-leading performance that is required to keep data-hungry processors fully used. IBM ESS 5000 is compatible with all IBM Elastic Storage Server models.

Fast time-to-value

IBM ESS 5000 combines IBM Spectrum Scale file management software with HDD storage for the ultimate in scale-out performance and simplicity to deliver up to 55 GBps of data throughput per IBM ESS 5000 system.

Operational efficiency

Containerized software installs and upgrades by using a powerful management GUI to minimize demands on IT staff time and expertise. Dense storage within a 4U106 or 5U92 storage enclosures means a small data center footprint.

Reliability

Software-defined IBM Spectrum Scale RAID erasure coding ensures data recovery while using less space than data replication. HDD rebuild impacts are mitigated. Worst case multiple HDD drive failure scenarios are resolved within minutes rather than hours or days. These mitigation technologies run without disrupting operations or data availability.

Deployment flexibility

The IBM ESS 5000 is available in a wide range of capacities, ranging up to petabytes per 42U rack. You can deploy it as a stand-alone system or scale out with extra IBM ESS 5000 systems, IBM ESS 3000 systems, or with previous generation IBM Elastic Storage Server systems.

1.4 License considerations

IBM ESS 5000 follows the same license model as the other IBM ESS products. The two available options are *Data Access Edition* (DAE) and *Data Management Edition* (DME).

IBM ESS uses capacity-based licensing, which means that a customer can connect as many clients as needed without extra license costs. For more information, see Capacity-based licensing.

For other types of configurations, contact IBM or your IBM Business Partner for license details.

For more information about IBM ESS 5000 pricing, see the IBM ESS 5000 announcement letter.

2

IBM Elastic Storage System 5000 architecture and technical overview

This chapter describes the architecture and provides an overview of IBM Elastic Storage System 5000 (IBM ESS 5000). It covers the following topics:

- 2.1, “Platform” on page 10
- 2.2, “Software enhancements” on page 18
- 2.3, “Enclosure overview: Sample reliability, availability, and serviceability enhancements” on page 24
- 2.4, “Reliability, availability, and serviceability features” on page 27
- 2.5, “Software-related RAS enhancements” on page 29
- 2.6, “Performance” on page 30
- 2.7, “GUI enhancements” on page 34

2.1 Platform

The IBM ESS 5000 is a high-capacity storage system that uses the IBM Spectrum Scale (IBM General Parallel File System (IBM GPFS)) file system. It combines the IBM Elastic Storage System 5105 on the IBM POWER9 architecture with the IBM Spectrum Scale software, which provides the clustered file system.

IBM ESS 5000 can be configured with the following hardware to create an enterprise-level solution:

- IBM ESS I/O node, protocol node, and IBM ESS Management Server (EMS) (5105-22E)
- Model 106 (5147-106) or Model 092 (5147-092) expansion enclosures
- 7965-S42 rack

The following types of servers are used in the IBM ESS 5000 system:

- EMS
- Protocol node
- I/O server

All these servers are based on the POWER9 architecture. All these server types have the same machine type and model (MTM) (5105-22E).

2.1.1 IBM ESS Management Server

The EMS manages and deploys the I/O servers and hosts the GUI. The specifications of the EMS server are as follows:

- One DD2.3 20 small cores, 190 W / 225 W, 2.5 GHz / 2.9 GHz
- 128 GB default memory, no non-volatile dual inline memory modules (NVDIMMs)
- No host bus adapters (HBAs)
- The same network interface card (NIC) / fabric options as the Network Shared Disk (NSD) server, which are C9, C6, C12, and C7
- Two SFF hard disk drives (HDDs), each with a capacity of 1.8 TB

Figure 2-1 shows the 5105-22E EMS.

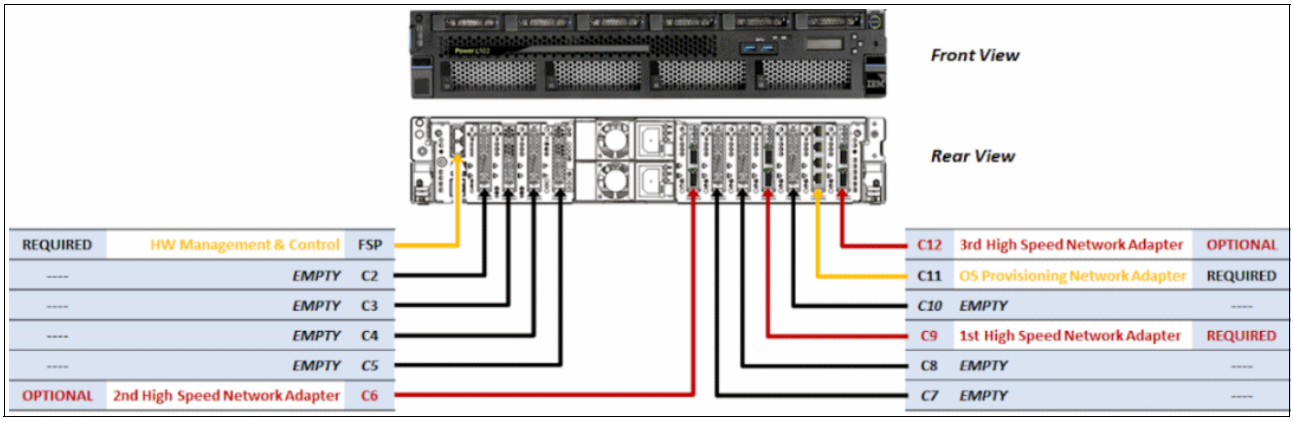


Figure 2-1 5105-22E IBM ESS Management Server

Table 2-1 shows the Management Node network adapters.

Table 2-1 PNSL101: Management Node network adapters

Plant order number	MTM	Serial number	Feature code	Quantity	Description
PNSL15	5105-22E	PN-10005	EC64	3	PCIe4 LP 2-Port 100 Gb EDR InfiniBand CAPI adapter
PNSL15	5105-22E	PN-10005	EL4M	1	PCIe2 LP 4-Port 1 GbE adapter

2.1.2 I/O node

The specifications of the I/O node are as follows:

- Processor: Two DD2.3 20 small cores, 190 W/ 225 W, 2.5 GHz / 2.9 GHz
- Memory:
  - Total of 32 DDR4 IS dual inline memory module (DIMM) slots
  - Six registered DIMMs (RDIMMs, also know as buffered DIMMs) per socket: 32 GB (default), 64 GB, or 128 GB at 2400 MHz, with 384 GB, 768 GB, or 1.5 TB capacities at 128 GBps
  - Two NVDIMMs per socket, 16 GB per at 2400 MHz logtip only, with 32 GB at 42.6 GBps interleaved per socket
  - Two small form factor (SFF) HDDs with a capacity of 1.8 TB each
- Storage:
  - EJ1F Solstice plus Fandango 8 SFF backplane
  - Mirrored SFF HDD, local boot only
  - Non-Volatile Memory Express (NVMe) not used
  - NVLink not used
- PCIe slots:
  - Five Gen3 X8 HBAs: Broadcom 9305-16e (FC ESA5) with C2, C6, C7, C8, and C12
  - Three Gen4 x16 CX5 adapters with C3, C9, and C4
  - One Gen3 x4 Austin Ethernet management with C11

Figure 2-2 shows the E 5105-22E IBM ESS Data Server.

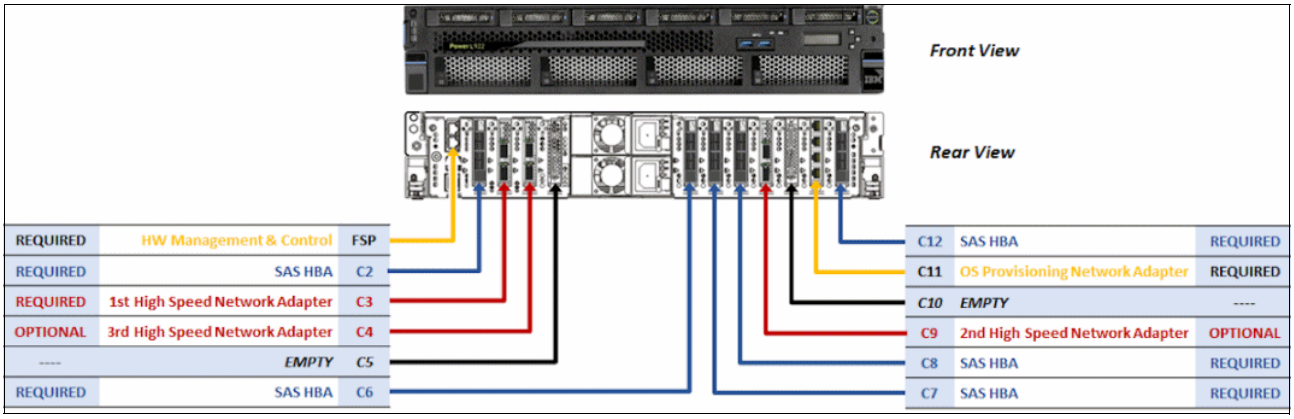


Figure 2-2 5105-22E IBM ESS Data Server

Table 2-2 shows the network adapters that are installed in each data server for this order (MFGNO PNSL101).

Table 2-2 Network adapters that are installed in each data server for this order (MFGNO PNSL101)

MTM	Feature code	Quantity	Description
5105-22E	EL4M	1	PCIe2 LP 4-port 1 GbE adapter
5105-22E	EC2T	2	PCIe3 LP 2-port 25/10 Gb NIC & RoCE SR/Cu adapter
5105-22E	EC67	1	PCIe4 LP 2-port 100 Gb RoCE EN LP adapter

Table 2-3 shows the serial-attached SCSI (SAS) adapters that are installed in each data server for this order (MFGNO PNSL101).

Table 2-3 SAS adapters that are installed in each data server for this order (MFGNO PNSL101)

MTM	Feature code	Quantity	Description
5105-22E	ESA5	5	LSI 9305-16 12 Gb HBA

Table 2-4 shows the list of data servers in this order (MFGNO PNSL101).

Table 2-4 List of data servers in this order (MFGNO PNSL101)

Plant order number	MTM	Serial number	Feature code	Quantity	Description
PNSL11	5105-22E	PN-10001	ESZY	1	IBM ESS Data Node Specify
PNSL12	5105-22E	PN-10002	ESZY	1	IBM ESS Data Node Specify

2.1.3 NVDIMMs

The IBM ESS 5000 I/O server node uses NVDIMMs for logtip to persist data across node restarts or power cycles. When power goes off, the contents of an NVDIMM is stored to persistent storage by using a Backup Power Module (IBM BPM). After the server is powered on, the data from persistent storage is restored to NVDIMM.

Figure 2-3 shows the top view of the I/O node with the IBM BPM modules and their cables for supporting NVDIMMs, for example, the P1-C22-T1-E1 IBM BPM module and P1-C22-T1 BPM cable for the NVDIMM In location P1-C22.

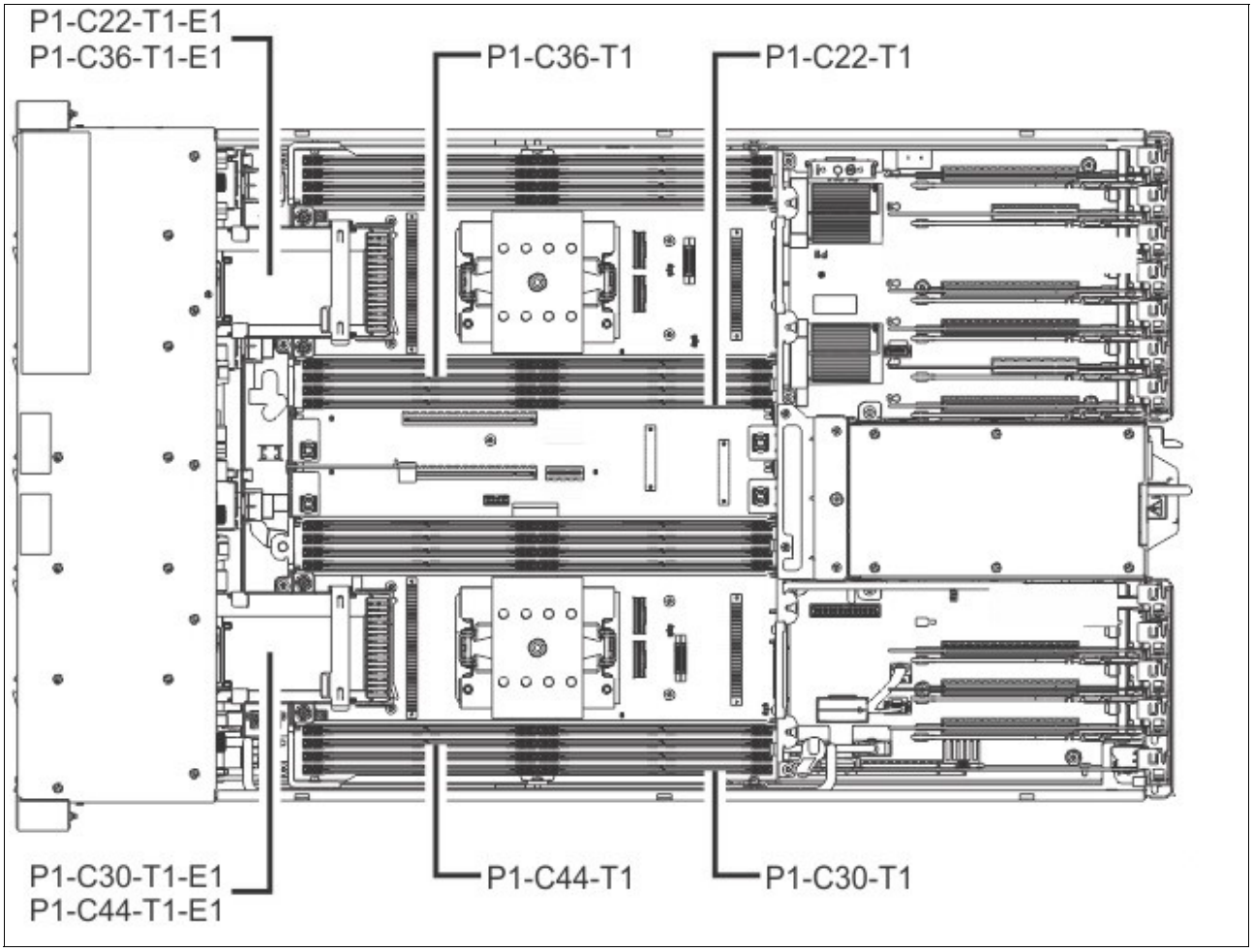


Figure 2-3 5105-22E IBM ESS Data Server top view with Backup Power Module (IBM BPM) and IBM BPM cable locations for the 5105-22E system

2.1.4 Protocol node

The IBM ESS access methods are like the ones for accessing an IBM Spectrum Scale cluster. Depending on the required configuration, the IBM ESS can be accessed by an IBM Spectrum Scale client, with a specific connector, or through dedicated protocol node servers. The specifications of the protocol node are as follows:

- Two DD2.3 20 small cores, 190 W / 225 W, 2.5 GHz / 2.9 GHz
- Default and minimum of 192 GB with no NVDIMMs
- Option for one DD3.2 proc with a minimum of 128 GB minimum memory
- Two SFF HDDs with a capacity 1.8 TB each
- Up to seven network adapters

2.1.5 IBM ESS 5000 variants

Based on the enclosure that is used in the IBM ESS 5000 system, the following variants of IBM ESS 5000 products are available:

- IBM ESS 5000 SC series
- IBM ESS 5000 SL series

IBM ESS 5000 SC series

The main features of the IBM ESS 5000 SC series are as follows:

- Uses the 5147-106 enclosures.
- Provides high capacity and performance.
- Uses up to six network adapters per IBM ESS system: 25G Ethernet, 100G Ethernet, and 100G InfiniBand.
- The available HDD options are 10 TB, 14 TB, and 16 TB.

Figure 2-4 shows the basic building block of the IBM ESS 5000 system with Model 106 expansion enclosures.

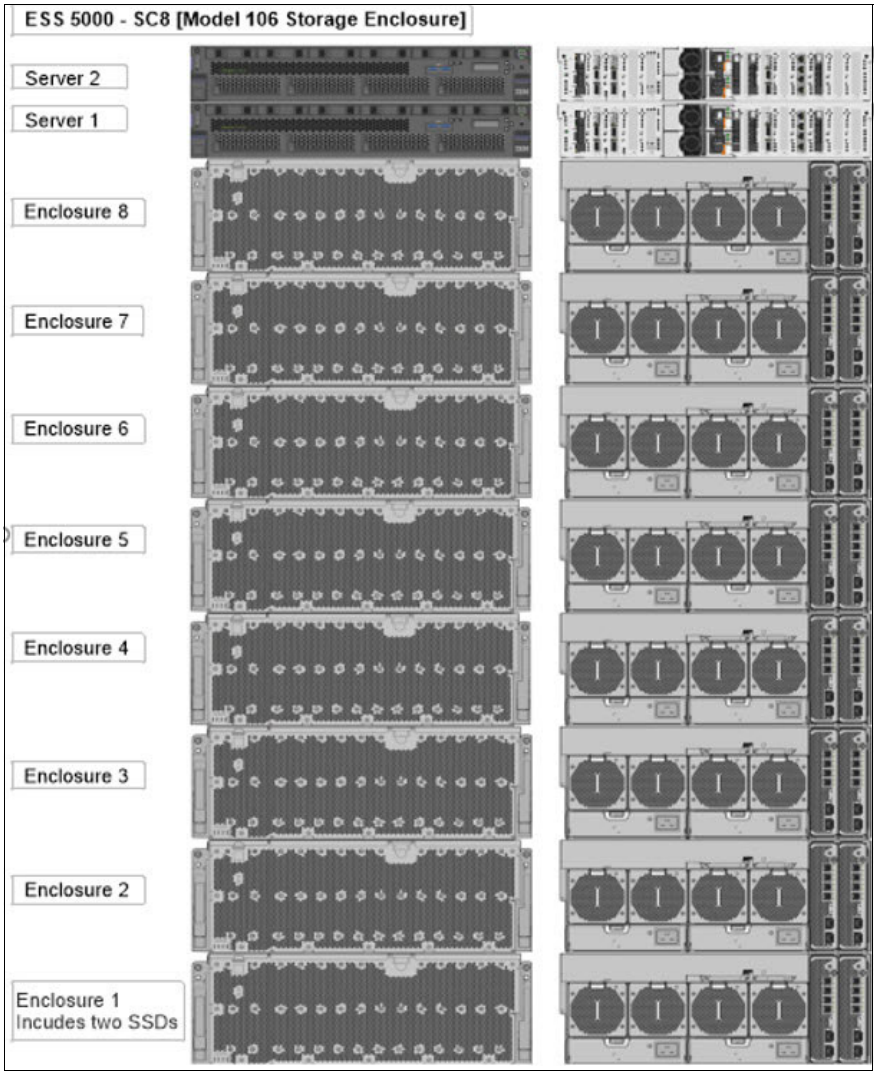


Figure 2-4 IBM ESS 5000 system with Model 106 expansion enclosures

2.1.6 IBM ESS 5000 SL series

The main features of the IBM ESS 5000 SL series are as follows:

- Uses the 5147-092 enclosures.
- Provides high capacity and performance.
- Uses up to six network adapters per IBM ESS system: 25G Ethernet, 100G Ethernet, and 100G InfiniBand.
- Four HDD options: 6 TB, 10 TB, 14 TB, and 16 TB.

Figure 2-5 shows the basic building block of the IBM ESS 5000 system with Model 092 expansion enclosures.

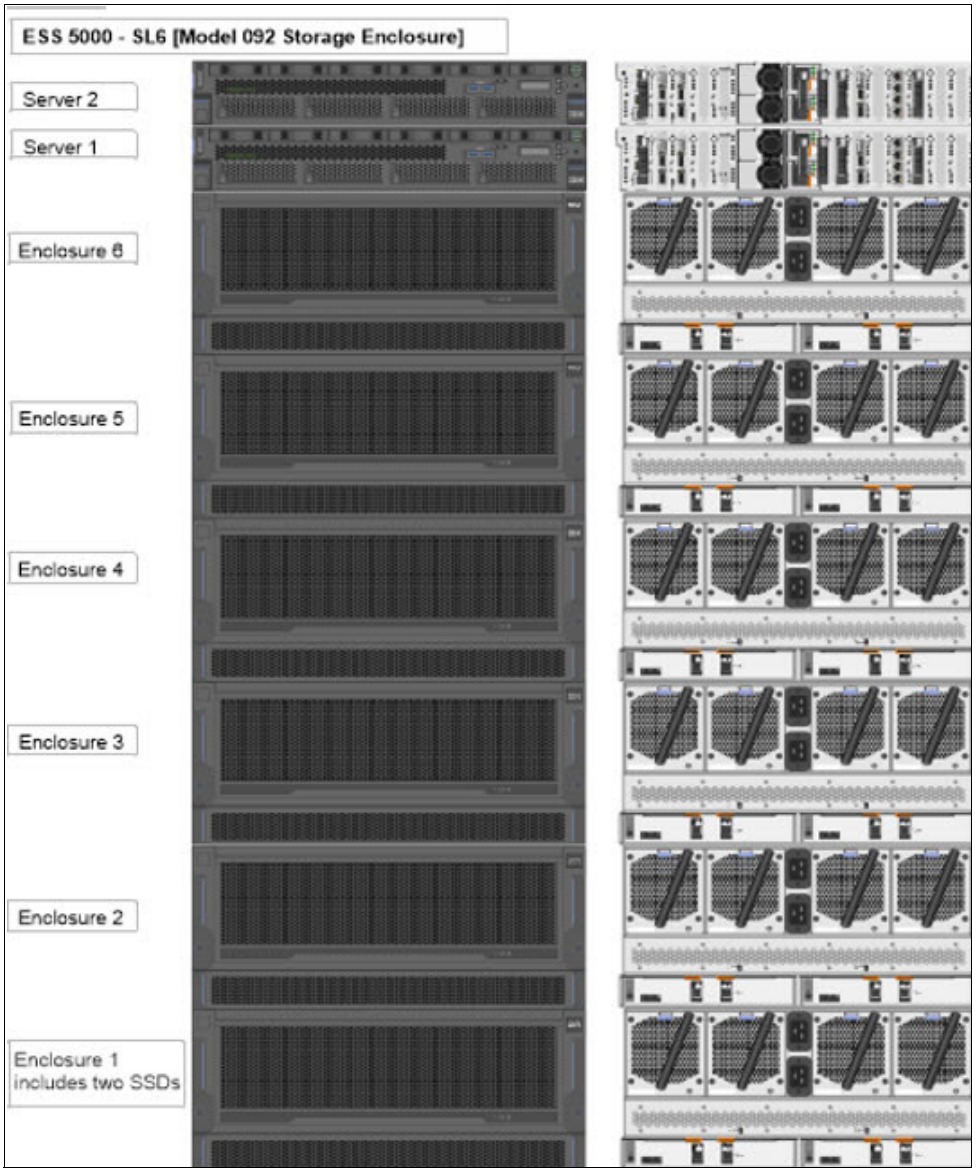


Figure 2-5 IBM ESS 5000 with Model 092 expansion enclosures

2.1.7 Enclosures

Enclosures are the storage expansion units that are mounted in the rack.

An IBM ESS 5000 system can have the following enclosures:

- IBM Elastic Storage System 5000 Expansion - Model 106 (5147-106): The IBM ESS 5000 system with this type of expansion enclosure is known as the IBM IBM ESS 5000 SC series.
- IBM Elastic Storage System 5000 Expansion - Model 092 (5147-092): The IBM ESS 5000 system with this type of expansion enclosure is known as the IBM ESS 5000 SL series.

IBM Elastic Storage System 5000 Expansion - Model 106 (5147-106)

The Model 106 has a 4U chassis. It holds up to 106 low profile (1-inch high) 3.5-inch form factor disk drive modules in a vertical orientation. Alternatively, disk slots can hold a low profile (5/8-inch high) 2.5-inch form factor disk with an adapter within the large form factor carrier. The enclosure is installed into a standard IBM rack, but needs extenders to allow the rear door to close.

Figure 2-6 shows the rear portion of the 5147-106 enclosure.

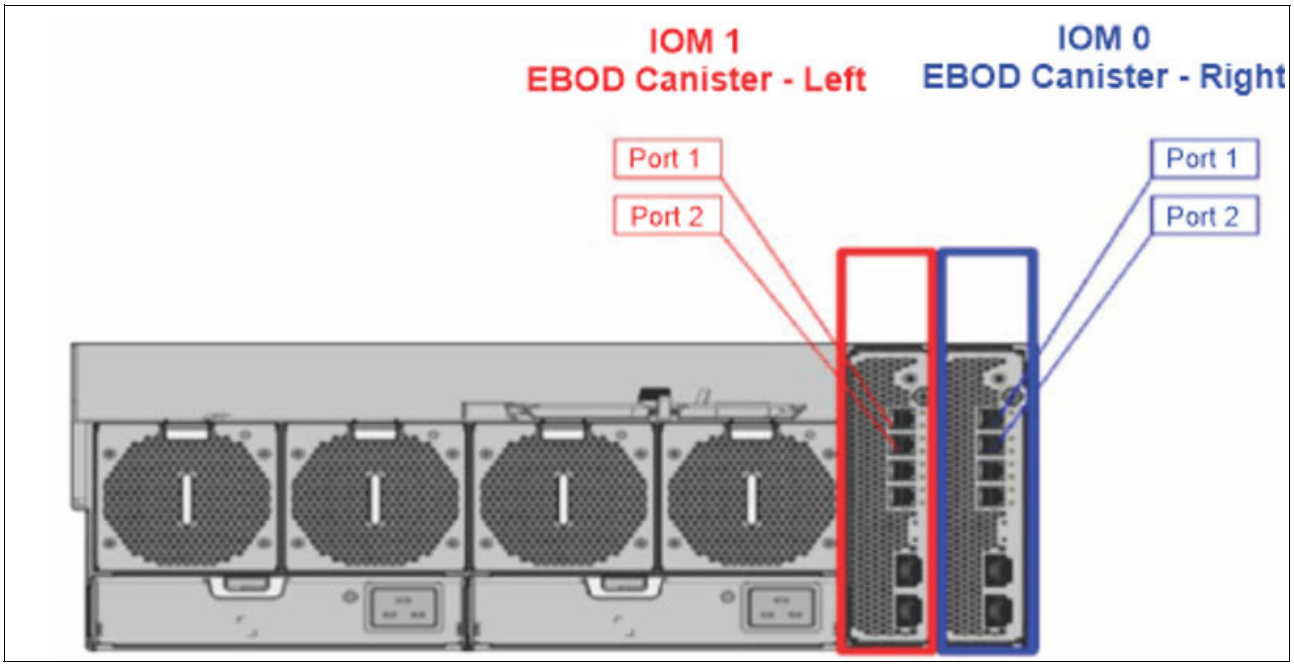


Figure 2-6 Model 106 expansion enclosure rear view

Figure 2-7 on page 17 shows disk locations in the 5147-106 expansion enclosure.

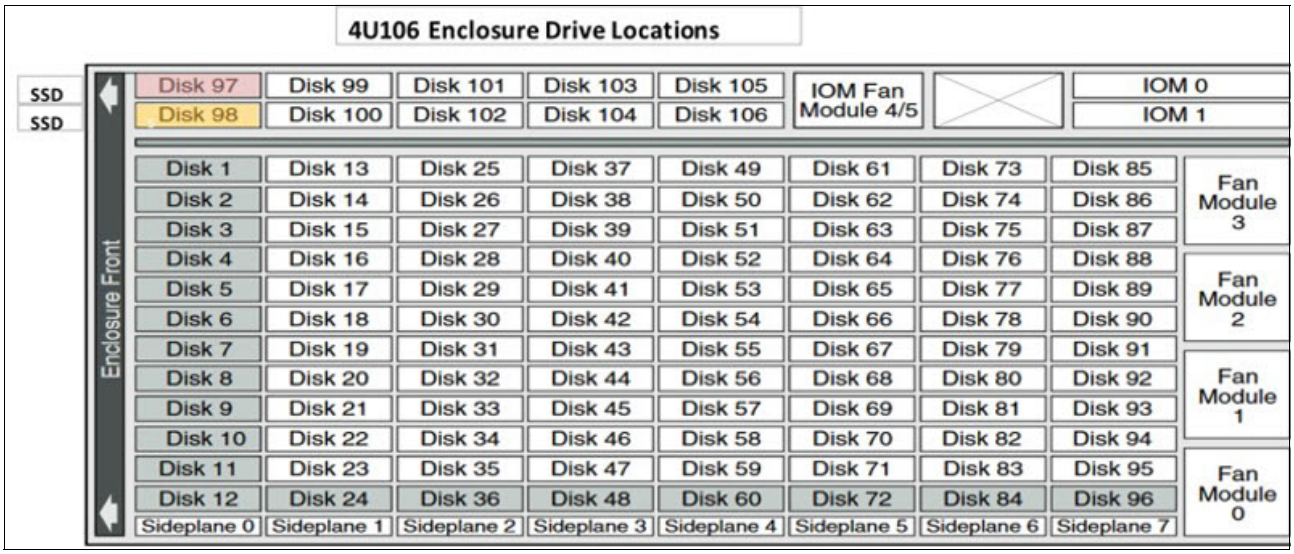


Figure 2-7 Model 106 expansion enclosure disk locations

IBM Elastic Storage System 5000 Expansion - Model 092 (5147-092)

The Model 092 enclosure is a high capacity and density expansion enclosure that is designed for use in cloud and enterprise environments. It holds up to ninety-two 3.5-inch SAS disk drives in a 5U, 19-inch rack mount enclosure.

Figure 2-8 shows the rear portion of the 5147-092 enclosure.

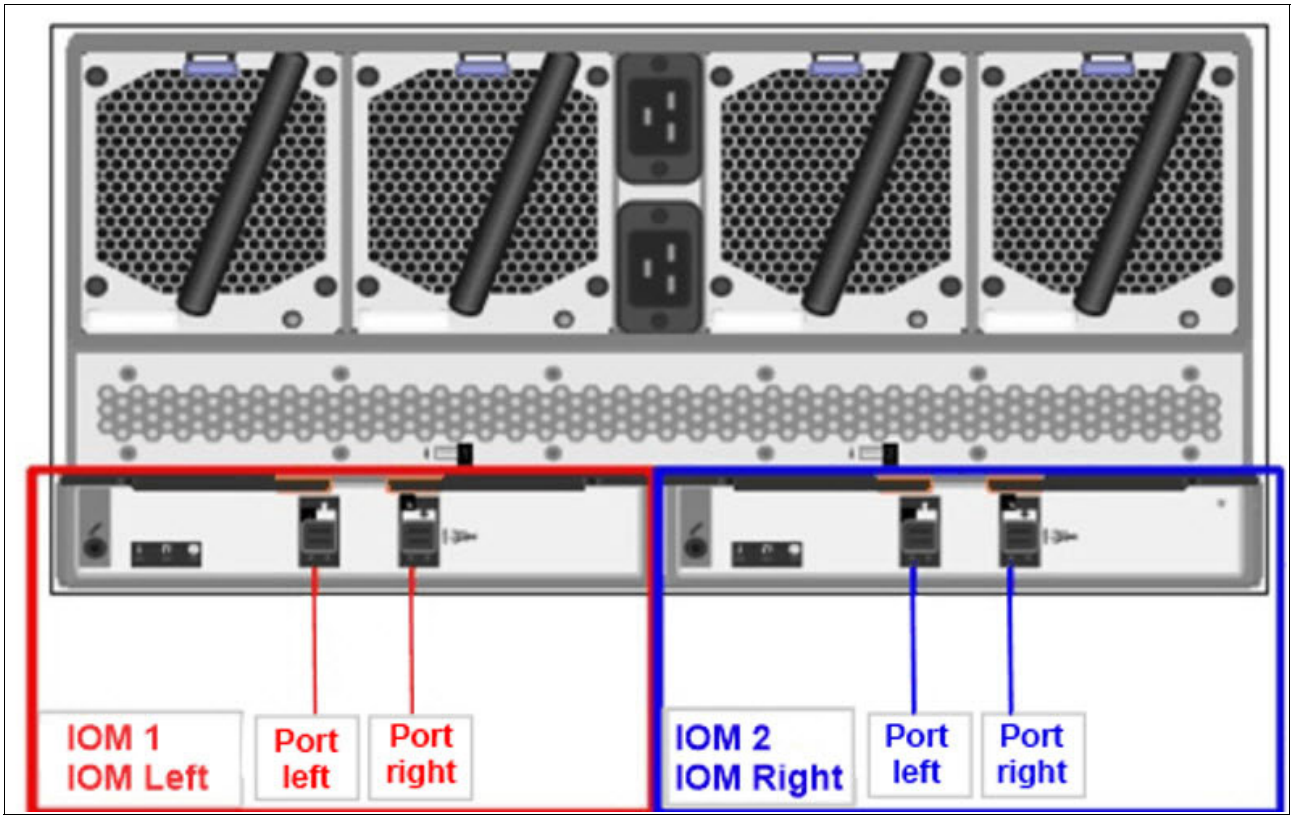


Figure 2-8 Model 092 expansion enclosure rear view

Figure 2-9 shows the disk locations in the Model 092 expansion enclosure.

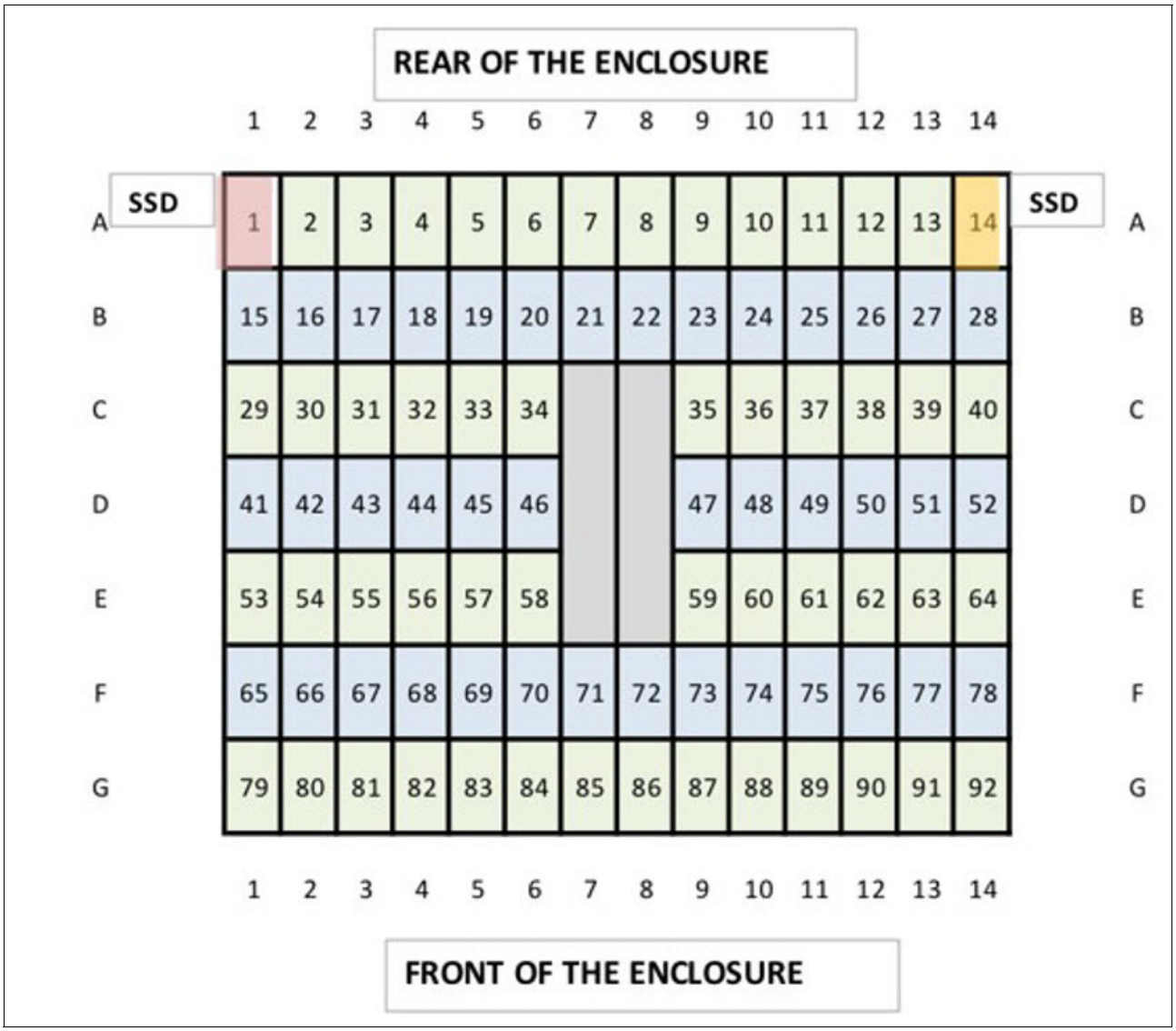


Figure 2-9 Model 092 expansion enclosure disk locations

2.2 Software enhancements

This section describes the software components that are used in the IBM ESS solution.

2.2.1 IBM ESS software solution stack overview

The IBM ESS solution provides an integrated and tested stack that bundles the operating system, adapter drivers, firmware, IBM Spectrum Scale software, management software, and installation scripts into a full IBM ESS software stack. This software solution stack is supported as an integrated IBM Spectrum Scale storage building block solution by IBM Service and Support.

Figure 2-10 on page 19 shows the IBM ESS 5000 high-level software architecture.

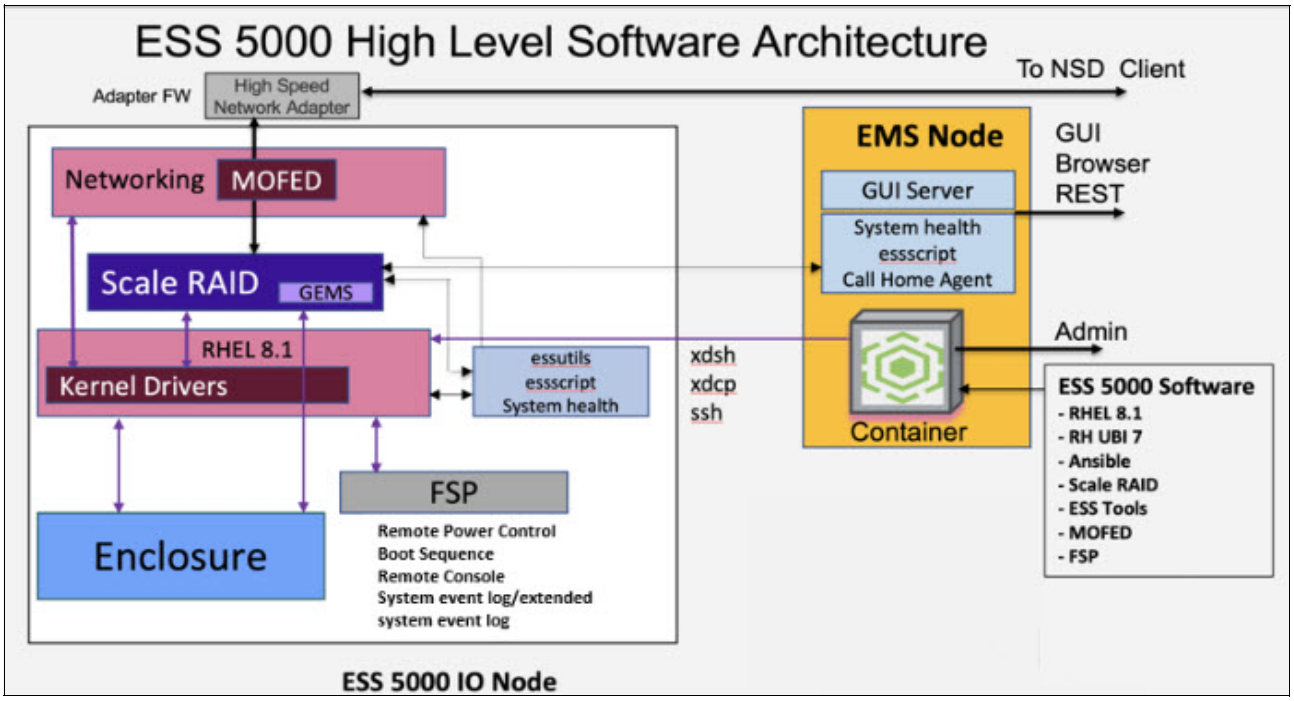


Figure 2-10 IBM ESS 5000 high-level software architecture

In the following sections, we briefly describe the following IBM ESS solution stack software components:

- Red Hat Enterprise Linux (RHEL) operating system.
- IBM Spectrum Scale high-performance parallel file system.
- IBM Spectrum Scale RAID software-defined storage (SDS).
- IBM ESS solution installation and management software, which includes, but is not limited to:
  - IBM ESS specific documentation for installation and upgrade scripts.
  - eXtreme Cluster Administration Toolkit (xCAT).
  - Other tools for the IBM Support Services Representative (IBM SSR) to use for installing IBM ESS, such as ESSUTILS.
  - The IBM ESS 5000 system features a container-based deployment model that focuses on ease of use.
- In addition, third-generation IBM ESS systems deploy a newer container-oriented management software stack in the EMS that includes Ansible playbooks for installation and orchestration.

IBM preinstalls this complete integrated and tested IBM ESS solution stack on the IBM ESS systems in IBM Manufacturing.

The IBM ESS solution stack levels are released as a version, release, and fix pack level.

For more information about the release levels of the IBM ESS software solution and the levels of the software components for that IBM ESS release level, see IBM Knowledge Center.

The IBM ESS solution stack components are periodically up-leveled, tested, and released as a new level of IBM ESS solution software. As a best practice, clients should plan to upgrade their IBM ESS to the current level of IBM ESS solution software stack at least once a year.

Operating system

The IBM ESS solution runs RHEL as the operating system on the IBM Spectrum Scale Data Servers.

Each IBM ESS solution release level integrates and tests a suitable current level of RHEL, including any necessary RHEL fixes and errata that are required for the successful operation of the IBM ESS solution stack. IBM regularly provides new IBM ESS solution release levels that incorporate newer levels of RHEL. These releases are provided often enough to ensure that a current level of RHEL is always available. Consider the following points:

- First-generation ESSs use RHEL Big Endian.
- Second generation and subsequent generation ESSs use RHEL Little Endian.
- All generations of IBM ESS can coexist in the same IBM Spectrum Scale cluster.

Embedded RHEL licensing starting with IBM ESS 5000 (including IBM ESS 3000)

Starting with the IBM ESS 5000 and including the IBM ESS 3000 and subsequent IBM ESS generations, IBM moved to an embedded RHEL license. This change greatly streamlines the client experience and makes install and deploy much easier. For this newest generation IBM ESS, clients no longer need a separate RHEL subscription for the RHEL on IBM ESS because IBM handles all RHEL subscriptions, licensing, and support.

IBM Spectrum Scale

In this section, we describe IBM Spectrum Scale, the high-performance parallel clustered file system software that is used in IBM ESS and is proven in enterprise-level high-performance computing (HPC) environments, university and research environments, and commercial business environments worldwide, including finance and healthcare. IBM Spectrum Scale runs on many of the world's largest supercomputers, including the 2018 Sierra and Summit supercomputers.

Formerly known as IBM GPFS, IBM Spectrum Scale is highly distributed, clustered file system software that provides high-speed (HS) concurrent data access to applications that run on multiple nodes and clusters. In addition to providing parallel high-performance file storage capabilities at petabyte scale, IBM Spectrum Scale provides tools for tiering, management, administration, and archiving of enterprise-level data. IBM Spectrum Scale is the IBM strategic SDS for enterprise big data, analytics, and artificial intelligence (AI) applications.

In the most common IBM Spectrum Scale deployment architecture, IBM Spectrum Scale data is accessed by IBM Spectrum Scale clients and users over a LAN network, accessing disk volumes that are known as NSDs that are attached to IBM Spectrum Scale nodes known as *NSD Data Servers*. In this paper, these nodes also are referred to as *Data Servers*.

IBM Spectrum Scale and IBM ESS

IBM Elastic Storage Systems and IBM ESS are pre-integrated, pre-tested IBM Spectrum Scale storage building blocks. All models of IBM ESS consist of a pair of servers with cross-connected attached IBM storage.

These servers run RHEL and IBM Spectrum Scale, and are defined as IBM Spectrum Scale NSD Data Servers. In all ESSs, the NSD Data Servers are cross-connected to provide failover and redundancy. If one of the NSD Data Servers fails, the IBM ESS fails over the storage and data to the other NSD Data Server, which ensures continued availability of the data in the IBM Spectrum Scale cluster.

IBM Spectrum Scale for IBM ESS licensing

All nodes in an IBM Spectrum Scale cluster run a copy of the IBM Spectrum Scale software. Instead of using the normal IBM Spectrum Scale software licensing, an IBM ESS typically is licensed for IBM Spectrum Scale by using the following specific IBM Spectrum Scale for IBM ESS IBM program IDs:

- 5765-DAE IBM Spectrum Scale for IBM ESS Data Access Edition (DAE)
- 5765-DME IBM Spectrum Scale for IBM ESS Data Management Edition (DME)

This specific IBM Program ID with a “Per Disk” metric is normally used for licensing IBM Spectrum Scale for IBM ESS software on an IBM Elastic Storage Server. IBM Spectrum Scale for IBM ESS software licenses include IBM Spectrum Scale RAID license entitlement. The license price for IBM Spectrum Scale for IBM ESS is tiered as solid-state drives (SSDs), and HDDs have different list prices per terabyte. You need to count only the number of SSDs or HDDs for the “Per Disk” metric.

An advantage of the IBM Spectrum Scale for IBM ESS “Per Disk” metric is that the size of the SSD or HDD does not affect the license list price. For example, if your IBM ESS model has 502 HDDs, your IBM Spectrum Scale for IBM ESS license list price is the same, regardless of whether you are specifying 6 TB HDDs or 14 TB HDDs.

IBM Spectrum Scale for IBM ESS licensing helps to contribute in building a complete IBM hardware and software solution by integrating the IBM Spectrum Scale and IBM ESS solution.

For more information about IBM Spectrum Scale and IBM Spectrum Scale for IBM ESS software licensing, options, and considerations (such as the use of IBM ESS in a Socket-licenses IBM Spectrum Scale cluster, or for IBM Spectrum Scale capacity licenses, or for using IBM ESS in an IBM Spectrum Scale Enterprise License Agreement environment), see the following resources:

- IBM Knowledge Center
- IBM Spectrum Scale: IBM Elastic Storage System - Licensing Information

2.2.2 IBM Spectrum Scale RAID

IBM Spectrum Scale RAID is an SDS controller that performs all of the storage controller functions that are normally associated with hardware storage controllers. IBM Spectrum Scale RAID integrates all high availability (HA) and functions of an advanced storage server into IBM Spectrum Scale SDS.

IBM Spectrum Scale RAID runs on the IBM ESS NSD data servers. IBM Spectrum Scale RAID provides sophisticated data placement and error correction algorithms to deliver high-levels of storage reliability, availability, and serviceability (RAS), and performance.

IBM Spectrum Scale RAID implements a declustered erasure code parity schema, distributing data, redundancy information, and spare space across all disks of the IBM ESS enclosures. With this approach, a significant improvement is realized on the application performance, and storage rebuild time overhead is reduced (disk failure recovery process) compared to conventional RAID controllers.

IBM Spectrum Scale RAID implements a large cache for performance by using memory on the IBM ESS NSD Data Servers. The large cache intelligently improves read and write performance, particularly for small block I/O operations.

IBM Spectrum Scale RAID mitigates performance impacts of storage rebuilds

If storage failures occur, IBM Spectrum Scale RAID reconstructs lost or erased stripes for I/O operations dynamically. By using the highly distributed erasure coding, IBM Spectrum Scale RAID mitigates the performance affects of storage failures.

IBM Spectrum Scale RAID end to end checksums

IBM Spectrum Scale RAID includes integrated end-to-end checksums that detect data corruption that might otherwise go undetected by a conventional storage controller. Unlike conventional storage controllers, IBM Spectrum Scale RAID is integrated with the IBM Spectrum Scale file system and performs end-to-end checksum comparison all the way to the IBM Spectrum Scale client code on the workstations. This feature ensures data integrity at a file system level, detecting and automatically correcting data corruption errors that might occur in conventional storage environments.

In an environment where a customer experienced excessive file system checks and suffered downtime to repair file systems, using IBM Spectrum Scale RAID end-to-end checksums mitigates file system check problems. This feature assures availability of data and removes application outages that are caused by file system checks.

IBM Spectrum Scale RAID disk hospital

One of the key features of IBM Spectrum Scale RAID is the *disk hospital*. This powerful function asynchronously diagnoses errors and faults in the IBM ESS storage media, down to the level of the individual drive and the individual performance of each drive. IBM Spectrum Scale RAID is fully aware of and tracks the performance of each individual drives because all drives do not perform equally. IBM Spectrum Scale RAID uses the individual performance history of each drive to make intelligent data allocation and data retrieval decisions.

Extensive health metrics down to the level of the individual drive are maintained by the disk hospital. Performance variation is continually monitored. If or when a disk metric exceeds a threshold, the storage media is marked for replacement according to the disk maintenance replacement policy for the declustered array.

As an example, the disk hospital features the following metrics:

- *relativePerformance*, which characterizes response times. Values are compared to the average speed. If the metric falls below a particular threshold, the hospital adds “slow” to the pdisk state, and the disk is prepared for replacement.
- *dataBadness*, which characterizes media errors (hard errors) and checksum errors. This disk is then marked for replacement.

IBM Spectrum Scale RAID commandless disk replacement

Another feature of IBM Spectrum Scale RAID is command-less disk replacement. If a drive fails or finds some errors, the disk hospital begins moving data off that drive. After data is drained, the disk hospital marks the drive as replaceable and turns the LED ON of the drive to indicate that the drive is ready for replacement. After a user identifies the bad drive that must be replaced, the user can remove the bad drive and insert a new drive. The disk hospital automatically identifies that a new drive is inserted in place of the bad drive and makes the new drive ready for use.

For more information about IBM Spectrum Scale RAID implementation and best practices, see the *Administering IBM Spectrum Scale RAID* manual at IBM Knowledge Center.

2.2.3 IBM ESS solution installation and management scripts

In this section, we provide an overview of the components of the IBM ESS solution installation and management scripts. This overview includes the following information:

- IBM ESS-specific documentation for installation and upgrade scripts.
- xCAT.
- IBM ESS 5000 contains a POWER9 EMS node and runs a Podman container. For more information, see *IBM ESS 3000 common setup instructions* at IBM Knowledge Center.
- IBM ESS Installation and Deployment Toolkit (ESSUTILS) for the IBM SSRs and administrators to use while installing and maintaining IBM ESS.

The IBM ESS 5000 enhances these installation and management tools with newer container-based methods, including Ansible playbooks for installation and orchestration.

For more information about these IBM ESS solution components, see the IBM ESS solution release-specific level information at IBM Knowledge Center.

Installing, upgrading, and administering guides

IBM provides manuals and documentation for deploying and administering IBM ESS, including the following publications:

- *Quick Deployment Guide*, which documents IBM ESS specific scripts for installing, deploying, and upgrading IBM ESS for experienced users.
- *IBM Spectrum Scale RAID Administration*, which focuses on administering IBM Spectrum Scale RAID on the IBM ESS.
- *Problem Determination Guide*, which provides more information about monitoring, troubleshooting, and maintenance procedures.

These documents can be accessed at IBM Knowledge Center.

eXtreme Cloud Administration Tool

The xCAT is open source distributed computing management software that was originally developed by IBM. It is commonly used for the deployment and administration of HPC Linux-based clusters. xCAT is included with each IBM ESS 5000 installation. It is used in installation scripts, but is not client visible.

For more information, see xCAT.

As part of every IBM ESS order, IBM includes the IBM Support Line for xCAT offering, which allows IBM Support to service any solution issues that are related to open source xCAT.

The IBM Program ID for this Support Line offering is 5641-CTx, where “x” is 1, 3, or 5 years.

ESSUTILS

ESSUTILS are IBM ESS installation and deployment toolkits that are primarily designed to facilitate IBM SSR hardware setup, installation, and deployment, and upgrade tasks. Any authorized IBM ESS system administrator also can use these tools. The ESSUTILS main menu is shown in Figure 2-11.

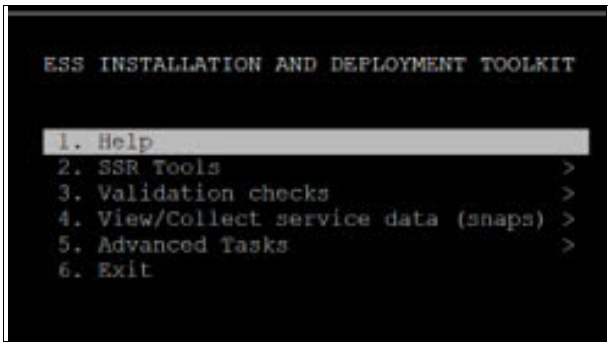


Figure 2-11 ESSUTILS main menu

ESSUTILS provides a set of task menus that are related to installation and deployment activities. When a task is selected from the menu, a command is issued to the system for that task. ESSUTILS can be run on the EMS node only.

For more information about ESSUTILS, see the *Quick Deployment Guide* and the *IBM ESS Command Reference*, which are available at IBM Knowledge Center.

2.3 Enclosure overview: Sample reliability, availability, and serviceability enhancements

IBM ESS 5000 is an extension of the Elastic Storage Server product family. IBM ESS 5000 is a high-capacity storage system that uses the IBM GPFS file system. It combines the IBM Elastic Storage System 5105 on the IBM POWER9 architecture with the IBM Spectrum Scale software, which provides the clustered file system. The IBM ESS 5000 MTMs include the 5105-22E server and 5147-106/5147-092 I/O expansion drawers.

IBM ESS 5000 is targeted at delivering the following key traits in an appliance-like customer experience:

- Disk hospital.
- Tightly coupled integrated unit.
- All hardware and software pieces are tightly coupled together.
- Everything is packaged together.
- Part of the overall IBM ESS product portfolio.
- Designed to coexist and interoperate with the overall IBM ESS and IBM Spectrum Scale infrastructure.

IBM ESS 5000 is a customer setup product with a combination of customer-replaceable units (CRUs) and field-replaceable units (FRUs).

2.3.1 Enclosure overview

The IBM ESS 5000 is available in two models, which are based on the model of the attached storage expansion enclosure:

- IBM ESS 5000 SC models use the 5147-106 storage enclosure.
- IBM ESS 5000 SL models use the 5147-092 storage enclosure.

IBM ESS enclosure 5147-106 (4U106)

An IBM ESS 5000 system with this type of expansion enclosure is known as the IBM ESS 5000 SC series. The IBM ESS enclosure 5147-106 supports a 4U (rack space) chassis. It holds up to 106 low profile (1 inch high) 3.5" form factor disk drive modules in a vertical orientation. Alternatively, disk slots can hold a low profile (5/8-inch high) 2.5" form factor disk with an adapter within the large form factor carrier.

IBM ESS enclosure 5147-092 (5U92)

The IBM ESS 5U92 storage enclosure introduces an architecture that reuses I/O modules, a common management API, and EBOD SBB 2.0 compatibility. This 5U enclosure is a dense storage device that delivers the highest bandwidth and scalability for the IBM ESS family to meet the needs of customers where capacity is paramount. It holds up to 92 low profile (1 inch high) 3.5" form factor disk drive modules in a vertical orientation. Alternatively, disk slots can hold a low profile (5/8-inch high) 2.5" form factor disk with an adapter within the large form factor carrier.

Figure 2-8 on page 17 shows the Model 092 expansion enclosure rear view.

Machine type and model and warranty

The MTM 5147 I/O Expansion units are rebranded I/O Expansion units. These units are installed by the IBM SSRs. The terms of the warranty are a 1-year 9x5 Next Business Day (NBD) plus 2-year warranty extension. Software installation and configuration are addressed by Lab Based Services (LBS) through a contract agreement. The repair strategy Tier 1 CRU is mandatory for the client unless there is a Warranty System Upgrade (WSU). Hardware upgrades are covered by the IBM SSRs. Firmware and microcode updates are the client's responsibility. Software stack updates are the responsibility of the client or LBS through a contract.

For the Power Servers MTMs (5148 IBM ESS Power Server, 5104 IBM ESS POWER9 LC Server, and 5105 IBM ESS POWER9 Scale Out), the terms are essentially the same. The IBM SSRs are responsible for installation. The terms of the warranty are a 1-year 9x5 NBD plus 2-year warranty extension. Software installation and configuration are addressed by LBS through a contract agreement. Hardware upgrades are covered by the IBM SSRs. Firmware and microcode updates are the client's responsibility. Software stack updates are the responsibility of the client or LBS through a contract. The repair strategy Tier 1 CRU is mandatory for the client unless there is a WSU for MTMs 5148 and 5105.

Table 2-5 shows the MTMs and warranties.

Table 2-5 Machine type and model and warranty

Item	5147 IBM ESS I/O Expansion	5148 IBM ESS Power Server	5105 IBM ESS POWER9 Scale Out
Warranty	3-year 9x5 NBD plus 2-yr warranty extension	3-year 9x5 NBD plus 2-year warranty extension	3-year 9x5 NBD plus 2-year warranty extension
Hardware installation	IBM SSR	IBM SSR	IBM SSR
Software installation and configuration	Client, IBM LBS (contract), or IBM Business Partner	Client, IBM LBS (contract), or IBM Business Partner	Client, IBM LBS (contract), or IBM Business Partner
Repair	On-site service repair limited (IOL) standard (Tier 1 CRU - client mandatory unless WSU)	IOL standard (Tier 1 CRU - client mandatory unless WSU)	IOL standard (Tier 1 CRU - client mandatory unless WSU)
Hardware upgrade	IBM SSR	IBM SSR	IBM SSR
Firmware and uCode updates	Client	Client	Client
Software stack updates	Client pr LBS (Contract) or IBM Business Partner	Client or LBS (Contract) or IBM Business Partner	Client or LBS (Contract) or IBM Business Partner

Components: FRU versus CRU

Here are the components of the 5U92 - 5147-092 I/O expansion drawer:

- CRUs:
  - 3.5" HDD carriers (supports 2.5" HDD)
  - Secondary expansion module (SEM)
  - Control panel board
  - Bezels
  - Power supplies
  - Power cables
  - Fan modules
  - Expansion canister
- FRUs:

  Main chassis with drive board and fan interface board (FIB)

Potential for SSD data loss after extended shutdown and best practices:

During the unusual events of 2020, some of our customers might be powering off their systems for extended periods. The Joint Electron Device Engineering Council (JEDEC) spec for Enterprise SSD drives requires that the drives retain data for a minimum of 3 months at 40 °C. So, after 3 months of a system being powered off in an environment that is at 40 °C or less, there is a potential of data loss or drive failures. This power-off time limitation is due to the physical characteristics of flash SSD media's gradual loss of electrical charge over an extended power-off period. There is a potential for data loss or flash cell damage, which might lead drive failures.

You should follow these best practices:

- Always maintain good backups, especially before extended shutdowns.
- A system (and its enclosed drives) should be powered on at least 1 week every 6 weeks of system power-off.
- Proper environmental control procedures should still be in place to ensure systems are experiencing less than 40 °C always, even if the systems are powered-off.
- If a system is being retired and used for another activity in the future, reformat all drives in the system. This action helps prevent SSD failure upon power-up, even if the data is not needed.

2.4 Reliability, availability, and serviceability features

The major RAS features are as follows:

- Call home
- Health and performance monitoring
- Disk hospital
- Concurrent maintenance
  - Failover
  - HDD/SSD
  - Redundant cooling
  - Redundant power supply: One plus one 1400 W power supply, 200 - 240 VAC
- Just a Bunch of Disks (JBOD) drives and select components
- Enclosure only with enclosure protection

2.4.1 Monitoring IBM ESS 5000 health

The **mmhealth** command provides a CLI for the health monitoring framework of IBM Spectrum Scale. This CLI can be used to query the state of different logical components, services, and events that are reported by the system.

Here are examples of services that are monitored by running **mmhealth**:

- Network (listed as NETWORK)
- File system (listed as FILESYSTEM)
- Node health (listed as NODE)
- IBM Spectrum Scale RAID (listed as NATIVE_RAID)

Here are examples of IBM ESS 5000 events:

- `expander_failed`: An enclosure expander failed.
- `enclosure_needsservice`: An enclosure must be serviced.
- `gnr_array_needsservice`: A declustered array needs service.
- `gnr_pdisk_missing`: A pdisk is marked missing.

All these health events are viewable from the IBM Spectrum Scale GUI by opening the Monitoring Events window. The GUI helps organize events by multiple occurrences, filters events by severity, marks notices read and unread, and lets the administrator address outstanding issues by running defined procedures from an **Actions** menu.

For more information about IBM ESS health monitoring, see IBM Knowledge Center.

2.4.2 Monitoring the IBM ESS performance

The IBM ESS 5000 is equipped with the suite of performance monitoring tools that come standard with all IBM Spectrum Scale versions. Performance can be monitored with either the **`mmperfmon`** or **`mmpmon`** commands through the GUI, or by integrating with the open source tool Grafana. These tools allow the administrator to examine key performance indicators and troubleshoot performance problems.

Note: In this section, we describe monitoring IBM ESS performance in IBM ESS 5000. In 2.6, “Performance” on page 30, we describe some performance enhancement tips.

The **`mmpmon`** command collects I/O performance statistics from the perspective of the IBM GPFS servicing client requests. It can track load distribution, volume, service times, and I/O patterns across one or more nodes.

IBM ESS 5000 V6.0.1 first released with an IBM Spectrum Scale version that is based on of Version 5.0.5.1, which allows performance monitoring to be configured and managed automatically by using the **`mmperfmon`** command. The **`mmperfmon`** command is used to monitor and collect more detailed performance data running at or below the IBM GPFS software stack, including the network, Network File System (NFS), Server Message Block (SMB), CPU usage, and memory usage. The **`mmperfmon`** command lets the user configure performance collector and sensor nodes and query individual performance metrics. A *sensor* is a component that collects a specific performance metric, and there are typically multiple sensors that are configured on a node. By default, sensors are started on every node. A *collector* is a node that aggregates data from a group of sensors, and there may be more than one collector that is specified.

The IBM Spectrum Scale GUI can interface with and manage IBM Spectrum Scale internal performance monitoring tools. The **`mmperfmon`** command should be used to configure the GUI node for performance monitoring. As part of this configuration, the GUI node must be specified as one of the clusters performance collector nodes. On IBM ESS 5000, the EMS node hosts the GUI service, meaning that the EMS node is specified as a performance collector.

For more information about configuring performance monitoring for the IBM ESS 5000 GUI, see IBM Knowledge Center.

Note: The **mmperfmon** command helps configure and query IBM Spectrum Scale internal performance monitoring tools, and **mmpmon** represents a user-driven performance command. The user of both these facilities should understand that the simultaneous operation of both tools might affect the output of each tool.

2.4.3 Physical disk health

As part of disk health monitoring, the disk hospital examines the relative performance and bit error rate of a disk. A disk that falls below the performance threshold has its pdisk state changed to *slow*, while a disk that exceeds the bit error rate (represented as *data badness*) has its pdisk state changed to *failing*. As part of the background diagnostic process in response to I/O errors, a disk might also enter various different pdisk states. The pdisk state changes produce **mmhealth** physical disk events that can be examined by using the **mmhealth** CLI or the IBM Spectrum Scale GUI

For more information about IBM ESS **mmhealth** events, see IBM Knowledge Center.

2.5 Software-related RAS enhancements

In this section, we describe the software-related RAS enhancements in the IBM ESS 5000.

2.5.1 Integrated Call Home

In the IBM ESS 5000 systems, beginning with IBM ESS V6.0.1 and later, call home events can be generated when a drive in an attached enclosure must be replaced. The IBM ESS 5000 can also generate call home events for other hardware-related events in the I/O server nodes that need service.

IBM ESS V6.0.1 and later automatically opens an IBM Service Request with service data, such as the location and field replaceable unit (FRU) number to perform the service task.

2.5.2 Software Call Home

IBM Electronic Service Agent (IBM ESA) for PowerLinux V4.5.5 and later can monitor the IBM ESS systems. The IBM ESA is preinstalled on the EMS node when the EMS node is shipped.

The rpm files for the IBM ESA are in the `/install/ess/otherpkgs/rhels8/ppc64le/ess/` directory.

The `esagent` rpm is also provided in the IBM ESS 5000 `binaries.iso` file in the container package. The ISO is mounted when **essmgr** is run to start the container. When mounted, the rpm file can be found in the `/install/ess/otherpkgs/rhels7/ppc64le/ess/` directory.

To verify that the `esagent` rpm is installed, run the following command:

```
rpm -qa | grep esagent
```

The command gives an output like the following one:

```
esagent.pLinux-4.5.5-0.noarch.rpm
```

The rpm should be installed during manufacturing. If it is not installed, run the following command:

```
cd /install/ess/otherpkgs/rhels8/ppc64le/ess/
rpm -ihv --nodeps esagent.pLinux-4.5.5-0.noarch.rpm
```

After the IBM ESA is installed, the IBM ESA portal can be reached by going to the following link:

```
https://<EMS or ip>:5024/esa
```

For example:

```
https://192.168.45.20:5024/esa
```

The IBM ESA uses port 5024 by default. It can be changed by using the IBM ESA CLI if needed. For more information about IBM ESA, see IBM Electronic Service Agent. On the Welcome window, log in to the IBM ESA GUI. If an untrusted site certificate warning is received, accept the certificate or click **Yes** to proceed to the IBM ESA GUI. You can get the context-sensitive help by selecting the **Help** option in the upper right.

After you log in, go to the Main Activate ESA window to run the activation wizard. The activation wizard requires valid contact, location, and connectivity information.

Electronic Service Agent configuration

Entities or systems that can generate events are called *endpoints*. The EMS, I/O Server Canisters, and attached enclosures can be endpoints in IBM ESS. Only enclosure endpoints can generate events, and the only event that is generated for call home is the disk replacement event.

Overview of a problem report

After the IBM ESA is activated and the endpoints for the nodes and enclosures are registered, they can send an event request to the IBM ESA to initiate a call home.

2.6 Performance

The IBM ESS 5000 uses the IBM POWER9 processor-based data servers, has 10 storage HBAs (SAS HBAs) to connect to storage enclosures, and supports up to 12 Ethernet or InfiniBand network port connections. The IBM ESS 5000 delivers up to 78% better sequential performance over the previous generation IBM ESS (IBM POWER8 processor-based). The measurements that were done in an IBM lab by using a 16 MiB file system and InfiniBand network with Remote Direct Memory Access (RDMA) enabled an achieved sequential read performance of 55 GBps and sequential write performance of 43 GBps.

Note: The performance measurements that are referenced here were made by using standard benchmarks in a controlled environment. The actual performance that is observed varies depending on the IBM ESS 5000 model (enclosure type and number of enclosures) along with other factors like the interconnection network, the configuration of client nodes, and the workload characteristics.

Some factors that are related to IBM ESS 5000 performance are network, tuning, and disk topology.

2.6.1 Network

From a network perspective, the first choice that must be made is to decide between configuring IBM Spectrum Scale to use Ethernet or InfiniBand. One of the desirable features of InfiniBand is its RDMA capability. With RDMA enabled, the communication between servers can bypass the operating system kernel, so the applications have lower latency and CPU utilization. With an Ethernet network that uses the TCP/IP protocol, the communications must go through the kernel stack, resulting in higher latencies than RDMA and reduced read and write bandwidths.

At the time of writing, using RDMA over Ethernet (RoCE) is not generally supported on the IBM ESS 5000 without a Request for Price Quotation (RPQ). If you are interested in a RoCE solution, consult with your sales representative, and a followup will be done to determine the viability of RoCE in your environment.

With Ethernet that uses the TCP/IP protocol, the bonding or link aggregation can be an important factor for performance. Most deployments use the Link Aggregation Control Protocol (LACP) (802.3ad) as the aggregation mode. The effective aggregate bandwidth depends on the load balancing across the interfaces. The switch load-balancing algorithm and hashing on the server side are responsible for balancing their respective outgoing traffic. In small clusters, it is possible the traffic across the bonded interfaces to be imbalanced. The load-balancing algorithm or the values being hashed might need to be changed to achieve better utilization. In clusters with many client and IBM ESS nodes, it is more likely that the traffic will be balanced across the bonded interfaces.

The network bandwidth can be assessed by using a tool like **`nsdperf`**. For an overview and usage instructions about this tool, see IBM Storage Community.

2.6.2 Tuning

IBM ESS performance also depends on the correct IBM Spectrum Scale RAID configuration, operating system, and network tuning. The **`essrun`** tool, which is used for deployment and also can be used for cluster and file system creation, and the **`mmvdisk server configure`** command, which is used during file system creation, automatically configure the IBM ESS tunables. But, sometimes the tunings might get changed accidentally due to an admin error, which might impact the performance. The tunings can be verified by using the following methods:

- The **`essinstallcheck`** command checks various aspects of the installation along with the IBM Spectrum Scale RAID configuration settings and tuned profile. For more information about how to run this command, see IBM Knowledge Center. Review the output carefully and address any issues.
- The IBM Spectrum Scale RAID configuration values can also be checked by using the new **`--verify`** option. The **`--verify`** option of the **`mmvdisk`** server configure command verifies whether the configuration settings for a **`mmvdisk`** server node class are as expected. The **`--verify`** option checks the real memory and server disk topology for each node in the node class, and checks whether the IBM Spectrum Scale RAID configuration attributes for the node class are set to the expected values for the memory and topology.

  The **`mmvdisk server configure --verify`** command can be used to check whether a new release of IBM Spectrum Scale RAID has updated best practice configuration values; to check whether the configuration values are affected by changes that were made to server memory or server disk topology; or to check whether accidental changes were made to the node class configuration. If an unexpected configuration is reported and the change was unintentional, the **`mmvdisk server configure --update`** command can be used to reset the node class to the correct configuration values.

Note: Some of the configuration values that are set independently of **mmvdisk** might be set to DEFAULT when an update runs.

Example 2-1 shows an example of the **mmvdisk server configure** command with the **--verify** option and its output.

Example 2-1 The mmvdisk server configure command with the --verify option

```
# mmvdisk server configure --nc ess5k_nc --verify
mmvdisk: Checking resources for specified nodes.
mmvdisk: Node class 'ess5k_nc' has 763 GiB total real memory per server.
mmvdisk: Node class 'ess5k_nc' has a paired recovery group disk topology.
mmvdisk: Node class 'ess5k_nc' has server disk topology 'ESS5K SC6 5-HBA'.
mmvdisk: Node class 'ess5k_nc' uses 'default.paired' recovery group
configuration.

daemon configuration attribute      expected value  configured value
----------------------------------  --------------  ---------------
pagepool                            573515602329    as expected
nsdRAIDTracks                       32K             as expected
nsdRAIDBufferPoolSizePct            80              as expected
nsdRAIDNonStealableBufPct           50              as expected
pagepoolMaxPhysMemPct               90              as expected
nspdQueues                          64              as expected
nsdRAIDBlockDeviceMaxSectorsKB      0               as expected
nsdRAIDBlockDeviceNrRequests        0               as expected
nsdRAIDBlockDeviceQueueDepth        0               as expected
nsdRAIDBlockDeviceScheduler         off             as expected
nsdRAIDFlusherFWLogHighWatermarkMB  1000            as expected
nsdRAIDSmallThreadRatio             2               as expected
nsdRAIDThreadsPerQueue              16              as expected
ignorePrefetchLUNCount              yes             as expected
maxBufferDescs                      2m              as expected
maxFilesToCache                     128k            as expected
maxMBpS                             30000           as expected
maxStatCache                        128k            as expected
nsdMaxWorkerThreads                 3842            as expected
nsdMinWorkerThreads                 3842            as expected
nsdMultiQueue                       256             as expected
numaMemoryInterleave                yes             as expected
panicOnIOHang                       yes             as expected
pitWorkerThreadsPerNode             32              as expected
workerThreads                       1024            as expected

mmvdisk: All configuration attribute values are as expected or customized.
```

- The tuned profile can also be verified by using the **tuned-adm** command, as shown in Example 2-2.

Example 2-2 The tuned-adm command

```
# tuned-adm active
Current active profile: scale
# tuned-adm verify
Verfication succeeded, current system settings match the preset profile.
See tuned log file (/var/log/tuned/tuned.log) for details.
```

2.6.3 Disk topology

The performance of an IBM ESS system can be impacted when there is a topology error in the disk subsystem configuration. It is useful to run **mmgetpdisktopology** along with **topsummary** when errors are suspected. The **mmgetpdisktopology** command gathers information about the disk subsystem on a IBM Spectrum Scale RAID server through operating system and device queries, and the output can be passed to an IBM Spectrum Scale RAID configuration script like **topsummary**, which prints a concise summary of the disk enclosures and their cabling along with any discrepancies.

Example 2-3 shows an example of the **mmgetpdisktopology** command with the **topsummary** option and its output.

Example 2-3 The mmgetpdisktopology command with the topsummary option

```
# mmgetpdisktopology > /tmp/top
# topsummary /tmp/top
GNR server: name ess5kaeth.gpfs.net arch ppc64le model 5105-22E serial 789953A
GNR enclosures found: 78T815F 78T8197 78T8199
Enclosure 78T815F (IBM 5147-106, number 1):
Enclosure 78T815F IOM L sg1012[5266][scsi5 port 4] IOM R sg677[5266][scsi4 port 4]
Enclosure 78T815F IOM sg1012 105 disks diskset "18676" IOM sg677 105 disks diskset "18676"
Enclosure 78T815F sees 105 disks (2 SSDs, 103 HDDs)

Enclosure 78T8197 (IBM 5147-106, number 2):
Enclosure 78T8197 IOM L sg1247[5266][scsi3 port 4] IOM R sg353[5266][scsi2 port 4]
Enclosure 78T8197 IOM sg1247 106 disks diskset "09245" IOM sg353 106 disks diskset "09245"
Enclosure 78T8197 sees 106 disks (0 SSDs, 106 HDDs)

Enclosure 78T8199 (IBM 5147-106, number 3):
Enclosure 78T8199 IOM L sg129[5266][scsi0 port 4] IOM R sg900[5266][scsi5 port 3]
Enclosure 78T8199 IOM sg129 106 disks diskset "56572" IOM sg900 106 disks diskset "56572"
Enclosure 78T8199 sees 106 disks (0 SSDs, 106 HDDs)

GNR server disk topology: ESS5K SC3 5-HBA (match: 99/100)
GNR configuration: 3 enclosures, 2 SSDs, 1 empty slot, 317 disks total, 2 NVRAM partitions
Location 78T815F-56 appears empty but should have an HDD

Slot C2 HBA model LSISAS3216 firmware[cli] 15.00.00.00 bios[cli] 08.31.00.00 uefi[cli] 15.00.00.00
Slot C2 HBA scsi5 U78D3.001.WZS09PK-P1-C2 [P3 78T8199 IOM R (sg900)] [P4 78T815F IOM L (sg1012)]
Slot C6 HBA model LSISAS3216 firmware[cli] 15.00.00.00 bios[cli] 08.31.00.00 uefi[cli] 15.00.00.00
Slot C6 HBA scsi4 U78D3.001.WZS09PK-P1-C6 [P4 78T815F IOM R (sg677)]
Slot C7 HBA model LSISAS3216 firmware[cli] 15.00.00.00 bios[cli] 08.31.00.00 uefi[cli] 15.00.00.00
Slot C7 HBA scsi3 U78D3.001.WZS09PK-P1-C7 [P4 78T8197 IOM L (sg1247)]
Slot C8 HBA model LSISAS3216 firmware[cli] 15.00.00.00 bios[cli] 08.31.00.00 uefi[cli] 15.00.00.00
Slot C8 HBA scsi2 U78D3.001.WZS09PK-P1-C8 [P4 78T8197 IOM R (sg353)]
Slot C12 HBA model LSISAS3216 firmware[cli] 15.00.00.00 bios[cli] 08.31.00.00 uefi[cli] 15.00.00.00
Slot C12 HBA scsi0 U78D3.001.WZS09PK-P1-C12 [P4 78T8199 IOM L (sg129)]
```

2.7 GUI enhancements

A GUI service runs on the EMS server. It can be used to monitor the health of the IBM ESS and to perform management tasks. This section provides a rough overview of the GUI and is by no means comprehensive.

To start or stop the GUI, run the **systemctl** command on the EMS server. Table 2-6 shows the **systemctl** command options.

Table 2-6 The systemctl command options

Command	Description
Start the GUI service.	**systemctl start gpfsgui**
Check the status of the GUI service.	**systemctl status gpfsgui**
Stop the GUI service.	**systemctl stop gpfsgui**

To access the GUI, enter the IP address or hostname of the EMS server in a web browser by using the secure https mode (`https://<IP or hostname of EMS>`).

2.7.1 GUI users

GUI users must be created before the GUI can be used. To grant special rights, roles are assigned to the users.

When the GUI is used for the first time, an initial user must be created by running the following command:

```
/usr/lpp/mmfs/gui/cli/mkuser <username> -g SecurityAdmin
```

Then, log in to the GUI with the new user and create more users by selecting **Services** → **GUI** → **Users**. By default, users are stored in an internal user repository. Alternatively, an external user repository can also be used by selecting **Services** → **GUI** → **External Authentication**.

2.7.2 System setup wizard

To set up the system, complete the following steps:

1. After logging in to the GUI for the first time, the system setup wizard starts, which looks up the system's information and performs several checks.

   Figure 2-12 on page 35 shows the Verify Storage window.

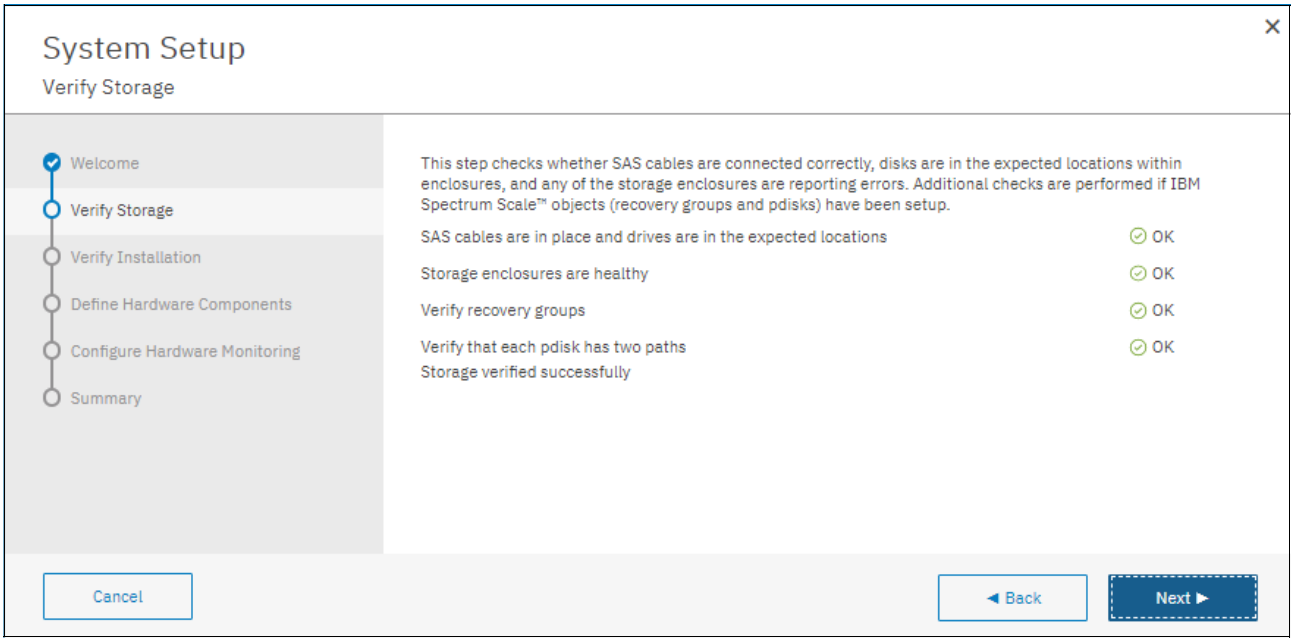


Figure 2-12 The System Setup wizard

2. In the Racks window, the racks where the IBM ESS 5000 systems are installed must be defined. Either choose a predefined rack type or click **Add new specification** in case none of the available rack types matches your rack. The selected rack type must have the same number of height units. A meaningful name can be specified for the racks to create.

   Figure 2-13 shows the Racks window.

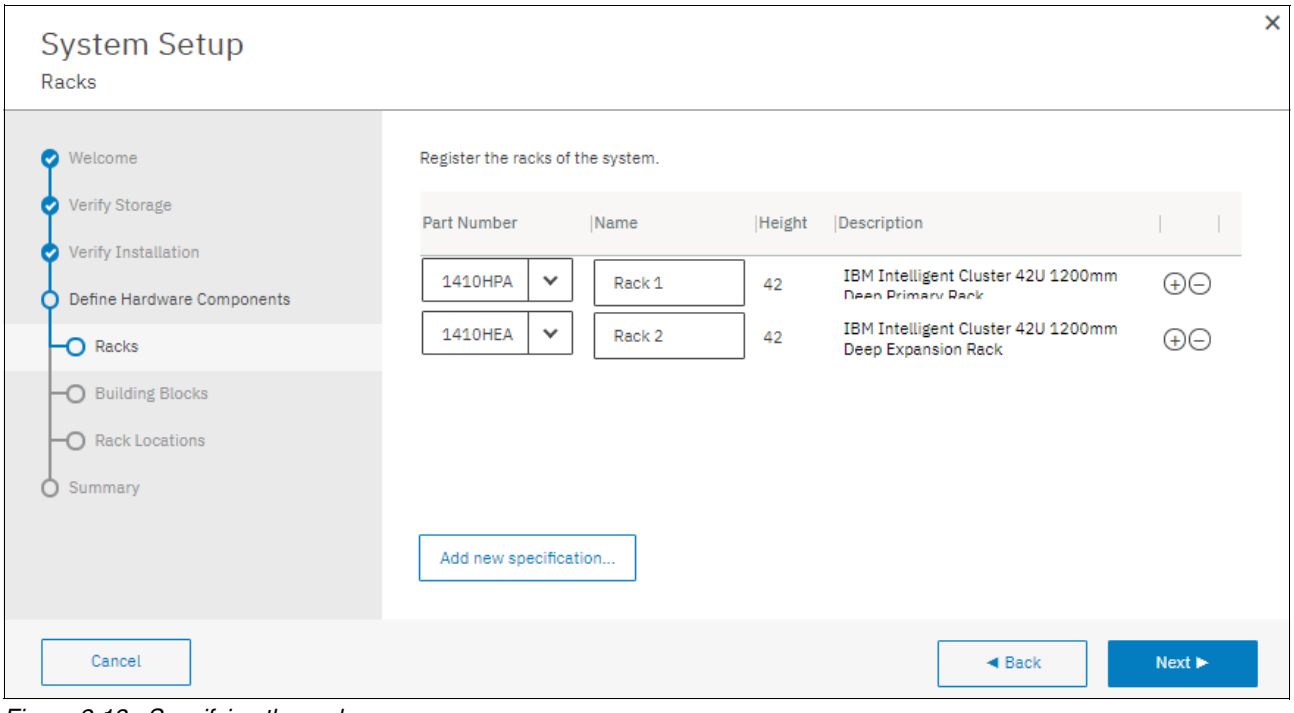


Figure 2-13 Specifying the racks

3. The Building Blocks window shows one row for each IBM ESS 5000 or other IBM ESS models. Assign names to each building block or go with the default.

   Figure 2-14 shows the Building Blocks window.

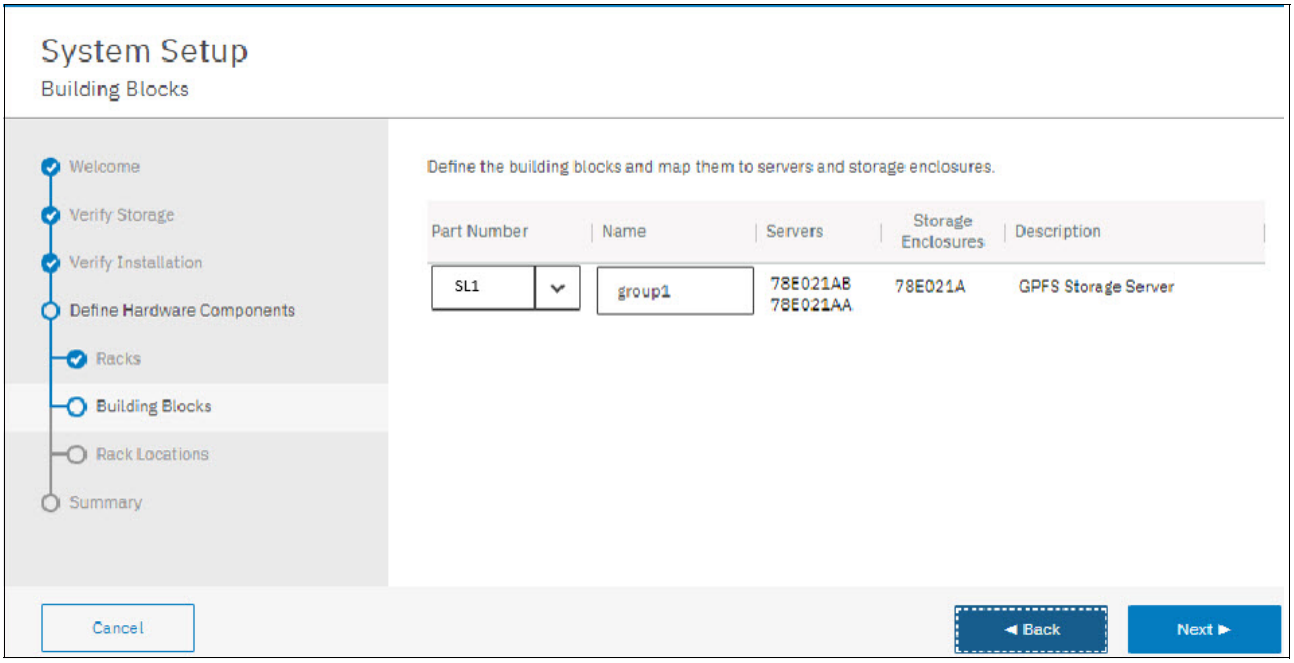


Figure 2-14 Defining building blocks

4. In the next step, the IBM ESS 5000 systems are assigned to the rack locations in which they are mounted.

   Figure 2-15 shows the Rack Locations window.

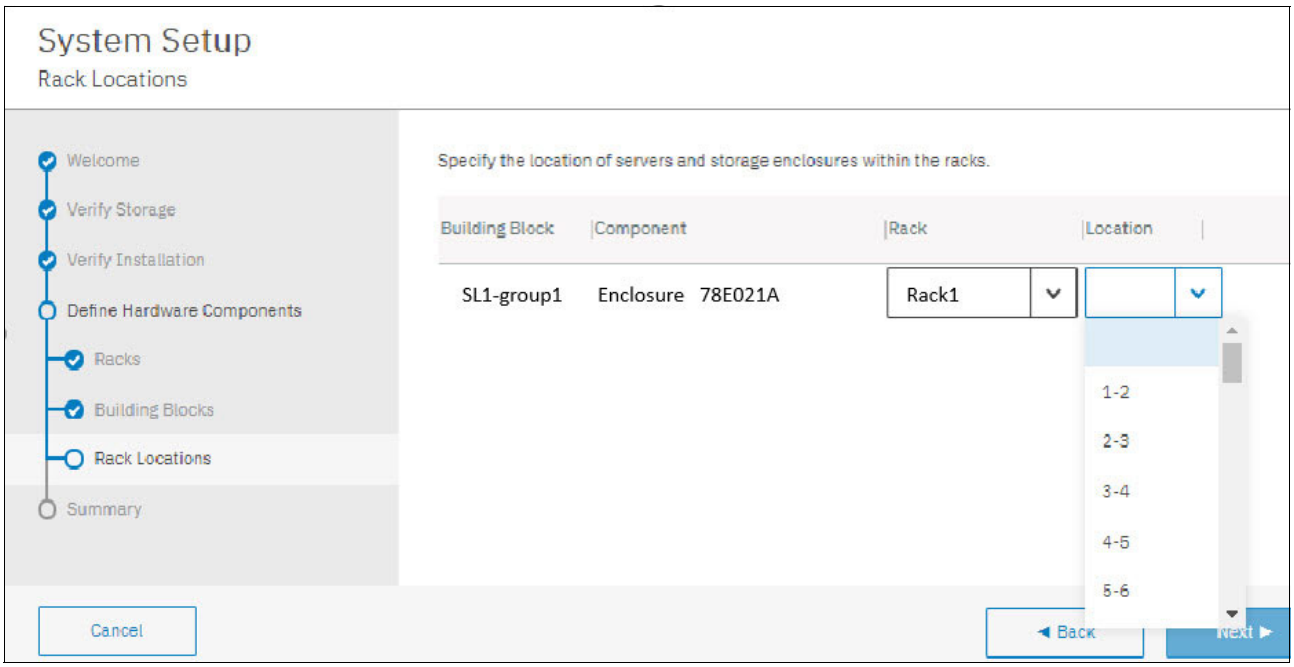


Figure 2-15 Assigning rack locations

5. The xCAT software is used to monitor the hardware of IBM POWER 8 processor-based servers. Therefore, the Configure Hardware Monitoring window appears only if IBM ESS 5000 systems are mixed with other IBM ESS models in one cluster. In a pure IBM ESS 5000 cluster, this step is not displayed because the IBM ESS 5000 is not monitored through xCAT. *POWER9 processor-based servers are detected by having the string "POWER9" in the component description.*

   Figure 2-16 shows the Configure Hardware Monitoring window.

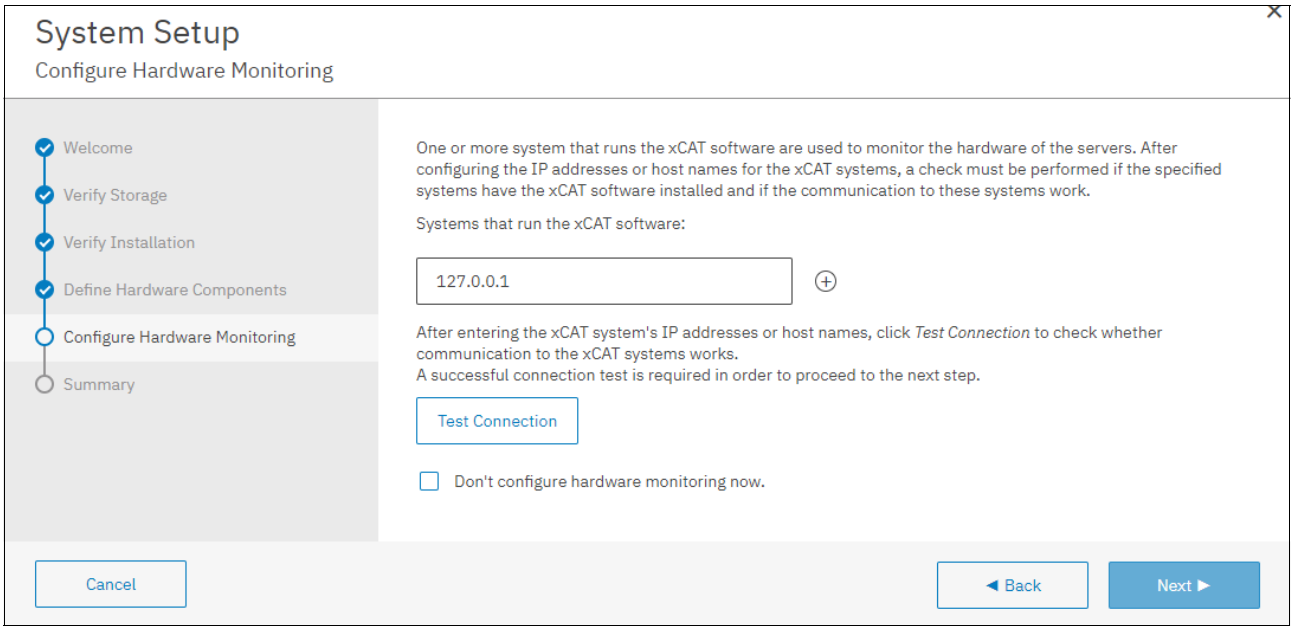


Figure 2-16 Configuring xCAT

6. By using the Configure Hardware Monitoring window, you can specify the IP address of the system where the xCAT software runs. The IP usually is 127.0.0.1 because xCAT runs on the EMS like the GUI. The connection to xCAT can also be configured at any later time by selecting **Monitoring** → **Hardware** and then clicking **Configure Hardware Monitoring**.

   Figure 2-17 shows the Configure Hardware Monitoring action.

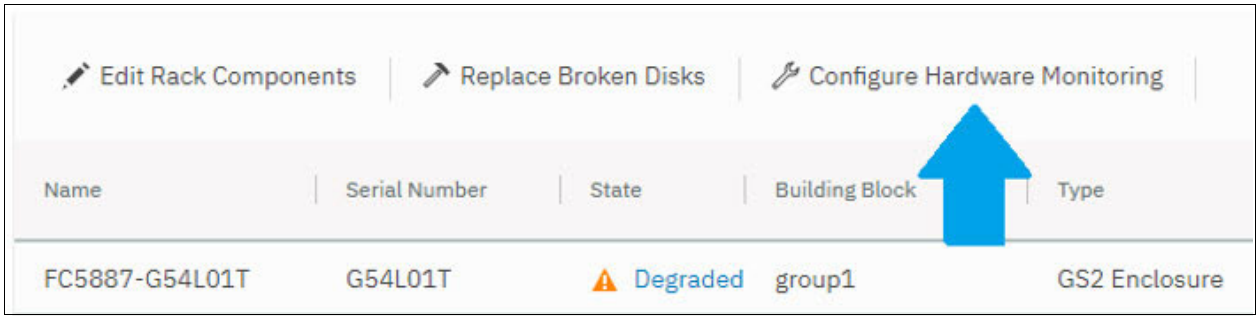


Figure 2-17 Configure Hardware Monitoring action

2.7.3 Using the GUI

After logging in to the GUI, the Overview window opens. This window provides a good view of all objects in the system and their health state. Clicking the numbers or links show a more detailed view.

Figure 2-18 shows the Overview window.

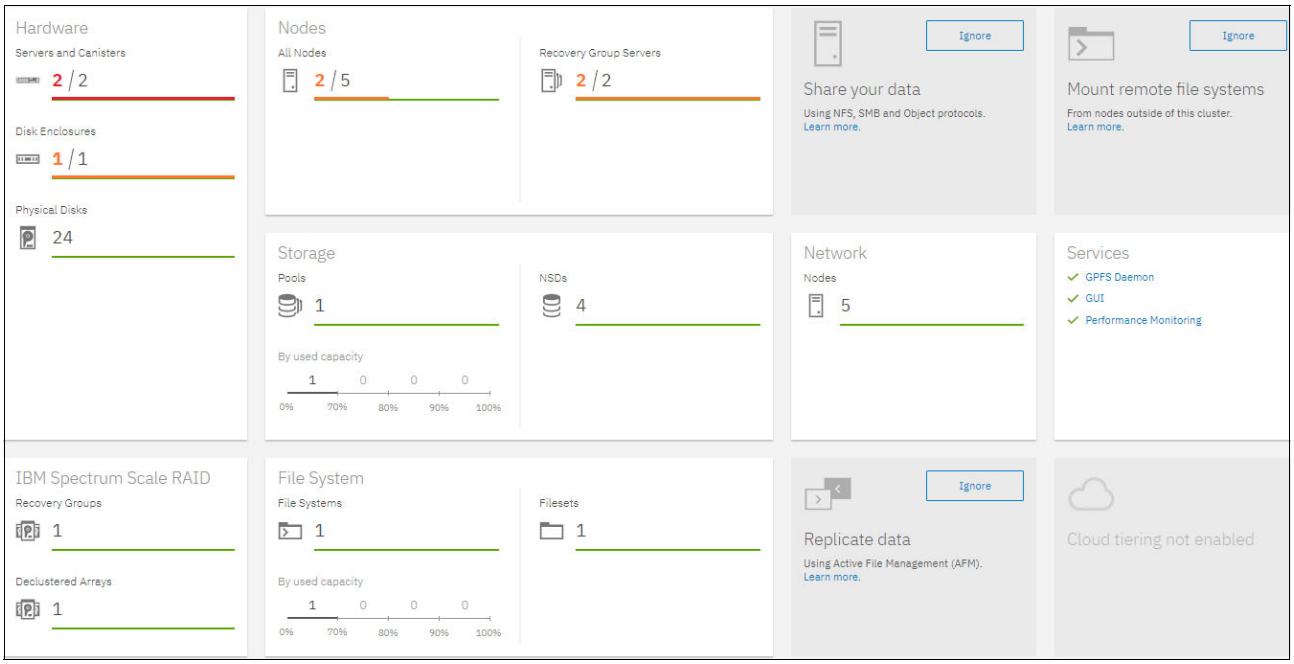


Figure 2-18 The Overview window

The header area of the GUI provides a quick view of the current health problems and tips for improvement. Additionally, there are links to some help resources.

Use the navigation menu on the left side of the GUI to go to other GUI windows (see Figure 2-19). Each GUI window has a unique URL that you can use to directly access the window, bookmark windows, and start the GUI in-context.

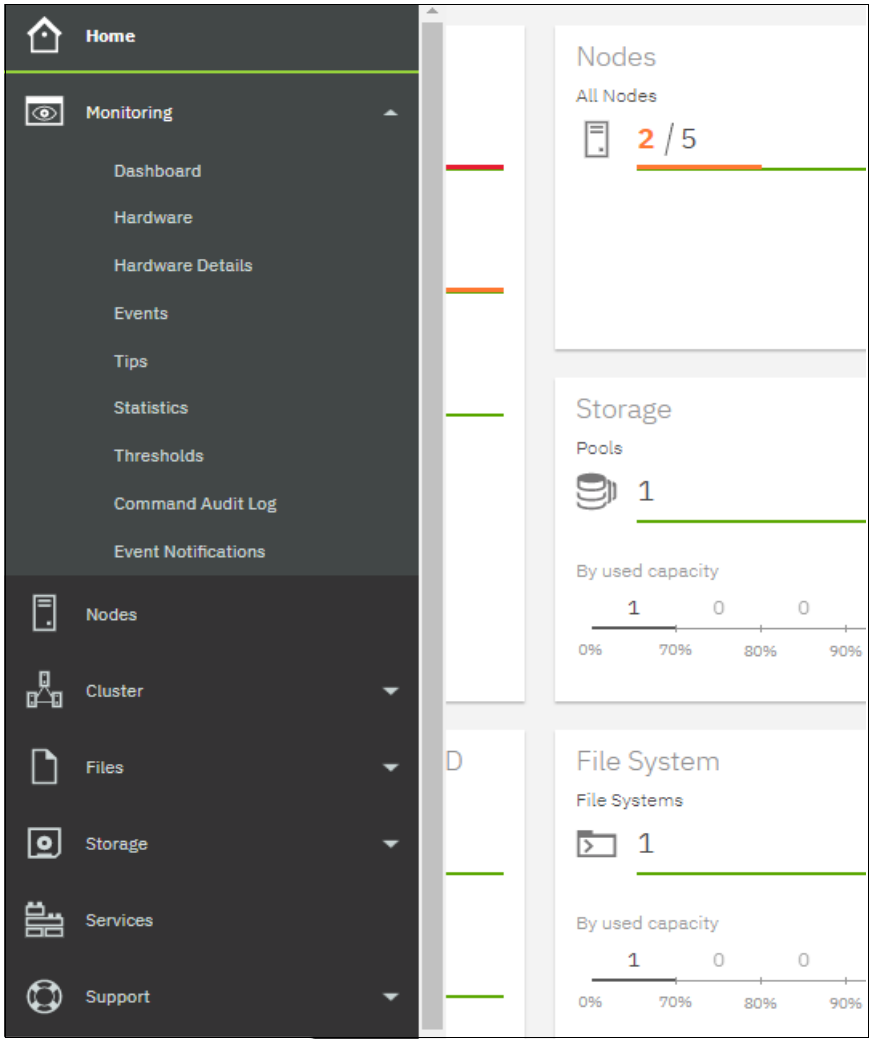


Figure 2-19 Navigation pane of the GUI

Some menus, such as Protocols, are displayed only when the related features, such as NFS, SMB, or AFM, are enabled.

Most tables that are shown in the GUI have columns that are hidden by default. Right-click the table header and select the columns to display, as shown in Figure 2-20.

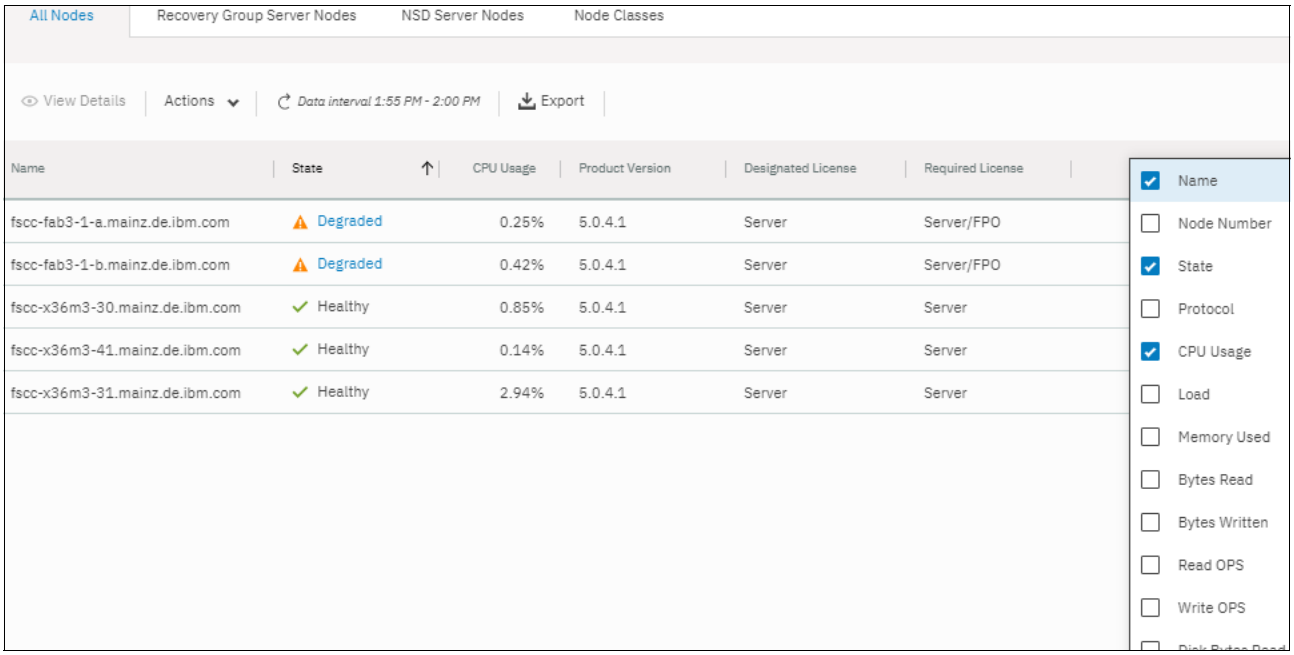


Figure 2-20 Showing and hiding table columns

The table values can be sorted by clicking one of the column headers. A little arrow in the table header indicates the sorting.

Double-click a table row to open a more detailed view of the selected item.

2.7.4 Monitoring of IBM ESS 5000 hardware

Selecting **Monitoring** → **Hardware** opens the IBM ESS 5000 enclosures within the racks. A table lists all enclosures and the related canisters, as shown in Figure 2-21.

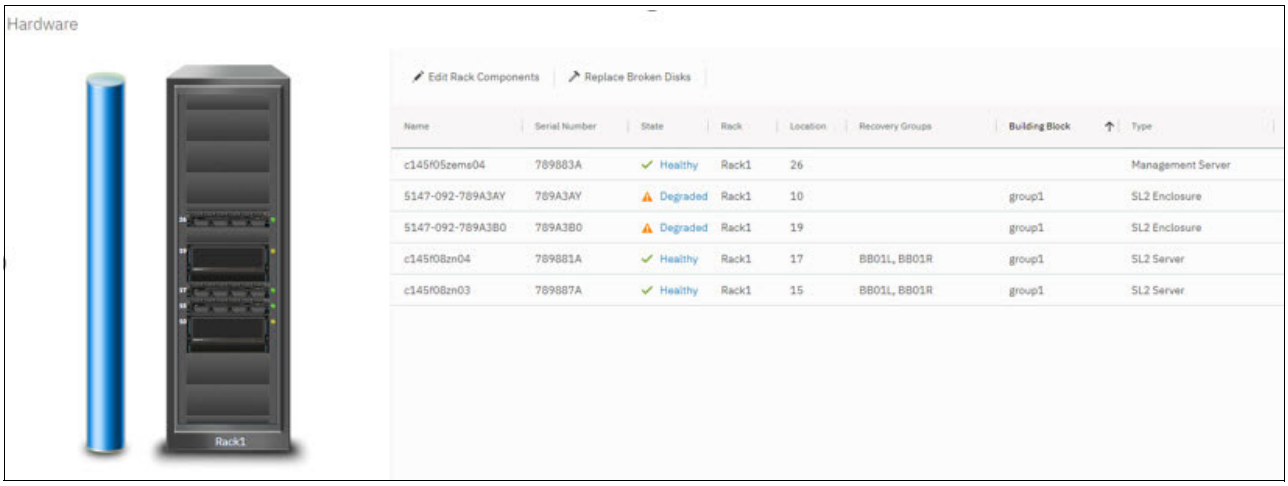


Figure 2-21 Hardware window with two IBM ESS 5000 systems

Use **Edit Rack Components** when IBM ESS enclosures or servers are added or removed, or if their rack location changed.

The **Replace Broken Disks** action starts a guided procedure to replace broken disks if there are any.

Click **Configure Hardware Monitoring** to specify the IP address of an xCAT server if it is used to monitor server hardware. The xCAT software is not used to monitor the canisters of the IBM ESS 5000. Therefore, this action is useful only when mixing the IBM ESS 5000 with other IBM ESS models in the same cluster, or to monitor the EMS server or POWER processor-based protocol servers. Any other servers are not monitored through the GUI.

Click the IBM ESS 5000 in the rack to see more information about the IBM ESS 5000, including the disks and the two canisters (Figure 2-22). Hover the mouse over the components, such as drives and power supplies, to see more information. Clicking components moves to a window with more detailed information. Broken disks are indicated with the color red, and you can use a menu (right-click) to replace the selected broken disk.

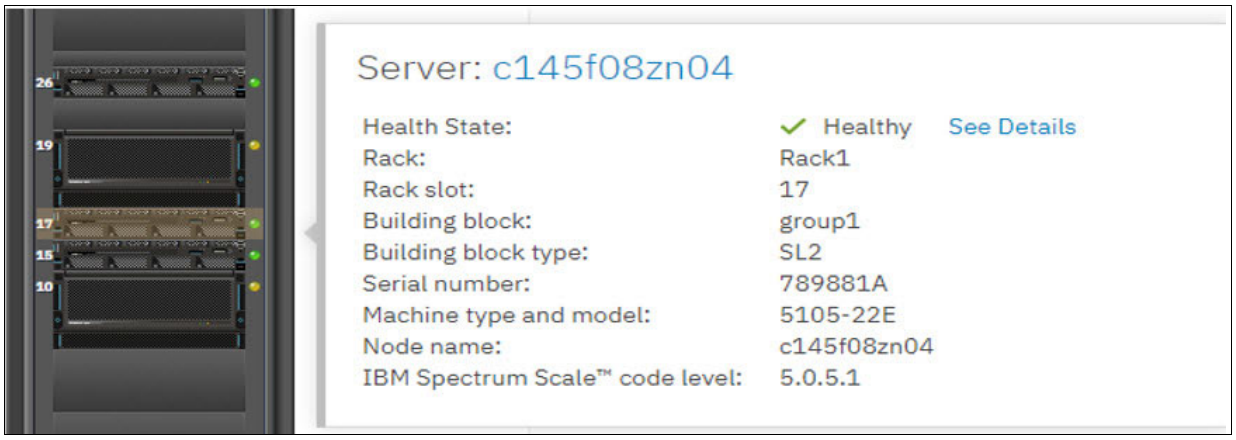


Figure 2-22 IBM ESS 5000 details in the Monitoring - Hardware window

If there is more than one rack, click the arrows that are on the left and the right side of the rack to switch to another rack.

The Hardware Details window can show more detailed information and the health states of the IBM ESS 5000 and its internal components, as shown in Figure 2-23.

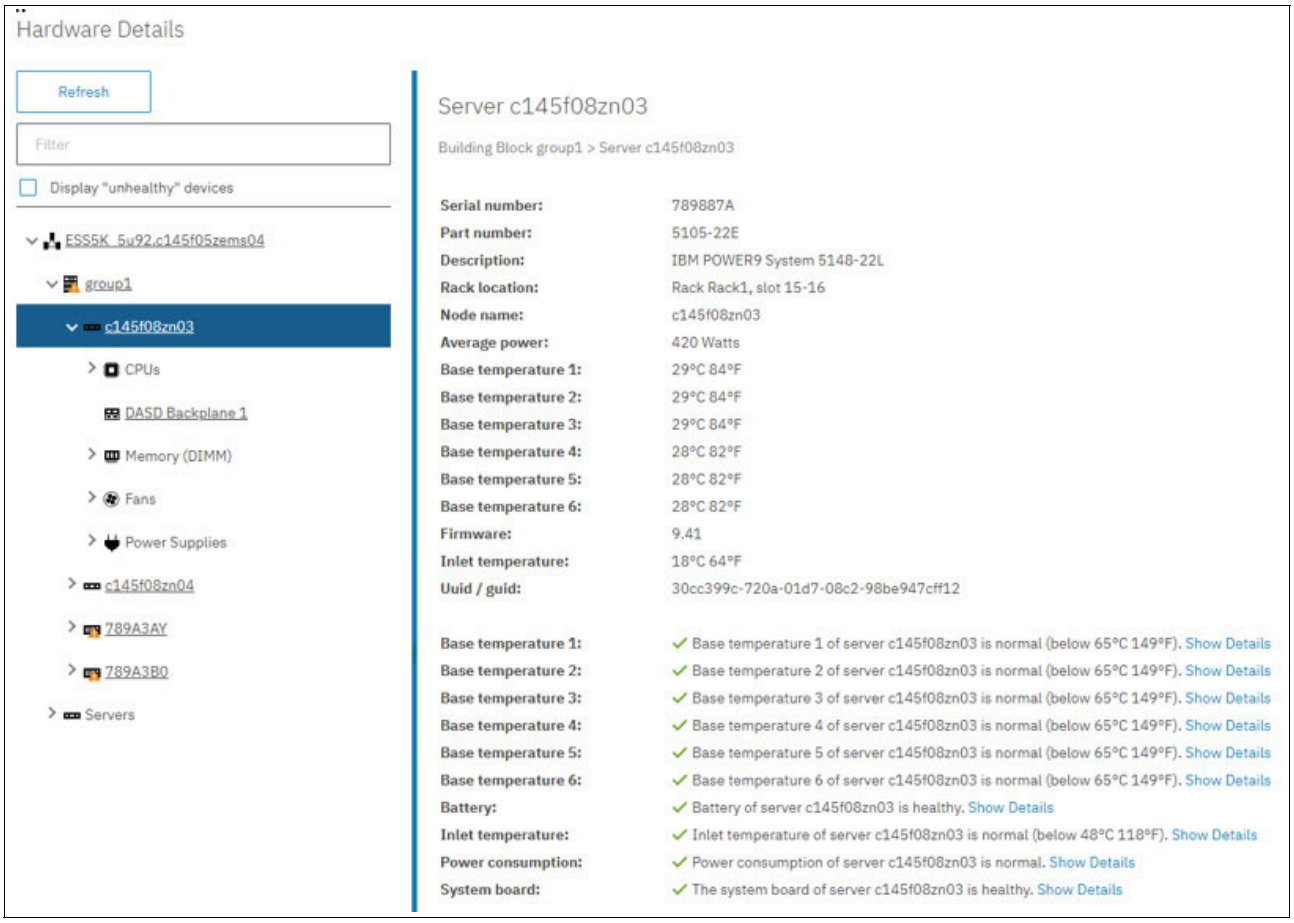


Figure 2-23 The Hardware Details window

This window allows the user to search for components by text, and filter the results to display only unhealthy hardware.

Click the **>** icon on the tree nodes to display the subsequent children, for example, to display all CPUs of the canister, as shown in Figure 2-24 on page 43.

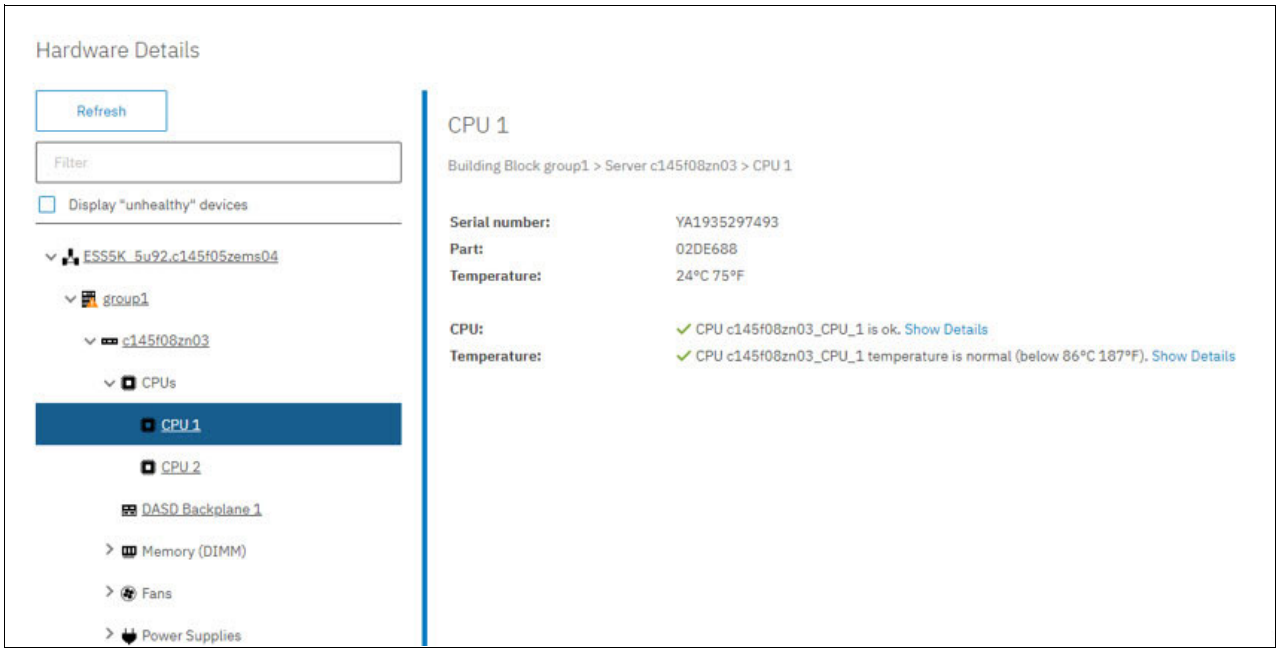


Figure 2-24 CPU details

Figure 2-25 shows the direct access storage device Backplane details.

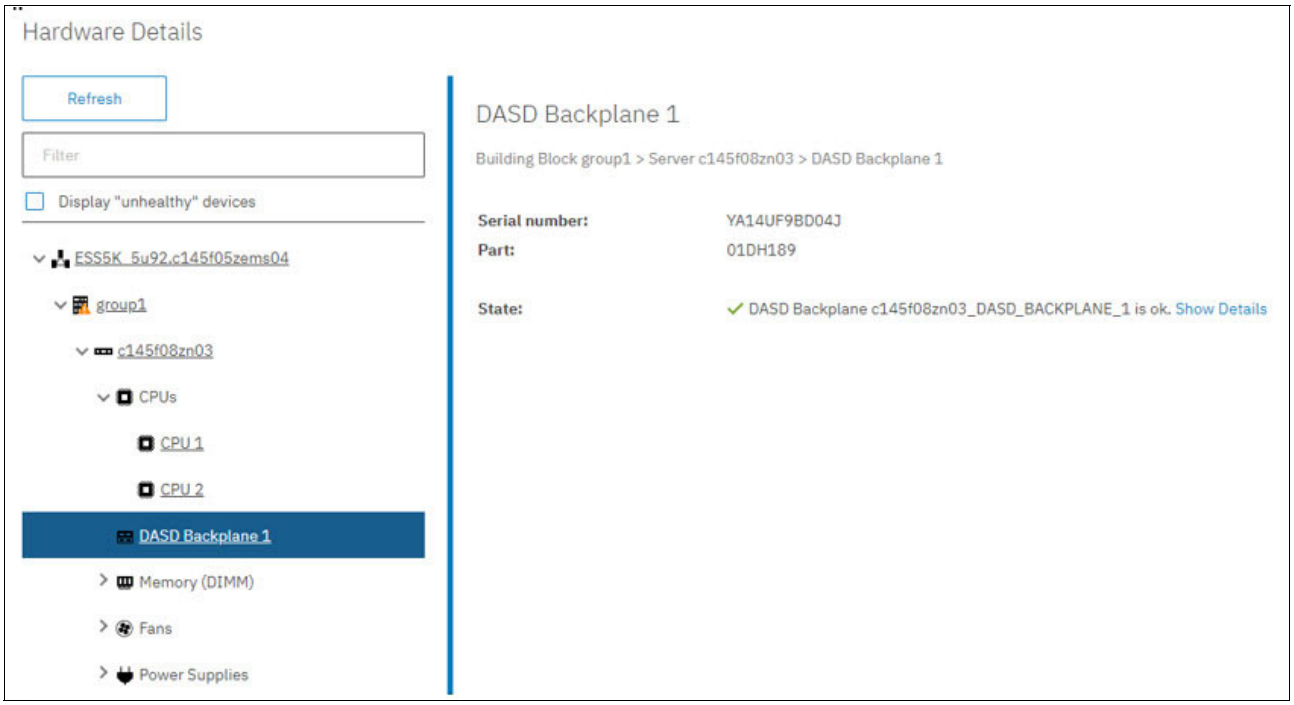


Figure 2-25 Direct access storage device Backplane details

Figure 2-26 shows the memory details.

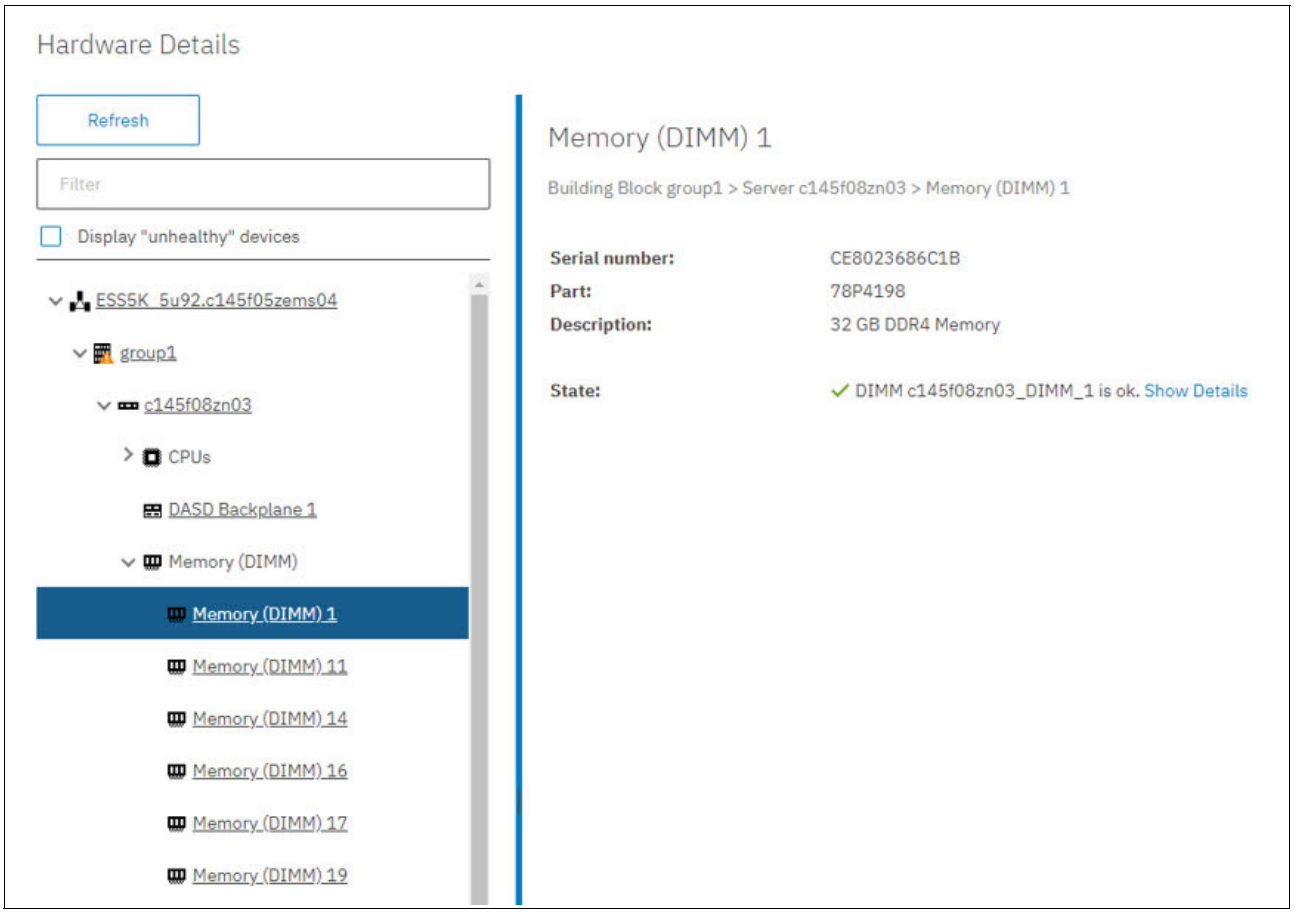


Figure 2-26 Memory details

Figure 2-27 on page 45 shows the fan details.

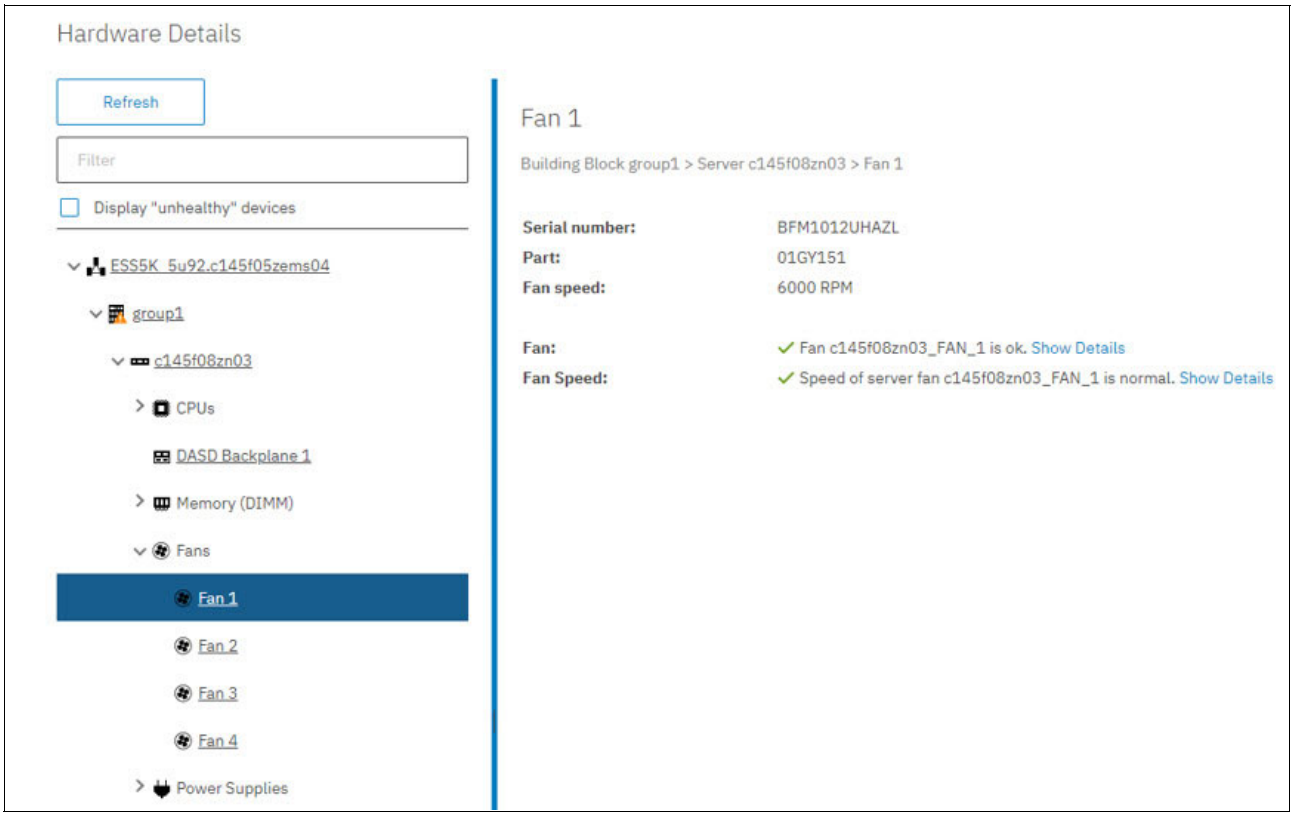


Figure 2-27 Fan details

Figure 2-28 shows the power supply details.

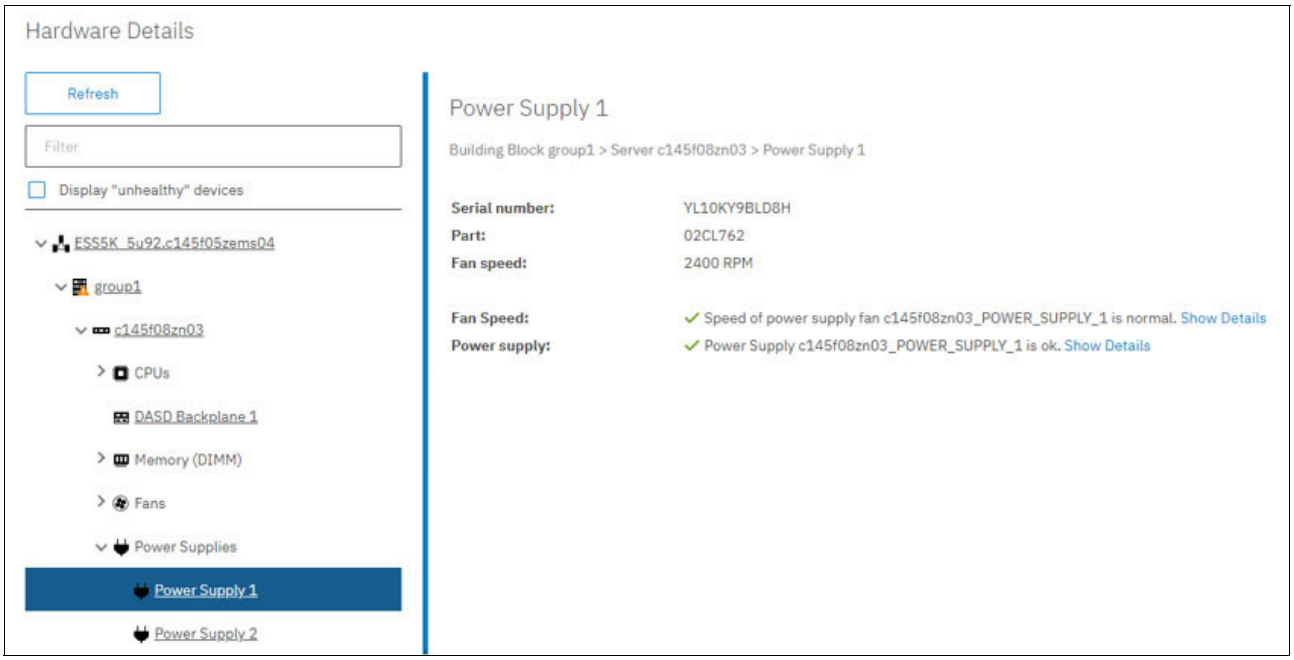


Figure 2-28 Power supply details

2.7.5 Storage

The **Storage** menu provides various views into storage, such as the physical disks, declustered arrays, recovery groups, virtual disks, NSDs, and storage pools. Figure 2-29 shows data for an IBM ESS 3000 (5141-AF8) enclosure.

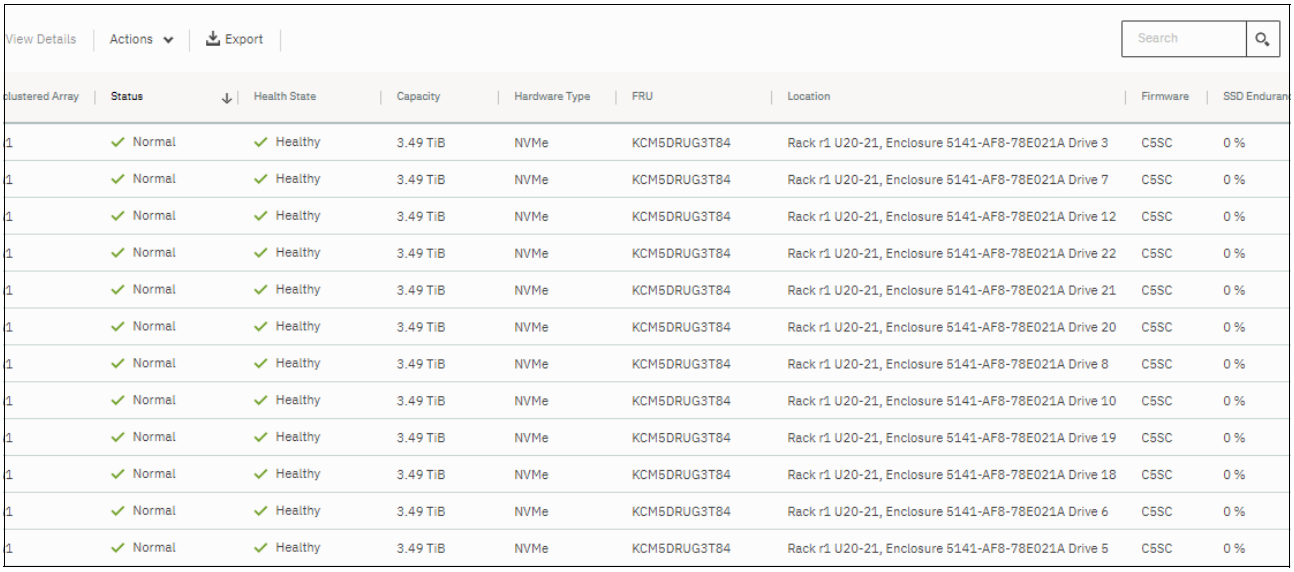


Figure 2-29 Storage menu

2.7.6 Event notification

The system can send emails and Simple Network Management Protocol (SNMP) notifications when new health events appear. Any combination of these notification methods can be used simultaneously. Select **Monitoring** → **Event Notifications** in the GUI to configure event notifications.

Sending emails

Select **Monitoring** → **Event Notifications** → **Email Server** to configure where emails should be sent. In addition to the email server, an email subject and the senders name can also be configured. By using the **Test Email** action, you can send a test email to an email address.

Figure 2-30 on page 47 shows the Email Server window.

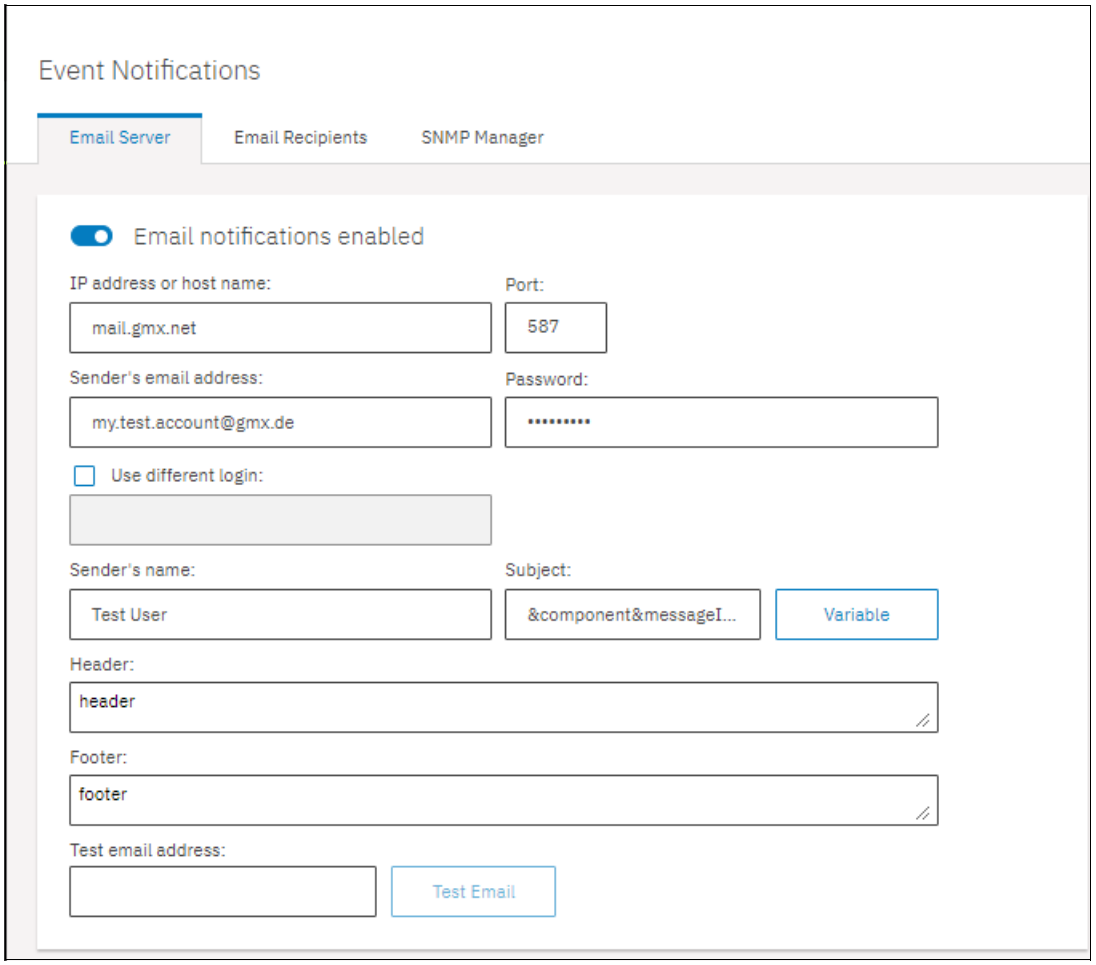


Figure 2-30 Configuring the email server

The emails can be sent to multiple email recipients, and you can define those emails by selecting **Monitoring** → **Event Notifications** → **Email Recipients**. For each recipient, you can select the components that receive emails, and the **For minimum severity level** (Tip, Info, Warning, or Error). Instead of receiving a separate email per event, a daily summary email can be sent. Another option is to receive a Daily Quota report.

Figure 2-31 shows the Create Email Recipient window.

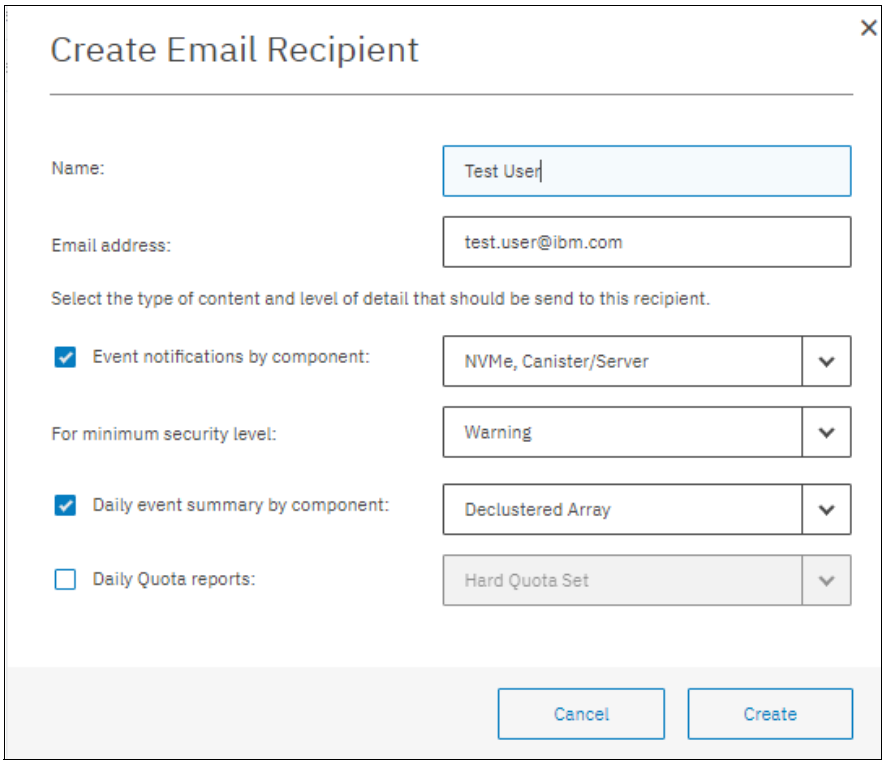


Figure 2-31 Create Email Recipient window

Sending SNMP notifications

Select **Monitoring** → **Event Notifications** → **SNMP Manager** to define one or more SNMP managers that receive an SNMP notification for each new event. As opposed to Email notification, no filters can be applied for the SNMP notification, and an SNMP notification is sent for any health event that occurs in the system.

The SNMP objects that are included in the event notifications are listed in Table 2-7.

Table 2-7 SNMP objects included in the event notifications

OID	Description	Example
.1.3.6.1.4.1.2.6.212.10.1.1	Cluster ID	317908494245422510
.1.3.6.1.4.1.2.6.212.10.1.2	Entity type	Drive Slot
.1.3.6.1.4.1.2.6.212.10.1.3	Entity name	SV44727220/DRV-1-6
.1.3.6.1.4.1.2.6.212.10.1.4	Component	Enclosure
.1.3.6.1.4.1.2.6.212.10.1.5	Severity	WARNING
.1.3.6.1.4.1.2.6.212.10.1.6	Date and time	17.10.2019 13:27:42.518
.1.3.6.1.4.1.2.6.212.10.1.7	Event name	drive_firmware_wrong
.1.3.6.1.4.1.2.6.212.10.1.8	Message	The firmware level of drive DRV-1-6 is wrong.
.1.3.6.1.4.1.2.6.212.10.1.9	Reporting node	gssoi2.spectrum

Example 2-4 shows an SNMP event notification that is sent when a performance monitoring sensor shuts down.

Example 2-4 Event notification

```
SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::enterprises.2.6.212.10.0.1
SNMPv2-SMI::enterprises.2.6.212.10.1.1 = STRING: "317908494245422510"
SNMPv2-SMI::enterprises.2.6.212.10.1.2 = STRING: "NODE"
SNMPv2-SMI::enterprises.2.6.212.10.1.3 = STRING: "gss-11"
SNMPv2-SMI::enterprises.2.6.212.10.1.4 = STRING: "PERFMON"
SNMPv2-SMI::enterprises.2.6.212.10.1.5 = STRING: "ERROR"
SNMPv2-SMI::enterprises.2.6.212.10.1.6 = STRING: "18.02.2016 12:46:44.839"
SNMPv2-SMI::enterprises.2.6.212.10.1.7 = STRING: "pmsensors_down"
SNMPv2-SMI::enterprises.2.6.212.10.1.8 = STRING: "pmsensors service should be started and is stopped"
SNMPv2-SMI::enterprises.2.6.212.10.1.9 = STRING: "gss-11"
```

The OID range `.1.3.6.1.4.1.2.6.212.10.0.1` denotes an IBM ESS GUI event notification (trap), and `.1.3.6.1.4.1.2.6.212.10.1.x` denotes IBM ESS GUI event notification parameters (objects).

The SNMP Management Information Base (MIB) file is available at the following location of each GUI node:

`/usr/lpp/mmfs/gui/IBM-SPECTRUM-SCALE-GUI-MIB.txt`

2.7.7 Dashboards

The Dashboard window provides an easy-to-read, single-page, and real-time user interface that provides a quick overview of the system performance.

There are some default dashboards that are included with the product. Users can further modify or delete the default dashboards to suit their requirements, and can create more dashboards. The same dashboards are available to all GUI users, so modifications are visible to all users.

A dashboard consists of several dashboard widgets that can be displayed within a chosen layout. There are widgets that are available to display performance metrics, system health events, file system capacity by file set, file sets with the largest growth rate in the last week, and timelines that correlate performance charts with health events.

Figure 2-32 shows the Dashboard window.

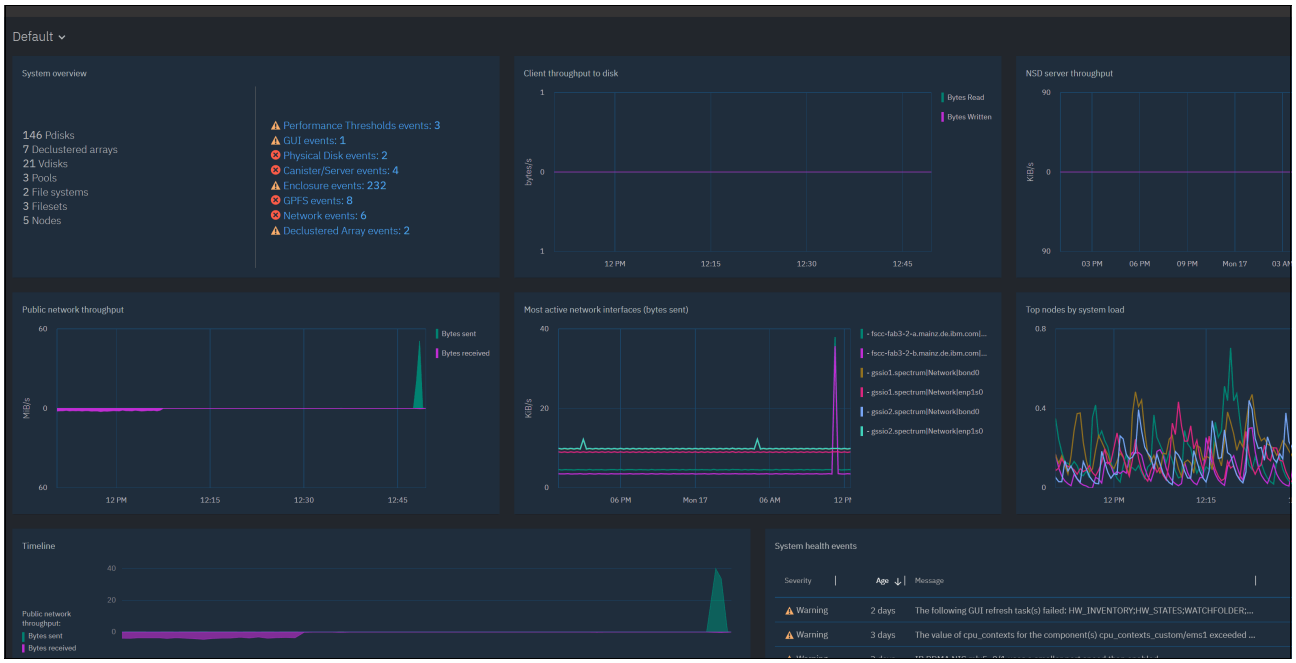


Figure 2-32 The Dashboard window

2.7.8 More information

The previous sections provided a rough overview of the GUI. For more information about the GUI, read *Monitoring and Managing the IBM Elastic Storage Server Using the GUI*, REDP-5471 and use the online help pages that are included within the GUI.

3

Planning considerations

This chapter provides planning considerations for installing the IBM Elastic Storage System 5000 (IBM ESS 5000). It includes the following sections:

- 3.1, “Planning” on page 52
- 3.2, “Stand-alone environment” on page 59
- 3.3, “Mixed environment” on page 71

3.1 Planning

When ordering an IBM ESS 5000, there are certain functional and non-functional requirements that must be fulfilled before the order can be made. The following sections describe these requirements.

3.1.1 Technical and Delivery Assessment

A Technical and Delivery Assessment (TDA) is an internal IBM process that includes a technical inspection of a completed solution design. Technical subject matter experts (SMEs) who were not involved in the solution design participate to answer the following questions:

- Will it work?
- Is the implementation sound?
- Will it meet customer requirements and expectations?

There are two TDA processes:

- The *presales TDA* is performed by the IBM Client Technical Specialist or an IBM Business Partner. This TDA can be done by using the FOS Design Engine tool, which can be found at IBM File Object Solution Design Studio.

  The IBM File Object Solution Design Engine leads the IBM Client Technical Specialist and IBM Business Partner team through collecting all requirements that are necessary to properly configure and size an IBM ESS 5000.

  You can find education for IBM personnel and IBM Business Partners about using the IBM File and Object Solution Design Engine at Learning Roadmaps.
- The *preinstallation TDA*, where SMEs evaluate the customer's readiness to install, implement, and support the proposed solution. The IMPACT tool can help with this process, and it can be found at IMPACT.

These TDA processes have assessment questions and baseline benchmarks that must be performed before the order can be fulfilled. Those tools are driven by IBM sales or resellers, so they can help and direct you regarding this process.

3.1.2 System planning worksheets

Customers are responsible for completing the system planning worksheets. Planning worksheets can help you identify important information that is needed when the system is installed and configured.

Then, the customer provides the worksheets to the IBM Support Services Representative (IBM SSR) when they install and configure the system.

Customers must complete the installation worksheet for the IBM SSR to start the installation. This process must be done by using the TDA process. The installation worksheet describes the IP addresses that the IBM SSR sets on each node.

The IBM SSR might need the following things for the IBM ESS 5000 installation:

- Ethernet cable
- USB-C (or USBA) to Ethernet dongle (This connector is required if your laptop does not have an Ethernet port.)

- Serial cable for POWER9 processor-based servers
- USB-C (or USBA) to serial dongle

If available, power on the HS switches and connect the Ethernet or InfiniBand cables to the adapters that are installed within the POWER9 processor-based servers. This step must be done before the customer can deploy the cluster, but it is not required for code 20. The low-speed Ethernet cables must be run to the proper locations on both switch and server sides before code 20 can begin (If the system came racked with the management switch, this step is done in manufacturing).

Customers must enter the following values so that the IBM SSR can perform the required networking tasks. If more than one IBM ESS 5000 node is being installed, you must add the corresponding rows.

The following values are best practices:

- Keep all management interfaces on `192.168.x.x/24` (netmask `255.255.255.0`).
- Keep all flexible service processor (FSP) (HMC1) interfaces on `10.0.0.x/24` (netmask `255.255.255.0`).

3.1.3 Networking hardware

This section covers the networks that are required for an IBM ESS building block to be able to operate. For more information about using the appropriate version, see at IBM Knowledge Center.

One IBM ESS building block is always composed of at least two POWER9 I/O nodes and one POWER9 EMS system. Further expansions of that system do not need an EMS system if the building block is being added to the same cluster as the previous one.

Note: Each cluster needs at least one EMS system. If the cluster is IBM ESS 5000 only, then that EMS system would be one POWER9 EMS system. As of this writing, the POWER9 EMS system can also manage IBM ESS 3000 systems. If in the same cluster you have other POWER8 IBM ESS based systems, then that cluster needs a second POWER8 EMS system.

However, I/O nodes and EMS systems are not enough to have a working environment. At a minimum, you need two non-routable, private netblocks (also called virtual local area networks (VLANs)) and at least one high-speed (HS) network to interconnect.

In Figure 3-1, you see a single POWER9 EMS managing one IBM ESS 5000 and one IBM ESS 3000 system.

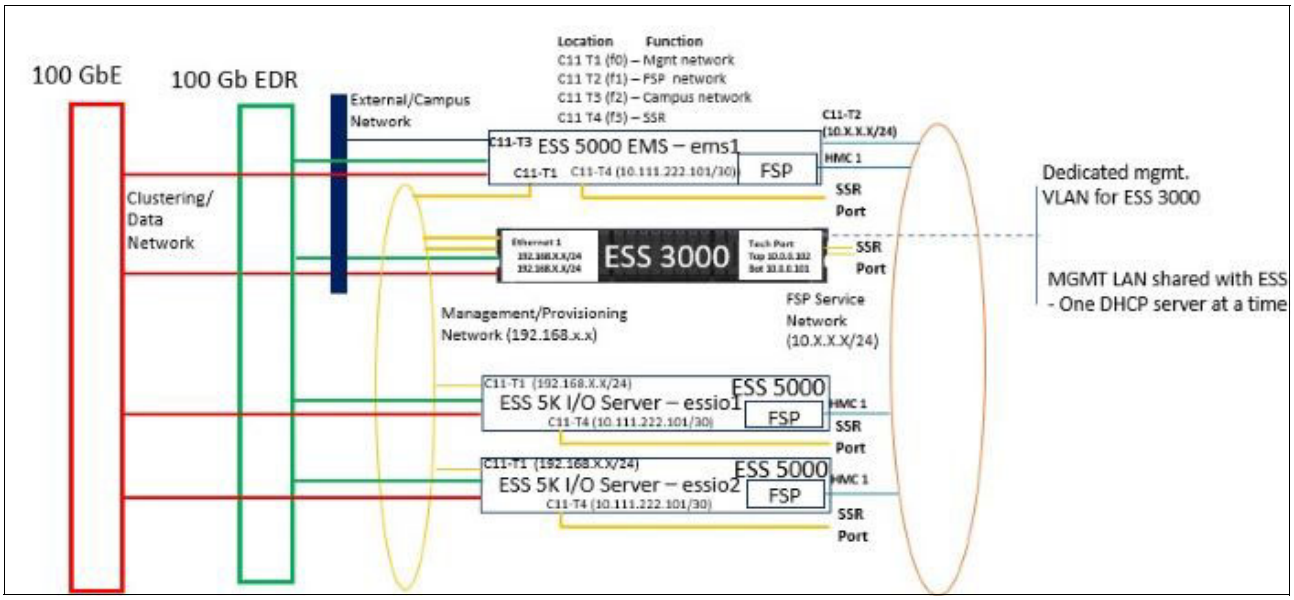


Figure 3-1 Single POWER9 EMS managing one IBM ESS 5000 and one IBM ESS 3000 system

Figure 3-1 shows two HS networks (the red one is 100 Gb Ethernet, and the green one is 100 Gb EDR). Only one network is required, but you can order IBM ESS systems with two networks. If you do so, all systems from the same cluster must use both networks.

You can also see the two non-routable, private networks (the yellow one is 1 Gb Ethernet, and the brown one is 1 Gb Ethernet). These networks are usually referred to as follows:

- Yellow as “Management”, “Provisioning,” and due to legacy reasons, “Xcat”
- Brown as “FSP”, and due to legacy reasons, “HMC”

You can also see the optional “Campus” network. Although it is not mandatory to have a campus network in the EMS, it is a best practice because call home, the GUI, and other services rely on that network to function.

Note: Although it is technically possible to connect I/O nodes to the campus network, we strongly discourage any other node than the EMS being connected to any campus network.

Finally, you can also see some references to the IBM SSR network. This port is a special one that is used only by an IBM SSR.

It is possible to order an initial IBM ESS building block (your first IBM ESS) without any hardware that provides any of these networks so that you can use your current hardware at your data center. However, whenever possible (as a best practice) order the network hardware as part of the IBM ESS building block.

Management network

The management network is a non-routable private network. It connects the EMS PCI card slot 11 (C11) port 4 (T1), which acts as a DHCP server to all I/O nodes on C11 T1, which act as DHCP clients.

Through this network, EMS and containers on the EMS manage the OS of the I/O nodes. This network cannot use VLAN tagging of any kind, so it must be configured as an access VLAN on the switch. You can choose any netblock that fits your needs, but as a best practice use a /24 block. If you have no preference, use the `192.168.45.0/24` block because it is the one that used in this paper.

Flexible service processor network

The flexible service processor (FSP) network is a non-routable private network. It connects the EMS C11 T2 with each of the I/O nodes out of band management ports that are labeled as “HMC 1".

EMS and the containers running on the EMS use this network to do FSP and baseboard management controller (BMC) operations on the physical servers, which include powering on and off the servers among many other operations. This network cannot use VLAN tagging of any kind, so it must be configured as an access VLAN on the switch. You can choose any netblock that fits your needs, but as a best practice use a /24 block. If you have no preference, use `172.16.0.0/24`.

Note: If you order the switches from IBM, both the Management and FSP network are provided by a Cumulus-based switch.

The Management or FSP network should not be part of the IBM Spectrum Scale cluster as a management or daemon network.

High-speed network

The HS network is where the IBM Spectrum Scale daemon and admin networks should be configured. Each one should be customer-dependent.

The IBM Spectrum Scale admin network has these characteristics:

- Used for the running of administrative commands.
- Requires TCP/IP.
- Can be the same network as the IBM Spectrum Scale daemon network or a different one.
- Establishes the reliability of IBM Spectrum Scale.

The IBM Spectrum Scale daemon network has these characteristics:

- It is used for communication between the mmfsd daemon of all nodes.
- Requires TCP/IP.
- In addition to TCP/IP, IBM Spectrum Scale can be optionally configured to use Remote Direct Memory Access (RDMA) for daemon communication. TCP/IP is still required if RDMA is enabled for daemon communication.
- Establishes the performance of IBM Spectrum Scale, as determined by its bandwidth, latency, and reliability of the IBM Spectrum Scale daemon network.

In cases where the HS is Ethernet based, unless you have good reasons not to do so, both the daemon and admin network should on the HS Ethernet network.

If you have InfiniBand networks, you can use Ethernet adapters on the HS if they are available, or you can use an IP over InfiniBand encapsulation.

Note: In the configuration scenario that is covered in this section, there are two IBM Mellanox provided switches and one single fabric with IP over InfiniBand for the admin network.

IBM SSR network port

The IBM SSR network port on the IBM ESS 5000 is on C11 port T4. That port should never be cabled or connected to any switch because it is *only* for IBM field engineers use. This port is configured on the `10.111.222.100/30` block.

Configuring the Cumulus switch for Management and FSP networks

If you order a Management or FSP switch from IBM on a racked setup, it should come preconfigured. But if you have a different switch or it is non-racked solution, set that Cumulus switch to fulfill the needs of the Management and FSP networks.

We assume that you have access to the Cumulus switch through a serial cable or management port.

Note: If you plan to configure central logging on the EMS, the management port of the 1 Gb Ethernet switch should be configured within the Management network block.

The configuration on the Cumulus switch is rather simple: It creates two independent access VLANs and one management IP to access the switch remotely through the IP protocol. The IBM configuration sets the upper row of ports as Management ports and the lower row of ports as FSP ports. For this example, we set the management switch IP address to `192.168.45.10`. To set the management IP address, run the commands that are shown in Example 3-1. Be careful if you are connected through that interface instead of a serial one because you might get disconnected if the IP address that you are setting is different that the one you are connected to already.

Example 3-1 Setting the management network IP address

```
root@cumulus:~# net add interface IP address 192.168.45.10/24
root@cumulus:~# net pending
root@cumulus:~# net commit
```

The contents in the `/etc/network/interfaces` file are shown in “Sample 1 Gb Ethernet configuration file” on page 104. After the file is saved in the switch file system, you must load the configuration so that it is enabled. Be careful not to log out yourself if you are connected through the IP management port. You should run the commands that are shown in Example 3-2.

Example 3-2 Reloading a Cumulus switch configuration

```
root@cumulus:/etc/network# ifreload -a
root@cumulus:/etc/network# ifquery -a
```

Now, you should have a fully configured Cumulus switch as it would have been configured by IBM manufacturing.

If your switch management IP is on the Management network, you can send the syslog events to the EMS by following the instructions at Enable Remote syslog.

Note: The syslog server is the EMS Management IP, and it listens to both TCP and UDP traffic. We strongly encourage the usage of UDP over TCP for syslog from the 1 Gb Cumulus switch.

Configuring the two InfiniBand switches for a high-speed network

In this scenario, we have two IBM 100 Gb EDR switches of 36 EDR ports each as part of an order of one IBM ESS 5000 building block (one POWER9 EMS and two POWER9 I/O nodes). We call them “switchA” (upper) and “switchB” (lower). The management IP address of the upper switch is `192.168.45.11/24`, and the IP address of the lower switch is `192.168.45.12/24`.

The EMS has one dual-port EDR card, and each I/O node has three dual-port EDR cards. Each port of the EMS is connected to port 1 on the switches. The upper port of each I/O node, one EDR card is connected to ports 2, 3, and 4 of the upper switch, and the lower port to ports 2, 3, and 4 of the lower switch. So, you have half the ports on each switch from each I/O node. The same is true for I/O node 2, but instead you use ports 5, 6, and 7 of each switch.

We are going to set a single InfiniBand fabric and run the Subnet Manager (SM) on the switches on a highly available (HA) configuration. We are also going to use RDMA for the IBM Spectrum Scale daemon network and IP over InfiniBand for the IBM Spectrum Scale admin network.

All the client (compute) nodes have two single-port EDR PCI cards, on which they connect one port to each switch for HA reasons. They can be connected with one port to each switch for ports 8 - 29, with a maximum number of 22 clients for this fabric design.

To interconnect each switch, use six EDR cables between the switches as inter-switch links (ISLs) because we have the same number of ISLs for one I/O node of the two, which ensures that we have a non-blocking connection from the clients to both I/O nodes. We use ports 30 - 36 for ISL.

As with the 1 Gb switch, we assume that you are already connected to the switches, preferably through a serial console.

Complete the following steps:

1. Set the “jump-start” configuration by setting the name and the management IP address. On each switch, you should run the commands that are shown in Example 3-3.

Example 3-3 Running jump-start on one switch

```
switchX [standalone: master] > enable
switchX [standalone: master] # conf terminal
switchX [standalone: master] (config) # configuration jump-start
```

Note: You can run the configuration wizard as many times as needed if you need to change any of the settings on the switch.

2. After you run the wizards on both switches, they now have unique names and management IP addresses in the same network block. Set the SM cluster on those switches. For this example, there is no more fabric than the two switches we configured. We set the virtual IP address of this SM cluster as `192.168.45.20/24` by logging in to switch A and running the commands that are shown in Example 3-4.

Example 3-4 Creating a HA subnet and adding a virtual IP address

```
switchA [standalone: master] > enable
switchA [standalone: master] # conf terminal
switchA [standalone: master] (config) # ib ha ESSIB-HA ip 192.168.45.20
255.255.255.0
switchA [ESSIB-HA: master] (config) # config write
```

3. Join switchB to the HA subnet by logging in to switchB and running the commands that are shown in Example 3-5. In this example, we show only the commands, not the switch format or the output.

Example 3-5 Creating a HA subnet and adding a virtual IP address

```
> enable
# conf terminal
# ib ha ESSIB-HA
# config write
```

4. Now, both switches are aware of each other, but SM is not yet configured. Log in to the virtual IP address `192.168.45.20`. (Do not use any of the switch management IP addresses.) Then, run the commands that are shown in Example 3-6.

Example 3-6 Creating a HA subnet and adding a virtual IP address

```
> enable
# conf terminal
# ib smnode switchA sm-priority 2
# ib smnode switchA enable
# ib smnode switchB sm-priority 1
# ib smnode switchB enable
# config write
```

Now, the SM is configured and should be running on switch B. You can see the status on the SM by running the **`show ib smnodes brief`** command.

Now, you should have a working InfiniBand fabric. To add client nodes by using IP over InfiniBand, see the configuration files examples in “Sample IP over InfiniBand configuration files” on page 108.

Note: For more information, see the IBM ESS 5000 documentation at IBM Knowledge Center.

3.2 Stand-alone environment

This section is about best practices for deploying and getting the most out of a stand-alone IBM ESS 5000.

A stand-alone IBM ESS 5000 unit, which is known as a building block, must at minimum consist of the following components:

- One EMS node in a 2U form factor.
- Two POWER9 IBM ESS 5000 nodes, each with a 2U form factor.
- A minimum of one external storage enclosure. The SC1 consists of a 106 disk enclosure with a 4U form factor, and the SL1 consists of a 92 disk enclosure in a 5U form factor.
- A 1 GbE or 10 GbE Network switch for a management network (1U).
- A 100 Gb HS InfiniBand or Ethernet network for internode communication (1U).

The EMS node acts as the administrative endpoint for your IBM ESS 5000 environment. It performs the following functions:

- Hosts the IBM Spectrum Scale GUI.
- Hosts Call Home services.
- Hosts system health and monitoring tools.
- Manages cluster configuration, file system creation, and software updates.
- Acts as a cluster quorum node.

The IBM ESS 5000 features a brand-new container-based deployment model that focuses on ease of use. The container runs on the EMS node. All the configurations tasks that were performed by the **`gssutils`** utility in IBM ESS are now implemented as Ansible playbooks that are run inside of the container. These playbooks are accessed by running the **`essrun`** command.

The **`essrun`** tool handles almost the entire deployment process, and it is used to install software, apply updates, and deploy the cluster and file system. Only minimal initial user input is required, and most of that is covered by the TDA process before setting up the system. The **`essrun`** tool automatically configures system tunables to get the most out of a single IBM ESS 5000 system. File system parameters and IBM Spectrum Scale RAID Erasure code selection can be customized from their defaults before file system creation.

For more information about deployment customization, see the IBM ESS 5000 Quick Deployment Guide at IBM Knowledge Center.

Here are some best practices:

- Refrain from running admin commands directly on the IBM ESS 5000 I/O servers. Use the EMS node instead.
- Do not mount the file system on the IBM ESS 5000 I/O servers because it uses more resources. The file system must be mounted on the EMS node for the GUI to function properly.

- To access the file system that is managed by the IBM ESS 5000 building block, you must use a remote cluster, IBM General Parallel File System (IBM GPFS) client nodes, or protocol nodes.
- On a single building block deployment, the I/O server nodes are specified as IBM GPFS cluster or file system manager nodes but the EMS node is not. Although the EMS node is considered the building block's primary management server, avoid specifying the EMS node as a manager node. The IBM GPFS management role is an internal designation that is used to manage the cluster and the file system, and it does not directly concern the function of the EMS node.

To access the storage that is managed by the IBM ESS building block, it is necessary to deploy IBM Spectrum Scale client nodes, set up a remote IBM Spectrum Scale cluster, or use IBM Spectrum Scale protocol nodes within your environment.

By purchasing a capacity-based license for either the Data Access Edition (DAE) or Data Management Edition (DME), you get an unlimited number of Network Shared Disk (NSD) client nodes on a separate remote mounted cluster at no additional cost if the remote cluster mounts only from a capacity-licensed cluster. Similarly, a capacity-based license incurs no additional charge for IBM Spectrum Scale server licenses that are needed for protocol nodes.

3.2.1 Small starter environment: Stand-alone IBM ESS 5000 system with IBM Spectrum Scale client nodes

If your workload is more performance-sensitive, then it might be prudent to deploy your stand-alone IBM ESS 5000 building block with a group of NSD client nodes. NSD client nodes can natively access IBM Spectrum Scale hosted storage over an HS network. In this configuration, the IBM ESS 5000 nodes act as NSD servers to provide access to a group of separate NSD client nodes. As opposed to a deployment with protocol nodes, this deployment requires installing IBM Spectrum Scale and the associated client licenses on all nodes that need to access the file systems that are hosted by the IBM ESS 5000.

When deploying an IBM ESS 5000 with NSD client nodes, it is a best practice to deploy the client nodes in a separate cluster and then remotely mount the file systems that are exported from the IBM ESS 5000 cluster. This configuration logically separates your storage nodes from your client workloads to allow for easier management. This configuration is shown in Figure 3-2 on page 61.

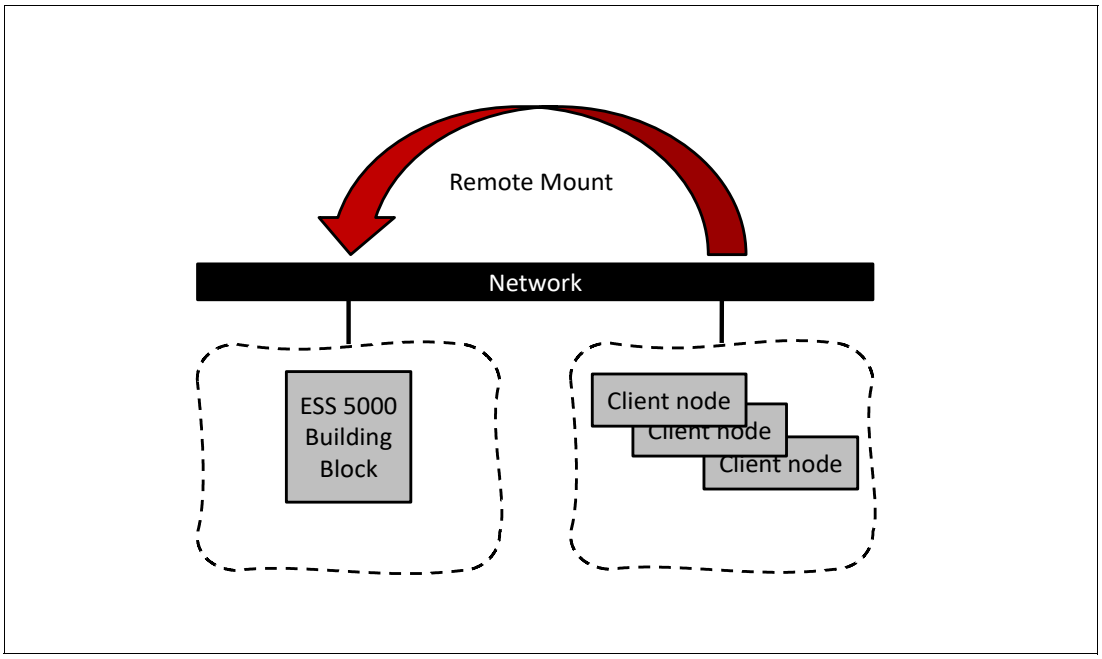


Figure 3-2 Deploying an IBM ESS 5000 with NSD client nodes

Note: Protocol node deployment is different than client node deployment. When deploying protocol nodes with the IBM ESS 5000, the protocol nodes are added to the existing IBM ESS 5000 cluster.

3.2.2 Installing and deploying a remote NSD client cluster

As a best practice, use the IBM Spectrum Scale installation toolkit by running the **`spectrumscale`** command when you create the NSD client cluster. The toolkit consists of four primary phases:

- User input and configuration phase
- Installation phase
- Deployment phase
- Upgrade phase

This section covers only the first two phases.

Prerequisites

Before you begin, make sure that the following conditions are met:

- An IBM ESS 5000 cluster with a building block is deployed, including file system creation.
- The client nodes are a supported architecture with the proper OS levels.
- All required packages are installed, software dependencies are met, kernel packages are installed, and base OS repositories are configured. The kernel packages should match the kernel the output of **`uname -a`** on each client node.
- Root login is enabled on the client nodes.
- Call home information is set up and available.
- Passwordless SSH is configured from the admin node to other nodes in the prospective client cluster.

- The networking is set up such that DNS is configured to resolve all short or long hostnames, or `/etc/hosts` is populated in a specific order.
- Ensure that network ports that are required for installation are open.
- Ensure that the **`LC_ALL`** and the **`LANGUAGE`** parameters are set to `en_US.UTF-8`.
- A copy of the self-extracting IBM Spectrum Scale installation package was obtained from IBM Fix Central. For new IBM ESS 5000 deployments, as a best practice use IBM Spectrum Scale V5.0.5 or later for the client cluster.

For more information about preparing to use the installation toolkit, see IBM Knowledge Center.

Configuration phase

To configure your system, complete the following steps:

1. Designate an installer node in the client cluster. For client nodes running Version 5.0.5.1, go to the `/usr/lpp/mmfs/5.0.5.x/installer` directory and run the following command:

   ```
   ./spectrumscale setup -s InstallNodeIP
   ```

   The **`-s`** flag specifies the IP address of a device on the installer node.

   Note: The installer node stores the configuration file that each node later retrieves. If you are using a node outside of the cluster to install the cluster, then you might have to use the **`-i`** flag to specify a private SSH key if passwordless SSH is not set up between the installer node and the rest of the cluster. The **`-i`** flag may be used if you are using a node inside the cluster to act as the installer node because it is a prerequisite that passwordless SSH is enabled. To simplify the installation process, as a best practice designate the installer node and the IBM GPFS admin node as the same node.

2. Populate the cluster definition file by running the **`./spectrumcale node add`** command:
   a. Add the admin node by running **`./spectrumscale node add -a nodename`**. The admin node acts as the coordinator node for installation, deployment, and upgrades by using the installation toolkit. On an IBM ESS 5000 cluster, the EMS fulfills this role, but a different node must be selected on the remote client cluster.
   b. Add the remaining client nodes by running **`./spectrumscale node add nodename`**. Using this method, the toolkit automatically selects quorum nodes based on the number of nodes in the cluster definition, and all nodes receive the manager node designation.
   c. To add a call home node to the client cluster, run **`./spectrumscale node add nodename -c`**. If a call home node is not specified, then one node from the cluster is assigned as the call home node. When finished, you can view the current cluster definition file by running **`./spectrumscale node list`**.

   For more information about defining the cluster topology, see IBM Knowledge Center.
3. Configure the cluster parameters before installing the cluster by running the **`./spectrumscale config gpfs`** command. To specify the cluster name, run **`./spectrumscale config gpfs -c myclustername.my.domain.name.com.`** If no cluster name is specified, then the toolkit uses the IBM GPFS admin node name as the cluster name. If the cluster name does not contain periods, then the cluster name is assumed to not be a fully qualified domain, and the cluster name domain inherits from the admin node domain.

Note: During the configuration phase, you can specify the ephemeral port range of the cluster nodes by running the **./spectrumscale config gpfs --ephemeral_port_range <low> <high>** command. IBM Spectrum Scale uses this range to dynamically create more sockets when exchanging data among nodes. The default port range that is set by the installation toolkit as of the release of IBM ESS 5000 is 60000 - 61000. Be sure that your firewall settings allow for open ports on the chosen ephemeral port range.

For more information about configuring the cluster with the installation toolkit, see IBM Knowledge Center.

Installation phase

Before moving on to the installation step, make sure to specify the call home information (you should have this information as a prerequisite) by running the following command:

```
./spectrumscale callhome config -n CustName -i CustID -e CustEmail -cn CustCountry
```

To install the cluster, run the **./spectrumscale install** command.

Note: Running the **install** command without arguments automatically starts an installation precheck step to validate the environment. To run the precheck by itself, run **./spectrumscale install -pr**. After the cluster is installed, your client cluster should be active.

For more information about the installation, see IBM Knowledge Center.

Postinstallation tuning best practices

After the client cluster is created in the installation phase, extra networking settings and performance settings can be applied. The **/usr/lpp/mmfs/samples/gss/gssClientConfig.sh**l script can be used to apply basic best practice settings for your client NSD cluster. Running this script with the **-D** option shows the configuration settings that it intends to set without setting them.

This script also attempts to configure your client nodes for RDMA access by setting the following **mmchconfig** parameter values if applicable:

- **verbsRdma enable**
- **verbsRdmaSend yes**
- **verbsPorts <active_verbs_ports>**

Enabling remote mounting for the remote NSD client cluster

You must set up remote cluster mounting between your IBM ESS 5000 and NSD client cluster to access your storage when deploying client nodes in a separate cluster. This process involves the following steps:

1. Generate private and public key pairs on the IBM ESS 5000 and NSD client cluster and enabling authorization.
2. Exchange public keys between the two clusters.
3. Grant remote access to file systems on the IBM ESS 5000 cluster by running the **mmauth** command.

4. Add the IBM ESS 5000 cluster to the NSD client cluster as a remote cluster by running the **mmremotecluster** command.
5. Add the remote file system from the IBM ESS 5000 cluster to the NSD client cluster by running the **mmremotefs** command.

Before starting, make sure the `gpfs.gskit` package is installed on all nodes of both clusters. To simplify the following example, we refer to the IBM ESS 5000 cluster name as "`ess5k.yourdomain.net`" and the NSD client cluster as "`client0.yourdomain.net`". The IBM ESS 5000 exports a single file system that is named "`fs1`". The hostnames of the IBM ESS 5000 I/O server nodes are "`essio1.yourdomain.net`" and "`essio2.yourdomain.net`".

Complete the following steps:

1. On both clusters, run the **mmauth** command to generate a private and public key pair:

   ```
   mmauth genkey new
   ```

   The public key can be found in "`/var/mmfs/ssl/id_rsa.pub`".
2. After the keys are updated, run the following command to enable authorization on both clusters:

   ```
   mmauth update . -l AUTHONLY
   ```
3. Exchange the public keys between the two clusters. To do so, copy the public keys of each cluster that is found in "`/var/mmfs/ssl/`" to the local storage.
4. After the keys are exchanged, grant the NSD cluster access to file system "`fs1`" by running the following commands on the IBM ESS 5000 cluster:

   ```
   mmauth add accessingCluster -k <path to NSD client cluster public key>
   mmauth grant client0.yourdomain.net -f fs0
   ```
5. On the NSD client cluster, add the IBM ESS 5000 as a remote cluster by running the following command:

   ```
   mmremotecluster add ess5k.yourdomain.net -n essio1.yourdomain.net,essio2.yourdomain.net
   ```
6. Add the remote file system `fs1` to the NSD client cluster as "`fs1_remote`". On the NSD client cluster, run the following command:

   ```
   mmremotefs add fs1_remote -f fs1 -C ess5k.yourdomain.net -T /fs1_remote
   ```
7. To verify this process, try mounting `fs1_remote` on the local NSD client cluster by running the following command:

   ```
   mmmount fs1_remote
   ```
8. If this command is successful, then `fs1_remote` should be mounted at local mount point `/fs1_remote`.

For more information, see IBM Knowledge Center.

3.2.3 Small starter environment: IBM ESS 5000 system with IBM Spectrum Scale protocol nodes

Cluster Export Services (CES) provide HA file and object services to an IBM Spectrum Scale cluster by using a Network File System (NFS), object, or Server Message Block (SMB) protocols. Because CES has specific hardware and software requirements, the code must be installed on nodes that are designated to run the CES software stack. These nodes are called *protocol nodes*.

Protocol nodes can be added to an IBM Spectrum Scale cluster containing an IBM ESS building block. They can also exist in non-IBM ESS IBM Spectrum Scale clusters. Protocol nodes are not attached to external storage, which means that SMB and NFS functions on protocol nodes can exist in clusters in which their storage is remotely mounted.

IBM ESS coexists in an IBM Spectrum Scale cluster with protocol nodes. IBM ESS has no effect on which type of protocol node can be used in an IBM Spectrum Scale cluster. However, from a planning perspective, IBM Spectrum Scale requires that all protocol nodes in an IBM Spectrum Scale cluster must be either all x86, POWER Little Endian, or POWER Big Endian within the same IBM Spectrum Scale cluster.

Starting with IBM ESS V5.3.1.1 in August 2018, IBM provided the option of ordering IBM Spectrum Scale protocol nodes based on the POWER8 5148-22L server. The POWER9 processor-based protocol server is 5105-22L. These protocol nodes are supported by the EMS and can be managed by the IBM ESS GUI and by IBM ESS installation tools, which provide greater ease of use and management capability for ordering a complete IBM Spectrum Scale, IBM ESS, and protocol node solution from IBM.

Supported protocol node configurations

The following sections describe the supported protocol node configurations.

Configuration 1: 5148-22L protocol nodes ordered and racked with a new 5148 IBM ESS (PPC64LE)

In this configuration, both a new 5148 IBM ESS and new 5105-22L protocol nodes are ordered and racked together. The EMS node, I/O server nodes, and protocol nodes have the OS, kernel, systemd, network manager, firmware, and OFED that is managed by the eXtreme Cluster Administration Toolkit (xCAT) running on the EMS. As a best practice, match IBM Spectrum Scale code levels between the IBM ESS and protocol nodes, but this practice is not mandatory.

Note: All protocol nodes in a cluster must be at the same code level.

Configuration 2: 5105-22L protocol nodes ordered stand-alone and added to an existing IBM ESS 5000

In this configuration, protocol nodes are ordered for attachment to a previously installed IBM ESS. The EMS node, I/O server nodes, and protocol nodes have the OS, kernel, systemd, network manager, firmware, and OFED kept in synchronization through the EMS. The EMS is used to manage and coordinate these levels. As a best practice, match IBM Spectrum Scale levels between the IBM ESS and protocol nodes, but this practice is not mandatory.

Note: All protocol nodes in a cluster must be at the same code level.

5105-22L protocol node hardware

With the IBM ESS 5.3.1.1 release, a protocol node feature code is introduced. This protocol node feature code allows the purchase of POWER9 processor-based nodes with a specific hardware configuration, which is tested and tuned by IBM to provide CES. The machine type and model (MTM) for protocol nodes is 5148-22L, and it has the following hardware configuration.

- 5105-22L POWER9 model.
- 2 x 20-core 2.9 Ghz POWER9 processors.
- 192 GB or greater memory.
- Up to five network adapters.
- As a best practice, plan for the GPFS admin or daemon network to use separate network adapters from the CES.

Figure 3-3 shows the IBM ESS (PPC64LE) with protocol nodes cabling diagram.

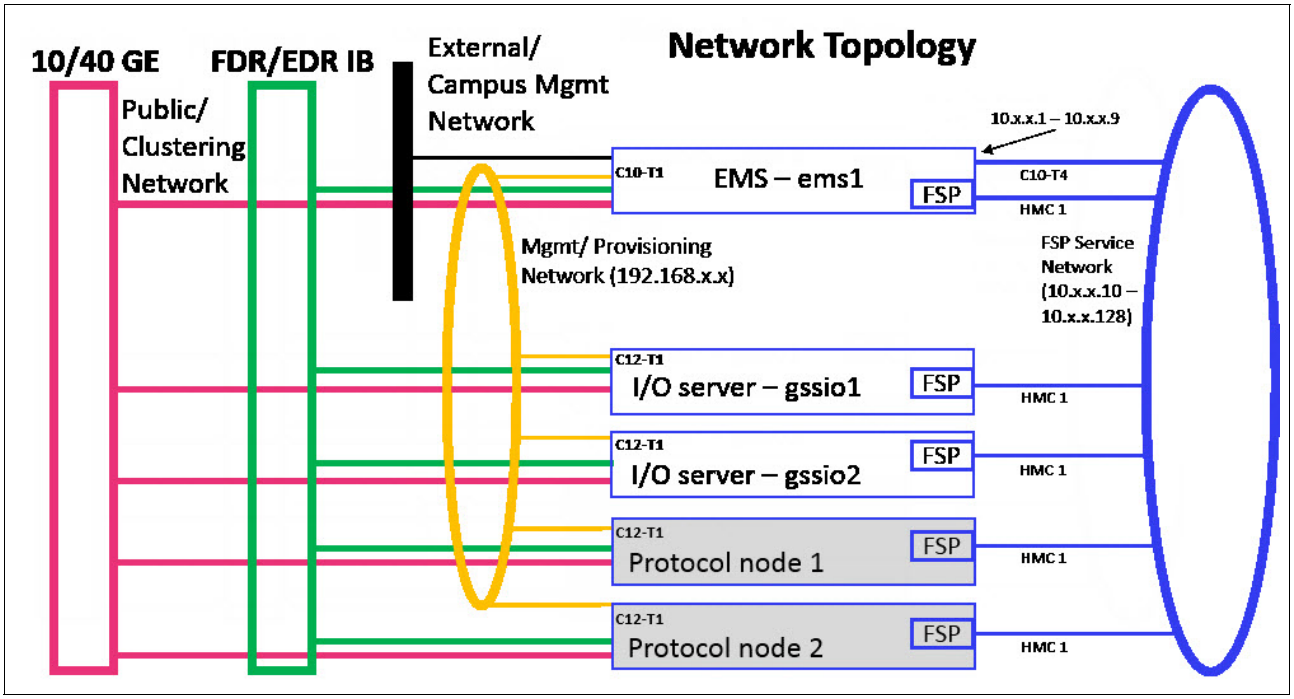


Figure 3-3 IBM ESS (PPC64LE) with protocol nodes cabling diagram

IBM ESS 5000 protocol node deployment by using the IBM Spectrum Scale installation toolkit

The following guidance is for adding a protocol node after storage deployment in an IBM ESS 5000 environment.

Note: The following instructions for protocol node deployment by using the installation toolkit are just an example scenario. For more information, see the following topics:

- “Installing IBM Spectrum Scale on Linux nodes with the installation toolkit” at IBM Knowledge Center.
- “Configuring the CES and protocol configuration” at IBM Knowledge Center.

The following prerequisites should be met:

- During file system creation, adequate space is available for the CES shared root file system.
- The IBM ESS 5000 container has the protocol node management IP addresses defined.
- The IBM ESS 5000 container has the CES IP addresses defined.

File system creation: Adequate space available for the CES shared root file system

In a default IBM ESS setup, you can use the Ansible based file system task to create the recovery groups, VDisk sets, and file system. By default, during this task, 100% of the available space might be consumed. If you plan to include protocol nodes in your setup, you must leave enough free space for the required CES shared root file system.

To adjust the amount of space that is consumed, use the **`--size`** flag, for example:

```
# essrun -G ess_ppc64le filesystem --suffix=-hs --size 80%
```

Running this command leaves approximately 20% space available for the CES shared root file system or more VDisks. If you are in a mixed IBM ESS 3000 and IBM ESS 5000 environment, you might not use the **`essrun`** file system task due to more complex storage pool requirements. In that case, when using **`mmvdisk`**, make sure that you leave adequate space for the CES shared root file system. The CES shared root file system requires around 20 GB of space for operation.

IBM ESS 5000 container: Protocol node management IP addresses defined

Before running the IBM ESS 5000 container, make sure to add the protocol node management IP addresses to `/etc/hosts`. These IP addresses are given to the IBM SSR through the TDA process and they are already set. The customer must define hostnames and add the IP addresses to the EMS node `/etc/hosts` file before running the container.

You also must define the HS IP addresses and hostnames. The IP addresses are set when running the Ansible network bonding task, but the hostnames and IP addresses must be defined in `/etc/ hosts` before the container starts. The HS hostnames must add a suffix of the management names. The IP addresses are user definable. Consult the network administrator for guidance.

For example:

```
# Protocol management IPs 192.168.45.23 prt1.localdomain prt1 192.168.45.24
prt2.localdomain prt2

# Protocol high-speed IPs 11.0.0.4 pr1-hs.localdomain prt1-hs 11.0.0.5
pr2-hs.localdomain prt2-hs
```

Note: `localdomain` is an example domain. The domain must be changed and also match that of the other nodes.

IBM ESS 5000 container: CES IP addresses defined

The final items that must be defined before starting the IBM ESS 5000 container are the CES IP addresses. The following example shows the usage of two IP addresses per node over the HS network. For best practices, see the IBM Spectrum Scale documentation at IBM Knowledge Center.

```
11.0.0.100 prt_ces1.localdomain prt_ces1
11.0.0.101 prt_ces2.localdomain prt_ces2
11.0.0.102 prt_ces3.localdomain prt_ces3
11.0.0.103 prt_ces4.localdomain prt_ces4
```

Instructions for deploying protocol nodes in an IBM ESS 5000 environment

The starting state in the example scenario is as follows:

- The IBM ESS storage is deployed and configured.
- Adequate space (approximately 20 GB) is available for the CES shared root file system.
- The protocol node required hostnames and IP addresses are defined on the EMS before starting the container.
- You are logged in from the IBM ESS 5000 container.

From the IBM ESS 5000 container, complete the following steps:

1. Ping the management IP addresses of the protocol nodes:

   ```
   # ping IPAdress1,...IPAddressN
   ```

   Each protocol node must respond to the **ping** test by indicating that they have an IP address set and it is on the same subnet as the container.

2. Run the config load task:

   ```
   # essrun -N prt1,prt2 config load -p RootPassword
   ```

   If you have more than one node, you can specify them in a comma-separated list.

3. Create network bonds.

   Note: Make sure that the nodes are connected to the HS switch before doing this step.

   ```
   # essrun -N prt1,prt2 network --suffix=-hs
   ```

4. Install the CES shared root file system:

   ```
   # essrun -G ess_ppc64le filesystem --suffix=-hs --ces
   ```

5. Log out of the container and run the SSH setup on the EMS node:

   a. Press Ctrl + p and then Ctrl + q to exit the container.

   b. Run the following commands for the SSH setup on the EMS node:

   ```
   # mkdir -p /root/pem_key
   # cp /root/.ssh/id_rsa /root/pem_key/id_rsa
   # ssh-keygen -p -N "" -m pem -f /root/pem_key/id_rsa (type yes after running
   this command) ./Spectrum_Scale_Data_Management-5.0.5.1-ppc64LE-Linux-install
   --silent
   # cd /usr/lpp/mmfs/5.0.5.1/installer/ ./spectrumscale setup -s
   EMSNodeHighSpeedIP -i /root/pem_key/id_rsa -st ess
   ```

   c. On the EMS node, find the installation package and run the installer:

   ```
   # ./ Spectrum_Scale_Data_Management-5.0.5.1-ppc64LE-Linux-install --silent
   ```

Note: Start `localrepo_AppStream` and `localrepo_BaseOs` in the protocol nodes before starting the installation. For configuring the repositories, run the **`essrun -G ces_ppc64le update -- offline`** command.

6. On the EMS node, complete the following steps:
   a. Change the directory to the installer directory:
      ```
      # cd /usr/lpp/mmfs/5.0.5.1/installer/
      ```
   b. List the current configuration:
      ```
      # ./spectrumscale node list
      ```
   c. Populate the current cluster configuration in the cluster definition file:
      ```
      # ./spectrumscale config populate -N EMSNodeHighSpeedName
      ```
   d. Designate the admin node:
      ```
      # ./spectrumscale node add EMSNodeHighSpeedIP -a
      ```
   e. Add the protocol node:
      ```
      # ./spectrumscale node add ProtocolNodeHighSpeedIP -p
      ```
   f. Run the installation precheck:
      ```
      # ./spectrumscale install -pr
      ```
   g. Regenerate the SSH keys:
      ```
      # ./spectrumscale setup -s EMSNodeHighSpeedIP -i /root/pem_key/id_rsa -st ess
      ```
   h. Set the port range:
      ```
      # ./spectrumscale config gpfs --ephemeral_port_range 60000-61000
      ```
   i. Run the installation procedure on the node:
      ```
      # ./spectrumscale install
      ```
   j. Configure the export IP pool:
      ```
      # ./spectrumscale config protocols -e CESIP1,CESIP2,...
      ```
   k. Set the CES shared root file system:
      ```
      # ./spectrumscale config protocols -f cesSharedRoot -m CESSharedRootMountPointLocation
      ```
   l. Enable protocols:
      ```
      # ./spectrumscale enable smb nfs hdfs
      ```
   m. Confirm the settings:
      ```
      #./spectrumscale node list
      ```
   n. Run the deployment precheck:
      ```
      # ./spectrumscale deploy --precheck
      ```
   o. Run the deployment procedure on the node:
      ```
      # ./spectrumscale deploy
      ```

Checking whether adding a deployment of a protocol node is successful

To check whether adding a deployment of a protocol node is successful, complete the following steps:

1. The **mmlscluster** command now shows that the node is added to the cluster:

   ```
   [root@c145f03zn04 ~]# mmlscluster
   ```

   Example 3-7 show the output of the command.

Example 3-7 IBM GPFS cluster information

```
GPFS cluster information
========================
  GPFS cluster name:         test01.gpfs.net
  GPFS cluster id:           10107750888975023342
  GPFS UID domain:           test01.gpfs.net
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
  Repository type:           CCR

 Node  Daemon node name            IP address  Admin node name             Designation
-------------------------------------------------------------------------------------
   1   c145f08zn01-hs.gpfs.net    11.0.0.2    c145f08zn01-hs.gpfs.net    quorum-manager-perfmon
   2   c145f08zn02-hs.gpfs.net    11.0.0.3    c145f08zn02-hs.gpfs.net    quorum-manager-perfmon
   3   c145f05zems06-hs.gpfs.net  11.0.0.1    c145f05zems06-hs.gpfs.net  quorum-perfmon
   4   c145f03zn04-hs.gpfs.net    11.0.0.4    c145f03zn04-hs.gpfs.net    perfmon
```

2. Check whether the CES IP is assigned by running the **mmlscluster --ces** command:

   ```
   [root@c145f03zn04 ~]# mmlscluster --ces
   ```

   Example 3-8 shows the output of the command.

Example 3-8 IBM GPFS cluster information

```
GPFS cluster information
========================
  GPFS cluster name:         test01.gpfs.net
  GPFS cluster id:           10107750888975023342

Cluster Export Services global parameters
-----------------------------------------
  Shared root directory:                /gpfs/fs5k
  Enabled Services:                     SMB NFS
  Log level:                            0
  Address distribution policy:          even-coverage

Node   Daemon node name            IP address       CES IP address list
-----------------------------------------------------------------------
   4   c145f03zn04-hs.gpfs.net     11.0.0.4         11.0.0.100 11.0.0.101
```

3. Check the health of the node by running the **mmhealth** command:

   ```
   [root@c145f03zn04 ~]# mmhealth node show
   ```

Example 3-9 shows the output of the command.

Example 3-9 Checking the health of the node

```
Node name:      c145f03zn04-hs.gpfs.net
Node status:    TIPS
Status Change:  29 min. ago

Component      Status        Status Change     Reasons
-------------------------------------------------------------------------------------------
GPFS           TIPS          29 min. ago       gpfs_pagepool_small
NETWORK        HEALTHY       29 min. ago       -
FILESYSTEM     HEALTHY       29 min. ago       -
CES            TIPS          2 min. ago        nfs_sensors_not_configured(NFSIO),
smb_sensors_not_configured(SMBGlobalStats, SMBStats)
PERFMON        HEALTHY       17 min. ago       -
THRESHOLD      HEALTHY       17 min. ago       -
```

4. To check the node information, run the **mmces** command:

   ```
   # mmces node list --verbose
   ```

   Example 3-10 shows the output of the command.

Example 3-10 Node information

```
Node Number   Node Name                Node Groups   Node Flags
------------ ------------------------ ------------ ------------
4             c145f03zn04-hs.gpfs.net                none
```

For more information about other command options of the **mmces** command, see IBM Knowledge Center.

3.3 Mixed environment

This section covers several installation scenarios for a mixed environment.

3.3.1 Adding an IBM ESS 5000 to an existing IBM ESS cluster

The following guidance is for adding a stand-alone IBM ESS 5000 building block into an existing IBM ESS cluster or into an existing cluster with an IBM Elastic Storage System 3000 (IBM ESS 3000).

Prerequisites and assumptions

Here are the prerequisites and assumptions:

- An existing IBM ESS or IBM ESS 3000 cluster is connected or reachable to the same HS network block.
- An existing IBM ESS or IBM ESS 3000 and new IBM ESS 5000 is connected or reachable to the same management low-speed network block.
- An IBM ESS 3000 is configured with a POWER8 EMS node and running a Version 6.0.1.0 Podman container. For more information, see “IBM ESS 3000 common setup instructions” at IBM Knowledge Center.

The IBM ESS 5000 contains a POWER9 EMS node that runs a Version 6.0.1.0 Podman container. For more information, see “IBM ESS 5000 Common installation Instructions” at IBM Knowledge Center.

- IBM ESS 5000 nodes are added to `/etc/hosts` and are common across POWER8 EMS and POWER9 EMS:
  - Low-speed names: Fully qualified domain names (FQDNs), short names, and IP addresses.
  - HS names: FQDNs, short names, and IP addresses (add a suffix for low-speed names).
- Hostname and domain are set in POWER9 EMS.
- The latest code for IBM ESS 3000 and IBM ESS 5000 is stored in `/home/deploy` on POWER8 and POWER9 EMS.
- The Linux root password is common across all of the nodes (Legacy, IBM ESS 3000, and IBM ESS 5000).

3.3.2 Adding an IBM ESS 5000 to an IBM ESS Legacy cluster

To add an IBM ESS 5000 to an IBM ESS Legacy cluster, complete the following steps:

1. Run **`config load`** within the IBM ESS 5000 container that is running in the POWER9 EMS to fix the SSH keys across all the nodes, as shown in Example 3-11.

Example 3-11 Running config load with the IBM ESS Legacy cluster

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N
ESS5000Node1,ESS5000Node2,GSSNode1,GSSNode2,ESS5000EMSNode,GSSEMSNode config
load -p RootPassword
```

2. Create bonds in an IBM ESS 5000 building block within the IBM ESS 5000 container running in the POWER9 EMS, as shown in Example 3-12.

Example 3-12 Creating network bonds

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N
ESS5000Node1,ESS5000Node2,ESS5000EMSNode network --suffix=Suffix
```

3. Add IBM ESS 5000 I/O nodes to an existing cluster from ESS5000Node1, as shown in Example 3-13.

Example 3-13 Adding IBM ESS 5000 nodes to the IBM ESS Legacy cluster

```
[root@ESS5000Node1~]# essaddnode -N ESS5000Node1,ESS5000Node2 --cluster-node
GSSEMSNode --nodetype ess5k --suffix=Suffix --accept-license --no-fw-update
```

4. Add an IBM ESS 5000 EMS node to an existing cluster from ESS5000Node1, as shown in Example 3-14.

Example 3-14 Adding an IBM ESS 5000 EMS to an IBM ESS Legacy cluster

```
[root@ESS5000Node1~]# essaddnode -N ESS5000EMSNode --cluster-node ESS5000Node1
--nodetype ems --suffix=Suffix --accept-license --no-fw-update
```

3.3.3 Adding an IBM ESS 5000 to an IBM ESS 3000 cluster

To add an IBM ESS 5000 to an IBM ESS 3000 cluster, complete the following steps:

1. Run **`config load`** within an IBM ESS 5000 container running in the POWER9 EMS to fix the SSH keys across all the nodes, as shown in Example 3-15.

Example 3-15 Running config load with IBM ESS 3000

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N
ESS5000Node1,ESS5000Node2,ESS5000EMSNode,
ESS3000Node1,ESS3000Node2,ESSP8EMSNode config load -p RootPassword
```

2. Create bonds in an IBM ESS 5000 building block within an IBM ESS 5000 container running in the POWER9 EMS, as shown in Example 3-16.

Example 3-16 Creating network bonds

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N
ESS5000Node1,ESS5000Node2,ESS5000EMSNode network --suffix=Suffix
```

3. Add IBM ESS 5000 I/O nodes to an existing IBM ESS 3000 cluster from within an IBM ESS 5000 container that is running in the POWER9 EMS, as shown in Example 3-17.

Example 3-17 Adding IBM ESS 5000 nodes to an IBM ESS 3000 cluster

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N ESS3000Node1 cluster
--add-nodes ESS5000Node1,ESS5000Node2 --suffix=Suffix
```

4. Add an IBM ESS 5000 EMS node to an existing IBM ESS 3000 cluster from within an IBM ESS 5000 container running in the POWER9 EMS, as shown in Example 3-18.

Example 3-18 Adding an IBM ESS 5000 EMS to an IBM ESS 3000 cluster

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N ESS5000Node1 cluster --add-ems
ESS500EMSNode --suffix=Suffix
```

3.3.4 Adding an IBM ESS 5000 system to a mixed IBM ESS Legacy and IBM ESS 3000 cluster

To add an IBM ESS 5000 to a mixed IBM ESS Legacy + IBM ESS 3000 cluster, complete the following steps:

1. Run **`config load`** within an IBM ESS 5000 container that is running in the POWER9 EMS to fix the SSH keys across all the nodes, as shown in Example 3-19.

Example 3-19 Running config load with IBM ESS Legacy + IBM ESS 3000

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N
ESS5000Node1,ESS5000Node2,ESS5000EMSNode,
ESS3000Node1,ESS3000Node2,GSSNode1,GSSNode2,GSSEMSNode config load -p
RootPassword
```

2. Create bonds in an IBM ESS 5000 building block within an IBM ESS 5000 container that is running in the POWER9 EMS, as shown in Example 3-20.

Example 3-20 Creating network bonds

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N
ESS5000Node1,ESS5000Node2,ESS5000EMSNode network --suffix=Suffix
```

3. Add IBM ESS 5000 I/O nodes to an existing IBM ESS Legacy + IBM ESS 3000 cluster from within the IBM ESS 5000 container that is running in the POWER9 EMS, as shown in Example 3-21.

Example 3-21 Adding IBM ESS 5000 nodes to the IBM ESS Legacy + IBM ESS 3000 cluster

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N ESS3000Node1 cluster
--add-nodes ESS5000Node1,ESS5000Node2 --suffix=Suffix
```

4. Add the IBM ESS 5000 EMS node to the existing IBM ESS Legacy + IBM ESS 3000 cluster from within the IBM ESS 5000 container that is running in the POWER9 EMS, as shown in Example 3-22.

Example 3-22 Adding the IBM ESS 5000 EMS to the IBM ESS Legacy + IBM ESS 3000 cluster

```
IBM ESS 5000 CONTAINER [root@cems0 /]# essrun -N ESS5000Node1 cluster --add-ems
ESS500EMSNode --suffix=Suffix
```

3.3.5 Scenario 1: Using IBM ESS 5000 for data NSDs for the existing file system

The scenario described here outlines how to add new IBM ESS 5000 data NSDs to an existing IBM ESS 3000 file system.

Note: There are many possible scenarios, but the two most likely are either adding IBM ESS 5000 NSDs to an existing 5000 file system (building-block addition) or adding IBM ESS 5000 NSDs to an existing IBM ESS 3000 file system.

Assumptions

In this scenario, we start with an existing IBM ESS 3000 file system that is deployed from a POWER8 EMS. The customer initially wanted a fast single pool (data and metadata), but as their business needs grew, it was quickly determined that a large amount of storage was needed for less-frequently accessed files.

The customer decided to purchase an IBM ESS 5000 building-block (for example purposes, an SL model) to extend the existing file system.

High-level plan

Here is the high-level plan:

1. TDA completed, worksheet filled out, and IBM SSR completes code 20.
2. Deploy the IBM ESS 5000 container on the POWER9 EMS.
3. Add the SL1 building-block to the cluster.
4. Use **mmvdisk** to create data NSDs (new failure group, and data pool).
5. Add data NSDs to the existing IBM ESS 3000 file system.
6. Set up a policy file to move data between pools as a best fit.
7. Run healthchecks.

8. Update compDB.
9. Update GUI.

Note: Links to the appropriate documentation and exact steps might be abbreviated.

TDA completed, worksheet filled out, and IBM SSR completes code 20

In this step, the customer has worked with technical sales to decide on the right solution. In our scenario, the customer wants more raw storage without sacrificing performance. They use the IBM ESS 5000 building-block as a new tier to offload less frequently used files from the IBM ESS 3000.

The customer completes the IBM ESS 5000 worksheet and schedules the IBM SSRs arrival. The IBM SSR uses the IBM ESS 5000 HWG to complete code 20. At the end, the IBM ESS is racked, powered, and cabled, the IP addresses are set, and the hardware cleaned.

For more information, see IBM Knowledge Center.

Deploying the IBM ESS 5000 container on the POWER9 EMS

The customer logs in to the EMS and completes the following steps:

1. Checks the version of IBM ESS 5000 that is installed. In our scenario, the IBM ESS was already updated to the latest level that is available on Fix Central. The EMS already had the latest IBM ESS 5000 .tar file in `/home/deploy` ready to use.
2. The customer finds the required package in `/home/deploy` and downloads the *IBM Elastic Storage System 5000 Version 6.0.1.1: Quick Deployment Guide*.

Before adding the SL to the cluster, complete the following steps:

1. Customer updates `/etc/hosts` by adding the new IBM ESS 5000:
   - FQDNs on the low-speed network.
   - FQDNs on the HS network (suffix of low speed).
   - FQDN of the container (over the management interface).
2. Customer stops the existing IBM ESS 3000 container and cleans up the network bridges.
3. Customer extracts the IBM ESS 5000 TGZ file and runs the binary files, which automate the following items:
   a. Extracts the .tar file contents.
   b. Runs **`essmkyml`**, which does the following actions:
      - Confirms the EMS FQDN.
      - Checks the container hostname.
      - Checks the container FSP IP address.
4. The container image is installed.
5. Container bridges (management and FSP) are created.
6. The container is started and entered.

Adding the SL1 building block to the cluster, creating NSDs, extending the file system, creating the policy file, updating compDB, and updating the GUI

Note: Example node names are:

- `ems1,ess5k1,ess5k2`
- `ess5k1-hs,ess5k2-hs`
- `ems1-hs`

Once they are inside the cluster, the customer runs the following example sequence.

1. Run **config load** to exchange the SSH keys:

   ```
   essrun -N ems1,ess5k1,ess5k2 config load -p <password>
   ```

2. Run the network bond setup:

   ```
   essrun -N ess5k1,ess5k2 network --suffix=-hs
   ```

3. Add nodes to the existing cluster:

   ```
   ssh ess5k1
   essaddnode -N ess5k1,ess5k2 --cluster-node <existing ess30000 node> --nodetype
   ess5k --accept-license
   ```

4. Create the **mmvdisk** nodeclass:

   ```
   mmvdisk nc create --node-class ess5k_ppc64le_mmvdisk -N ess5k1-hs,ess5k2-hs
   ```

5. Configure the **mmvdisk** nodeclass:

   ```
   mmvdisk server configure --nc ess5k_ppc64le_mmvdisk --recycle one
   ```

6. Create Recovery Groups.

7. Define a vdiskset:

   ```
   mmvdisk vs define --vs vs_fs5k_1 --rg ess5k_rg1,ess5k_rg2 code 8+2p --bs 16M
   --ss 80% --nsd-usage dataOnly --sp data
   ```

   Note: We assume that the original IBM ESS 3000 file system has a 16 MB blocksize.

8. Create a vdiskset:

   ```
   mmvdisk vs create --vs vs_fs5k_1
   ```

9. Add the vdiskset to the existing IBM ESS 3000 file system:

   ```
   mmvdisk fs add --file-system fs3k --vdisk-set vs_fs5k_1
   ```

   Note: We assume that the original IBM ESS 3000 file system name was `fs3k`.

10. Add a policy file, which can be used to move data from the system pool to the data pool when thresholds are met.

    Note: You can also use the GUI to define policies. For more information, see IBM Knowledge Center.

    The following example rule ingests the writes on the IBM ESS 3000 and moves the data to the IBM ESS 5000 when it reaches 75% capacity on the IBM ESS 3000:

    a. Add a callback for automatic movement of data between the pools, as shown in Figure 3-4.

```
mmaddcallback MIGRATION --command /usr/lpp/mmfs/bin/mmstartpolicy --event
lowDiskSpace,noDiskSpace --parms "%eventName %fsName"
```

Figure 3-4 Adding a callback for automatic movement of data between the pools

    b. Write the policy into a file with the content that is shown in Figure 3-5 on page 77.

```
RULE 'clean_system' MIGRATE FROM POOL 'system' THRESHOLD(75,25) WEIGHT(KB_ALLOCATED)
POOL 'data'
```

Figure 3-5 Writing the policy into a file

> **Important:** You must understand the implications of this rule before applying it in your system. When capacity on the IBM ESS 3000 reaches 75%, it migrates files (larger ones first) out of the system pool to the data pool until the capacity reaches 25%.

11. Run healthchecks:

```
.mmhealth node show -a
gnrhealthcheck
```

12. On the EMS, update compDB:

```
mmaddcompspec default --replace
```

Now, add the IBM ESS 5000 nodes to the pmsensors list and use the **Edit rack components** option in the GUI to slot the new nodes into the frame.

3.3.6 Scenario 2: Using the IBM ESS 5000 to create a file system

Creating a file system with IBM ESS 5000 requires the IBM ESS 5000 container to be running.

After the container is running and the cluster and recovery groups are created, the user can create the file system by running the **essrun** command:

```
$ essrun -G ess_ppc64le filesystem --suffix=-hs
```

> **Note:** By default, this command attempts to use all the available space. If you need to create multiple file systems or a CES shared root file system for protocol nodes, consider using less space.
>
> For example:
>
> ```
> $ essrun -G ess_ppc64le filesystem --suffix=-hs --size 80%
> ```

For CES deployment, the IBM ESS 5000 system should have a CES file system. To create the CES file system, run the following command:

```
$ essrun -G ess_ppc64le filesystem --suffix=-hs --name cesSharedRoot --ces
```

> **Note:** A CES and other file systems can coexist on the same IBM ESS cluster.

4

Use cases

This chapter reviews the major use cases for IBM Spectrum Scale and IBM Elastic Storage System 5000 (IBM ESS 5000).

This chapter covers the following topics:

- 4.1, “IBM ESS 5000 use cases overview” on page 80
- 4.2, “Large capacity and data lake workloads” on page 81
- 4.3, “Analytics and high-performance workloads” on page 86

4.1 IBM ESS 5000 use cases overview

There are three main use case segments for IBM Spectrum Scale and Elastic Storage Server. This chapter describes each of these segments:

- Enterprise data lakes and specific industry applications

  The IBM ESS 5000 high scalability and high-performance hard disk drive (HDD) storage is used for IBM Spectrum Scale clusters that are used in enterprise data lakes and data oceans. Using IBM Spectrum Scale high performance and high scalability, enterprises can build petabyte-level common enterprise data platforms to underpin enterprise central repositories of unstructured file and analytics data. IBM ESS 5000 provides the HDD storage that has enterprise levels of support, and reliability, availability, and serviceability (RAS) with extreme performance. Specific industry applications that might be supported include SAP and serial-attached SCSI (SAS) Grid.

  These use cases are described in 4.2, “Large capacity and data lake workloads” on page 81.

- Storage for high-performance computing (HPC), artificial intelligence (AI), analytics, and high-performance workloads

  The IBM ESS 5000 provides high-performance HDD models for IBM Spectrum Scale clusters running data-intensive HPC, big data and AI applications, and big data analytics (such as Hadoop or Cloudera).

  These use cases are described in 4.3, “Analytics and high-performance workloads” on page 86.

- Data optimization and resiliency

  IBM Spectrum Scale and IBM ESS 5000 are ideal solutions for enterprise data optimization and data management. These data optimization use cases are infrastructure prerequisites for the use case categories in the two previous bullets in this list. Any enterprise data lake, industry apps, or storage for AI and analytics applications require the cost-effective data management and data optimization use cases that IBM Spectrum Scale and IBM ESS 5000 provide, including:

  – Archive (which includes cyberresiliency, that is, “air gap”).
  – Information Lifecycle Management (ILM), which is tiering of data in an enterprise data lake).
  – Backup and restore (high-speed (HS) backup and HS restore for data protection).

  These use cases are described in 4.4, “Data optimization and resiliency” on page 93.

All three of these use case segments apply across all industries and many different scenarios. All these use cases are in production usage today with IBM ESS.

Tip: For more information and an overview of IBM Spectrum Scale fundamentals and how IBM ESS fits into a IBM Spectrum Scale system, see *Introduction Guide to the IBM Elastic Storage System*, REDP-5253 at http://www.redbooks.ibm.com/abstracts/redp5253.html.

4.2 Large capacity and data lake workloads

In this section, we review use cases for large capacity and enterprise data lakes.

The IBM ESS 5000 high scalability and high-performance HDD storage is used for IBM Spectrum Scale clusters that are used in enterprise data lakes and data oceans. Using IBM Spectrum Scale high performance and high scalability, enterprises can build petabyte-level common enterprise data platforms to underpin enterprise central repositories of unstructured file and analytics data. IBM ESS 5000 provides the HDD storage that has enterprise levels of support, and RAS with extreme performance. Specific industry applications that might be supported include SAP and serial-attached SCSI (SAS) Grid.

4.2.1 Large capacity use cases

IBM Spectrum Scale can store and manage hundreds of petabytes of data and billions of files, all in a common global namespace. The IBM ESS 5000 systems are IBM Spectrum Scale storage building blocks.

Thus, IBM ESS 5000 performance and scalability grows in a linear fashion by simply adding more IBM ESS 5000 storage building blocks. IBM Spectrum Scale automatically handles the distribution and management of the files across one or multiple IBM ESS 5000 systems. Different-sized IBM ESS building blocks can be grouped into tiers of storage in which the data is automatically managed by the IBM Spectrum Scale policy engine.

Figure 4-1 shows an example of a IBM Spectrum Scale storage system with one IBM ESS 3000 and two IBM ESS 5000 systems.

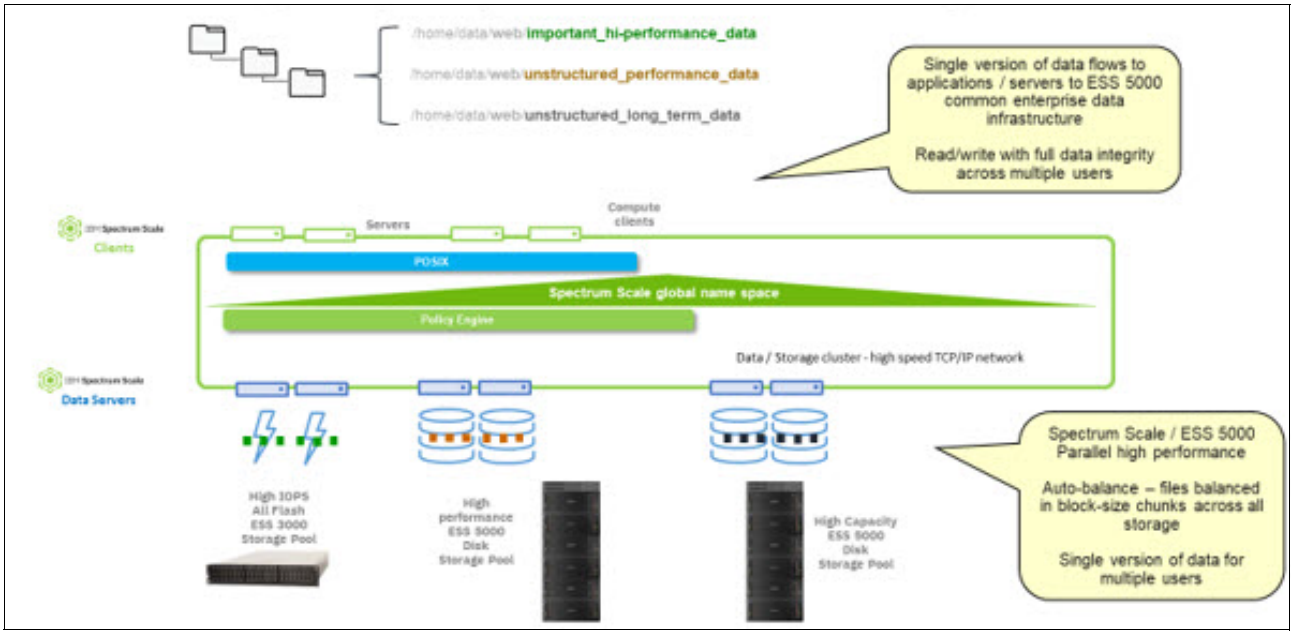


Figure 4-1 IBM Spectrum Scale storage system with one IBM ESS 3000 and two IBM ESS 5000 systems

Figure 4-1 shows an IBM Spectrum Scale IBM ESS 5000 common enterprise data infrastructure. IBM Spectrum Scale provides full read/write data integrity across multiple users and multiple IBM ESS 5000 systems. IBM Spectrum Scale automatically distributes file data across one or multiple IBM ESS 5000 systems in a storage pool, which provides parallel high performance for users and load balances the workload.

Because of the IBM Spectrum Scale parallel architecture, the IBM ESS 5000 capacity can be dynamically expanded. Figure 4-2 shows an example of dynamically added extra IBM ESS 5000 systems to each of the two HDD IBM Spectrum Scale storage pools.

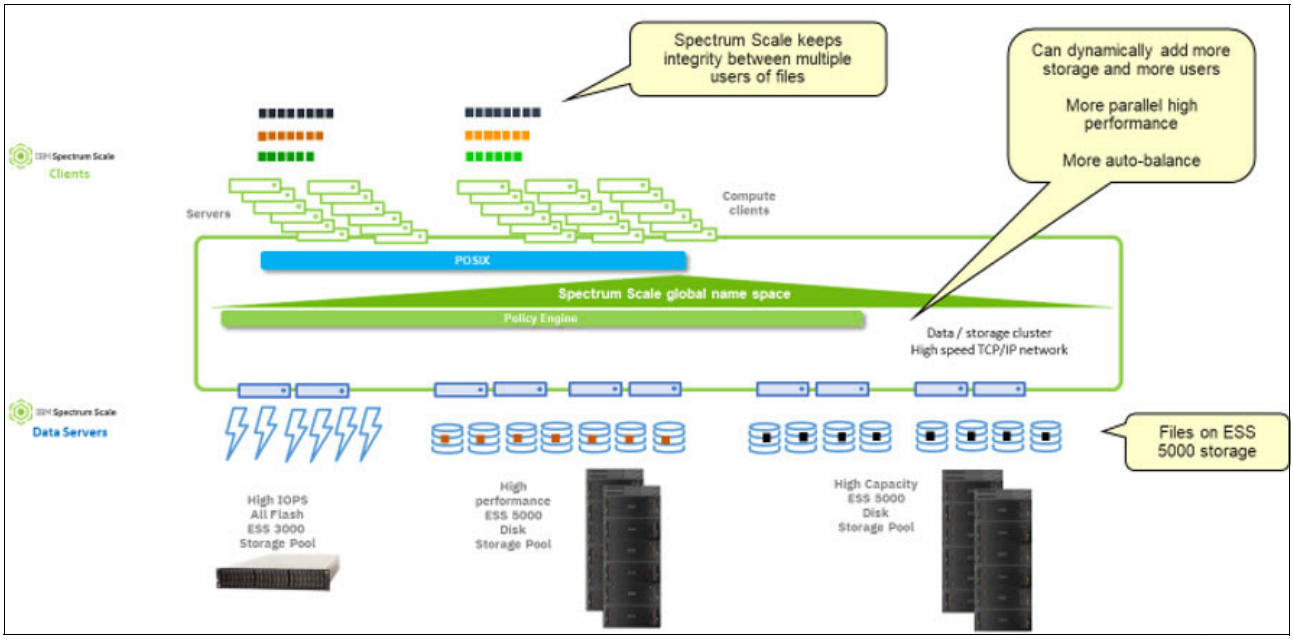


Figure 4-2 IBM ESS 5000 large capacity workloads: Scalable growth 1

By adding more IBM ESS 5000 capacity, we also added more POWER9 NSD servers and HDDs. IBM Spectrum Scale automatically increases the parallelism and continues to load balance the workload across all NSD servers and HDDs in the storage pool. We can continue to add more IBM ESS 5000 systems, as shown in Figure 4-3.

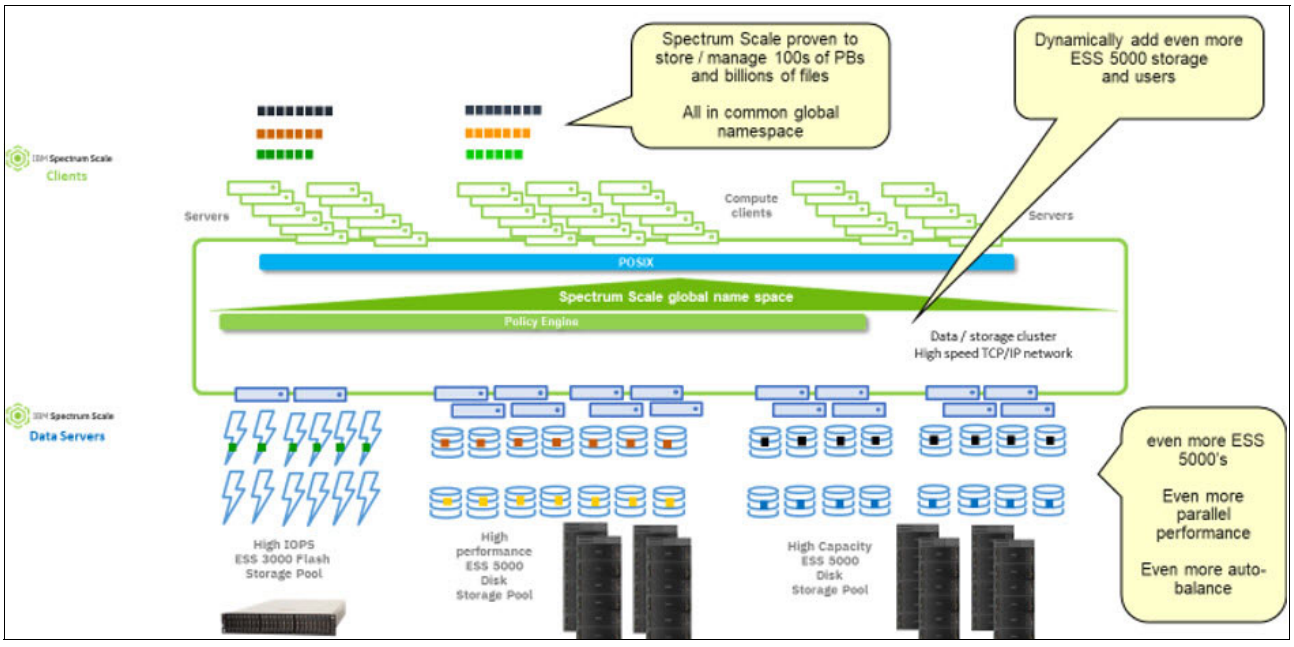


Figure 4-3 IBM ESS 5000 large capacity workloads: Scalable growth 2

IBM Spectrum Scale and IBM ESS 5000 are designed to provide linear performance and scalability by using the architecture of IBM Spectrum Scale. IBM Spectrum Scale has many proven instances of IBM Spectrum Scale and IBM ESS running hundreds or thousands of nodes in enterprise production usage, and tens or thousands of petabytes of IBM Spectrum Scale storage.

4.2.2 Data lake use case

Section 4.2.1, “Large capacity use cases” on page 81 showed how IBM Spectrum Scale and the IBM ESS 5000 are used together to provide an enterprise-scale large capacity storage system. This section describes how that use case enables enterprise data lakes. In fact, IBM Spectrum Scale provides functions, scalability, performance, data resiliency, and data protection to provide a common enterprise data platform.

A data lake is a *common enterprise data platform*, which is a central enterprise data repository that is used to provide the following items:

- Single copy of data: High performance with full data integrity access by multiple users.
- Unified data access: Access a single copy of data through multiple protocols, including POSIX, NFS, Server Message Block (SMB), Object, or Hadoop.
- Data management: Ability to manage data at petabyte scale with automated policies.

The key technology components that must work together as an integrated solution to provide a common enterprise data platform are listed in Figure 4-4.

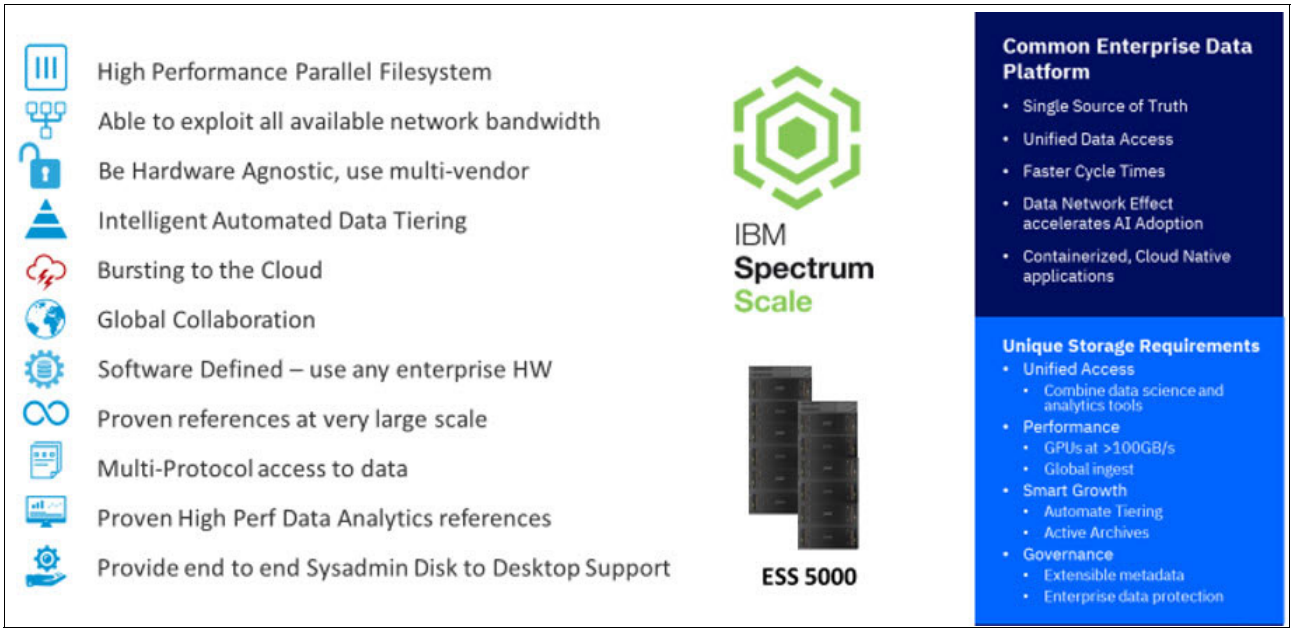


Figure 4-4 Common Enterprise Data Platform that is provided by IBM Spectrum Scale and IBM ESS

IBM Spectrum Scale and the IBM ESS 5000 provide each of these components and features that are required for a common enterprise data platform:

- A high-performance parallel file system.
- Ability to use all available network bandwidth for throughput.
- Hardware-neutral so that they can coexist with vendor products.
- Intelligent automated data tiering.
- Bursting to the cloud.

- Provide global collaboration.
- Be software defined: Can coexist with other enterprise storage hardware.
- Proven references at large scale.
- Multi-protocol access to data.
- Proven high-performance data analytics references.
- Provide end to end enterprise service and support.

Figure 4-5 show an IBM Spectrum Scale and IBM ESS 5000 enterprise data lake enterprise common data platform.

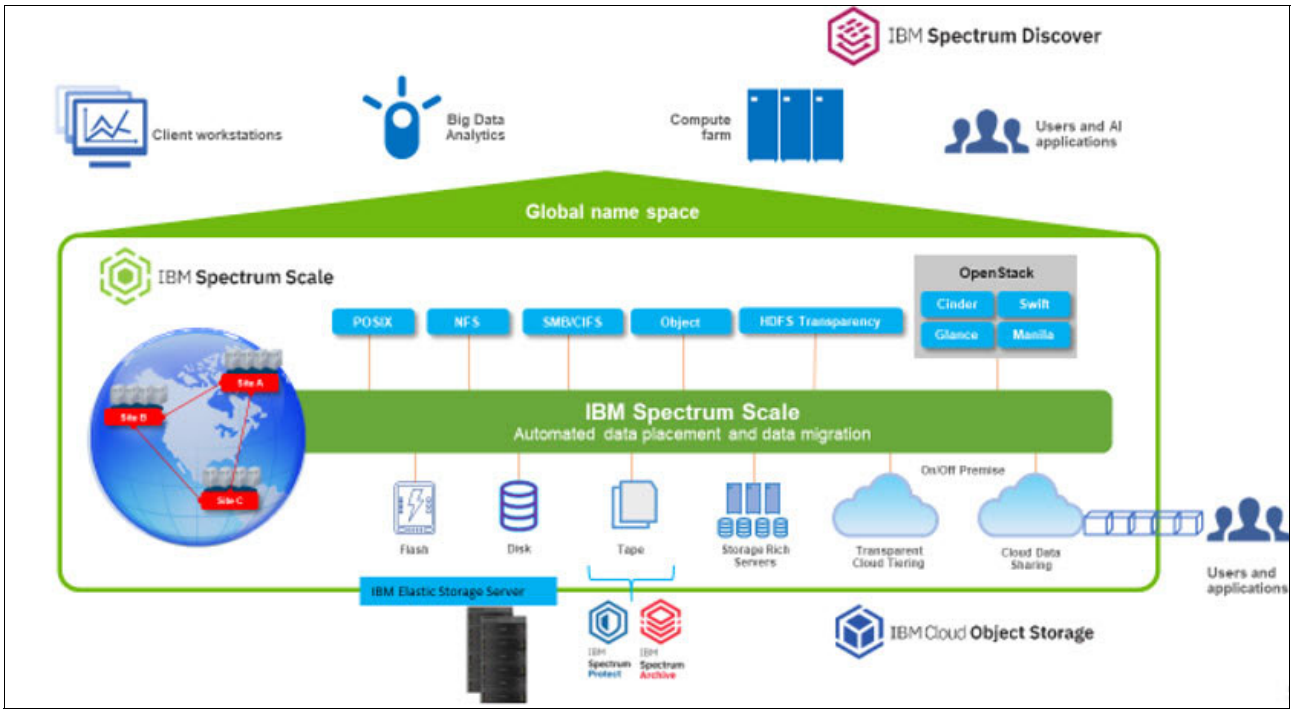


Figure 4-5 Diagram of an IBM Spectrum Scale and IBM ESS 5000 enterprise data lake enterprise common data platform

From a technical perspective, we can physically diagram the many functions that IBM Spectrum Scale and IBM ESS 5000 provide to the enterprise data lake, as shown in Figure 4-6 on page 85.

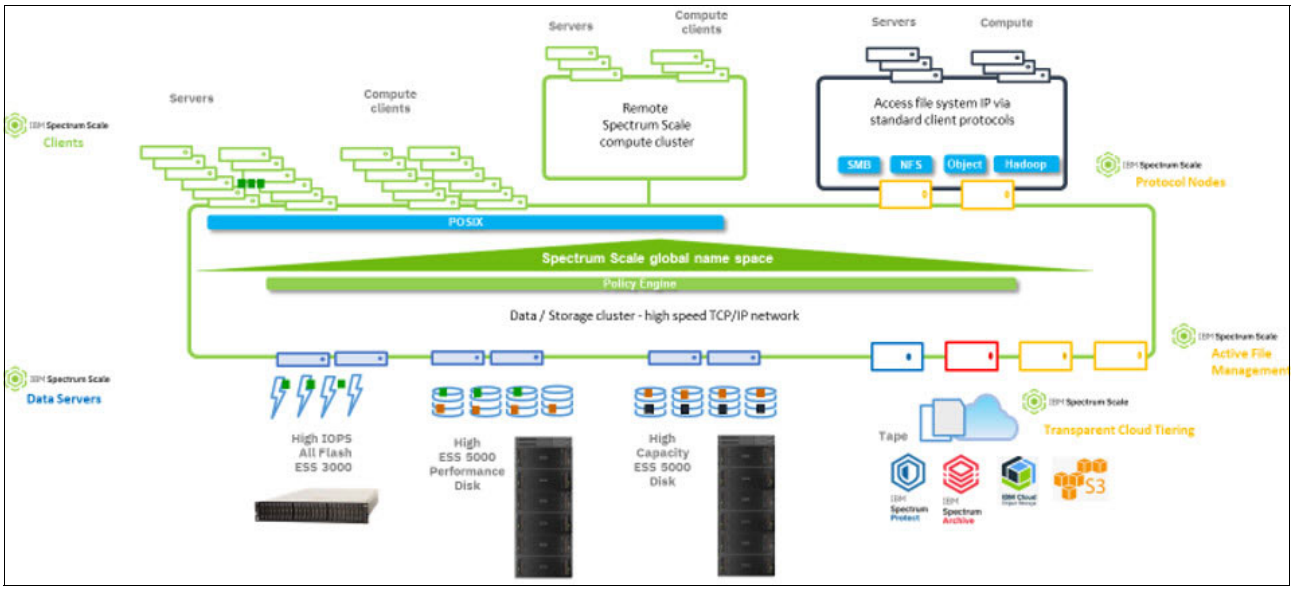


Figure 4-6 Data lake use case

In this use case, IBM Spectrum Scale and IBM ESS provide the following items to the enterprise data lake:

- High-performance parallel file system shared storage platform: Provides an end-to-end collaborative common enterprise data platform for big data analytics and AI workflows.
- Simplified management: Supports modernizing and combining workflows on a single common enterprise data platform:
  - A single global namespace can share content over high-performance and house networks.
  - A global namespace can span multiple IBM Spectrum Scale instances and multiple sites.
  - Significant reduction in file duplication.
- Intelligent automatic tiering: Intelligent automatic tiering of data internally between flash drives and HDD, and externally to tape, object, and cloud resources:
  - Delivers cost-effective data economics by automatically managing and tiering data to different classes of storage assets.
  - Reserves HS storage for work in progress, and moves everything else transparently to lower-cost archive storage.
  - Maintains visibility and access to migrated and archived content.
- Multi-protocol access to data:
  - Supports industry standard network file sharing protocols.
  - SMB, NFS, S3 Object, and Hadoop Distributed File System (HDFS).
  - High-performance parallel file system for extreme performance for real-time enterprise analytics.

- Global collaboration with IBM Spectrum Scale Active File Management:
  - File system caching and a single namespace view across geographically distributed remote sites.
  - No costly WAN acceleration required.
  - Extend collaborative workflows to wherever best suits the work required and talent available.
- Proven limitless scale:
  - Single namespace for data, physically tiered across flash, disk, tape, object, and new technology.
  - Seamless expansion and upgrades.
  - Automated management reduces administration burden.
- Unified IBM services and support:
  - System level enterprise 24x7 service and support.
  - Direct access to IBM Spectrum Scale expertise.
  - Single point of contact for all support issues.

The following sections review in more detail how IBM Spectrum Scale and IBM ESS can provide the physical data optimization and resiliency to the enterprise large capacity workloads and enterprise data lake use cases.

4.3 Analytics and high-performance workloads

IBM Spectrum Scale is flexible and scalable software-defined file storage for analytics workloads. Enterprises around the globe deploy IBM Spectrum Scale to form large data lakes and content repositories to perform HPC and analytics workloads. It can scale performance and capacity without bottlenecks. IBM Spectrum Scale solves the challenge of explosive growth of unstructured data against a flat IT budget. IBM Spectrum Scale provides unified file and object software-defined storage (SDS) for high-performance, large-scale workloads, and it can be deployed on-premises or in the cloud.

IBM Spectrum Scale is POSIX compatible, so it supports various applications and workloads. Traditional systems and analytics systems use and share data that is hosted on IBM Spectrum Scale file systems. Hadoop and Spark services can use a storage system to save IT costs because no special-purpose storage is required to perform the analytics. IBM Spectrum Scale features a rich set of enterprise-level data management and protection features that include snapshots, ILM, compression, and encryption, all which provide more value than traditional analytic systems do.

For more information, see *IBM Spectrum Scale: Big Data and Analytics Solution Brief*, REDP-5397, found at:

http://www.redbooks.ibm.com/abstracts/redp5397.html

There are four main use cases for the analytics and high-performance workloads segment:

- HPC and data-intensive technical computing
- Big data and analytics with Hadoop
- Storage for AI: Machine learning and deep learning (ML/DL)
- Genomics medicine workloads in IBM Spectrum Scale

We describe these use cases in the next sections.

4.3.1 HPC and data-intensive technical computing

Having laid a common enterprise data platform foundation as described in 4.2.2, “Data lake use case” on page 83, it is now straightforward to run an enterprise data-intensive technical computing strategy. All the enterprise data that might be needed tor operate data-intensive technical computing can be sourced and managed by using this platform.

Here are the customer requests, IBM solution, and customer benefits for this use case:

- Customer requests: The representative customer requests for data-intensive technical computing are as follows:
  - High-performance sequential throughput for leading-edge and emerging Industry analytics applications, such as oil, gas, healthcare, genomics, physics research, defense, intelligence, and electronic design automation (EDA).
  - Up to a multi-petabyte data scale, low cost, extreme performance, and data manageability.
- IBM solution: The IBM solution for these customer requests is as follows:
  - IBM Spectrum Scale software providing a high-performance enterprise shared data infrastructure that is accessible through multiple protocols.
  - An Elastic Storage Server as a packaged storage solution for petascale supercomputing.
- Customer benefits: The customer benefits are as follows:
  - Combining diverse structured and unstructured data types for high-performance, zero latency analytics for business insight.
  - Premium performance, premium scale, and premium manageability.
  - IBM ESS with IBM Spectrum Scale RAID providing premium data reliability, premium data integrity, and consistent persistent performance.

Figure 4-7 shows the variety of data-intensive computing environments that are in production usage today on IBM Spectrum Scale.

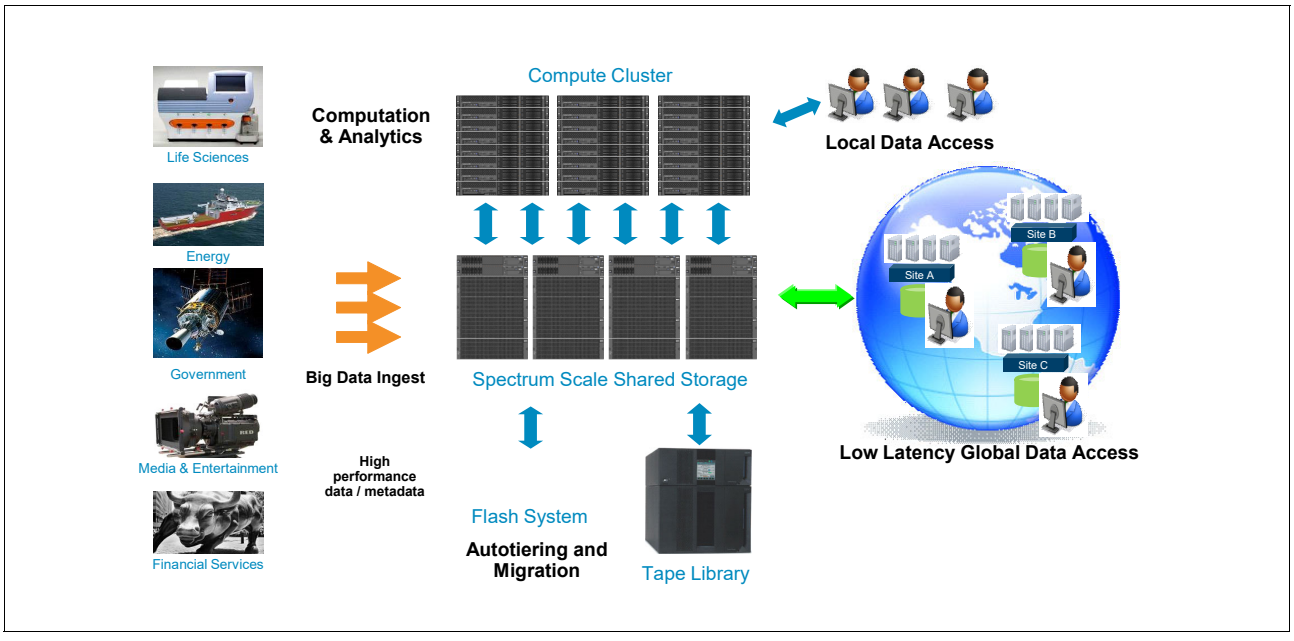


Figure 4-7 HPC and data-intensive computing

Figure 4-8 shows what the variety of data-intensive computing environments all have in common: They must be scalable and cost-effective, and receive high performance from their data and common enterprise data platform to meet their wide variety of technical computing objectives.

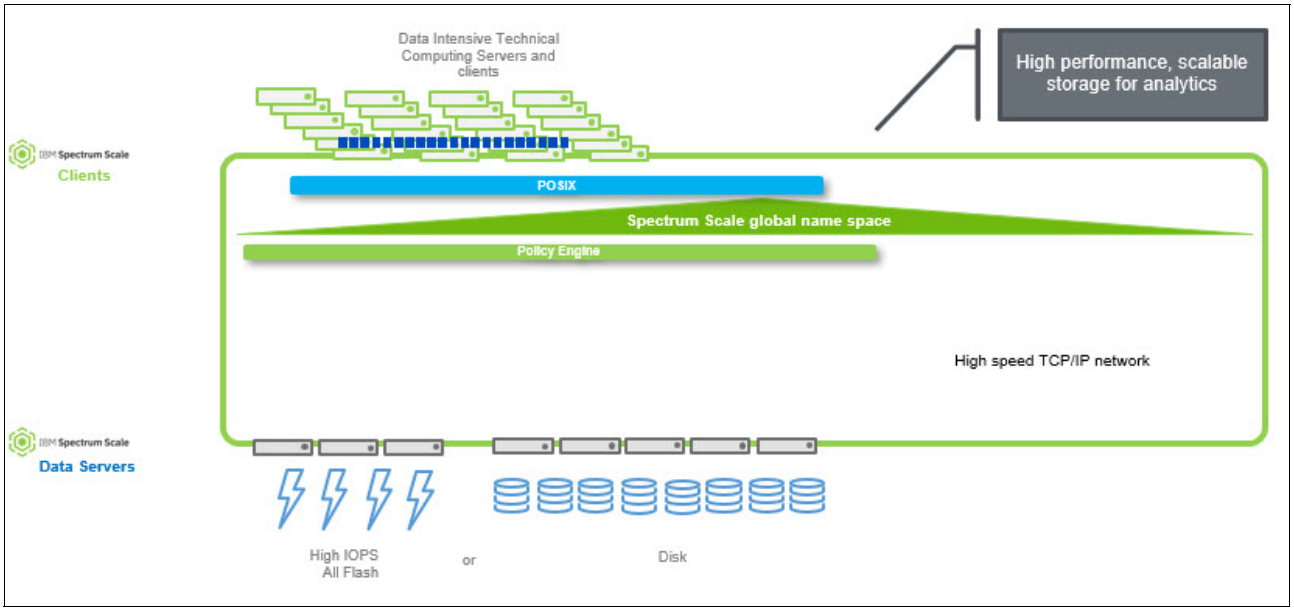


Figure 4-8 Data-intensive technical computing

4.3.2 Data and analytics with Hadoop

Here are the customer requests, IBM solution, and customer benefits for this use case:

- Customer requests: The representative customer requests for data and analytics with Hadoop are as follows:
  - Hadoop or HDFS environments usually require time-consuming onloads and offloads of data to and from other data sources.
  - Hadoop or HDFS setups might be inefficient in terms of data copies with sometimes up to eight copies of data.
  - Lack of redundancy and enterprise RAS.
  - Usability and ease-of use.
- IBM solution: The IBM solution for these customer requests is as follows:
  - Instead of HDFS being the only data repository, use the IBM Spectrum Scale common enterprise data platform for Hadoop applications.
  - By using this platform, you integrate existing Hadoop nodes into one common architecture.
- Customer benefits: The customer benefits are as follows:
  - One file system with a global namespace for easy scalability and efficient metadata mechanisms mean fast analysis on large volumes of data and a shorter time to better results.
  - Integrated backup and automated policy-driven tiering means less administrative expense.
  - No reinvestment is needed.

- IBM proven RAS.
- No need to change APIs or program interfaces (Java binding), and no hidden costs.
- Lower investment in raw storage capacity because it has more efficient data protection (including distributed RAID) than traditional Hadoop.

Figure 4-9 shows Hadoop in an enterprise workflow with IBM Spectrum Scale.

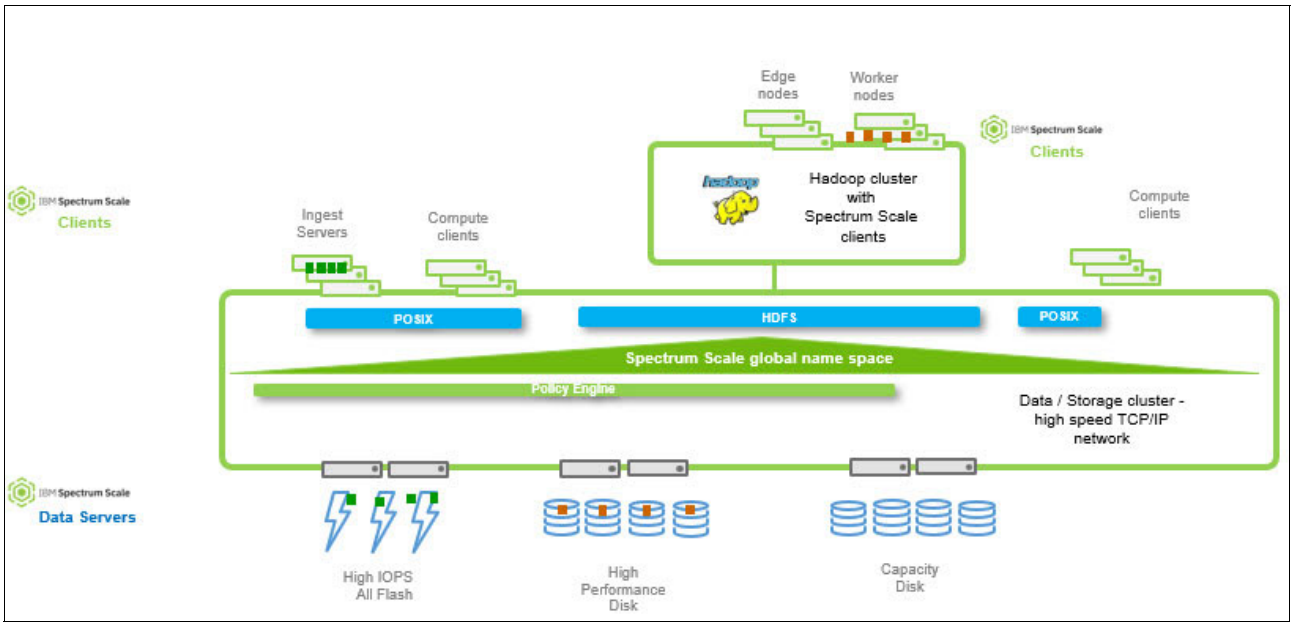


Figure 4-9 Hadoop in an enterprise workflow with IBM Spectrum Scale

Figure 4-10 shows the IBM Spectrum Scale and IBM ESS with Hadoop technical diagram. For more information about this architecture, see *Hortonworks Data Platform on IBM Power with IBM Elastic Storage Server: Reference Architecture and Design Version 1.0*.

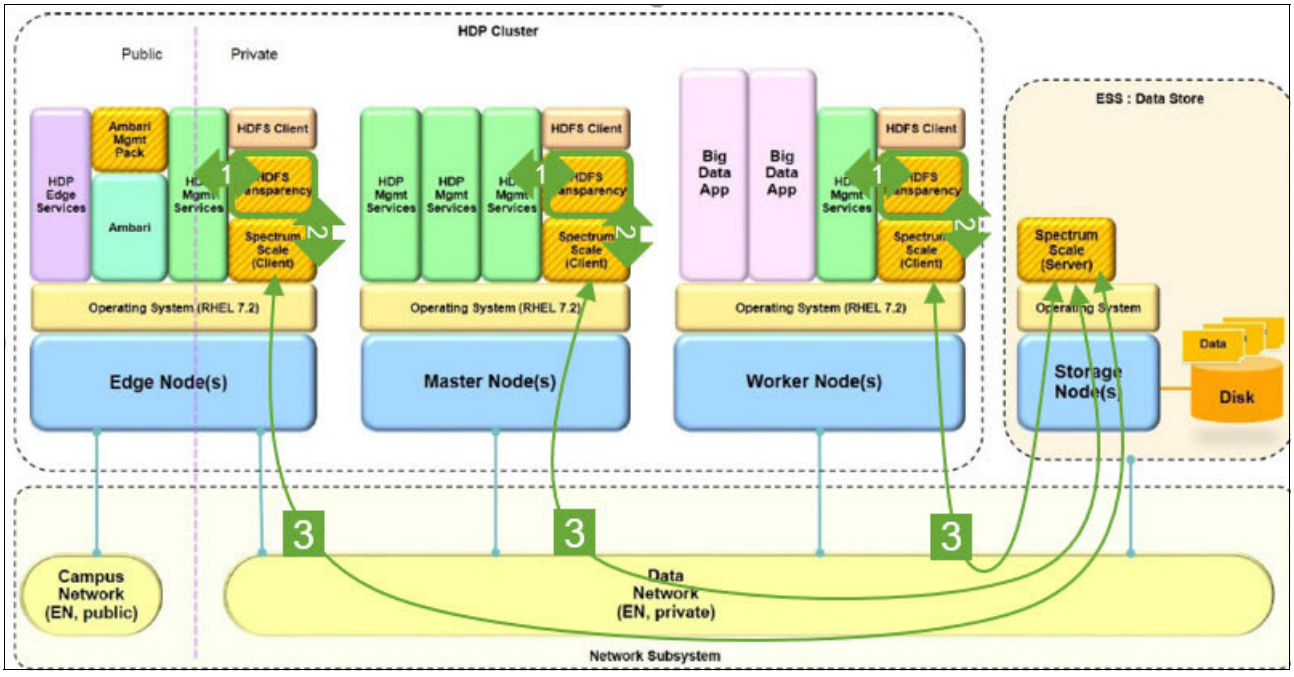


Figure 4-10 IBM Spectrum Scale and IBM ESS with Hadoop technical diagram

4.3.3 Storage for AI: Machine learning and deep learning

Data is the fuel for AI, and AI cannot exist without an information architecture. The best AI is built on a foundation of data that is collected and organized as carefully as it is analyzed, and then infused into the business. Organizations are challenged with gaining insights from their data for many reasons. Data silos make it difficult to get a holistic view of all your information, which limits the value of AI. An infrastructure that was not built for AI is not flexible enough to respond to new demands without adding complexity.

Every successful AI project goes through a multi-step process that starts with having the correct data and progresses to using AI, as shown in Figure 4-11.

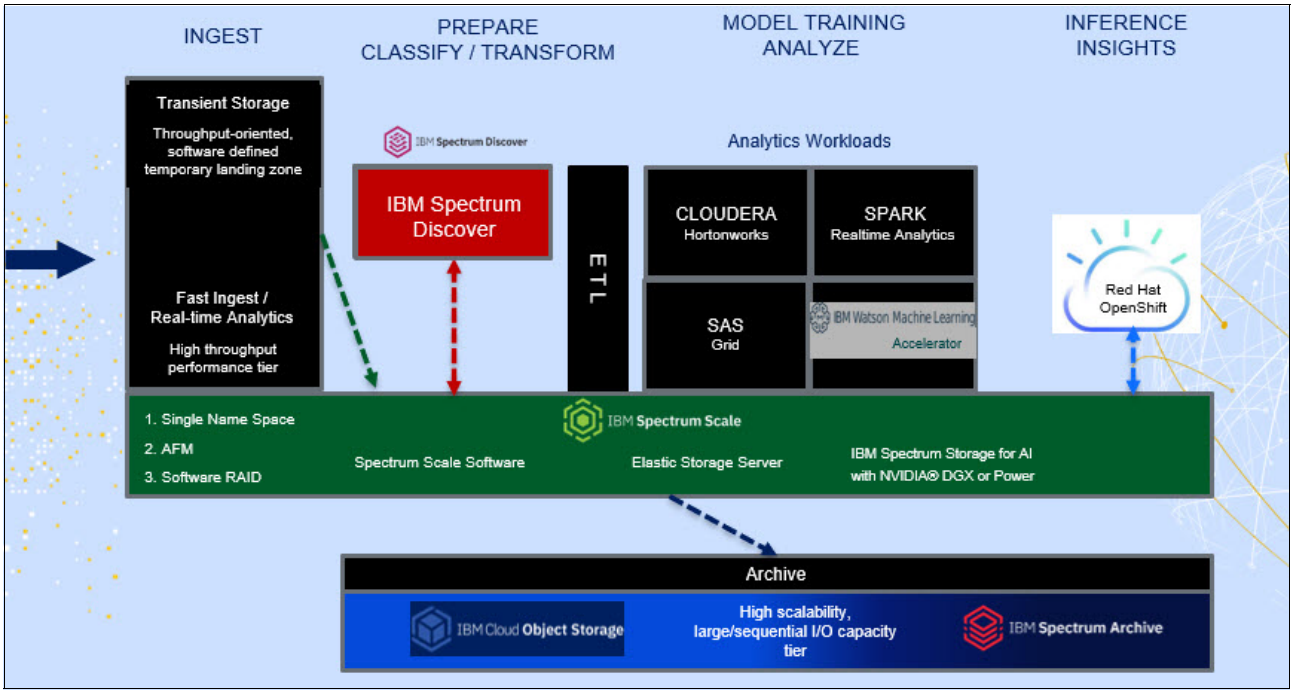


Figure 4-11 IBM Spectrum Storage for AI: The fastest path from ingest to insights

Data goes through the following steps in this architecture:

1. Data comes into the infrastructure from multiple sources out on the edge.
2. The data goes into the first phase of the workflow ingest, which ingests the data into your storage. Ingest or data collection benefits from the flexibility of SDS at the edge, and demands high throughput.
3. The next phase is classify and transform, where the analyst prepares and further identifies the data that ultimately is used in their model.
4. Data from one or more sources is extracted and then copied to the data lake. During this part of the process, the analyst further tags the metadata for a more accurate data set.
5. For the analyst to train their model, they need a scalable shared infrastructure where their multiple analytics workloads can run.
6. After the model run completes, the output must be shared and then archived for potential future use.

IBM ESS High Performance Tier of storage to keep AI Data Pipelines and GPUs running at peak performance

This section shows how the IBM ESS High Performance Tier of storage can keep AI Data Pipelines and GPUs running at peak performance. Figure 4-12 shows the recognized AI Data Pipeline: Ingest, Organize, Analyze, and ML/DL.

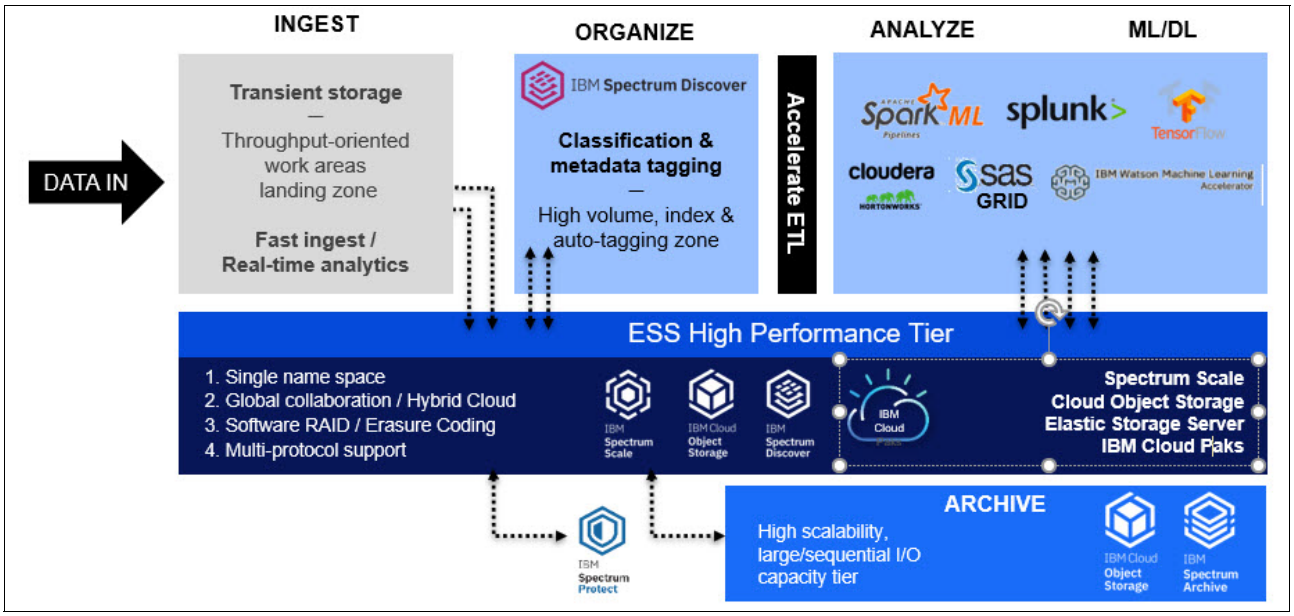


Figure 4-12 AI Data Pipeline: Ingest, Organize, Analyze, and ML/DL

By adding an IBM ESS High Performance Tier to the existing data pipeline, you can *accelerate* the speed at which ingest can happen, the speed at which IBM Spectrum Discover can do classification and metadata tagging, and the speed at which analysis and models of ML/DL can be done.

IBM Spectrum Scale provides a central data lake, which is data repository to source the data to move efficiently through this pipeline, and you also can accelerate the essential backups of this environment. The IBM ESS High Performance Tier accelerates the IBM Spectrum Protect metadata operations and scanning. You can use this high-performance tier as the staging area to drive data out to the archive layers, such as IBM Spectrum Archive or IBM Cloud Object Storage

NVIDIA DGX and IBM ESS reference architecture for AI and ML workloads

NVIDIA and IBM created a reference architecture for NVIDIA DGX and IBM ESS that work on AI and ML workloads. For more information about this reference architecture, see *IBM Spectrum Storage for AI with NVIDIA DGX Systems: Proven Infrastructure Solution for AI workloads*.

Together, NVIDIA and IBM provide an integrated, individually scalable compute and storage solution with end-to-end parallel throughput from flash to GPU for accelerated DL training and inference. The scalable infrastructure solution integrates the NVIDIA DGX-1 systems and NVIDIA DGX-2 systems with IBM Spectrum Scale file storage software, which powers the IBM Elastic Storage Server family of storage systems that includes the new IBM ESS 5000.

The reference architecture covers the linear growth of the AI or ML system of both GPU workloads on the NVIDIA DGX systems. It also demonstrates the linear growth capabilities of 40 GBps per IBM ESS 3000 unit for read random workloads. In the market for AI and ML workloads, systems other than NVIDIA DGX are available, and all can benefit from the outstanding performance capabilities of the IBM ESS 5000 system and IBM Power Systems with GPUs, such as the IBM Power System AC922 system.

Figure 4-13 shows the Storage for AI with NVIDIA DGX scenario.

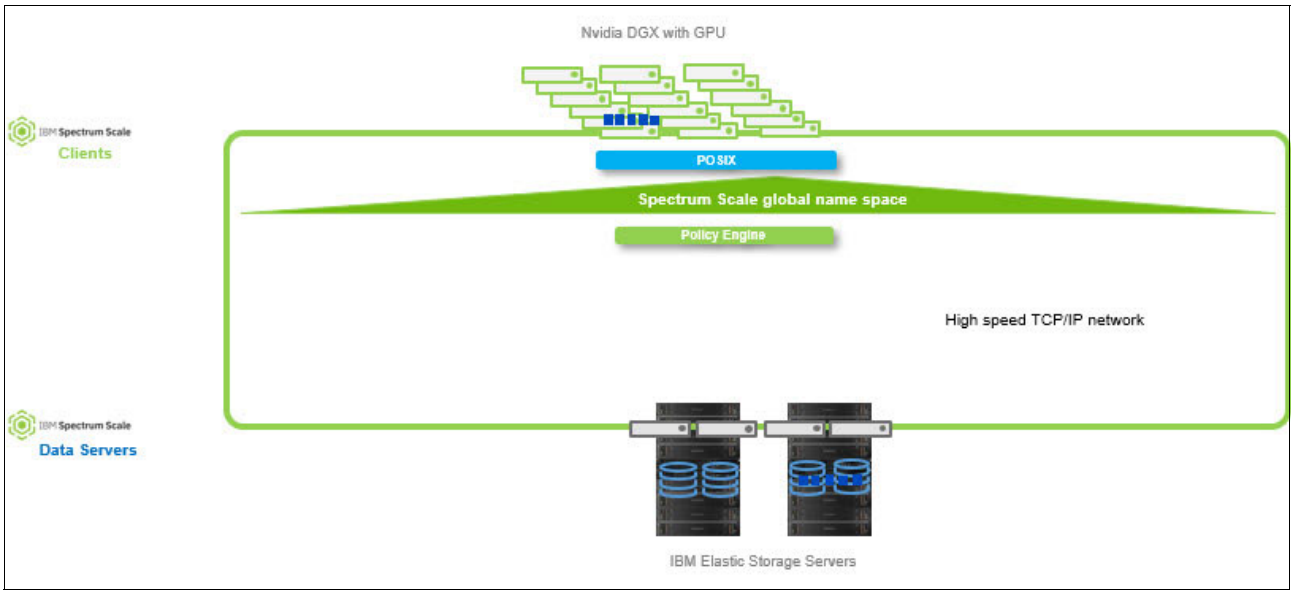


Figure 4-13 Storage for AI with NVIDIA DGX

Genomics medicine workloads in IBM Spectrum Scale

IT administrators, physicians, data scientists, researchers, bioinformaticians, and other professionals who are involved in the genomics workflow need the correct foundation to achieve their research objectives efficiently. Concurrently, they want to improve patient care and outcomes. Thus, you must understand the different stages of the genomics workload and the key characteristics of it.

Advanced genomics medicine customers are outgrowing network-attached storage (NAS). The move from a traditional NAS system or a modern scale-out NAS system to a parallel file system like IBM Spectrum Scale requires a new set of skills. Thus, IBM Spectrum Scale Blueprint for Genomics Medicine Workloads must provide basic background information. It must also offer optional professional services to help customers successfully migrate to the new infrastructure.

For more information, see *IBM Power Systems S922, S914, and S924 Technical Overview and Introduction*, REDP-5497, found at:

`http://www.redbooks.ibm.com/abstracts/redp5479.html`

4.4 Data optimization and resiliency

IBM Spectrum Scale storage with IBM Elastic Storage Server provides the required high-performance tiering (HPT) capability to produce more tiers of storage in an enterprise data lake, and high-performance data ingest, data management, and data archival in data centers.

IBM ESS provides the data infrastructure foundation that is required for data optimization and resiliency, which includes use cases such as:

- Data optimization:
  - Archive
  - ILM
- Resiliency:
  - Backup and restore

IBM ESS is an ideal solution for enterprise data optimization and data management. Typically, these data optimization use cases are infrastructure prerequisites for cost-effective data management in data lakes, industry applications, and *storage for AI and analytics* environments. These use cases are archive (which includes the *cyberresiliency air gap*), (ILM, which is tiering of data in an enterprise data lake), and backup and restore (HS backup and HS restore requirements).

Note: *Air gap* refers to the physical isolation of systems or networks to avoid widespread corruption of data due to malware infection, system failures, or human error.

These three use case segments can be applied across all industries and many different scenarios.

IBM ESS 5000 extends existing IBM ESS environments by providing the following functions:

- ILM, more tiers of storage, and an enterprise data lake (a single store for all enterprise data)
- High-performance data ingest and storage in data centers

Figure 4-14 shows data optimization and resiliency in an IBM ESS 5000 environment.

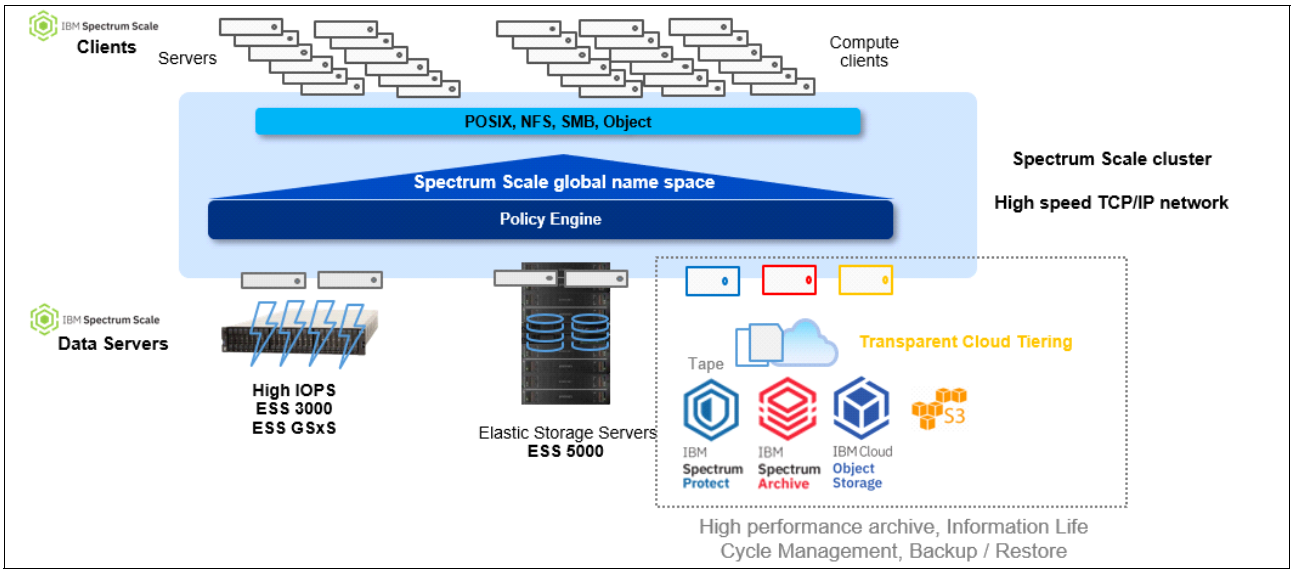


Figure 4-14 Data optimization and resiliency

4.4.1 Archive use case

Figure 4-15 on page 95 shows archiving IBM Spectrum Scale IBM ESS data to tape.

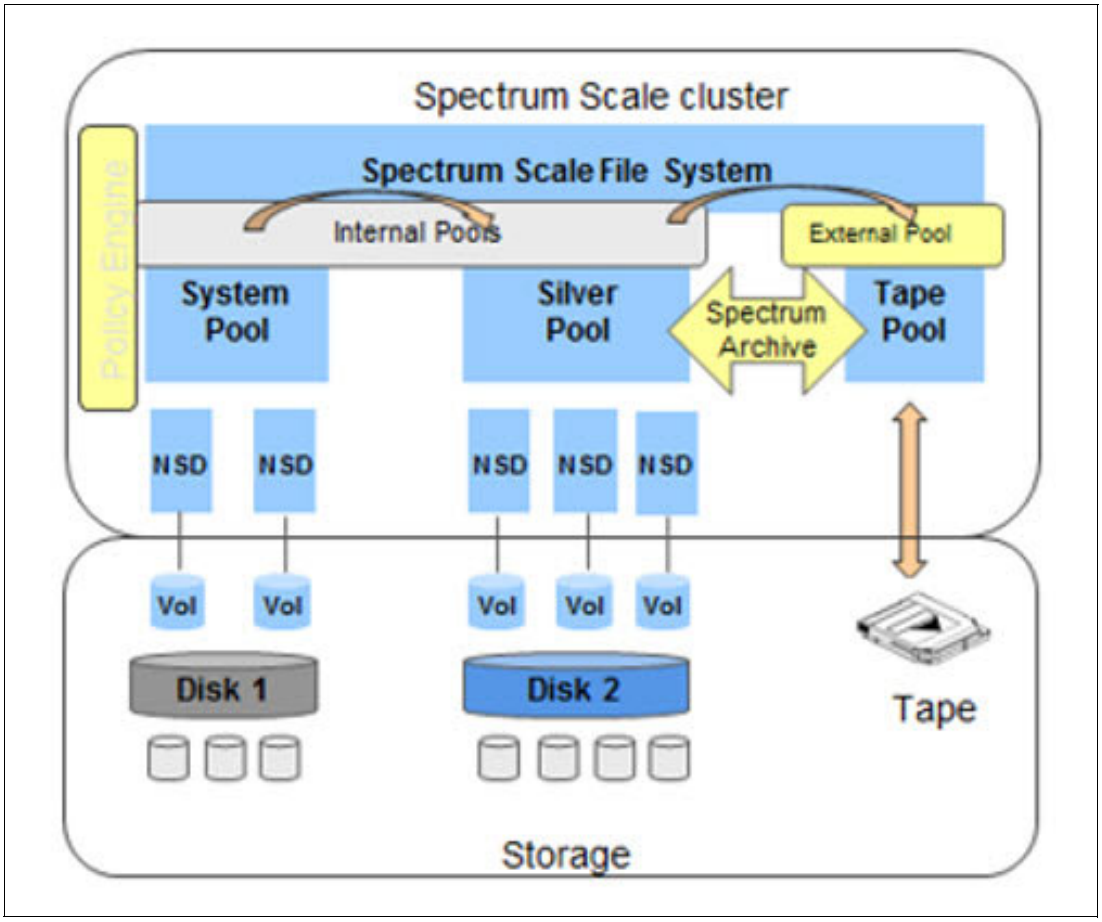


Figure 4-15 Archiving IBM Spectrum Scale IBM ESS data to tape

Note: Disk 1 and Disk 2 in Figure 4-15 refer to IBM ESS 5000 storage.

Here are the characteristics of this use case:

- IBM Spectrum Scale storage such as IBM ESS enables transparent migration of data among flash and HDD storage tiers.
- IBM Spectrum Archive or IBM Spectrum Protect can be configured as a tape tier.
- IBM Spectrum Scale policies can be used to migrate automatically files from disk to tape.
- After migration, the file remains visible in the IBM Spectrum Scale file system and can be quickly recovered

Archiving with IBM Spectrum Protect

Figure 4-16 shows the archive use case with IBM Spectrum Protect.

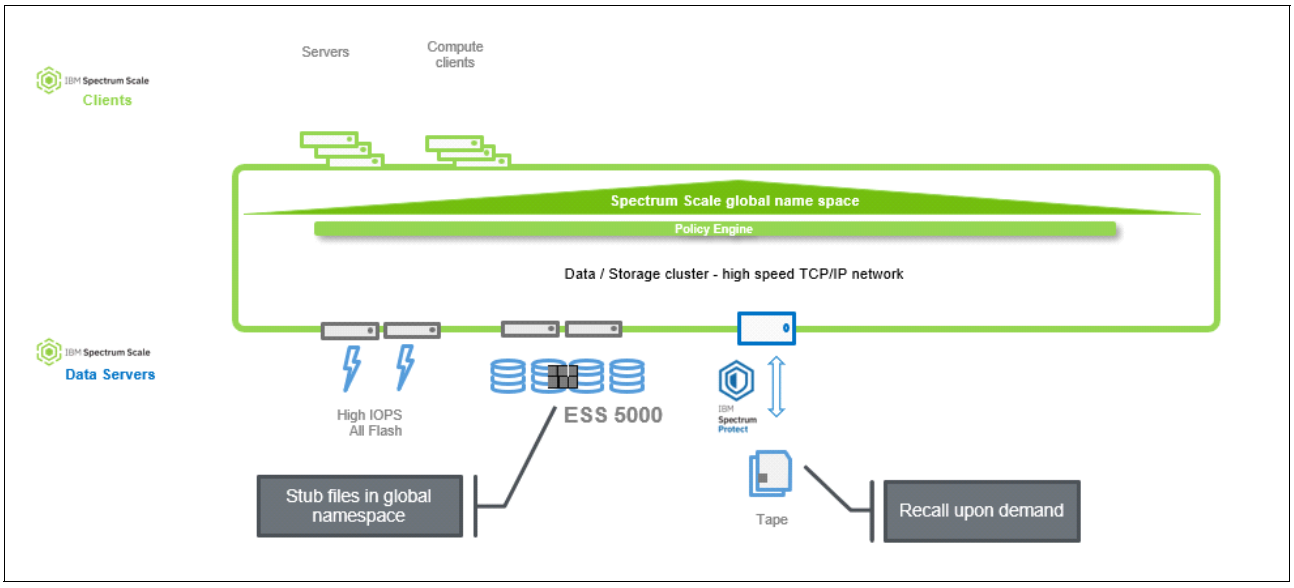


Figure 4-16 Archive with IBM Spectrum Protect

IBM Spectrum Protect interacts with the IBM Spectrum Scale file system on the IBM ESS, and IBM Spectrum Protect provides Hierarchical Storage Management (HSM) for IBM Spectrum Scale data and migrate files to the external storage, usually a tape drive or tape library.

When files are migrated by IBM Spectrum Protect, a stub file remains in the file system, and the files can be recalled on demand or can be recalled by using policies.

When the file is recalled, a copy of the file remains in the archive.

Archiving with IBM Spectrum Archive

Figure 4-17 shows the archive use case with IBM Spectrum Archive.

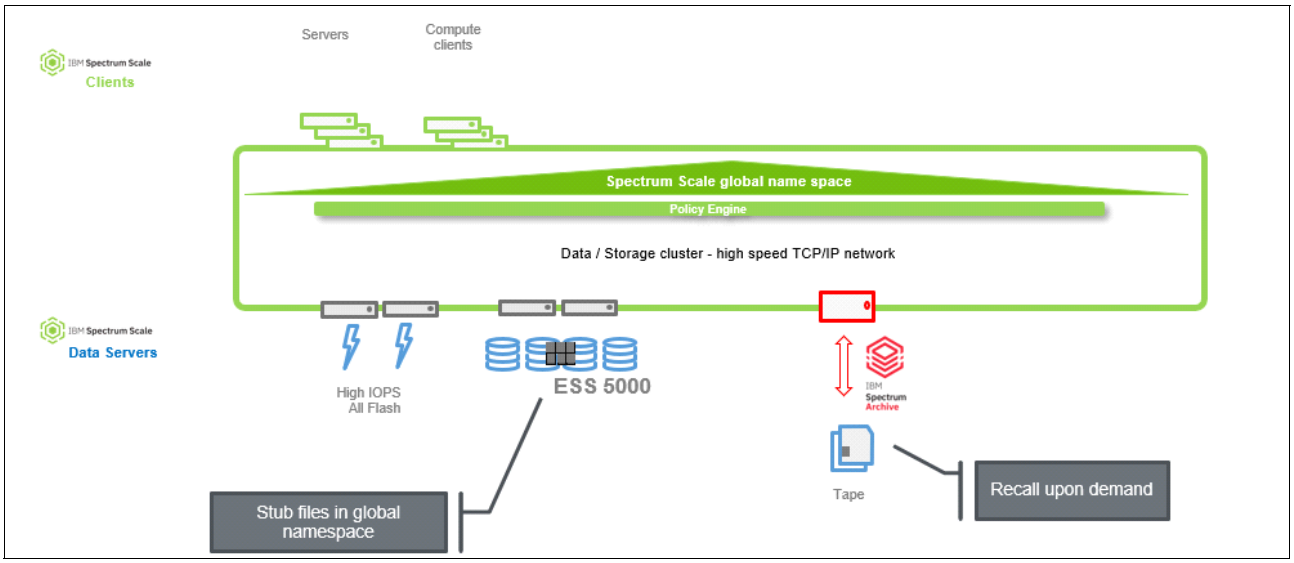


Figure 4-17 Archiving with IBM Spectrum Archive

IBM Spectrum Archive (a different IBM product from IBM Spectrum Protect) interacts with IBM Spectrum Scale storage such as IBM ESS and provides HSM on IBM Spectrum Scale data to migrate the file to the external storage, usually a tape drive or tape library.

When the file is migrated by IBM Spectrum Archive, a stub file remains in IBM Spectrum Scale, and the file can be recalled on demand or can be recalled by using policies.

When the file is recalled, a copy of the file remains in the archive.

IBM Spectrum Archive might be preferable in circumstances where heightened transparency of the file's true location is wanted versus the IBM Spectrum Protect method.

Archiving with Active File Management <-> Cloud Object Server

Archiving with Active File Management <-> Cloud Object Server (also know as AFM Object) is a new function with IBM Spectrum Scale V5.1. It is an enhancement to IBM Spectrum Scale Active File Management function.

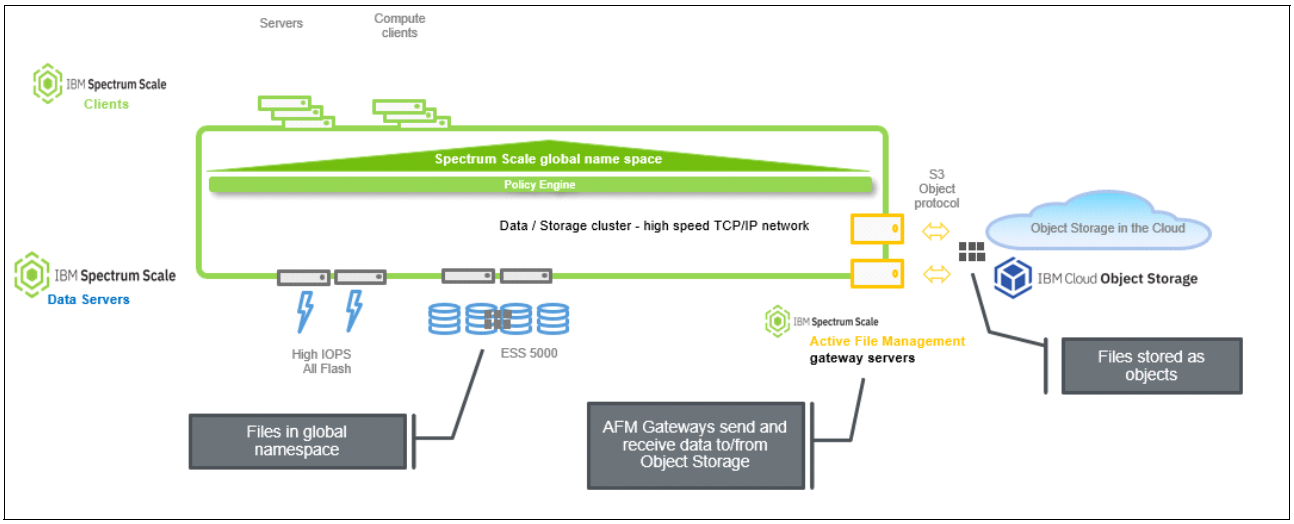


Figure 4-18 AFM Object for archiving or data acceleration

AFM Object for High-Performance Tiering or Data Accelerator for Analytics and AI adds to AFM the ability to cache IBM Spectrum Scale data to and from any object server that supports the S3 Object protocol.

High-performance IBM Spectrum Scale storage can now use AFM caching and HS buffering to do active archiving directly to and from an object storage, which provides data access capabilities between IBM Spectrum Scale and object storage and accelerates data workflows that help machine learning (ML), GPU, and deep learning (DL) applications that must process unstructured data by using GPUs on object storage. Currently, the AFM Object provides caching for files that uses the NFS protocol or IBM Spectrum Scale as remote servers.

With AFM Object, a user can associate an IBM Spectrum Scale file set with a bucket of objects on a cloud object server or other servers that support the S3 protocol. With the file protocol, the file metadata and data are cached in the AFM file set. With this enhancement, AFM Object provides the more sophisticated mapping that is required to associate file semantics with objects. The file set can cache the metadata only or both metadata and data.

There are two main ways to use AFM Object:

- Cache a subset of objects for an application that uses the data for a while and produces some output that is uploaded back to the cloud object server. This use case is the *data acceleration* use case. In this mode, not all file system operations are enabled, mainly for performance reasons.
- Use the full file system mode. In this mode, the application is presented with cloud object server as though it is an extension of the file system. In this mode, performance is not as good as in the first use case due to more communication with the cloud object server.

This section assumes that the reader knows about IBM Spectrum Scale and the AFM feature. For more information about AFM, see IBM Knowledge Center.

Use cases for AFM Object

Figure 4-19 shows use cases for AFM Object.

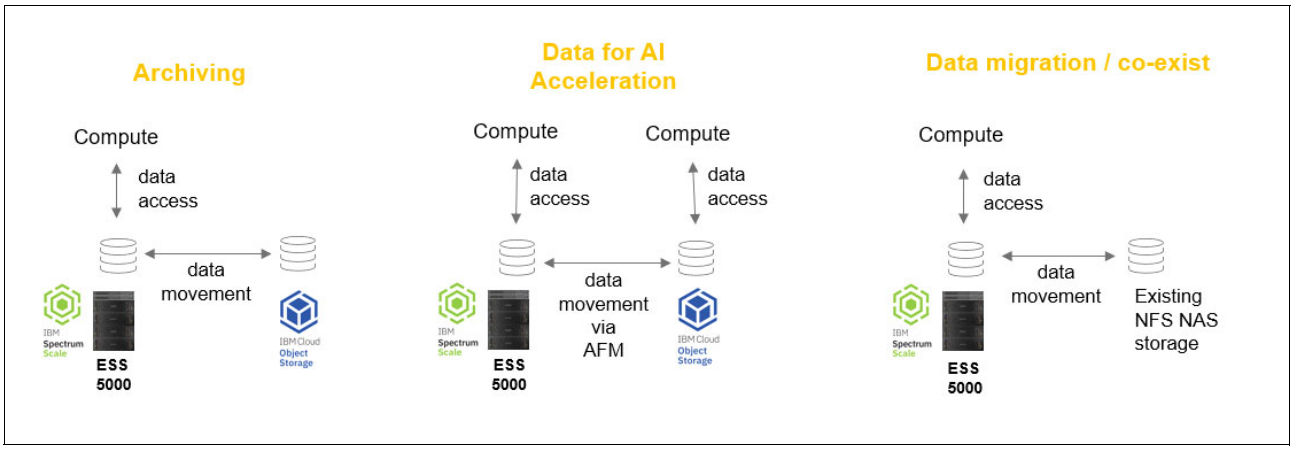


Figure 4-19 Use cases for AFM Object

AFM Object provides new integration capabilities for file and object data. This enhancement enables IBM Spectrum Scale and IBM ESS 5000 to integrate and accelerate data access among IBM Spectrum Scale and external data sources within the enterprise, and to help mitigate *data silos*, which can cause excessive manual copying of data among different business processes.

There are two major use cases for AFM Object:

- Active Archive: Archive data from IBM Spectrum Scale and IBM ESS 5000 to low-cost object storage, with fast recall on demand.
- Data for AI Acceleration: Optimize and automate data movement between IBM Spectrum Scale and IBM ESS 5000 file storage and the enterprise's on-premises or off-premises object storage.

There is a third use case for AFM with IBM Spectrum Scale and IBM ESS 5000 that uses the AFM ability to access standard NFS storage. It has the following functions:

- Migrate data to IBM Spectrum Scale from NFS NAS storage easily for consolidation and lowering overall storage costs.
- Co-exist with existing NAS storage. IBM Spectrum Scale and IBM ESS 5000 connect to and read/write data to and from existing NAS NFS storage.

4.4.2 Information Lifecycle Management

IBM Spectrum Scale storage such as IBM ESS provides an integrated infrastructure for unstructured data management and ILM.

Figure 4-20 shows ILM.

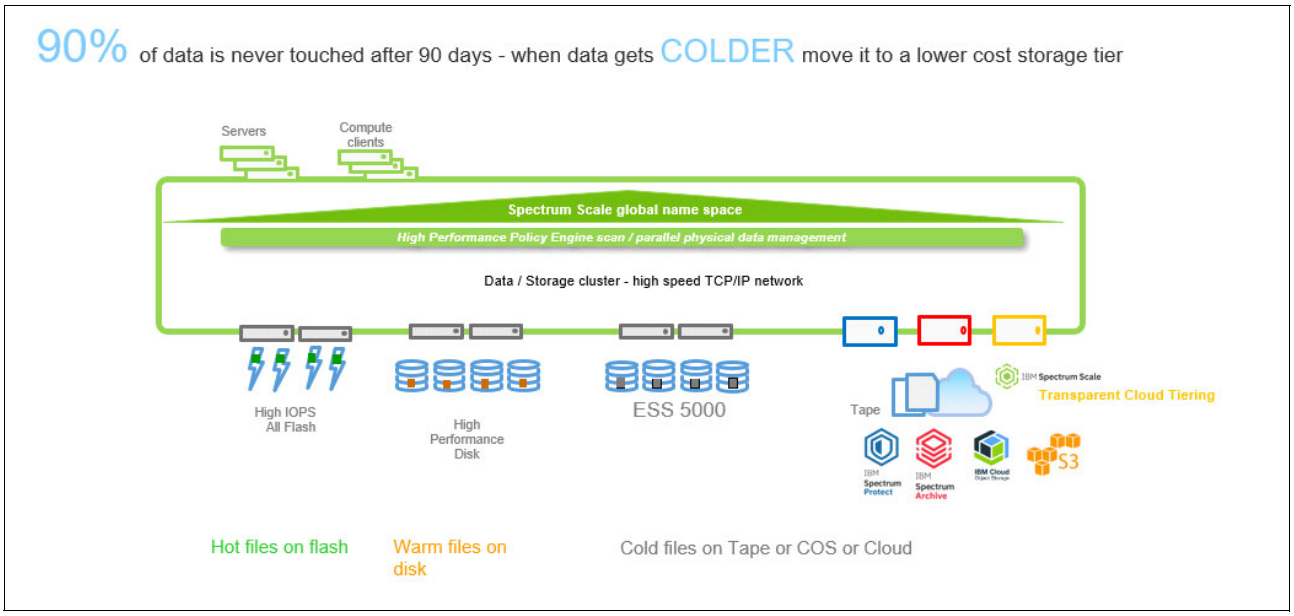


Figure 4-20 Information Lifecycle Management

IBM Spectrum Scale ILM is cost-effective and automatically moves data in the common enterprise data platform to the optimum storage tier. This action helps to optimize costs and can increase access times. Typically, you want hot files to be on flash, warm files on disk, and cold files (which studies show often make up 90% of enterprise data) on tape, IBM Cloud® Object Storage, or the cloud.

IBM Spectrum Scale Policy Engine follows the rules that are given to it. The Policy Engine performs an efficient HS scan of the IBM Spectrum Scale data. This Policy Engine scan is fast, efficient, and capable of running in short periods even when there are petabytes of storage and millions or billions of files.

4.4.3 Resiliency

Resiliency is key to the survival of a digital business.

As companies become digital businesses, they quickly realize that IT resiliency is no longer optional, but a fundamental business requirement. A few of these IT resiliency items include:

- De-stage cold data to an external storage pool (typically tape).
- Release storage capacity in the primary dataspace to create significant disk and flash capacity savings.
- Applications can accommodate a longer time for recall from external storage pools or media.
- Provide physical isolation of systems or networks to avoid widespread corruption of data due to malware infection, system failures, or human error to maximize security.

The cost of a service outage, which can result in missed business opportunities and stalled productivity can be enormous. Even a short outage can cost millions of dollars.

The impact of an outage varies greatly depending on the business. Some businesses might be able to survive an outage of a few hours or a day, while for others, even getting poor performance for several seconds can impact the bottom line.

IT resiliency ensures that the system is running, but it also ensures that entire IT infrastructure stack (compute, operating system, middleware, application, network, and storage layers) should consistently be available to make certain that services and data are accessible. In addition, transactions must be completed successfully and on time with good performance. All regulatory compliance and security requirements should be met.

Failures and cyberattacks do occur, and maintenance eventually must be applied to the IT infrastructure. The only question is; “What will occur and when?” Having immediately available backup through redundant IT components and automated recovery can greatly reduce the duration of outages. Without these measures, you must first identify and fix the problem, and then manually restart the affected hardware and software components.

Redundancy and the technologies that use it are perhaps the most prevalent mechanism to ensure resiliency, which is one of many key strengths of IBM Spectrum Scale and IBM ESS.

IBM Spectrum Scale and IBM ESS are key components of the IBM Storage for Cyber Resiliency and Modern Data Protection solutions. The built-in features of IBM Spectrum Scale, such as pervasive encryption, immutability, and data protection (backup and recovery, snapshots, and replications) contribute to the overall cyber resilience of an enterprise. Working with other IBM Spectrum Storage offerings such as IBM Spectrum Protect, IBM Spectrum Protect Plus, and IBM Spectrum Archive, IBM Spectrum Scale provides the storage foundation to build a resilient enterprise. For example, you can take IBM Spectrum Scale data on IBM ESS and rapidly back it up by using IBM Spectrum Protect, and IBM Spectrum Archive can be used to archive IBM Spectrum Scale data on IBM ESS to tape.

Backup and restore

IBM Elastic Storage System systems centralize data to provide robust HS backup and restore capabilities, including petabyte-scale data integrity:

- Faster backup and restore of business data based on parallel access to either IBM Spectrum Protect or IBM Spectrum Archive.
- More cost-effective and higher utilization of the backup storage target.
- Lower operational load on the IT resource team (automated, policy driven backups, restores, and automated data tiering of backup data).
- Seamless migration of data without downtime, and lower operational risks.
- All the data across all the tiers including tape is available in one global namespace, which makes data recall (recovery) and compliance-driven data discovery fast and efficient.

The ability of IBM Elastic Storage Server to increase linearly throughput and scalability enables this use case to scale without limits. You can add as many IBM Spectrum Scale data servers and backup capacity and performance as the budget and the network allows.

Figure 4-21 shows the backup and restore use case.

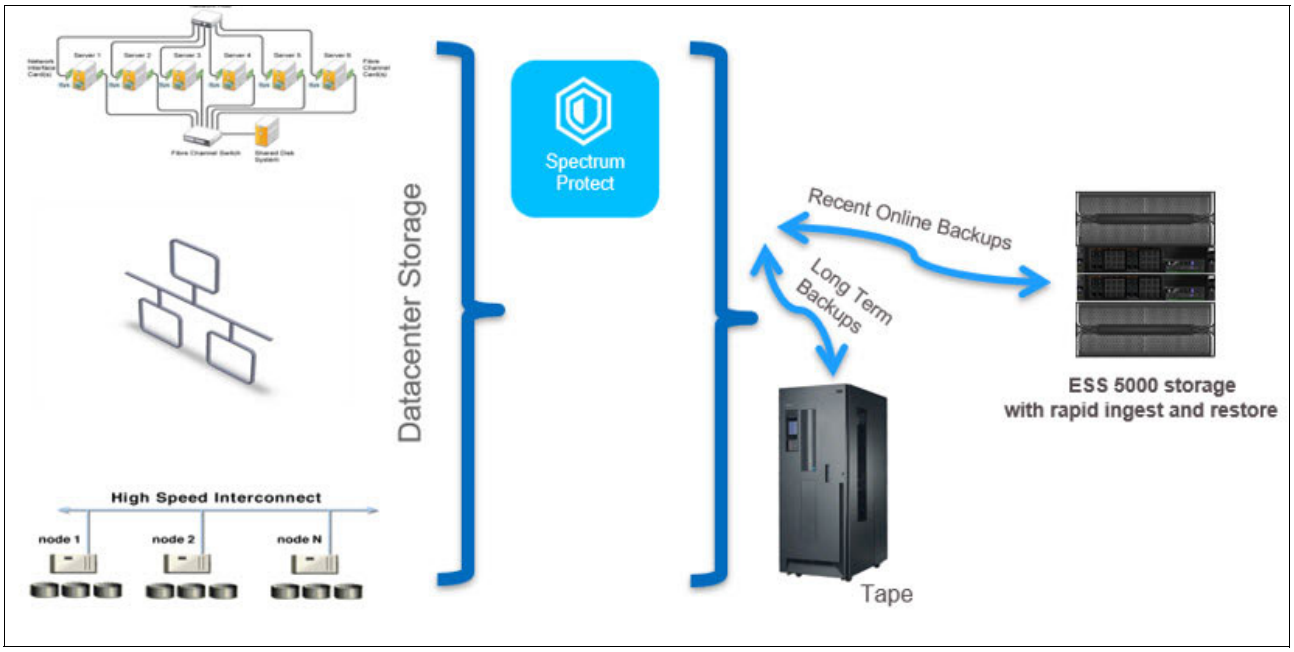


Figure 4-21 Backup and Restore use case

Other features in IBM Spectrum Scale can help improve the resiliency of your data:

- Using replication and failure groups, which are described in IBM Knowledge Center.
- Choosing the VDisk RAID codes based on IBM General Parallel File System (IBM GPFS) file system usage and data requirements, which are described in IBM Knowledge Center.

A

Sample configuration files

This appendix provides the following configuration files:

- “Sample 1 Gb Ethernet configuration file” on page 104
- “Sample IP over InfiniBand configuration files” on page 108

Sample 1 Gb Ethernet configuration file

Example A-1 shows a sample 1 Gb Ethernet configuration file.

Example: A-1 Sample /etc/network/interfaces file

```
source /etc/network/interfaces.d/*.intf
# The loopback network interface
auto lo
iface lo inet loopback
# The primary network interface
auto eth0
#iface eth0 inet dhcp
iface eth0
    address 192.168.45.10/24
    #gateway 192.168.45.1
# EVEN Ports/Lower ports PVID 101 for FSP network
auto swp2
iface swp2
   bridge-access 101

auto swp4
iface swp4
   bridge-access 101

auto swp6
iface swp6
   bridge-access 101
auto swp8
iface swp8
   bridge-access 101

auto swp10
iface swp10
   bridge-access 101

auto swp12
iface swp12
   bridge-access 101

auto swp14
iface swp14
   bridge-access 101

auto swp16
iface swp16
   bridge-access 101

auto swp18
iface swp18
   ridge-access 101

auto swp20
iface swp20
   bridge-access 101
```

```
auto swp22
iface swp22
   bridge-access 101

auto swp24
iface swp24
   bridge-access 101

auto swp26
iface swp26
   bridge-access 101

auto swp28
iface swp28
   bridge-access 101

auto swp30
iface swp30
   bridge-access 101
auto swp32
iface swp32
   bridge-access 101

auto swp34
iface swp34
   bridge-access 101

auto swp36
iface swp36
   bridge-access 101

auto swp38
iface swp38
   bridge-access 101

auto swp40
iface swp40
   bridge-access 101

auto swp42
iface swp42
   bridge-access 101

auto swp44
iface swp44
   bridge-access 101

auto swp46
iface swp46
   bridge-access 101

auto swp48
iface swp48
   bridge-access 101
```

```
# ODD Ports/Upper ports PVID 102 for Management network
auto swp1
iface swp1
   bridge-access 102

auto swp3
iface swp3
   bridge-access 102

auto swp5
iface swp5
   bridge-access 102

auto swp7
iface swp7
   bridge-access 102

auto swp9
iface swp9
   bridge-access 102

auto swp11
iface swp11
   bridge-access 102

auto swp13
iface swp13
   bridge-access 102

auto swp15
iface swp15
   bridge-access 102

auto swp17
iface swp17
   bridge-access 102

auto swp19
iface swp19
   ridge-access 102

auto swp21
iface swp21
   bridge-access 102

auto swp23
iface swp23
   bridge-access 102

auto swp25
iface swp25
   bridge-access 102

auto swp27
```

```
iface swp27
   bridge-access 102

auto swp29
iface swp29
   bridge-access 102

auto swp31
iface swp31
   bridge-access 102

auto swp33
iface swp33
   bridge-access 102

auto swp35
iface swp35
   bridge-access 102

auto swp37
iface swp37
   bridge-access 102

auto swp39
iface swp39
   bridge-access 102

auto swp41
iface swp41
   bridge-access 102

auto swp43
iface swp43
   bridge-access 102

auto swp45
iface swp45
   bridge-access 102

auto swp47
iface swp47
   bridge-access 102

auto bridge
iface bridge
bridge-vlan-aware yes

bridge-ports glob swp1-48
bridge-pvid 101
bridge-pvid 102
bridge-stp off
```

Sample IP over InfiniBand configuration files

On the Red Hat operating system, the following three files must be modified by using the following content. You must replace the IP addresses with the actual IP addresses of your environment.

Note: Always refer to the vendor documentation about how to set up IP over InfiniBand. These configuration files that are shown here are for reference only.

We assume that the high-speed (HS) network is `10.0.11.0/24`.

Example A-2 shows a sample `/etc/sysconfig/network-scripts/ifcfg-bond1` file.

Example: A-2 Sample /etc/sysconfig/network-scripts/ifcfg-bond1 file

```
DEVICE=bond1
IPADDR=10.0.11.XX
NETWORK=10.0.11.0
NETMASK=255.255.255.0
BROADCAST=10.0.11.255
USERCTL=no
BOOTPROTO=none
ONBOOT=yes
NM_CONTROLLED=yes
BONDING_OPTS="mode=active-backup miimon=100 updelay=100 downdelay=100"
MTU=2044
```

Example A-3 shows a sample `/etc/sysconfig/network-scripts/ifcfg-ib0` file.

Example: A-3 Sample /etc/sysconfig/network-scripts/ifcfg-ib0 file

```
DEVICE=ib0
TYPE=InfiniBand
NM_CONTROLLED=yes
ONBOOT=yes
USERCTL=no
MASTER=bond1
SLAVE=yes
BOOTPROTO=none
```

Example A-4 shows a sample `/etc/sysconfig/network-scripts/ifcfg-ib1` file.

Example: A-4 Sample /etc/sysconfig/network-scripts/ifcfg-ib1 file

```
DEVICE=ib1
TYPE=InfiniBand
NM_CONTROLLED=yes
ONBOOT=yes
USERCTL=no
MASTER=bond1
SLAVE=yes
BOOTPROTO=none
```

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *Introduction Guide to the IBM Elastic Storage System*, REDP-5253
- *Monitoring and Managing the IBM Elastic Storage Server Using the GUI*, REDP-5471
- *SAP HANA and ESS: A Winning Combination*, REDP-5436

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- IBM ESS 5000 IBM Knowledge Center:

  https://www.ibm.com/support/knowledgecenter/SSZL24_5K_6.0.1/ess5000_601_welcome.html

- IBM Spectrum Scale V 5.0.5 Planning Considerations:

  https://www.ibm.com/support/knowledgecenter/en/STXKQY_5.0.5/com.ibm.spectrum.scale.v5r05.doc/bl1in_PlanningForIBMSpectrumScale.htm

- Licensing on IBM Spectrum Scale

  https://www.ibm.com/support/knowledgecenter/en/STXKQY_5.0.5/com.ibm.spectrum.scale.v5r05.doc/bl1ins_capacitylicense.htm

- The **mmvdisk** command reference:

  https://www.ibm.com/support/knowledgecenter/en/SSYSP8_5.3.5/com.ibm.spectrum.scale.raid.v5r04.adm.doc/bl8adm_mmvdisk.htm

- Using IBM Cloud Object Storage with IBM Spectrum Scale:

  https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WUS12361USEN

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

Implementation Guide for IBM Elastic Storage System 5000

(0.2”spine)
0.17”<->0.473”
90<->249 pages

SG24-8498-00

ISBN 0738459224

Printed in U.S.A.