

IBM Software-Defined Storage Guide

Larry Coyne

Joe Dain

Eric Forestier

Patrizia Guaitani

Robert Haas

Christopher D. Maestas

Antoine Maille

Tony Pearson

Brian Sherman

Christopher Vollmar



Storage





International Technical Support Organization

IBM Software-Defined Storage Guide

July 2018

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

Third Edition (July 2018)

© Copyright International Business Machines Corporation 2016, 2018. All rights reserved.

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

Contents

Notices	vii
Trademarks	viii
 Preface	 ix
Authors	ix
Now you can become a published author, too	xi
Comments welcome	xii
Stay connected to IBM Redbooks	xii
 Chapter 1. Why software-defined storage	 1
1.1 Introduction to software-defined architecture	2
1.2 Introduction to software-defined storage	3
1.3 Introduction to Cognitive Storage Management	7
 Chapter 2. Software-defined storage	 9
2.1 Introduction to SDS	10
2.2 SDS overview	11
2.2.1 SDS supports emerging as well as traditional IT consumption models	12
2.2.2 Required SDS capabilities	14
2.2.3 SDS Functions	15
2.3 SDS Data-access protocols	16
2.3.1 Block I/O	16
2.3.2 File I/O	17
2.3.3 Object Storage	17
2.4 SDS reference architecture	17
2.5 Ransomware Considerations	18
 Chapter 3. IBM SDS product offerings	 19
3.1 SDS architecture	20
3.1.1 SDS control plane	23
3.1.2 SDS data plane	23
3.2 IBM Spectrum Connect	23
3.2.1 VMware integration	23
3.2.2 PowerShell	25
3.2.3 Containers and Kubernetes	25
3.3 IBM Spectrum Control	26
3.3.1 Key capabilities	26
3.3.2 Benefits	26
3.3.3 IBM Data and Storage Management Solutions features	27
3.3.4 IBM Spectrum Control Standard Edition	28
3.3.5 IBM Spectrum Control Advanced Edition	28
3.4 IBM Virtual Storage Center	31
3.5 IBM Storage Insights	31
3.5.1 IBM Storage Insights	32
3.5.2 IBM Storage Insights Pro	33
3.6 IBM Copy Services Manager	34

3.7 IBM Spectrum Protect	35
3.7.1 Key capabilities	36
3.7.2 Benefits	36
3.7.3 Backup and recovery	37
3.7.4 Tool set	37
3.7.5 IBM Spectrum Protect Operations Center	38
3.7.6 IBM Spectrum Protect cloud architectures	39
3.7.7 Cloud Tiering	40
3.7.8 IBM Spectrum Protect Plus	40
3.7.9 IBM Spectrum Protect for Virtual Environments	41
3.8 IBM Spectrum Protect Snapshot	42
3.9 IBM Spectrum Copy Data Management	43
3.10 Block, file, and object storage	47
3.10.1 Block storage	47
3.10.2 File storage	47
3.10.3 Object storage	47
3.11 IBM Block Storage solutions	48
3.11.1 IBM Spectrum Virtualize	48
3.11.2 IBM Spectrum Access Blueprint	52
3.11.3 IBM Spectrum Accelerate	54
3.11.4 IBM XIV Storage System Gen3	59
3.11.5 IBM FlashSystem A9000 and A9000R	59
3.12 IBM File Storage solutions	59
3.12.1 IBM Spectrum Scale	59
3.12.2 IBM Spectrum NAS	69
3.12.3 IBM Spectrum Archive	70
3.13 IBM Object Storage solutions	77
3.13.1 IBM Cloud Object Storage	77
3.14 IBM storage support of OpenStack components	90
3.14.1 Cinder	90
3.14.2 Swift	90
3.14.3 Manila	90
3.14.4 IBM SDS products that include interfaces to OpenStack component	91
Chapter 4. IBM Storage Systems for SDS	93
4.1 Integration with SDS	94
4.2 IBM FlashSystem family	94
4.2.1 FlashSystem benefits	95
4.2.2 IBM FlashCore technology	95
4.2.3 FlashSystem 900	96
4.2.4 IBM FlashSystem V9000	97
4.2.5 IBM FlashSystem A9000/A9000R	97
4.3 IBM TS4500 and TS3500 tape libraries	98
4.3.1 TS4500 tape library	99
4.3.2 TS3500 tape library	100
4.4 IBM DS8880	101
4.5 IBM Storwize family	104
4.5.1 Storwize V5000 Gen2 and V5000F Gen2	104
4.5.2 Storwize V7000 Gen2+ and V7000F Gen2+	105
4.5.3 IBM SAN Volume Controller as an SDS appliance	106
4.5.4 Ongoing IBM contributions to OpenStack Cinder for the Storwize family	107
4.6 IBM Elastic Storage Server	109
4.7 VersaStack for Hybrid Cloud	110

Chapter 5. Use cases	113
5.1 Secure file and object storage for backups	114
5.2 Hybrid Cloud telecommunications storage optimization project	116
5.3 Government sector transformation to private cloud	118
5.4 Cloud service provider use case with SAP HANA	119
5.5 Media and entertainment company Hybrid Cloud	121
5.6 Life sciences and genomics	124
5.7 Oil and gas solutions	126
5.8 Automobile analytics with Hortonworks	127
5.9 Ensuring an insurance company's SAS Grid performance	129
Chapter 6. Next steps	133
Related publications	139
IBM Redbooks	139
Online resources	140
Help from IBM	142

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.

Accesser®	HyperSwap®	Linear Tape File System™
AIX®	IBM®	MicroLatency®
Aspera®	IBM Cloud™	Power Systems™
Bluemix®	IBM Elastic Storage™	POWER8®
Cleversafe®	IBM FlashSystem®	PureFlex®
Cognos®	IBM SmartCloud®	Real-time Compression™
Db2®	IBM Spectrum™	Redbooks®
DB2®	IBM Spectrum Accelerate™	Redpaper™
developerWorks®	IBM Spectrum Archive™	Redbooks (logo)  ®
DS8000®	IBM Spectrum Control™	Slicestor®
Easy Tier®	IBM Spectrum Protect™	Storwize®
ECKD™	IBM Spectrum Scale™	System Storage®
FASP®	IBM Spectrum Storage™	Tivoli®
FlashCopy®	IBM Spectrum Virtualize™	WebSphere®
GPFS™	IBM Z®	XIV®
HiperSockets™	Insight®	z Systems®

The following terms are trademarks of other companies:

Veracity, are trademarks or registered trademarks of Merge Healthcare Inc., an IBM Company.

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

Linear Tape-Open, LTO, Ultrium, the LTO Logo and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

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

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

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

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

Preface

Today, new business models in the marketplace coexist with traditional ones and their well-established IT architectures. They generate new business needs and new IT requirements that can only be satisfied by new service models and new technological approaches. These changes are reshaping traditional IT concepts.

Cloud in its three main variants (Public, Hybrid, and Private) represents the major and most viable answer to those IT requirements, and software-defined infrastructure (SDI) is its major technological enabler.

IBM® technology, with its rich and complete set of storage hardware and software products, supports SDI both in an open standard framework and in other vendors' environments. IBM services are able to deliver solutions to the customers with their extensive knowledge of the topic and the experiences gained in partnership with clients.

This IBM Redpaper™ publication focuses on software-defined storage (SDS) and IBM Storage Systems product offerings for software-defined environments (SDEs). It also provides use case examples across various industries that cover different client needs, proposed solutions, and results. This paper can help you to understand current organizational capabilities and challenges, and to identify specific business objectives to be achieved by implementing an SDS solution in your enterprise.

Authors

This paper was produced by a team of specialists from around the world working at the IBM Client Center Montpellier, France.

Larry Coyne is a Project Leader at the International Technical Support Organization, Tucson Arizona Center. He has 34 years of IBM experience, with 23 in IBM storage software management. He holds degrees in Software Engineering from the University of Texas at El Paso and Project Management from George Washington University. His areas of expertise include client relationship management, quality assurance, development management, and support management for IBM Storage Management Software.

Joe Dain is a Senior Engineer and Master Inventor in Tucson, Arizona and works in the Storage and Software-defined Infrastructure CTO Office. He joined IBM in 2003 with a BS in Electrical Engineering. His areas of expertise include backup, restore, disaster recovery, object storage, data reduction techniques (such as data deduplication and compression), and emerging storage technology trends. He is on his fourteenth IBM invention plateau with over 60 patents issued and pending worldwide, including 22 high-value patents.

Eric Forestier is a certified presales IT Specialist at the IBM Client Center in Montpellier, France. He is part of the IBM Systems organization, supporting IBM Sales for storage opportunities through demonstrations and Business Partner enablements. More specifically, he supports IBM Cloud™ Object Storage, IBM Spectrum™ products including IBM Spectrum Accelerate™, and VersaStack demonstrations. He has been in IBM for 33 years. Before this assignment, he spent 8 years as an IBM PureFlex® and System x IT Specialist, enabling ISV's for the Telco industry. Formerly, he spent 16 years as a software developer in the IBM Network Division working in various areas, such as networking, telephony, and Internet, then 7 years as a presales IT Specialist in the IBM Software Group, enabling IBM Partners to include the pervasive computing assets into their solutions. He authored several IBM Redbooks® publications, including on IBM WebSphere®, IBM Spectrum Accelerate, and IBM XIV®. He holds a Master's degree in Engineering from the Ecole Centrale de Lyon in France.

Patrizia Guaitani is an Executive Infrastructure Architect and Manager of Client Technical Specialist based in Milan, Italy. Patrizia has more than 30 years of IT infrastructure experience encompassing virtualization, high availability, cloud, and hybrid cloud projects. She is mainly focused on helping customers build solutions using the IBM Software-defined Infrastructure offering, and develop private and hybrid storage cloud solutions using the IBM Spectrum Storage™ family and Converged Infrastructure solutions.

Robert Haas is the Department Head for Cloud and Computing Infrastructure at IBM Research - Zurich since 2018. From 2015 to 2017, he was CTO for Storage in Europe, accelerating our storage innovation efforts in supporting cloud environments. From 2012 to 2014 he was on assignment with the Corporate Strategy team in Armonk, focusing on the software-defined technical strategy for Storage. Prior to this, he had launched and managed the Storage Systems Research Group at IBM Research - Zurich, specializing in cloud storage, security, Flash enablement, and tape storage software. During his 20-year career at IBM he has contributed to over 60 patents, standard specifications, and scientific publications. He received his Ph.D. degree from the Swiss Federal Institute of Technology in Zurich, and an MBA at Warwick Business School in the UK.

Christopher D. Maestas is a World Wide Solution Architect for IBM File and Object Storage Solutions, with over 20 years of experience in deploying and designing IT systems for clients in various spaces including HPC. He holds a degree in Computer Science from the University of New Mexico. He works with IBM clients in any industry across the globe in ensuring that access to data is protected and highly performing in their respective environments. He has experience scaling performance and availability with a variety of file systems technologies, including running it on network technologies like Ethernet, Omnipath, and InfiniBand. He has developed benchmark frameworks to test out systems for reliability and validate research performance data. He has also led global enablement sessions online and face-to-face, discussing how best to position mature technologies like IBM Spectrum Scale™ and emerging technologies like Cloud Object Storage and IBM Spectrum NAS.

Antoine Maille is an IBM Certified Architect expert. Since 2002, he has been involved in planning and leading large distributed environments' infrastructure projects. Initially, he worked as the benchmark manager responsible for testing and qualifying new products in real customer contexts. Currently, Antoine is one of the leaders of the storage design center at the IBM Client Center in Montpellier, France.

Tony Pearson is a Master Inventor and Senior Software Engineer in the IBM Tucson Executive Briefing Center, and is a subject matter expert for all IBM storage hardware and software solutions. He has worked on IBM Storage for more than 30 years, with 19 patents for storage solutions and technologies. He is known for sharing his extensive knowledge about IBM System Storage® in his *Inside System Storage* blog, one of the most popular blogs on IBM developerWorks®. He has a bachelor's degree in Computing Engineering and master's degree in Electrical Engineering, both from the University of Arizona.

Brian Sherman is an IBM Distinguished Engineer with over thirty years' experience as an IT Specialist since joining IBM from McMaster University in 1985 with a Mathematics and Computer Science Degree. Brian has been involved in Storage since joining IBM, and has held various storage-related roles including level 2 software support, storage implementation services, and branch Systems Engineer in the Public and Financial Sectors. Brian currently is the technical lead for Software-defined Storage (SDS), IBM Spectrum Storage Family, IBM DS8000®, and XIV/A9000 in the World Wide Advanced Technical Skills (ATS) organization. He also develops and provides worldwide technical education on new Storage Hardware and Software product launches, and participates on several Storage Product Development teams.

Christopher Vollmar is a Storage Solutions Specialist and Storage Client Technical Specialist based in Toronto, Ontario, Canada with the IBM Systems Group. Christopher is currently focused on helping customers build storage solutions using the IBM Spectrum Storage/Software-defined Storage family. He is also focused on helping customers develop private and hybrid storage cloud solutions using the IBM Spectrum Storage family and Converged Infrastructure solutions. Christopher has worked for IBM for over 14 years across a number of different areas of the IBM business, including System Integrators and System x. He has spent the past 7 years working with the IBM System Storage team in Toronto as a Client Technical Specialist and Solution Specialist. Christopher holds an honours degree in Political Science from York University and holds an IT Specialist-Storage certification from The Open Group.

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

Ann Lund

International Technical Support Organization

Erwan Auffray
Marc Bouzigues
Olivier Fraimbault
Benoit Granier
Joelle Haman
Thierry Huche
Hubert Lacaze
Marc Lapierre
Khanh Ngo
Christine O'Sullivan
Gauthier Siri
Olivier Vallod

IBM Systems

Thanks to the authors of the previous edition of this paper:

Larry Coyne
Mathias Defiebre
Phil Gilmer
Shivaramakrishnan Gopalakrishnan
Mikael Lindström
Brandon Mann
Gianluca Perilli
Mladen Portak

Now you can become a published author, too

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us.

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

- Use the online **Contact us** review Redbooks form:

ibm.com/redbooks

- Send your comments in an email:

redbooks@us.ibm.com

- Mail your comments:

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

Stay connected to IBM Redbooks

- Find us on Facebook:

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

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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



Why software-defined storage

This chapter describes the business context for software-defined storage (SDS) within a larger IT software-defined environment. Benefits to IT infrastructures and their supported enterprises are described at a strategic level in support of key initiative areas that include Cognitive, Cloud, Analytics, Mobile, Social, and Security and traditional computing.

This chapter includes the following sections:

- ▶ Introduction to software-defined architecture
- ▶ Introduction to software-defined storage
- ▶ Introduction to Cognitive Storage Management

1.1 Introduction to software-defined architecture

Fundamentally, software-defined architecture is an IT advancement that enables automation of infrastructure configuration, supporting rapid deployment and aligned to real-time application requirements. These are long-standing goals of IT open systems optimization. Software-defined (SD) architecture is an evolving technology that has been made feasible by the abstraction of infrastructure component interfaces delivered through the virtualization of server, storage, and network infrastructures.

Although SD is finding widespread application in cloud implementations, it provides substantial agility and utilization improvements across IT environments, particularly those with rapidly changing application infrastructure support requirements. This is becoming even more evident in container-based environments such as Docker. Applications are launched, scaled up and down, and removed at a much faster pace and with far higher densities than what is seen in environments using virtual machines, for instance.

SD architecture targets new business models using tighter interactions with customers who use new workloads like cognitive, big data, analytics, social business, and mobile. SD architecture also encompasses traditional IT business workloads like enterprise resource planning (ERP), human resources (HR), and customer relationship management (CRM) systems that continue to be important within an integrated infrastructure. Customer-interaction-oriented models are often referred to as *Systems of Engagement* (SoE), whereas traditional back-end structures are called *Systems of Record* (SoR). Together they form *Systems of Insight* (SoI), as shown in Figure 1-1.

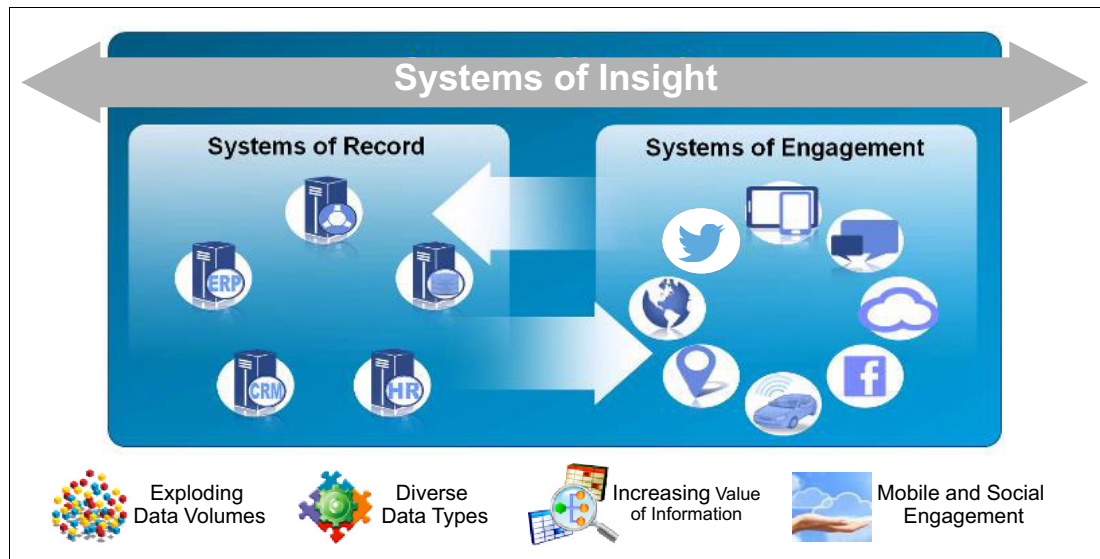


Figure 1-1 Systems of Insight

Supporting Systems of Insight workloads is driving IT to reshape itself to accommodate these new business needs while integrating them with traditional applications and infrastructure. Enabling IT's capability to facilitate these increasingly challenging business requirements is the primary goal of software-defined architecture.

1.2 Introduction to software-defined storage

A fundamental inhibitor to fully realize IT optimization has been achieving integrated systems automation, which is often referred to as *autonomic computing*. This goal has been difficult primarily due to the lack of adequately standardized component interfaces to support a unified infrastructure automation approach. The high levels of customization required involve labor-intensive and time-consuming work, which results in higher costs and affects the ability of IT to rapidly react to new and changing business application requirements.

The primary business reasons for implementing SDS fall into these broad categories, which have interrelated characteristics and benefits:

- ▶ Facilitate IT automation to improve business and IT agility at lower cost
- ▶ Optimized systems administration and control to allow effective and efficient resource utilization that lowers cost and supports business requirements
- ▶ Ease of deployment and redeployment of infrastructure resources
- ▶ Performance tuning with optimal alignment of available resources to application requirements
- ▶ Capacity planning simplification with larger storage pools supporting multiple service levels
- ▶ Enable advanced application deployment of Systems of Insight using Systems of Engagement and Systems of Record:
 - Cognitive
 - Cloud
 - Analytics
 - Mobile
 - Social
 - Security
- ▶ Simplified architecture to reduce specialized components and skills requirements
- ▶ Virtually limitless elastic data scaling
- ▶ Support for block, file, and object data types

These business reasons for implementing SDS result from the combination of business pressures, new applications, data growth, and hybrid cloud that are challenging traditional storage approaches. There is a need to free data from hardware constraints, and realign with new business processes and applications.

IBM suggests a framework based on a software-defined architecture that encompasses orchestrators and Software-Defined Infrastructure (SDI), including software-defined storage. Such a framework creates and implements optimized IT infrastructures that can help enterprises attain competitive advantage by delivering higher value and profitability through speed and efficiency in provisioning IT services.

Most enterprise IT architectures already use virtualization to manage growth and improve agility. Containers can be seen as a further optimization to virtualization, calling for a new generation of orchestration methods. Virtualization and the abstraction of IT component interfaces that it provides enable integrated software definition across the entire infrastructure.

Standardized software interfaces support the automation of infrastructure administrative tasks, such as configuration, monitoring, and provisioning, so that IT organizations can respond to changing application and business requirements immediately.

For any enterprise today, it is increasingly critical to quickly and efficiently deliver resources in support of not only traditional business workloads, but also to enable cognitive, cloud, big data and analytics, mobile, and social-business services. The SD approach helps an enterprise fulfill its business requirements, and respond to business requests faster and more effectively.

SDS is a key component that supports the software-defined infrastructure framework within the overall software-defined environment, along with software-defined compute and software-defined network constructs, as shown in Figure 1-2. Although each of these constructs could be used separately, substantial synergy and value results from an integrated approach, which most organizations should adopt.

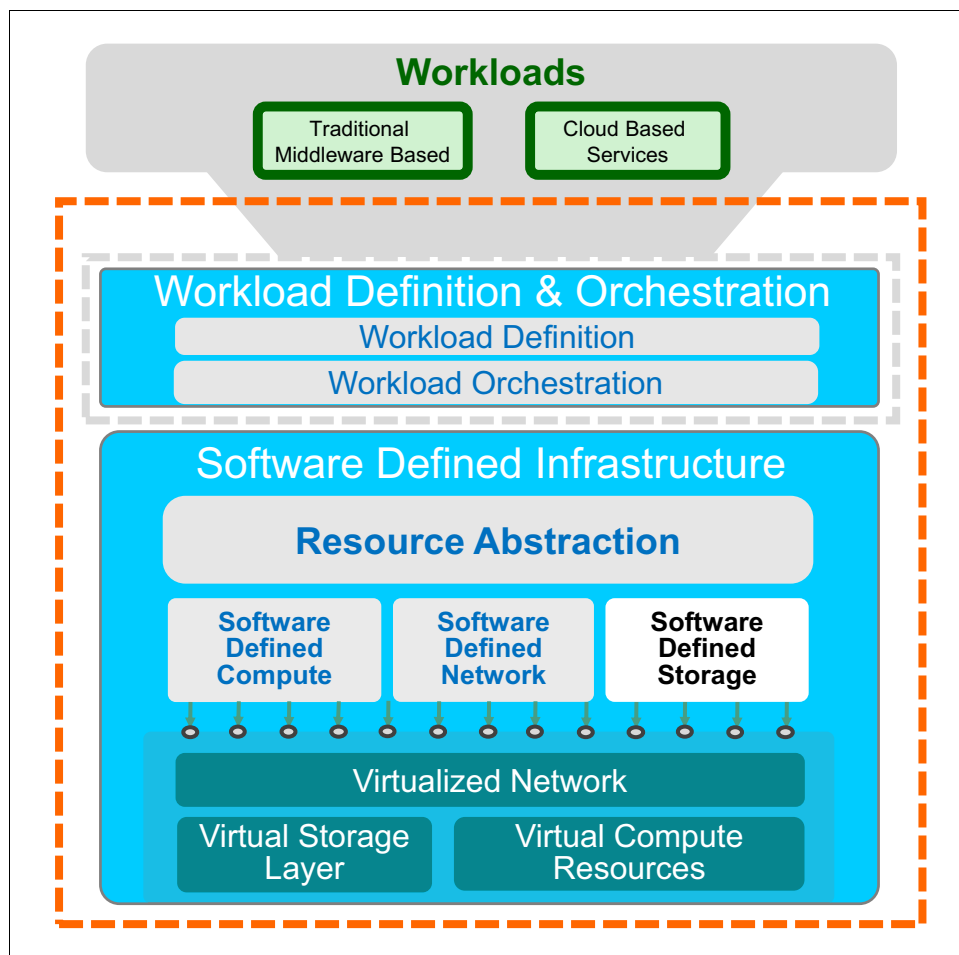


Figure 1-2 Software-defined architecture framework

A software-defined architecture consists of the following key elements:

- ▶ Workload definition and orchestration
- ▶ Software-defined infrastructure (SDI)

A software-defined infrastructure is a key component of an SDE, providing abstraction of physical resources through these concepts:

- ▶ Software-defined compute
- ▶ Software-defined network
- ▶ Software-defined storage

These SDI resources must be dynamically configurable to support the workloads through application programming interfaces (APIs).

SD drives efficiency by optimizing the connections between workloads and resources based on the needs of the business. Workloads are defined and orchestrated based on usage patterns, and resources are managed and deployed according to business rules and policies. IT Enterprises employing an SD approach have several core advantages over processes that have traditionally been handled manually.

The fundamental idea of SD consists of automatically orchestrating infrastructure resources (compute, storage, networks, and management) to programmatically (that is, by using software programming) meet workload requirements in real time. From this point of view, these are some of the major attributes of SD:

- ▶ **Agility:** IT resource customers expect to use infrastructure resources on demand based on immediate business requirements. The agility of IT resource allocation and use needs to be made near instantaneous to support new workloads.
- ▶ **Standardization:** Today consumers are less interested in the specific infrastructure components and are more concerned with ensuring the appropriate service level characteristics needed for their applications are in place. SDI brings uniformity by automating, standardizing, and integrating IT infrastructures.
- ▶ **Provisioning and Orchestration:** Rather than building unique infrastructure systems of server, network, and storage components for applications, IT providers need to configure pools of resources and put them together in a way that can be dynamically delivered programmatically (by using software) with service-level-oriented interfaces appropriate to IT consumers.

Of course, hardware is required within the SDI to provide resources to support server, storage, and network infrastructure needs. The essential characteristic requirement for an SDI is that these hardware components be dynamically configurable to support real-time service level requirements. As an example, SDS can take advantage of disaggregated architectures that allow the most flexible composition of hardware resources, relying on technologies such as NVMe over Fabrics to connect compute resources to storage resources across a fabric with very low additional latency.

SD by itself does not provide infrastructure aligned to business IT service level requirements unless the proper software definable components are in place. High performance, availability, and security service levels require software definable components that can be configured to meet these business requirements.

Similarly, lower level (for example, best effort) service levels should generally be configured with software definable component costs aligned to these business requirements. SD architectures that need to support varied service levels still require appropriate performance and capacity planning across higher performance components, and differentiation of availability requirements for cost optimization.

SD supports the optimization of infrastructure service levels to available component resources in real time. Implementing an SDI framework supports the transformation from static infrastructure into dynamic, continually optimized, workload-aware, and virtualized resources that enable line-of-business users to better use IT as needed, facilitating far greater business agility.

The deployment velocity requirements of Systems of Engagement demand that this new interaction between the consumer and the infrastructure provider define workloads in a way that enables the infrastructure to dynamically respond to the needs of those workloads.

Analytics processing, for example, typically needs to rapidly access data, efficiently process that data, and then release resources when the analysis completes. SD is an ideal IT infrastructure implementation approach in this scenario. Similarly, SD supports efficient deployment of rapidly growing and dynamically evolving transactional applications that support the rapidly increasing number of mobile devices IT now manages.

SD's value is even more apparent in hybrid scenarios, such as Social Analytics employed in sentiment analysis used to find trends in customer opinions and thinking, or developing a macro-level understanding of worldwide events to create opportunity out of raw, uncorrelated, or disparate data. Without SD, the ability to react in a timely fashion to new workload requirements is limited, which inhibits IT's ability to meet the dynamic infrastructure requirement for these workloads. Inhibited applications are often delayed and less effectively deployed, resulting in under-realized or missed business opportunities.

Although SD provides a compelling framework to address the challenge of dynamic infrastructure configuration and provisioning, important architectural design factors must still be addressed. These factors range from network data latency to infrastructure component interoperability within a specific software-defined implementation. These are some of the key considerations:

- ▶ Data locality and latency, and network bandwidth and data scale considerations, can be more readily addressed within SDI, but certainly must be properly planned for in the design process.
- ▶ Infrastructure diversity (in terms of public, hybrid, and private cloud) and legacy infrastructure support are essential benefits of SD implementations, but still require careful planning around security, availability, and relative cost parameters.
- ▶ Component selection, fit for purpose, and web-scale (custom commoditization) require attention to relative cost/performance, especially in the years of SD transition ahead.
- ▶ Application and component interoperability, and open standards conformity, typically drive reductions of component and administrative support cost, but there are design scenarios where custom or vendor-proprietary solutions continue to make business sense through SD transition and likely beyond.
- ▶ Transition and legacy support must be handled in an evolutionary way to minimize business and customer impacts.

The business value of SD is too great to ignore, and is maximized when these design parameters are given proper consideration in planning and deploying software-defined infrastructures. IBM is investing heavily in developing offerings across the spectrum of the software-defined architecture:

- ▶ Building block components
- ▶ Integrated cloud offerings
- ▶ Implementing and supporting open API standards and architectures to help businesses achieve improved agility and competitiveness with outstanding customer satisfaction

Figure 1-3 shows the building blocks required to support new IT infrastructure with a highlight of SDS, which is the focus of this paper.

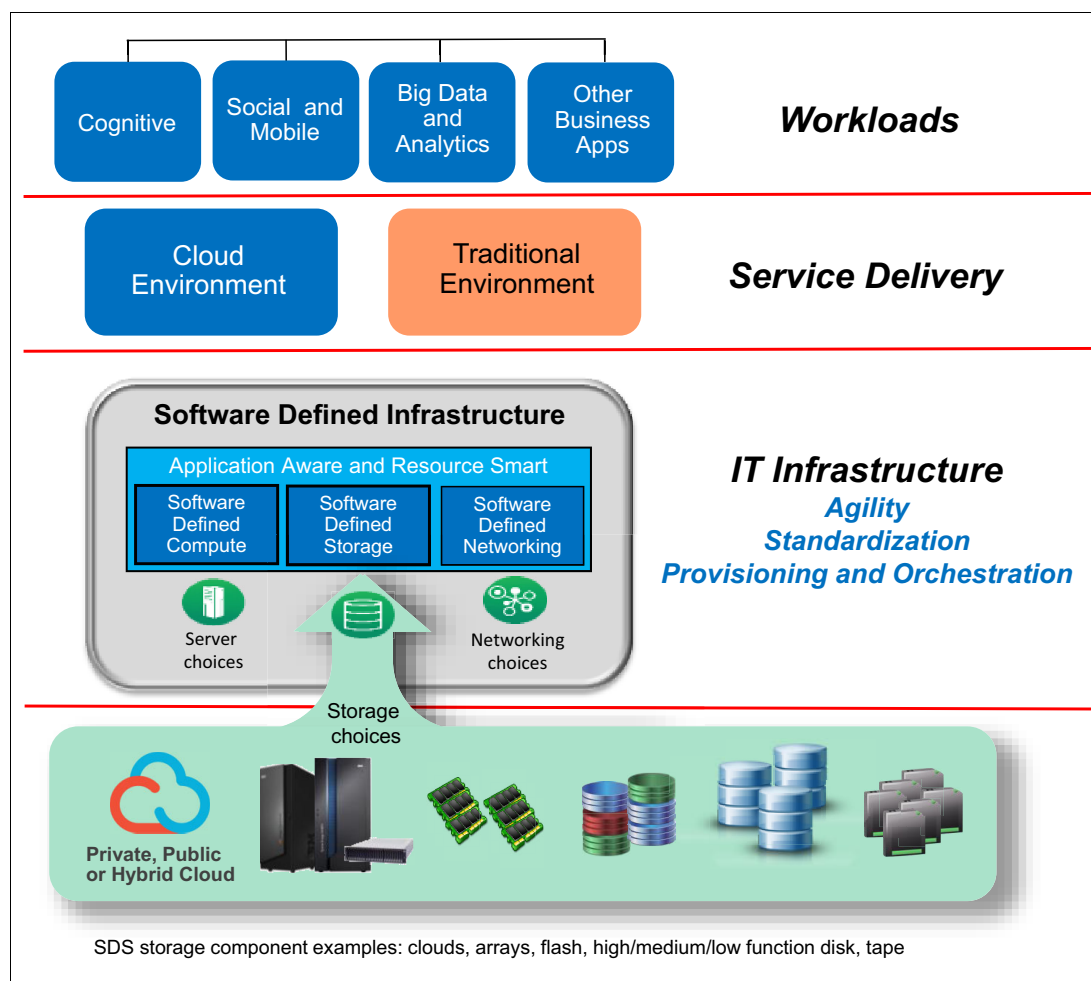


Figure 1-3 SDS building blocks of SDI for support of new IT business requirements

1.3 Introduction to Cognitive Storage Management

Cognitive systems are about exploring data and finding new correlations and new context in that data to provide new solutions. Cognitive computing is becoming a new industry. As businesses race to deliver the next generation of personalized services, cognitive and cloud computing are playing a pivotal role in remaking IT infrastructure for the cognitive era.

Infrastructure now must capture, secure, store and distribute data to meet the demands of these new cognitive services. Clients in every industry are using data, advanced analytics and, increasingly, cognitive technologies to differentiate their cloud-based processes, services, and experiences to innovate and create business value.

Often, big data characteristics are defined by the “five V’s”:

- ▶ Variety
- ▶ Volume
- ▶ Velocity
- ▶ Veracity
- ▶ Visibility

Big data requires innovative forms of information processing to draw insights, automate processes, and assist in decision making.

Big data can be structured data that corresponds to a formal pattern, such as traditional data sets and databases. Also, big data includes semi-structured and unstructured formats, such as word processing documents, videos, images, audio, presentations, social media interactions, streams, web pages, and many other kinds of content. Unstructured data is not contained in a regular database and is growing exponentially, making up the majority of all the data in the world.

Cognitive computing enables people to create a new kind of value by finding answers and insights that are locked away in the volumes of data. Cognitive systems can in turn be used to enhance storage management, for instance to deliver cognitive support and health insights.

Furthermore, by taking advantage of the vast insights extracted as metadata from the data itself, as well as from how that data is being used, cognitive systems can contribute to perform, for instance, value-based placement of data and predictive caching. This is key to handle increasingly large storage deployments that call for more intelligent automation, taking advantage of the variety of information sources handled by a cognitive storage management system.



Software-defined storage

This chapter provides an overview of how IT storage infrastructures can be designed with programmable software-defined storage (SDS) components into systems controlled through software, enabling dynamic assignment of IT resources. Resources can then be assigned based on application workload requirements with best-available resources aligned to business-requirements-based service level policies. Additionally, the benefits of how storage infrastructure can be provisioned, configured, reconfigured, and deprovisioned through SDS to optimize and use IT resources based on real-time business needs are illustrated.

This chapter includes the following sections:

- ▶ Introduction to SDS
- ▶ SDS overview
- ▶ SDS Data-access protocols
- ▶ SDS reference architecture

2.1 Introduction to SDS

What exactly does Software-Defined Storage refer to? Most IT storage systems today are already based in either software, microcode, or both.

SDS in today's business context refers to IT storage that goes beyond typical array interfaces (for example, command-line interface and graphical user interface) to operate within a higher architectural construct. SDS supports overall IT architectural definition, configuration, and operations, often referred to as *software-defined infrastructure* (SDI). The greatest value and versatility of this approach is the standardized programming interfaces applied across a heterogeneous, multivendor IT infrastructure.

IDC, an industry analysis firm, has defined three criteria for software-defined storage, as shown in Figure 2-1 on page 11:

- ▶ Runs on industry-standard hardware
- ▶ Offers a full suite of storage services
- ▶ Embraces multiple storage options

SDS runs on industry-standard hardware, and does not rely on proprietary adapters, ASIC, FPGA, or other co-processors. In this case, standard hardware includes x86-based, as well as OpenPOWER servers.

SDS offers a full suite of storage services. These services include data footprint reduction technologies (such as compression, data deduplication, and thin provisioning) and copy services (such as point-in-time copies or remote distance mirroring). Software that merely re-routes I/O traffic across to different devices is not SDS.

SDS embraces multiple storage options. SDS can be sold as software-only by one vendor, enabling customers to purchase industry-standard servers from a variety of other suppliers. SDS can also be pre-installed on industry-standard servers and sold as a pre-built system or appliance. In addition, SDS can run in private or public cloud configurations, with a storage utility license that charges only for the amount of storage capacity or I/O bandwidth actually used.

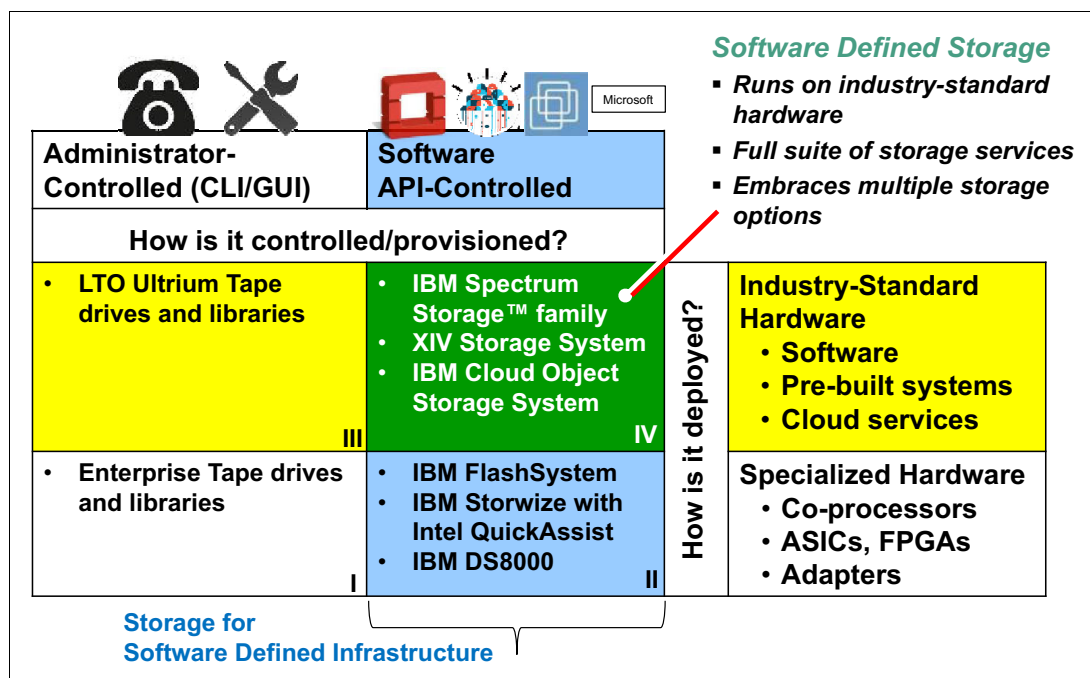


Figure 2-1 Storage for software-defined infrastructure

2.2 SDS overview

It is important to recognize that the SDS and the larger SDI architectural concepts are relatively new and still evolving technological approaches to supporting software-defined architectures. These approaches are not all fully developed into mature, straightforward implementations. Therefore, this document provides guidance in planning SDS deployments in an incremental and evolutionary way that optimizes value while avoiding potential operational disruptions or dead-end IT architecture investments.

SDS has been defined in different ways for different users and technology providers, but to fully realize the potential of this technology, true SDS implementations incorporate the following characteristics:

- ▶ Programmable interfaces to support dynamic storage configuration and management
- ▶ Support for block, file, and object storage interfaces
- ▶ Dynamic service level configuration
- ▶ Enterprise scale storage virtualization and pooling
- ▶ Generic data storage and control components

SDS is a new storage architecture for a wide variety of data storage requirements based on a set of loosely coupled software and hardware components dynamically configurable to meet customers' workload requirements. It is a model that encompasses traditional workloads (systems of record) and newer types of workloads (systems of engagement), and is optimized for interoperability across hardware and software platforms. SDS provides greater storage infrastructure flexibility to share resources while maintaining the required service levels and enabling customers to better use data for greater business insights.

SDS delivers software-based storage services to an SDI through the following capabilities:

- ▶ Storage virtualization
- ▶ Automated policy-driven administration for storage management functions
- ▶ Analytics and optimization
- ▶ Backup and copy management
- ▶ Integration and API services
- ▶ Security
- ▶ Massive scale-out architecture
- ▶ Cloud accessibility

Although point SDS solutions are available, it is important to recognize an enterprise-wide SDS implementation will not generally be realized by installing one product offering, and will not be software only. Software and hardware products and their specific features must be orchestrated to meet specific customer workload requirements in an enterprise. Most IT organizations want an evolutionary transition path into SDS and SDI to gain experience, avoid risk, and preserve existing infrastructure investments.

Loosely coupled means that each product brings its own advantages in terms of new features and functions without having to redesign all of the storage infrastructure to be able to integrate them.

Meet customer's workload requirements presumes that documented or implied SLAs are established between the storage provider and the users/customers, and that metrics are collected and reports delivered to validate that SLAs are met.

SDS creates smarter interactions between workloads and storage by using the following techniques:

- ▶ Exposing storage capabilities for the workloads to dynamically provision storage with the most suitable (architecturally fit for purpose with required service level capabilities) characteristics
- ▶ Introducing new operations and concepts between workloads and storage to help storage better adapt to the needs of workloads (Applied Analytics)
- ▶ Moving some of the storage functions closer to the workloads to use higher-level knowledge and lower cost (capability to provide storage functions at the server rather than rely solely on storage array features)

2.2.1 SDS supports emerging as well as traditional IT consumption models

The fundamental emerging market that SD addresses is the so-called digital consumption model, which is aligned to the digital economy. This paper outlines traditional consumption models to compare and contrast with new and emerging ones, and how the SD framework can provide more effective and efficient support. This is in the context of other consumption models, such as legacy dedicated, managed services, and newer and in many cases overlapping consumption models such as IaaS, SaaS, and PaaS cloud models.

The term *consumption model* was borrowed from economics to describe how resources are purchased and used by consumers. In this context, the relative perspective of the various types of consumers is key. Although this is not a new concept, it figures prominently in the basis for SDS as a means of addressing IT requirements that arise from the emerging digital economy consumption models where flexibility and rapid deployment (agility) are essential.

From glass house mainframes to departmental computing through client/server and most recently cloud computing, a central theme has been to optimize the costs of data processing.

Is it ultimately more cost-effective (including monetized risk) for an organization to create its own data processing environment or to procure external capability?

Architectural criteria have undergone significant shifts and realignments from dedicated, customized, and consolidated to shared, commoditized, and distributed, depending on business requirements and technological capabilities, balanced with relative costs. Ultimately users have never been concerned with the underlying structure by which the information they consume is designed. Their concern is with content, format, ease of use, and accessibility.

SDS supports the SDI goal of technical agility in supporting new workloads across the cloud, mobile, social, and analytics infrastructure spaces with the necessary security for data storage by using these techniques:

- ▶ **Automation:** Realization of autonomic data storage capabilities for deployment, provisioning, monitoring, reconfiguration, performance management, and capacity planning.
- ▶ **Virtualization:** The near-universal abstraction of functionality across underlying components is the primary enabler of the software-defined model, making systems integration and configuration of infrastructure components possible by using software programming interfaces.
- ▶ **Programmatically administered:** Storage is deployed, configured, and administered by using open programming standard APIs, enabling policy-based automation of infrastructure storage resources.

SDS enables new storage consumption models that, in some respects, resemble vending machines. Customers (data storage consumers) see the products (storage service catalog) that they want to buy, and decide which one meets their requirements (service levels). The customer then decides how much is needed, inserts the coins (chargeback/pay-per-use), and presses the dispense button.

The vending machine starts a series of actions (provisioning orchestration) to provide the customer with the chosen products. The release of the products is immediate (real-time). Beyond this analogy though, SDS analytically monitors the product usage and adjusts service level and capacity based on real-time application needs and available resources.

Figure 2-2 shows the new storage consumption model.



Figure 2-2 SDS control plane versus data plane vending machine analogy

Importantly, this hypothetical storage capacity vending system is able to measure (analytics and optimization) the available products (storage services) and when they are below a predefined threshold. It can then send alerts to be refilled with the required service level capacity. The access method (service catalog and orchestration) and the products (storage services) exposed by the vending-machine (SDS infrastructure) can be tailored to various customers.

New products can be added as required by business needs. The dispenser (SDS system) is responsible to ensure its products' quality. Only the customers who have access to the vending-machine's main panel can buy them (security).

2.2.2 Required SDS capabilities

Depending on the business scenario, SDS needs to support SDI with storage services characterized by some fundamental capabilities:

- ▶ Optimal workload allocation: SDS must be able to support applications that deliver optimal performance and store data according to relative business importance. Therefore, SDS must be able to orchestrate the movement of data and workloads across storage service level tiers to achieve agreed-upon performance, availability, retention, and security. These requirements are typically defined in SLAs.
- ▶ Data Retention and Archive: Robust SDS should also support archival capability for long-term data retention in line with business requirements to support regulatory requirements, as well as effective data rationalization (purge expired data).

- ▶ **Agility and Scalability:** As mentioned, the ability to rapidly deploy IT support of new business initiatives lies at the core of SD framework. Therefore, SDS must ensure that storage services can scale according to business needs and be agile enough to respond within the requested time frame. This feature requires the following capabilities:
 - The ability to provide metrics about the storage infrastructure usage and to report on those metrics
 - Tiered capacity by service level, and allocate and reclaim capacity based on current business requirements
 - Fault tolerance as appropriate and rapid troubleshooting of the infrastructure
 - The ability to provision storage on demand, with support for various consumption models including cloud
- ▶ **Flexible Data Access:** SDS must be able to provide the access methods and protocols that are required by business applications, enabling data sharing and multi-tenancy across data centers if needed.

2.2.3 SDS Functions

To support these capabilities, SDS-enabled components must include the following integrated, API-controlled functions when needed to support business application service levels:

- ▶ **Storage virtualization:** Enables the creation of unique pools of storage capacity starting from heterogeneous disk storage arrays and devices. Advanced function capabilities like transparent data migration, tiering, thin-provisioning, compression, and local/remote copy services can be software-defined or use array capabilities, depending on the disk storage vendor and its native technology. These features can be leveraged and integrated by the orchestrator to provide the SDS capabilities.
- ▶ **Policy automation:** Enables automation of storage provisioning with definition of storage policies that are correlated to automation workflows based on the features provided by the virtualized storage capacity, and the measurements and reports provided by the analytics tools.
- ▶ **Analytics and Optimization:** Provides the metrics to measure the storage performances and the capacity usage, and the tools to report on them and compare them to the required service levels. Also ensures optimization of the storage capacity usage based on the collected metrics and policies aligned to the automation workflows. Metrics can include chargeback data as required.
- ▶ **Availability, Backup, and Copy Management:** Physical and logical integrity of data within the storage infrastructure that provides backup/restore capability, and local/remote copies for test or disaster recovery purposes.
- ▶ **Integration and API services:** All of the features that are provided by SDS must be integrated, not only internally but also with the other functions available in the SDI and with the business applications. The most common way of doing that is to communicate with the other infrastructure services by using specific interfaces called application programming interfaces (APIs).
- ▶ **Security:** Enables and ensures secure access to the data by authorized persons. Common features to ensure this functionality are the encryption of data at rest, SAN zoning, LUN masking, Access Control Lists, Role-based Authentication Control (RBAC), and the usage of enhanced attributes in the object storage environment that enable a per-object access granularity.

- ▶ **Massive scale-out architecture:** Support for big data and analytics is a key driver for SDS, with the capability for rapid deployment (and recovery/redeployment) of massive storage capacity for these and other Systems of Insight.
- ▶ **Cloud accessibility:** SDS must be able to support cloud implementation, connect to the cloud, and eventually use storage cloud services from other providers. SDS can offer storage cloud services to customers who want to access them in a “cloud” fashion. It should also expose a storage service catalog and support storage self-provisioning to business requirements.

2.3 SDS Data-access protocols

It is fundamental to know the way that data will be accessed, and particularly the protocol that the customer wants or needs to use. This section covers the main standard data access protocols that can be used in an SDS environment, although this is not a complete list.

Figure 2-3 shows how in traditional IT environments there were essentially two ways of accessing data:

- ▶ The Block I/O protocol that is the closest to the physical HW
- ▶ The File I/O Object Storage access methods

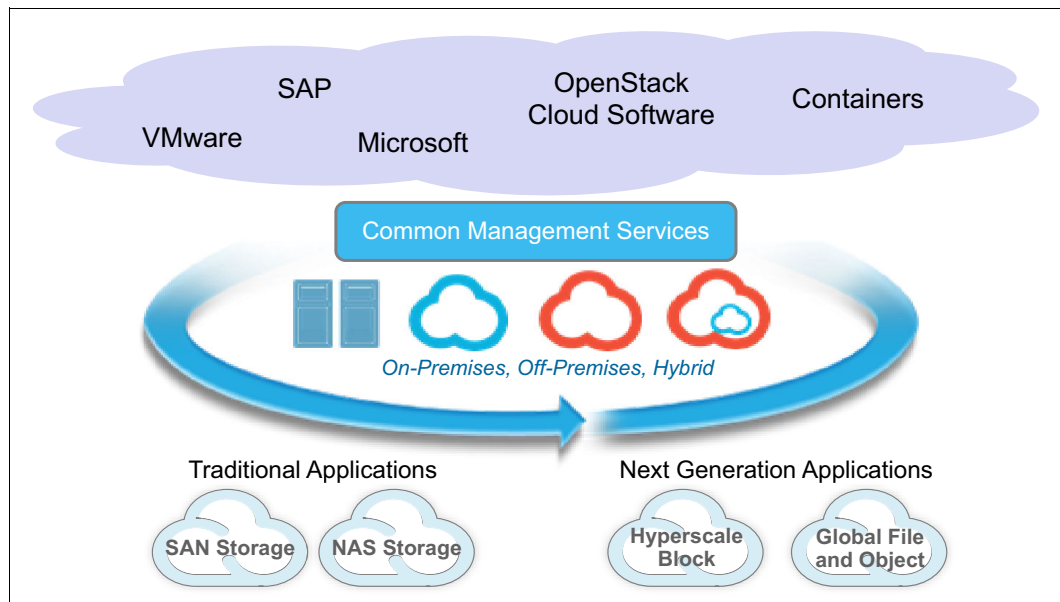


Figure 2-3 Data types in an SDS environment

2.3.1 Block I/O

The term *block I/O* is traditionally used to represent all local disk I/O on a computer. This term is also used to describe I/O over the Fibre Channel SAN, because the I/O in this environment is driven by the operating system. This type of I/O occurs at a logical layer below the layer of a file system. The operating system uses file system drivers that manage the file system. It is these drivers that manipulate the file system at the block I/O or disk level.

Generally, applications only use file I/O commands to manipulate bytes of data. They issue commands to the operating system to access and manipulate their files. The operating system then translates those I/O requests into lower-level commands that use block I/O to manipulate the file system and the application data file.

2.3.2 File I/O

In order for an application to access a file, it needs to know the name of the file and its location (file path). If the name of the file is not predetermined, a user can browse a location or storage container to locate the file or create one.

An application that performs file I/O is generally expected to know how data is organized within a file, or has access to a description that provides that information. The descriptor can also be another file, such as an XML file, or be embedded within the data file at a predetermined location. This is typical of documents that are created in Microsoft Word, Microsoft Excel, and many other document types, as well as in databases. It is also possible for the file to contain only data, and the descriptor is stored within the application. In any case, if an application can get access to its data correctly, it is then able to process that data.

2.3.3 Object Storage

The most recent approach is called *Object Storage*, which is a way of accessing data seen as objects described by a richer set of metadata when compared to a normal File System implementation. This way of accessing data can be used with normal applications, but it is especially beneficial when dealing with large quantities of unstructured data. An object consists of the following parts:

User Data	Data coming from the user application.
Metadata	Information about the data that includes system properties and custom, user-defined attributes, which improves data sharing and search among applications.
Object ID	The unique identification code of the Object.

Objects are accessed using HTTP protocols, including Amazon S3 or OpenStack Swift protocols.

2.4 SDS reference architecture

SDS is one the three main components of the new SDI architecture. The main characteristic of the SDS reference architecture, similar to the software-defined networking (SDN) environment, is the separation of the storage functions into two main layers:

- ▶ **SDS control plane:** The control plane is a software layer that manages the virtualized storage resources. It provides the high-level functions that are needed by the customer to run the business workload and enable optimized, flexible, scalable, and rapid provisioning storage infrastructure capacity. These capabilities span functions like policy automation, analytics and optimization, backup and copy management, security, and integration with the API services, including other cloud provider services.
- ▶ **SDS data plane:** The data plane encompasses the infrastructure where data is processed. It consists of all basic storage management functions, such as virtualization, RAID protection, tiering, copy services (remote, local, synchronous, asynchronous, and point-in-time), encryption, compression, and data deduplication that can be requested by the control plane. The data plane is the interface to the hardware infrastructure where the data is stored. It provides a complete range of data access possibilities and spans traditional access methods, such as block I/O (for example, iSCSI), File I/O (NFS, SMB, or Hadoop Distributed File System (HDFS)), and object-storage.

Figure 2-4 shows the SDS capabilities of the control plane and data plane.

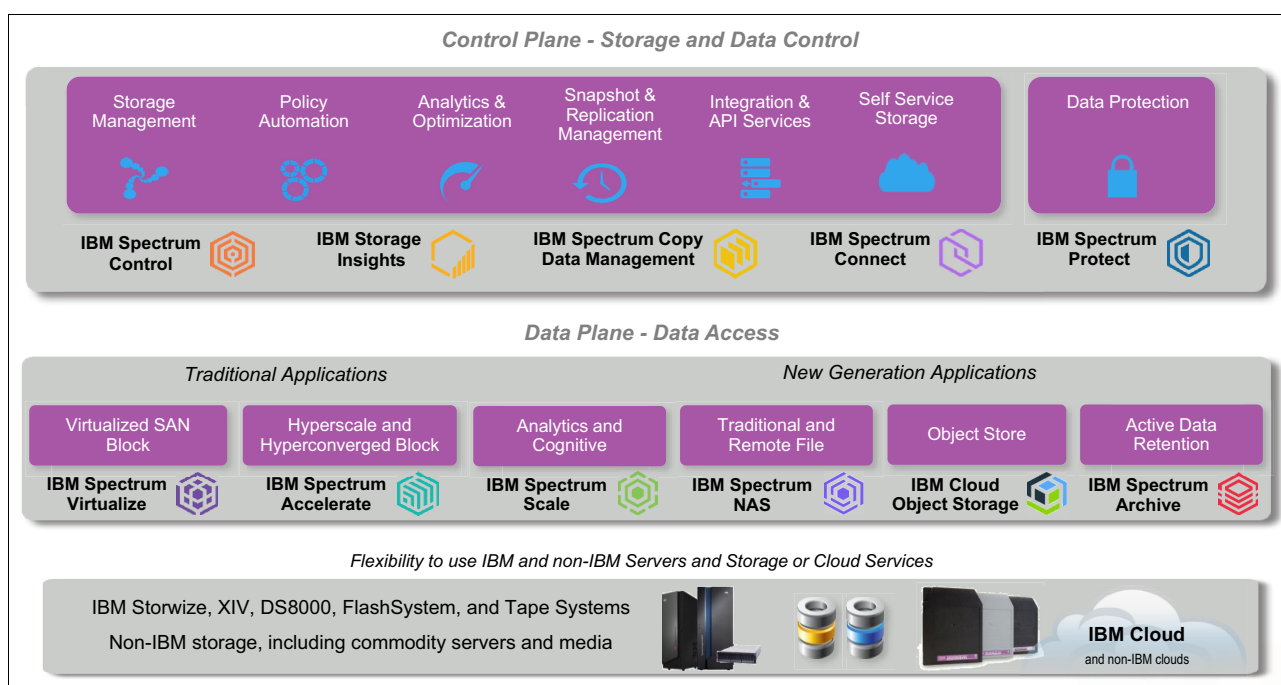


Figure 2-4 SDS capabilities through the control plane and data plane

SDS provides the agility, control, and efficiency needed to meet rapidly changing business requirements by dynamically optimizing infrastructure capabilities to application service level requirements. New business requirements are the driving force for the emergence of this new IT storage infrastructure architecture.

SDS is built with standardized software APIs to provide organizations the underlying capabilities to support applications aligned to the digital consumption model of the digital economy. It also provides a compelling value proposition for optimizing traditional workloads within this consolidated architectural construct.

2.5 Ransomware Considerations

Ransomware has reached high levels of attention due to the dramatic damages that it can cause by locking away user data. Security is among the various functions that SDS has to provide, and SDS can play a critical role in preventing damages caused by ransomware in particular. Because of encryption, auditing, and virus detection, abnormal accesses to data can be detected and if needed interrupted. Furthermore, copies of data can be protected through “air gap” measures that physically separate them, such as exporting off-site copies that can’t be accessed online anymore.



IBM SDS product offerings

This chapter gives an overview of IBM software-defined storage (SDS) products offerings with a focus on the IBM Spectrum Storage family of products and their capabilities and benefits.

The products that are discussed are organized by the management capabilities and the type of solution they provide, whether block, file, or object.

This chapter includes the following sections:

- ▶ SDS architecture
- ▶ IBM Spectrum Connect
- ▶ IBM Spectrum Control
- ▶ IBM Virtual Storage Center
- ▶ IBM Storage Insights
- ▶ IBM Copy Services Manager
- ▶ IBM Spectrum Protect
- ▶ IBM Spectrum Protect Snapshot
- ▶ IBM Spectrum Copy Data Management
- ▶ Block, file, and object storage
- ▶ IBM Block Storage solutions
- ▶ IBM File Storage solutions
- ▶ IBM Object Storage solutions
- ▶ IBM storage support of OpenStack components

3.1 SDS architecture

The Spectrum Storage family of solutions are depicted by their primary function as either a Control Plane or Data Plane solution.

Figure 3-1 shows the IBM SDS architecture with a mapping of the IBM Spectrum Storage family of products across the SDS control plane and data plane.

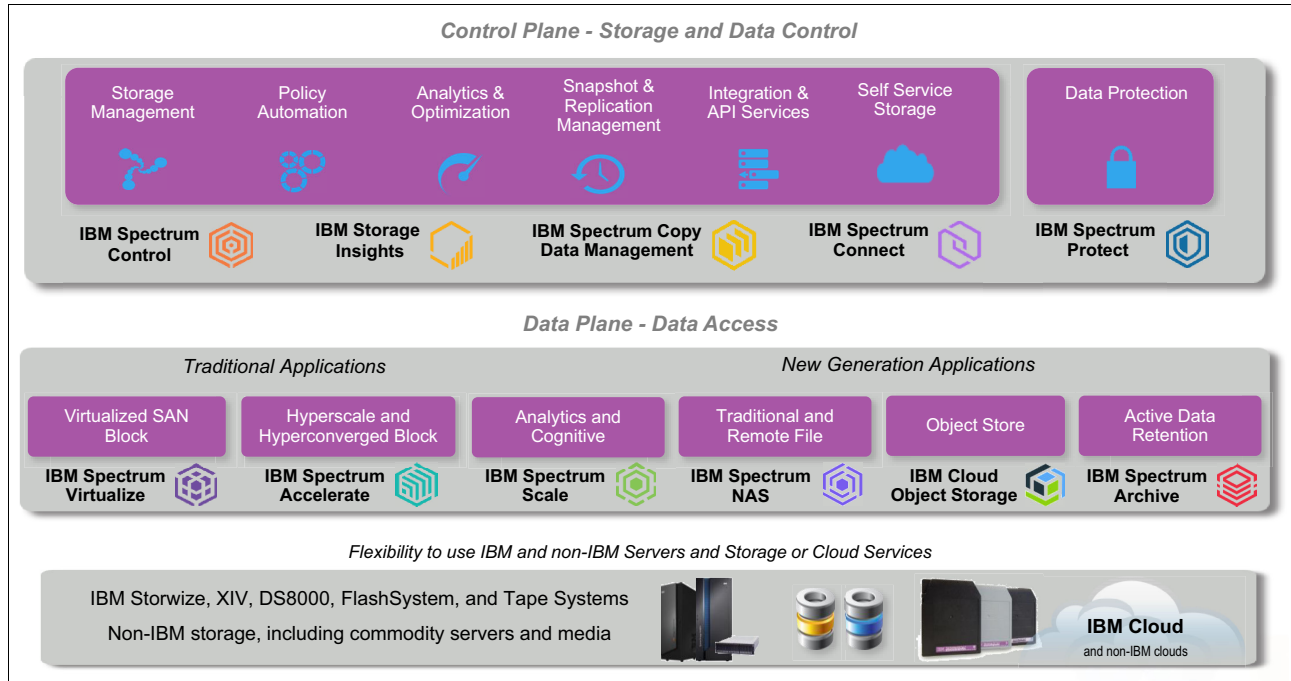


















Figure 3-1 IBM Spectrum Storage family mapped to SDS Control Plane and Data Plane

Table 3-1 is an overview of the IBM Spectrum Storage family high-level descriptions with the products that provide the functions.

Table 3-1 IBM Spectrum Storage Family descriptions

	IBM Spectrum Storage family member	Description
SDS Control Plane		
	IBM Spectrum Connect	Simplifies multi-cloud deployment across your IBM enterprise storage systems. Formally IBM Spectrum Control™ Base. Facilitates API connections for components, such as VMWare and Containers.
	IBM Spectrum Control	Automated control and optimization of storage and data infrastructure.
	IBM Storage Insights	Analytics-driven, storage resource management solution that is delivered from the cloud in a SaaS model. Provides for Cognitive Storage Management capabilities for clients with IBM Storage.
	IBM Copy Services Manager	Automated control and optimization of storage replication features.
	IBM Spectrum Protect™	Optimized data protection for client data through backup and restore capabilities.
	IBM Spectrum Protect Plus	Data protection for Virtual Machines.
	IBM Spectrum Protect Snapshot	Integrated application-aware point-in-time copies.
	IBM Spectrum Copy Data Management	Automate creation and use of copy data snapshots, vaults, clones, and replicas on existing storage infrastructure.

	IBM Spectrum Storage family member	Description
SDS Data Plane		
	IBM Spectrum Virtualize™	Core SAN Volume Controller function is virtualization that frees client data from IT boundaries.
	IBM Spectrum Accelerate	Enterprise storage for cloud that is deployed in minutes instead of months.
	IBM Spectrum Scale	Storage scalability to yottabytes and across geographical boundaries.
	IBM Spectrum NAS	Remote and departmental file share.
	IBM Cloud Object Storage	Object storage solution that delivers scalability across multiple geographies.
	IBM Spectrum Archive™	Enables long-term storage of low activity data.
	IBM Spectrum Storage Suite	A single software license for all your changing software-defined storage needs. Straightforward per-TB pricing for the entire IBM Spectrum Storage suite. Includes IBM Spectrum Accelerate, IBM Spectrum Scale, IBM Spectrum Protect, IBM Spectrum Control, IBM Spectrum Scale, IBM Spectrum Archive, and IBM Cloud Object Storage.
IBM Spectrum Storage Solutions		
	IBM Spectrum Access Blueprint	Enterprise data management and protection in hybrid and multi-cloud environments.

3.1.1 SDS control plane

The control plane is a software layer that manages the virtualized storage resources. It provides all the high-level functions that are needed by the customer to run the business workload and enable optimized, flexible, scalable, and rapid provisioning storage infrastructure capacity. These capabilities span many functions, such as storage virtualization, policies automation, analytics and optimization, backup and copy management, security, and integration with the API services, including other cloud provider services.

3.1.2 SDS data plane

The data plane encompasses the infrastructure where data is processed. It consists of all basic storage management functions such as virtualization, RAID protection, tiering, copy services (remote, local, synchronous, asynchronous, and point-in-time), encryption, compression, and data deduplication that can be started and managed by the control plane. The data plane is the interface to the hardware infrastructure where the data is stored. It provides a complete range of data access possibilities, spanning traditional access methods like block I/O (for example, Fibre Channel or iSCSI) and File I/O (POSIX compliant), to object-storage and Hadoop Distributed File System (HDFS).

3.2 IBM Spectrum Connect



IBM Spectrum Connect is a centralized server system that consolidates a range of IBM storage provisioning, automation, and monitoring solutions through a unified server platform.

IBM Spectrum Connect provides support for the following functions:

- ▶ VMWare environments with IBM Storage integration for advanced features and APIs
- ▶ Container orchestration with Kubernetes including IBM Cloud Private
- ▶ PowerShell command-lets for provisioning and managing IBM Storage systems

3.2.1 VMware integration

As shown in Figure 3-2, IBM Spectrum Connect provides a single-server back-end location and enables centralized management of IBM storage resources for the use of independent software vendor (ISV) platforms and frameworks. These frameworks currently include VMware vCenter Server, VMware vSphere Web Client, and VMware vSphere Storage APIs for Storage Awareness (VASA). IBM Spectrum Connect is available for no extra fee to storage-licensed clients.

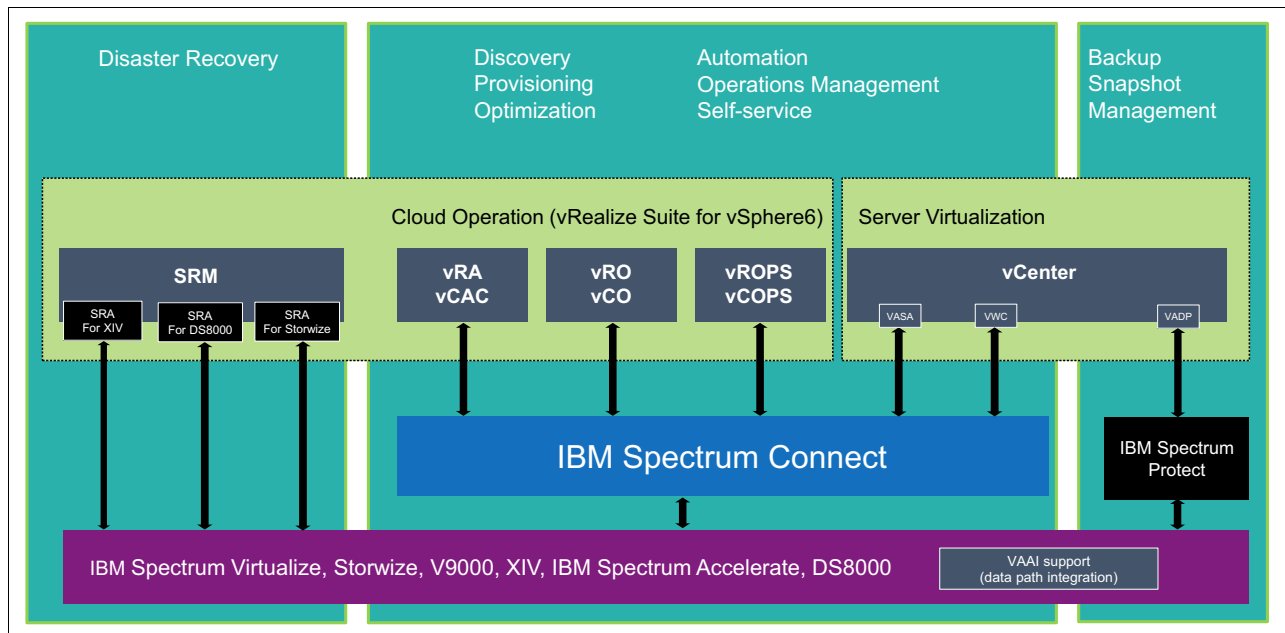


Figure 3-2 IBM Spectrum Connect

As shown in Figure 3-2, IBM Spectrum Connect is not in the data path. IBM Spectrum Connect runs in the control plane, as shown in Figure 3-1 on page 20. IBM Spectrum Connect provides integration between IBM Block Storage and VMware. Clients use IBM Spectrum Connect if they are using, or plan to use, the VMware Web Client (VWC), VMware Virtual Volumes (VVOL), or the vRealize Automation Suite from VMware.

IBM Spectrum Connect provides common services, such as authentication, high availability (HA), and storage configuration for IBM Block Storage in homogeneous and heterogeneous multiple target environments. IBM Spectrum Connect manages IBM XIV Storage System, A9000, A9000R, IBM DS8000 series, IBM SAN Volume Controller, the IBM Storwize® family, and third-party storage subsystems.

IBM Storage connectivity to VMware through IBM Spectrum Connect is shown in Figure 3-3.

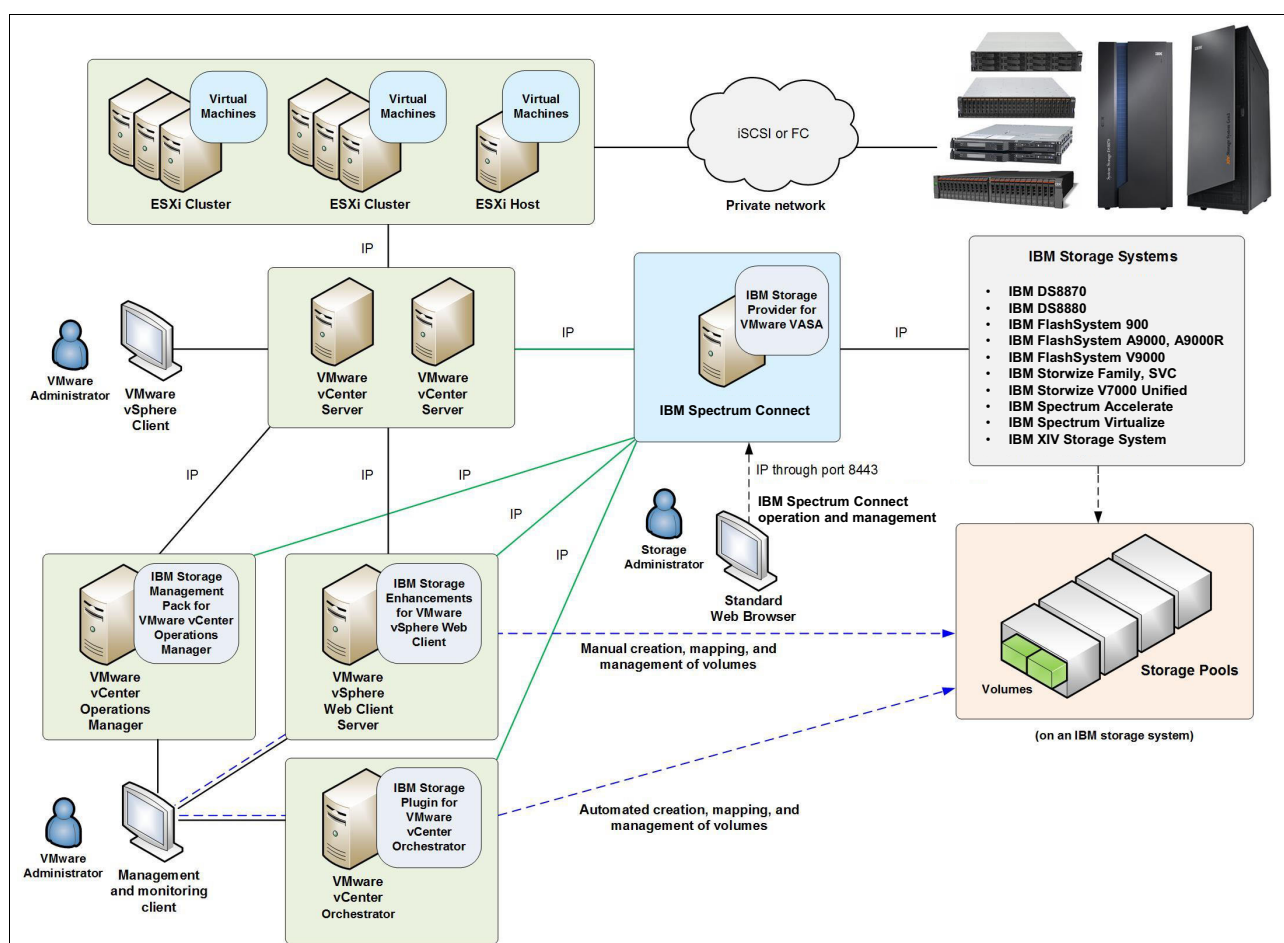


Figure 3-3 IBM Storage connectivity to VMware through IBM Spectrum Connect

3.2.2 PowerShell

IBM developed multiple PowerShell command-lets (cmdlets) for provisioning and managing IBM storage systems through trusted PowerShell commands to the devices. These command-lets are included in the IBM Storage Automation Plug-in for PowerShell, which is deployed on a PowerShell host that uses Spectrum Control as the common user interface. The capabilities can also be used with PowerCLI to automate storage-related tasks for Microsoft environments that are managed in VMware vSphere.

3.2.3 Containers and Kubernetes

IBM Spectrum Connect enables container orchestration with Kubernetes, including IBM Cloud Private solutions. It simplifies the provisioning of storage for containers by defining policies by SLA or by workload. It supports multiple and variable IBM storage systems as a single interface. It also provides better storage visibility to improve troubleshooting in containerized environments. For more information about the integration with Containers, see “This section describes IBM block storage solutions.” on page 48.

For more information about IBM Spectrum Connect, see this website:

<https://www.ibm.com/us-en/marketplace/spectrum-connect>

3.3 IBM Spectrum Control



IBM Spectrum Control provides efficient infrastructure management for virtualized, cloud, and software-defined storage by reducing the complexity that is associated with managing multivendor infrastructures. It also helps businesses optimize provisioning, capacity, availability, protection, reporting, and management for today's business applications without having to replace existing storage infrastructure. With support for block, file, and object workloads, IBM Spectrum Control enables administrators to provide efficient management for heterogeneous storage environments.

3.3.1 Key capabilities

IBM Spectrum Control helps organizations transition to new workloads and updated storage infrastructures by providing the following advantages to significantly reduce total cost of ownership:

- ▶ A single management console that supports IBM Spectrum Virtualize, IBM Spectrum Accelerate, IBM Cloud Object Storage, and IBM Spectrum Scale environments, enabling holistic management of physical and virtual block, file, and object systems storage environments.
- ▶ Insights that offer advanced, detailed metrics for storage configurations, performance, and tiered capacity in an intuitive web-based user interface with customizable dashboards so that the most important information is always accessible.
- ▶ Performance monitoring views that enable quick and efficient troubleshooting during an issue with simple threshold configuration and fault alerting for HA.

3.3.2 Benefits

IBM Spectrum Control can help reduce the administrative complexity of managing a heterogeneous storage environment, improve capacity forecasting, and reduce the amount of time spent troubleshooting performance-related issues. IBM Spectrum Control provides the following key values:

- ▶ Transparent mobility across storage tiers and devices for IBM Spectrum Virtualize based designs
- ▶ Centralized management that offers visibility to block, file, and object workloads and control and automation of block storage volumes

The IBM Spectrum Control dashboard window, in which all of the managed resources in a data center are presented in an aggregated view, is shown in Figure 3-4.

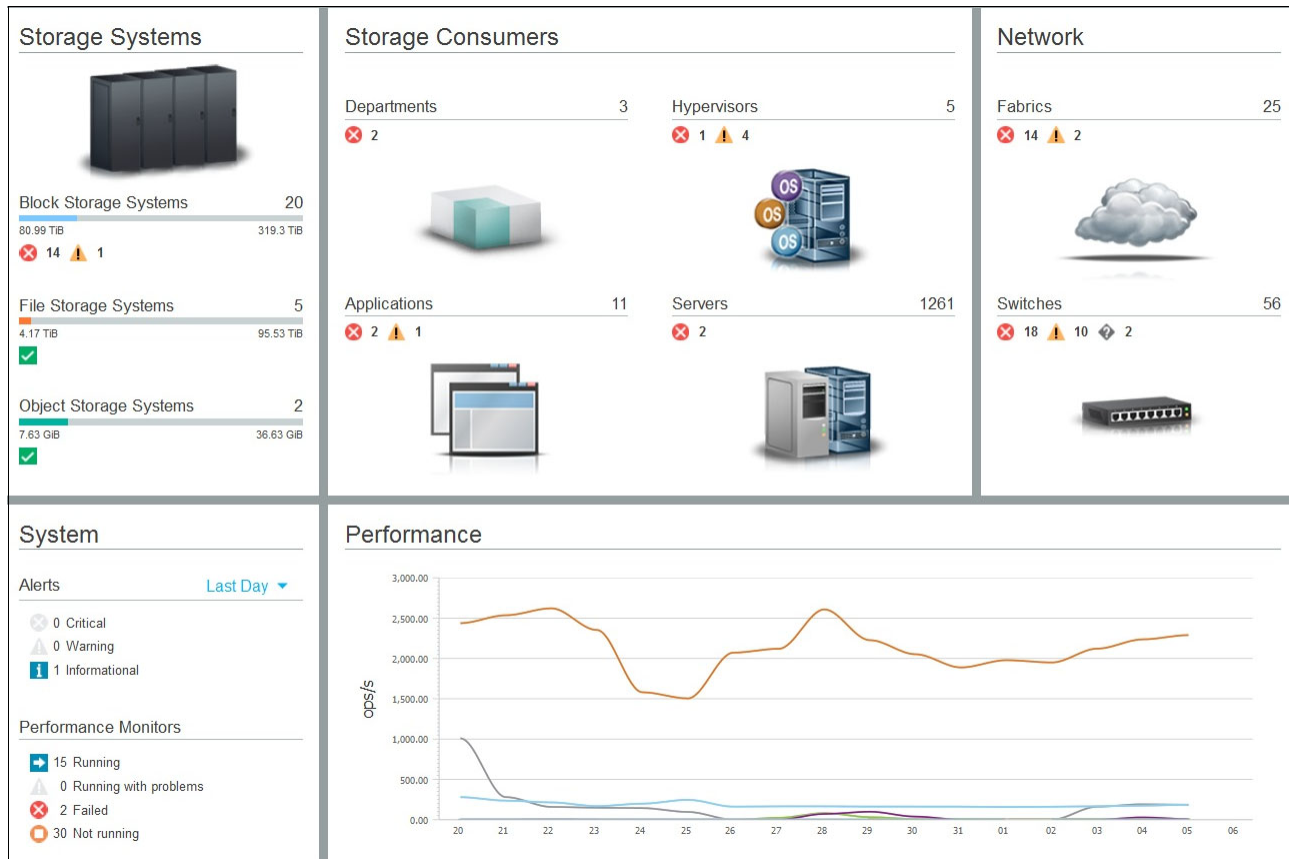


Figure 3-4 Single dashboard for monitoring all storage components

3.3.3 IBM Data and Storage Management Solutions features

IBM Spectrum Control solutions provide improved visibility, simplified administration, and greater scalability. This section describes the features of the specific products that provide the functions for IBM Spectrum Control.

Note: The Management Layer of VSC is now called IBM Spectrum Control Advanced Edition.

The IBM Spectrum Control offerings are shown in Figure 3-5.

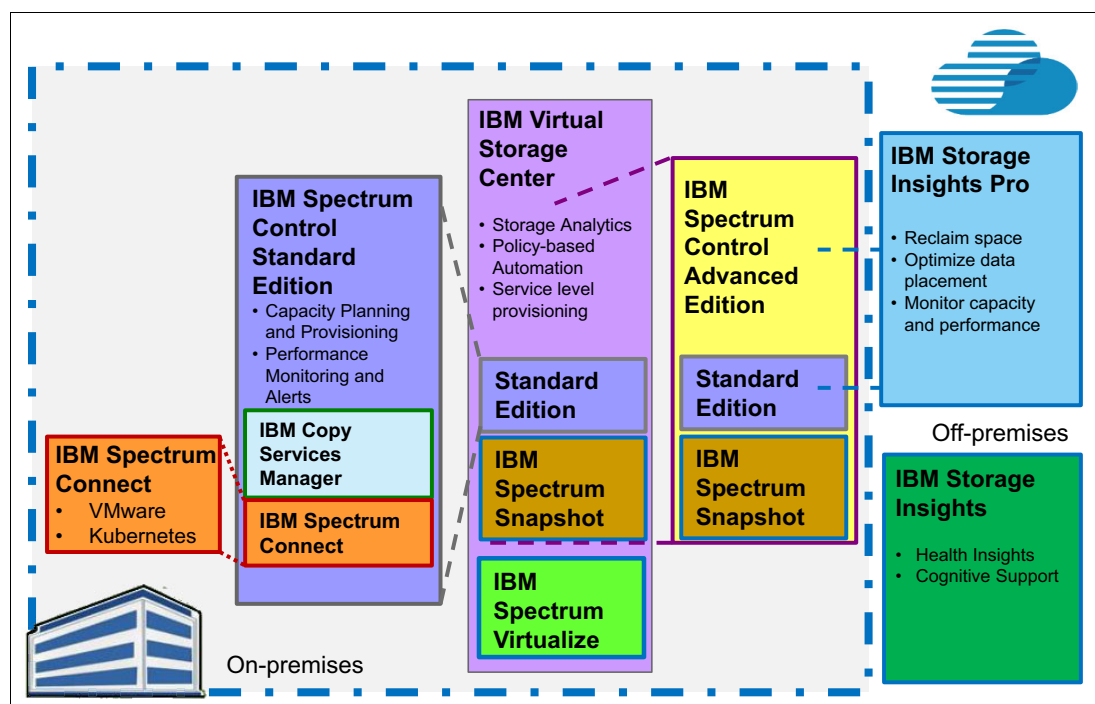


Figure 3-5 IBM Spectrum Control offerings

3.3.4 IBM Spectrum Control Standard Edition

IBM Spectrum Control Standard Edition is designed to provide storage infrastructure and data management capabilities for traditional and software-defined storage environments. IBM Spectrum Control Standard Edition includes the following primary features:

- ▶ Capacity visualization and management
- ▶ Performance reporting and troubleshooting
- ▶ Health and performance alerting
- ▶ Data Path view
- ▶ Department and Application grouping
- ▶ Hypervisor integration with VMware
- ▶ Management of IBM Replication features with Copy Services Manager (CSM)

3.3.5 IBM Spectrum Control Advanced Edition

IBM Spectrum Control Advanced Edition includes all of the features of IBM Spectrum Control Standard Edition, and adds the following advanced capabilities:

- ▶ Tiered storage optimization with intelligent analytics for IBM Spectrum Virtualize
- ▶ Service catalog with policy-based provisioning
- ▶ Self-service provisioning with restricted use logins
- ▶ Application-based snapshot management from IBM Spectrum Protect™ Snapshot

Advanced Edition includes the following built-in efficiency features that help users avoid complicated integration issues or the need to purchase add-ons or extra licenses:

- ▶ **Simplified user experience:** Virtual Storage Center provides an advanced GUI where administrators can perform common tasks consistently over multiple storage systems, including those systems from different vendors. The IBM storage GUI enables simplified storage provisioning with intelligent presets and embedded best practices, and integrated context-sensitive performance management.
- ▶ **Near-instant, application-aware backup and restore:** To reduce downtime in high-availability virtual environments, critical applications such as mission critical databases or executive email requiring near-instant backups must have little or no impact on application performance. Application-aware snapshot backups can be performed frequently throughout the day to reduce the risk of data loss. Virtual Storage Center simplifies administration and recovery from snapshot backups through the inclusion of IBM Spectrum Protect Snapshot.

IBM Spectrum Protect Snapshot, previously known as IBM Tivoli® Storage FlashCopy® Manager, is designed to deliver data protection for business-critical applications through integrated application snapshot backup and restore capabilities. These capabilities are achieved through the utilization of advanced storage-specific hardware snapshot technology to help create a high-performance, low-impact, application data protection solution. It is designed for easy installation, configuration, and deployment, and integrates with various traditional storage systems and software-defined storage environments.

- ▶ **IBM Tiered Storage Optimizer:** Virtual Storage Center uses performance metrics, advanced analytics, and automation to enable storage optimization on a large scale. Self-optimizing storage adapts automatically to workload changes to optimize application performance, eliminating most manual tuning efforts. It can optimize storage volumes across different storage systems and virtual machine vendors. The Tiered Storage Optimizer feature can reduce the unit cost of storage by as much as 50 percent, based on deployment results in a large IBM data center.

IBM Spectrum Control Advanced Edition is data and storage management software for managing heterogeneous storage infrastructures. It helps to improve visibility, control, and automation for data and storage infrastructures. Organizations with multiple storage systems can simplify storage provisioning, performance management, and data replication.

IBM Spectrum Control Advanced Edition simplifies the following data and storage management processes:

- ▶ A single console for managing all types of data on disk, flash, file, and object storage systems.
- ▶ Simplified visual administration tools that include an advanced web-based user interface, the ability to see servers and connected hypervisors (such as VMware) and IBM Cognos® Business Intelligence with pre-designed reports.
- ▶ Storage and device management to give you fast deployment with agent-less device management.
- ▶ Intelligent presets that improve provisioning consistency and control.
- ▶ Integrated performance management features end-to-end views that include devices, SAN fabrics, and storage systems. The server-centric view of storage infrastructure enables fast troubleshooting.
- ▶ Data replication management that enables you to have remote mirror, snapshot, and copy management, and supports Windows, Linux, UNIX, and IBM z® Systems® data with the inclusion of IBM Spectrum Copy Services Manager.

IBM Spectrum Control enables multi-platform storage virtualization and data and storage management. It supports most storage systems and devices by using the Storage Networking Industry Association (SNIA) Storage Management Initiative Specification (SMI-S), versions 1.0.2, 1.1, and 1.5 and later.

Hardware and software interoperability information is provided on the IBM Support Portal for IBM Spectrum Control. For more information about the interoperability matrix, see this website:

<http://www.ibm.com/support/docview.wss?uid=swg27047049>

Advanced Edition enables you to adapt to the dynamic storage needs of your applications by providing storage virtualization, automation, and integration for cloud environments, including the following features:

- ▶ OpenStack cloud application provisioning: Advanced Edition includes an OpenStack Cinder volume driver that enables automated provisioning using any of the heterogeneous storage systems that are controlled by IBM Cloud Orchestrator or Virtual Storage Center. OpenStack cloud applications can access multiple storage tiers and services without adding complexity.
- ▶ Self-service portal: Advanced Edition can provide provisioning automation for self-service storage portals, which enables immediate responses to service requests while eliminating manual administration tasks.
- ▶ Pay-per-use invoicing: Advanced Edition now includes a native chargeback tool. This tool allows customers to create chargeback or showback reports from the native GUI, or work with more advanced reporting as part of the embedded Cognos engine that is also included for building custom reports.

IBM Cognos-based reporting helps create and integrate custom reports about capacity, performance, and utilization. IBM Spectrum Control provides better reporting and analytics with no extra cost through integration with Cognos reporting and modeling.

Some reporting is included. Novice users can rapidly create reports with the intuitive drag function. Data abstraction and ad hoc reporting makes it easy to create high-quality reports and charts. You can easily change the scaling and select sections for both reporting and charting. Reports can be generated on schedule or on demand in multiple distribution formats, including email.

IBM Spectrum Control provides better user management and integration with external user repositories, such as Microsoft Active Directory. Enhanced management for virtual environments provides enhanced reporting for virtual servers (VMware). Tiered Storage Optimization provides integration with the existing storage optimizer and storage tiering reporting. Tiered Storage Optimization is policy-driven information lifecycle management (ILM) that uses virtualization technology to provide recommendations for storage relocation.

It provides recommendations for workload migration based on user-defined policy that is based on file system level data, performance, and capacity utilization. This feature ensures that only the highest performing workloads are allocated to the most expensive storage.

IBM Spectrum Control in a VMware environment

IBM Spectrum Control supports the following functions in VMware:

- ▶ Acts as a Control Plane for the storage reporting and provides for the capability to provision Spectrum Virtualize based architectures.
- ▶ Provides a view into the connected VMWare servers and their virtual machines.

IBM Spectrum Control in an OpenStack environment

The IBM Spectrum Control OpenStack Cinder driver enables your OpenStack-powered cloud environment to use your IBM Spectrum Control installation for block storage provisioning.

IBM Spectrum Control provides block storage provisioning capabilities that a storage administrator can use to define the properties and characteristics of storage volumes within a particular service class. For example, a block storage service class can define RAID levels, tiers of storage, and various other storage characteristics.

3.4 IBM Virtual Storage Center

Organizations need to spend less of their IT budgets on storage capacity and storage administration so that they can spend more on new, revenue-generating initiatives. Virtual Storage Center (VSC) delivers an end-to-end view of storage with the ability to virtualize Fibre Channel block storage infrastructures, helping you manage your data with more confidence with improved storage utilization and management efficiency.

It combines IBM Spectrum Control Advanced features with IBM Spectrum Virtualize capabilities to deliver an integrated infrastructure to transform your block storage into an agile, efficient, and economical business resource.

IBM Virtual Storage Center is a virtualization platform and a management solution for cloud-based and software-defined storage. It is an offering that combines both IBM Spectrum Control Advanced Edition with IBM Spectrum Virtualize, including SAN Volume Controller, members of the IBM Storwize family, and FlashSystem V9000. VSC helps organizations transition to new workloads and update storage infrastructures.

It enables organizations to monitor, automate, and analyze storage. It delivers provisioning, capacity management, storage tier optimization, and reporting. VSC helps standardize processes without replacing existing storage systems, and can also significantly reduce IT costs by making storage more user and application oriented.

Cloud computing is all about agility. Storage for clouds needs to be as flexible and service-oriented as the applications it supports. IBM Virtual Storage Center can virtualize existing storage into a private storage cloud with no “rip and replace” required.

3.5 IBM Storage Insights



IBM Storage Insights provides for a storage resource management solution that is delivered in an SaaS format from the IBM Cloud. This delivery allows the versions and updates to be automatically managed by IBM and enable customers to use the platforms for insight and knowledge of their storage environment. Two components can build off each other; IBM Storage Insights is available to all IBM Storage customers running storage platforms that are under support. It provides for an integrated support experience, short-term monitoring, and the ability for cognitive insight into storage health.

IBM Storage Insights Pro is a for fee-based offering that is integrated into the same management pane as IBM Storage Insights. It provides storage performance reporting, insight into usage, and the ability to monitor use at the Application or Department level.

3.5.1 IBM Storage Insights

The success of a business is intertwined with its IT performance. As IT environments become increasingly complex, the technology that is supposed to help is now beyond what humans alone can manage. To be successful, enterprises must rethink how they use technology to give them more power than ever before.

IBM Storage Insights provides Cognitive Storage Management capabilities for clients with IBM Storage. This new support capability automates data access and increases insight into storage health, performance, and capacity. It is designed for these clients to enjoy faster resolution of issues with minimal effort, HA and the confidence of services delivered from the IBM Cloud (see Figure 3-6).

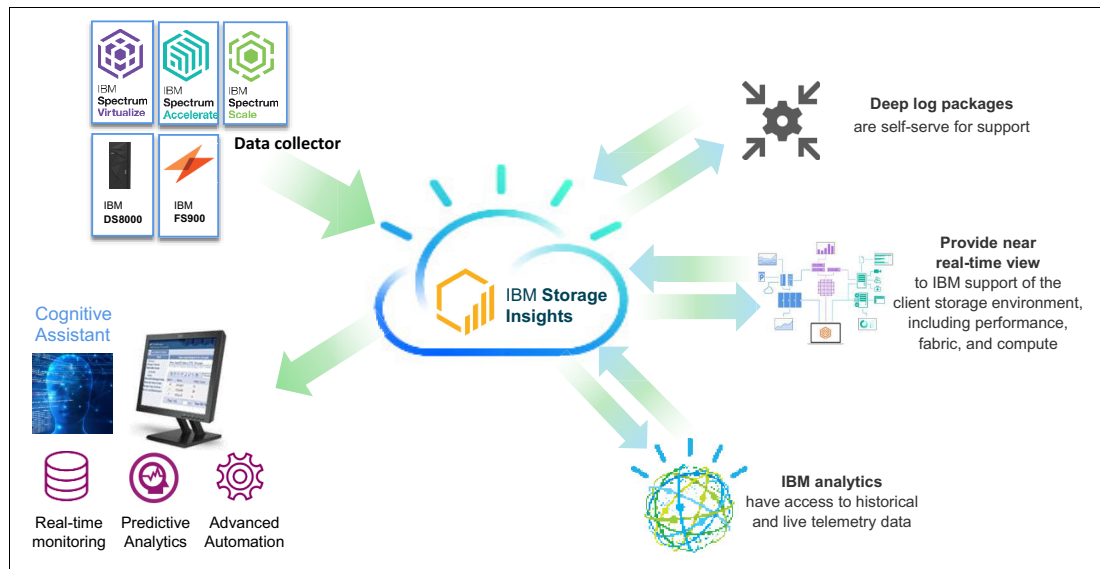


Figure 3-6 IBM Storage Insights

The platform can use IBM's Data Lake and Knowledge Base that includes curated data that is based on 30+ years of operations experience. This rich foundation provides valuable knowledge and insights that cognitive capabilities can mine.

By using operational data and applying artificial intelligence (or cognitive technology) to deliver actionable insights and drive automation are at the core of this transformation. Partnering people with cognitive technologies allows enterprises to autonomously run and optimize their IT environments based on business needs.

This ability, in turn, enables cognitive insights for faster, data-driven decisions and autonomous management and governance of IT operations. The result is the delivery of higher service quality by anticipating problems, reducing errors, and responding more quickly to incidents and service requests (see Figure 3-7).

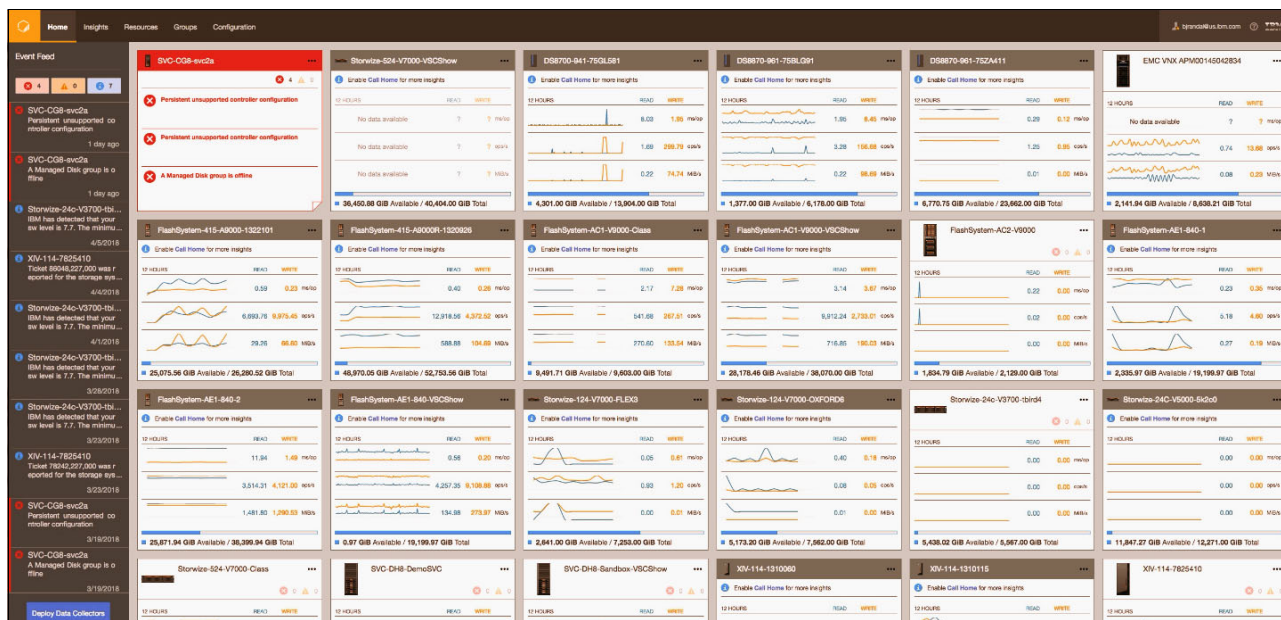


Figure 3-7 IBM Storage Insights Dashboard

3.5.2 IBM Storage Insights Pro

IBM Storage Insights Pro, formally called IBM Spectrum Control Storage Insights, is an analytics-driven, storage resource management solution that is delivered over the cloud. The solution uses cloud technology to provide visibility into on-premises storage with the goal of helping clients optimize their storage environments in today's data-intensive world. This software as a solution (SaaS) runs on IBM Cloud can deploy in as little as 5 minutes and show actionable insights in 30 minutes.

IBM Storage Insights Pro is a cloud data and storage management service that is deployed in a secure and reliable cloud infrastructure that provides the following features:

- ▶ Accurately identify and categorize storage assets
- ▶ Monitor capacity and performance from the storage consumer's view, including server, application, and department-level views
- ▶ Increase capacity forecasting precision by using historical growth metrics
- ▶ Reclaim unused storage to delay future purchases and improve utilization
- ▶ Optimize data placement based on historical usage patterns that can help lower the cost

An example of IBM the Storage Insights Pro dashboard is shown in Figure 3-8.



Figure 3-8 IBM Storage Insights Pro Dashboard

For more information about IBM Storage Insights Pro, see the following websites:

- ▶ <http://www.ibm.com/systems/storage/spectrum/insights>
- ▶ <http://www.ibm.com/marketplace/cloud/analytics-driven-data-management/us/en-us>

IBM Storage Insights can also be used with IBM Spectrum Control. For more information, see “IBM Spectrum Control offerings” on page 28.

3.6 IBM Copy Services Manager



IBM CSM replication management tool set (formerly in IBM Tivoli Storage Productivity Center) is included in IBM Spectrum Control. This replication management solution delivers central control of your replication environment by using simplified and automated complex replication tasks.

By using the CSM functions within IBM Spectrum Control, you can coordinate copy services on IBM Storage, including DS8000, DS6000TM, SAN Volume Controller, Storwize V7000, IBM Spectrum Accelerate, and XIV. You can also help prevent errors and increase system continuity by using source and target volume matching, site awareness, disaster recovery testing, and standby management. Copy services include IBM FlashCopy, Metro Mirror, Global Mirror, and Metro Global Mirror.

You can use Copy Services Manager to complete the following data replication tasks and help reduce the downtime of critical applications:

- ▶ Plan for replication when you are provisioning storage
- ▶ Keep data on multiple related volumes consistent across storage systems during a planned or unplanned outage
- ▶ Monitor and track replication operations

- ▶ Automate the mapping of source volumes to target volumes
- ▶ Practice disaster recovery procedures

The IBM Copy Services Manager family of products consists of the following products:

- ▶ Copy Services Manager provides HA and disaster recovery for multiple sites
- ▶ Copy Services Manager for IBM z Systems provides HA and disaster recovery for multiple sites
- ▶ Copy Services Manager Basic Edition for z Systems provides HA for a single site if a disk storage system failure occurs

The Copy Services Manager overview window is shown in Figure 3-9.

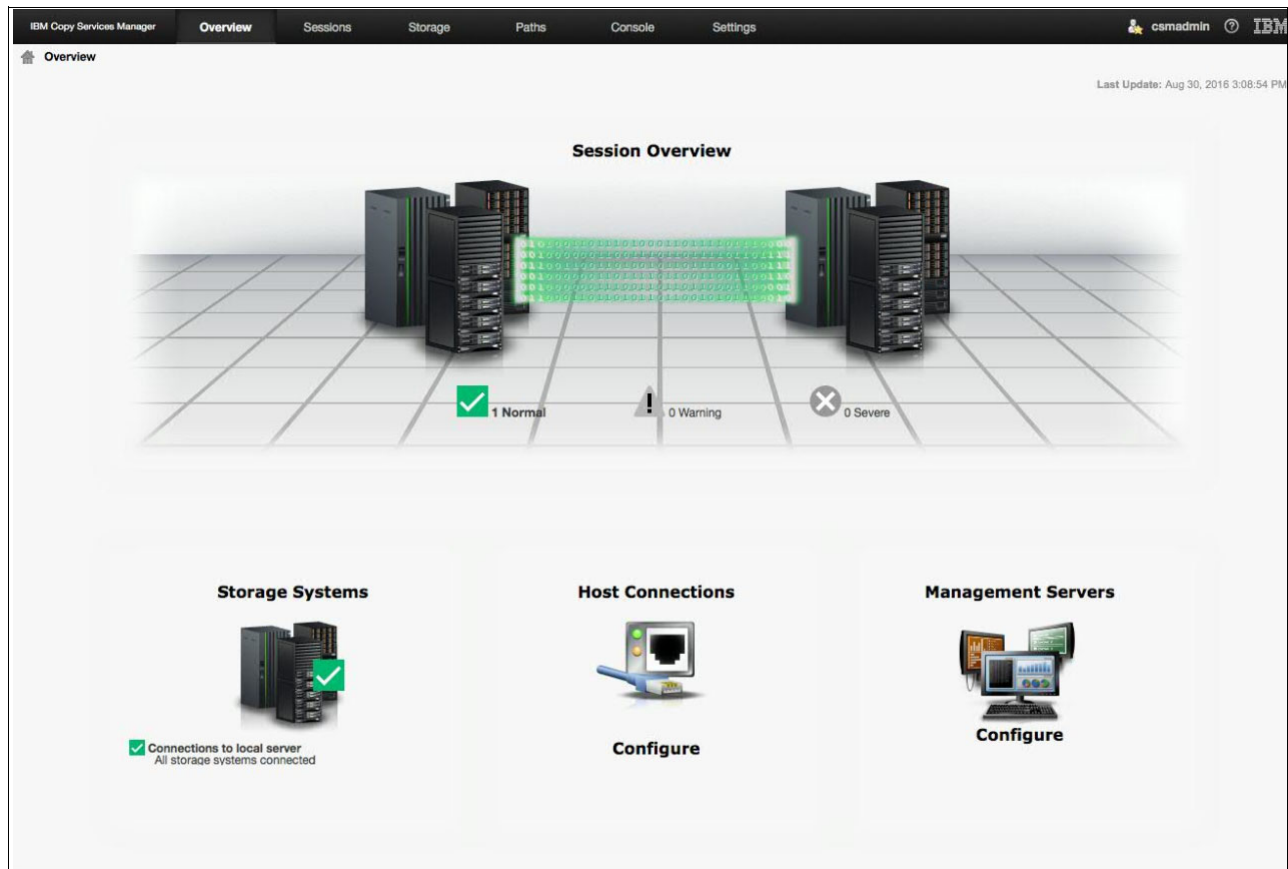


Figure 3-9 Copy Services Manager Overview window

3.7 IBM Spectrum Protect



IBM Spectrum Protect is an intuitive, intelligent, and transparent software that provides a set of product features that allow you to design adaptive and comprehensive data protection solutions. It is a comprehensive data protection and recovery solution for virtual, physical, and cloud data. IBM Spectrum Protect provides backup, snapshot, archive, recovery, space management, bare machine recovery, and disaster recovery capabilities.

3.7.1 Key capabilities

IBM Spectrum Protect features the following capabilities:

- ▶ Protects virtual, physical, and cloud data with one solution
- ▶ Reduces backup and recovery infrastructure costs
- ▶ Delivers greater visualization and administrator productivity
- ▶ Simplifies backups by consolidating administration tasks
- ▶ Space Management moves less active data to less expensive storage, such as tape or cloud
- ▶ Provides long-term data archive for data retention, such as for compliance with government regulations

3.7.2 Benefits

IBM Spectrum Protect includes the following benefits:

- ▶ Application-aware and VM-aware data protection for any size organization
- ▶ Simplified administration
- ▶ Built-in efficiency features: Data deduplication, compression, and incremental ‘forever’ backup
- ▶ Integrated multi-site replication and disaster recovery
- ▶ Multi-site data availability with active-active replication-based architecture and heterogeneous storage flexibility using disk, tape, or cloud

Whatever your data type and infrastructure size, IBM Spectrum Protect scales from a small environment that consists of 10 - 20 machines to a large environment with thousands of machines to protect. The software product consists of the following basic functional components:

- ▶ IBM Spectrum Protect server with IBM Db2® database engine

The IBM Spectrum Protect server provides backup, archive, and space management services to the IBM Spectrum Protect clients, and manages the storage repository. The storage repository can be implemented in a hierarchy of storage pools by using any combination of supported media and storage devices. These devices can be directly connected to the IBM Spectrum Protect server system or be accessible through a SAN or be cloud storage accessible using TCP/IP.

- ▶ IBM Spectrum Protect clients with application programming interfaces (APIs)

IBM Spectrum Protect enables data protection from failures and other errors by storing backup, archive, space management, and “bare-metal” restore data, and compliance and disaster-recovery data in a hierarchy of auxiliary storage. IBM Spectrum Protect can help protect computers that run various operating systems, on various hardware platforms and connected together through the Internet, wide area networks (WANs), local area networks (LANs), or storage area networks (SANs).

It uses web-based management, intelligent data move-and-store techniques, and comprehensive policy-based automation that work together to increase data protection and potentially decrease time and administration costs.

The progressive incremental methods that are used by IBM Spectrum Protect back up only new or changed versions of files, greatly reducing data redundancy, network bandwidth, and storage pool consumption as compared to traditional methods.

3.7.3 Backup and recovery

Despite rapid data growth, data protection and retention systems are expected to maintain service levels and data governance policies. Data has become integral to business decision-making and basic operations, from production to sales and customer management. Data protection and retention are core capabilities for their role in risk mitigation and for the amount of data involved.

The storage environment offers the following functions that improve the efficiency and effectiveness of data protection and retention:

- ▶ **Backup and recovery:** Provides cost-effective and efficient backup and restore capabilities, improving the performance, reliability, and recovery of data that is aligned to business required service levels. Backups protect current data, and are unlikely to be accessed unless data is lost or corrupted.
- ▶ **Archiving:** Stores data that includes long-term data retention requirements for compliance or business purposes by providing secure and cost effective solutions with automated process for retention policies and data migration to different storage media.
- ▶ **Node Replication:** Ensures uninterrupted access to data for critical business systems, reducing the risk of downtime by providing the capability to fail over transparently and as instantaneously as possible to an active copy of the data.

Optimizing all of these areas helps an organization deliver better services with reduced application downtime. Data protection and retention, archiving, and node replication can improve business agility by ensuring that applications have the correct data when needed, while inactive data is stored in the correct places for the correct length of time.

3.7.4 Tool set

IBM Spectrum Protect is a family of tools that helps manage and control the “information explosion” by delivering a single point of control and administration for storage management needs. It provides a wide range of data protection, recovery management, movement, retention, reporting, and monitoring capabilities by using policy-based automation.

Products: For an updated list of the available products in the IBM Spectrum Protect family, see the following website:

<https://www.ibm.com/us-en/marketplace/data-protection-and-recovery>

For more information about the most recent releases, see IBM Spectrum Protect Knowledge Center:

https://www.ibm.com/support/knowledgecenter/en/SSEQVQ_8.1.0/tsm/welcome.html

The main features, functions, and benefits that are offered by the IBM Spectrum Protect family are listed in Table 3-2.

Table 3-2 Main features, functions, and benefits of IBM Spectrum Protect

Feature	Function	Benefits
Backup and recovery management	Intelligent backups and restores using a progressive incremental backup and restore strategy, where only new and used files are backed up	Centralized protection based on smart-move and smart-store technology, which leads to faster backups and restores with fewer network and storage resources needed
Hierarchical storage management	Policy-based management of file backup and archiving	Ability to automate critical processes related to the media on which data is stored while reducing storage media and administrative costs associated with managing data
Archive management	Managed archives	Ability to easily protect and manage documents that need to be kept for a designated length of time
Advanced data reduction	Combines incremental backup, source inline, and target data deduplication, compression, and tape management to provide data reduction	Reduces the costs of data storage, environmental requirements, and administration

3.7.5 IBM Spectrum Protect Operations Center

IBM Spectrum Protect Operations Center is a graphical user interface (GUI), with new features (as shown in Figure 3-10). It provides an advanced visualization dashboard, built-in analytics, and integrated workflow automation features that dramatically simplify backup administration.

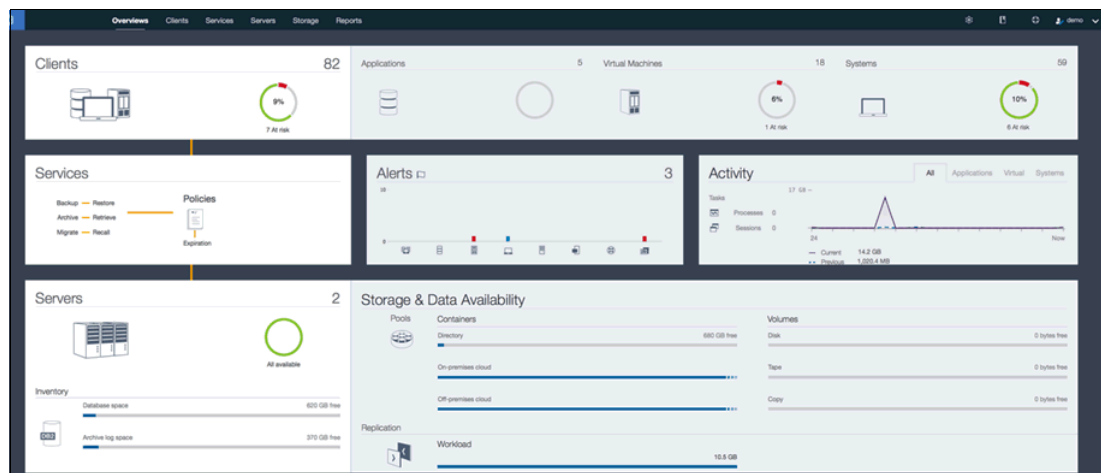


Figure 3-10 IBM Spectrum Protect Operations Center

3.7.6 IBM Spectrum Protect cloud architectures

IBM Spectrum Protect has multiple cloud architectures to meet various requirements. Figure 3-11 shows several IBM Spectrum Protect cloud architectures for storing IBM Spectrum Protect cloud-container storage pools.

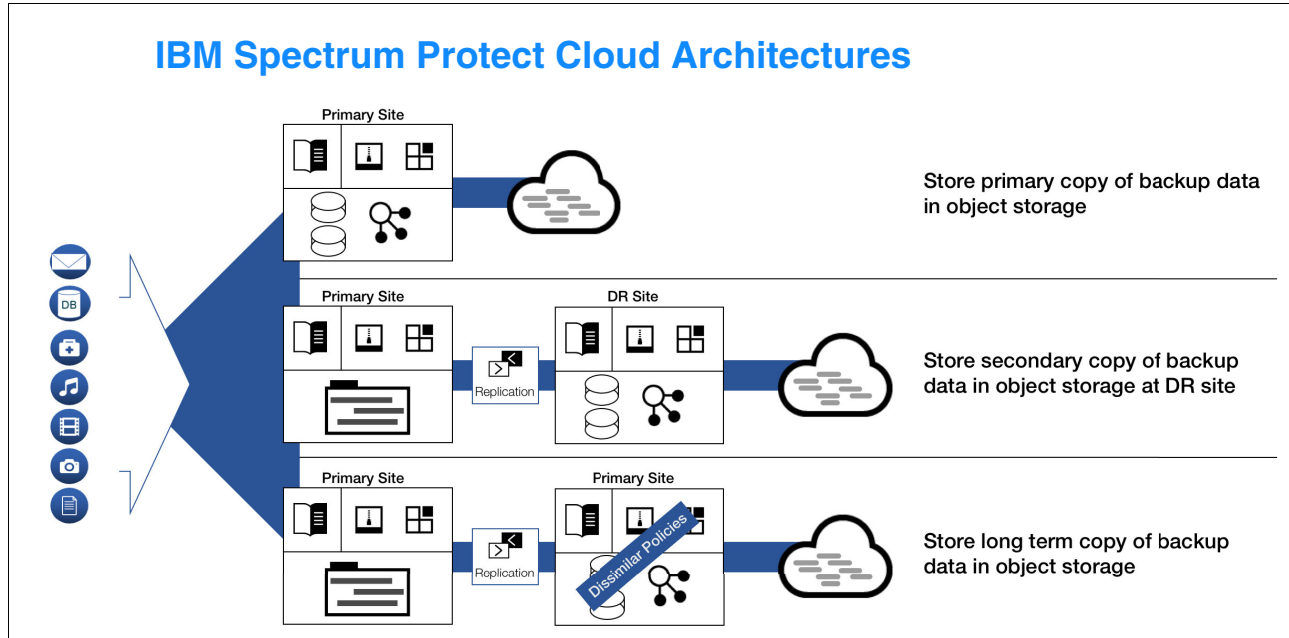


Figure 3-11 IBM Spectrum Protect Cloud Architectures

IBM Spectrum Protect supports the following cloud providers:

- ▶ IBM Cloud
- ▶ Amazon Web Services (Amazon Simple Storage Service S3)
- ▶ Microsoft Azure (Blob Storage)

IBM Spectrum Protect also supports IBM Cloud Object Storage (Cleversafe®) as an on-premise storage system that is configured by using dedicated hardware.

Several third-party vendors, including Scality RING and EMC Elastic Cloud Storage, validated their object storage hardware and software for use with IBM Spectrum Protect. For more information about these third-party devices, see this website:

<http://www.ibm.com/support/docview.wss?uid=swg22000915>

IBM Spectrum Protect can also protect data that is hosted in an OpenStack environment, and use the OpenStack (Swift) environment as a repository for backup and archive objects.

Data privacy considerations

Although the security of sensitive data is always a concern, data that you store off-premises in a cloud computing system should be considered particularly vulnerable. Data can be intercepted during transmission, or a weakness of the cloud computing system might be used to gain access to the data.

To guard against these threats, define a cloud-container storage pool to be encrypted. When you do, the server encrypts data before it is sent to the storage pool. After data is retrieved from the storage pool, the server decrypts it so is understandable and usable again.

Your data is protected from eavesdropping and unauthorized access when it is outside your network because it can be understood only when it is back on premises.

3.7.7 Cloud Tiering

Beginning with version 8.1.3 IBM Spectrum Protect, IBM Spectrum Protect allows tiering of data from directory container storage pools to cloud container storage pools, as shown in Figure 3-12.

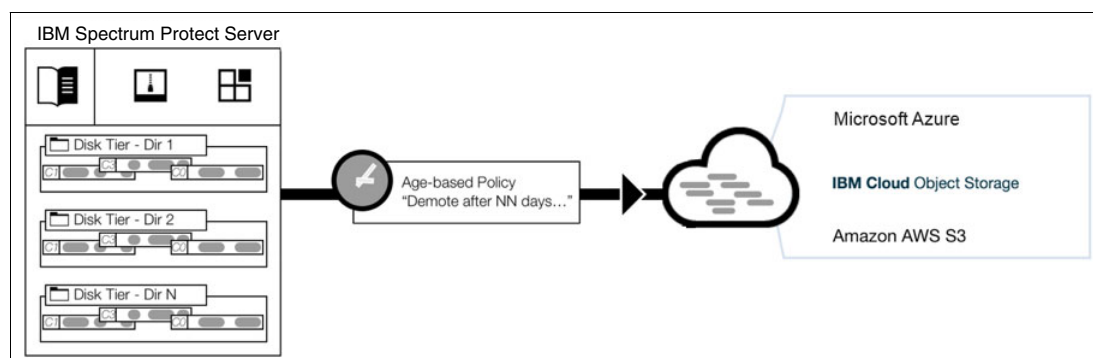


Figure 3-12 IBM Spectrum Protect cloud tiering

With cloud tiering, data is stored on block storage device for quick ingest and operational recovery. A storage rule keeps data on disk for a specified number of days, after which it is migrated off to cloud storage. When restoring data, the IBM Spectrum Protect server automatically restores data from wherever it is stored. The tiering feature supports on-premises and off-premises implementations of cloud storage.

3.7.8 IBM Spectrum Protect Plus



IBM Spectrum Protect Plus is a data protection and availability solution for virtual environments that can be deployed in minutes and protect your environment within an hour. It simplifies data protection, whether data is hosted in physical, virtual, software-defined or cloud environments. It can be implemented as a stand-alone solution or integrate with your IBM Spectrum Protect environment to off-load copies for long-term storage and data governance with scale and efficiency.

IBM Spectrum Protect Plus uses APIs and incremental-forever data copy technology to create backup copies of Hyper-V and VMware virtual machines. It stores these copies as addressable snapshot images on a vSnap server and optionally offloads them to IBM Spectrum Protect. IBM Spectrum Protect Plus can support all of the tape and cloud environments that are supported by IBM Spectrum Protect.

IBM Spectrum Protect Plus creates and maintains a global catalog of all copies of data and optionally indexes files. When the need to recover arises, this global catalog enables the administrator to quickly search and identify what they want to recover instead of browsing through hundreds of objects and recovery points.

IBM Spectrum Protect Plus provides instant access and restore from the catalog so that an administrator can restore the organization's operations in a matter of minutes and enables multiple use cases. These key use cases include Data Protection, Disaster Recovery (DR), Development and Test (Dev/Test), and Business Analytics.

Installation and Configuration

Because it is delivered as a Virtual Machine Image, IBM Spectrum Protect Plus is extremely simple to deploy. Install into your virtualization environment and access the dashboard by using a web browser.

Immediately available SLAs provide secure self-service management for defining schedules and protecting virtual machines. A simple to use dashboard (see Figure 3-13) provides information protection status, SLA compliance, VM Sprawl, and storage usage.

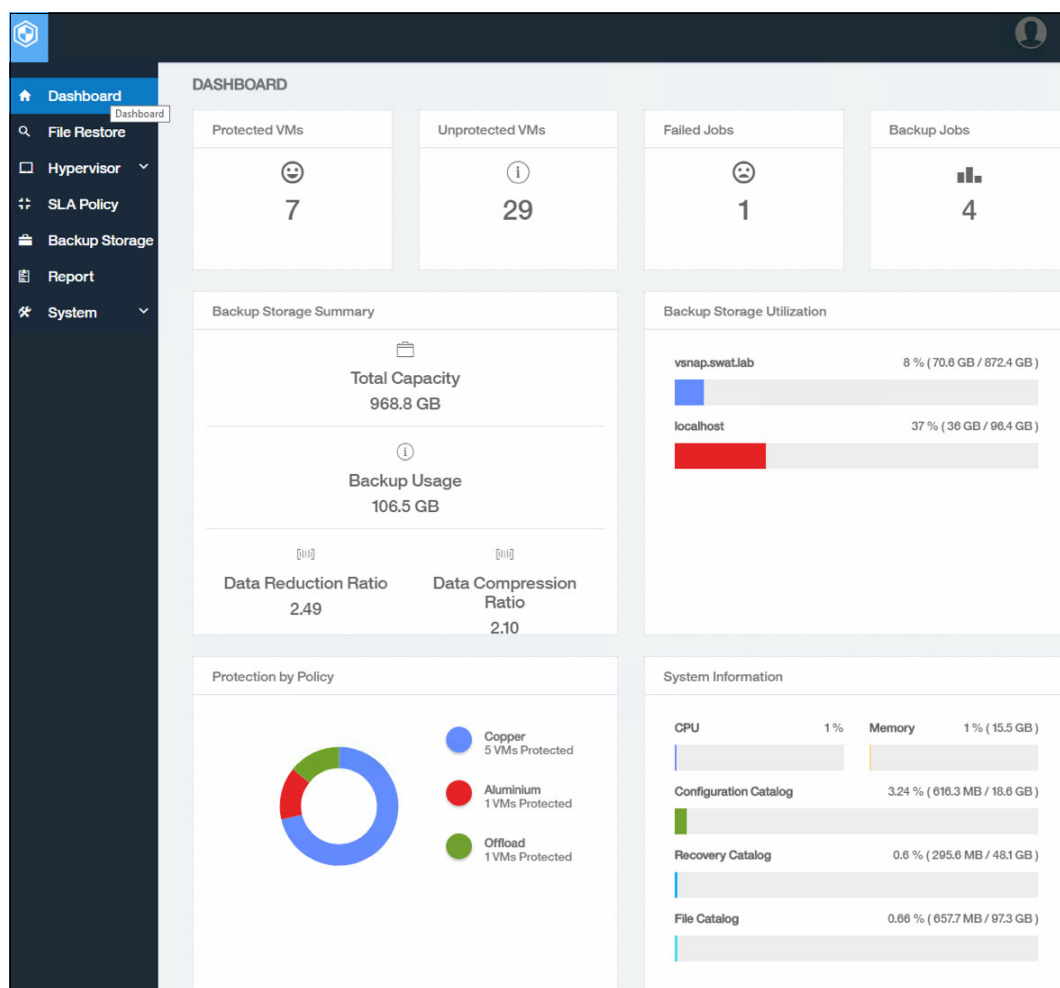


Figure 3-13 IBM Spectrum Protect Dashboard

3.7.9 IBM Spectrum Protect for Virtual Environments

IBM Spectrum Protect for Virtual Environments simplifies data protection for virtual and cloud environments. It protects VMware and Microsoft Hyper-V virtual machines by offloading backup workloads to a centralized IBM Spectrum Protect server for safe keeping. Administrators can create backup policies or restore virtual machines with a few clicks.

IBM Spectrum Protect for Virtual Environments enables your organization to protect data without the need for a traditional backup window. It allows you to reliably and confidently safeguard the massive amounts of information that virtual machines generate.

IBM Spectrum Protect for Virtual Environments provides the following benefits:

- ▶ Improves efficiency with data deduplication, incremental “forever” backup, and other advanced IBM technology to help reduce costs.
- ▶ Simplifies backups and restores for VMware with an easy-to-use interface that you can access from within VMware vCenter or vCloud Director.
- ▶ Enables VMware vCloud Director and OpenStack cloud backups.
- ▶ Enables faster, more frequent snapshots for your most critical virtual machines.
- ▶ Flexible recovery and copy options from image-level backups give you the ability to perform recovery at the file, mailbox, database object, volume, or VM image level by using a single backup of a VMware image.

Eliminates processor usage that is caused by optimized virtual machine backup by supporting VMware vStorage APIs for Data Protection and Microsoft Hyper-V technology, which simplifies and optimizes data protection.

3.8 IBM Spectrum Protect Snapshot



In today's business world, where application servers are operational 24 hours a day, the data on these servers must be fully protected. You cannot afford to lose any data, but you also cannot afford to stop these critical systems for hours so you can protect the data adequately. As the amount of data that needs protecting continues to grow exponentially and the need to keep the downtime associated with backup to an absolute minimum, IT processes are at their breaking point. Data volume snapshot technologies such as IBM Spectrum Protect Snapshot can help minimize the effect caused by backups and provide near instant restore capabilities.

Although many storage systems are now equipped with volume snapshot tools, these hardware-based snapshot technologies provide only “crash consistent” copies of data. Many business critical applications, including those that rely on a relational database, need an extra snapshot process to ensure that all parts of a data transaction are flushed from memory and committed to disk before the snapshot. This process is necessary to ensure that you have a usable, consistent copy of the data.

IBM Spectrum Protect Snapshot helps deliver the highest levels of protection for mission-critical IBM Db2, SAP, Oracle, Microsoft Exchange, and Microsoft SQL Server applications by using integrated, application-aware snapshot backup and restore capabilities. This protection is achieved by using advanced IBM storage hardware snapshot technology to create a high-performance, low-impact application data protection solution.

The snapshots that are captured by IBM Spectrum Protect Snapshot can be retained as backups on local disk. With optional integration with IBM Spectrum Protect, customers can use the full range of advanced data protection and data reduction capabilities, such as data deduplication, progressive incremental backup, hierarchical storage management, and centrally managed policy-based administration, as shown in Figure 3-14 on page 43.

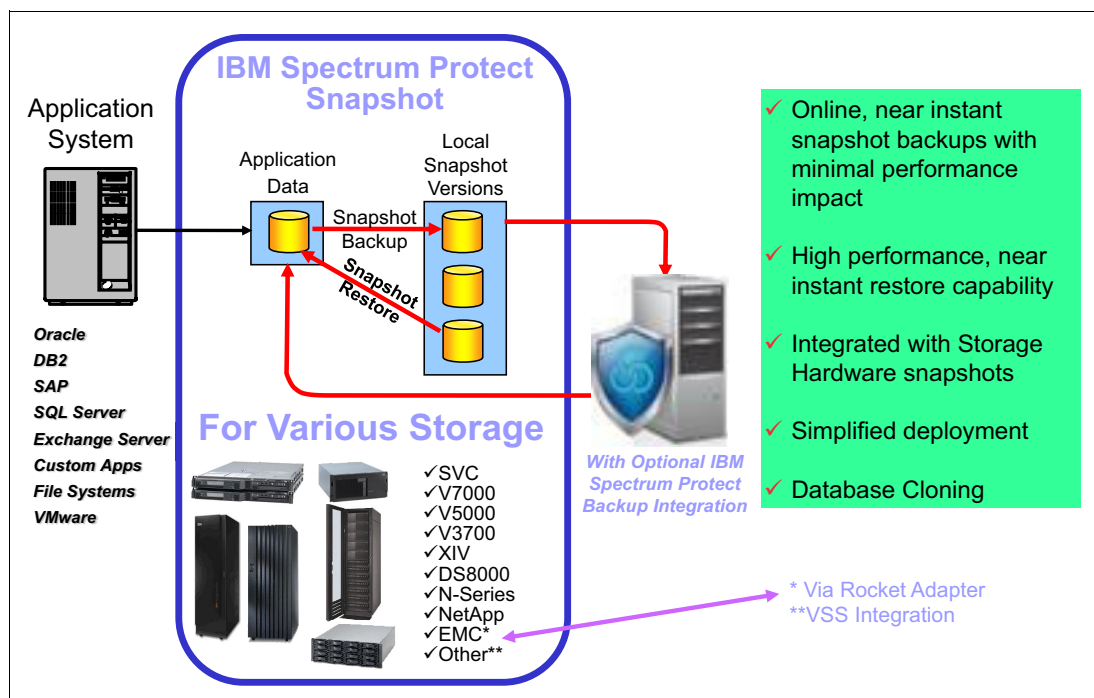


Figure 3-14 IBM Spectrum Protect Snapshot storage snapshot capabilities

Because a snapshot operation typically takes much less time than the time for a tape backup, the window during which the application must be aware of a backup can be reduced. This advantage facilitates more frequent backups, which can reduce the time that is spent performing forward recovery through transaction logs, increase the flexibility of backup scheduling, and ease administration.

Application availability is also significantly improved due to the reduction of the load on the production servers. IBM Spectrum Protect Snapshot uses storage snapshot capabilities to provide high speed, low impact, application-integrated backup and restore functions for the supported application and storage environments.

Automated policy-based management of multiple snapshot backup versions, together with a simple and guided installation and configuration process, provide an easy way to use and quick to deploy data protection solution that enables the most stringent database recovery time requirements to be met.

For more information, see this website:

<https://ibm.biz/BdZgV3>

3.9 IBM Spectrum Copy Data Management



IBM Spectrum Copy Data Management makes copies available to data consumers when and where they need them, without creating unnecessary copies or leaving unused copies on valuable storage. It catalogs copy data from across local, hybrid cloud, and off-site cloud infrastructure, identifies duplicates, and compares copy requests to existing copies. This process ensures that the minimum number of copies are created to service business requirements.

Data consumers can use the self-service portal to create the copies they need when they need them, creating business agility. Copy processes and work flows are automated to ensure consistency and reduce complexity. IBM Spectrum Copy Data Management rapidly deploys as an agentless VM as it helps manage snapshot and FlashCopy images made to support DevOps, data protection, disaster recovery, and Hybrid Cloud computing environments.

This member of the IBM Spectrum Storage family automates the creation and catalogs the copy data on existing storage infrastructure, such as snapshots, vaults, clones, and replicas. One of the key use cases centers around use with Oracle, Microsoft SQL server, and other databases that are often copied to support application development, testing, and data protection.

The IBM Spectrum Copy Data Management software is an IT modernization technology that focuses on using existing data in a manner that is efficient, automated, scalable, and easy to use to improve data access. IBM Spectrum Copy Data Management (Figure 3-15), with IBM storage arrays, delivers in-place copy data management that modernizes IT processes and enables key use cases with existing infrastructure.

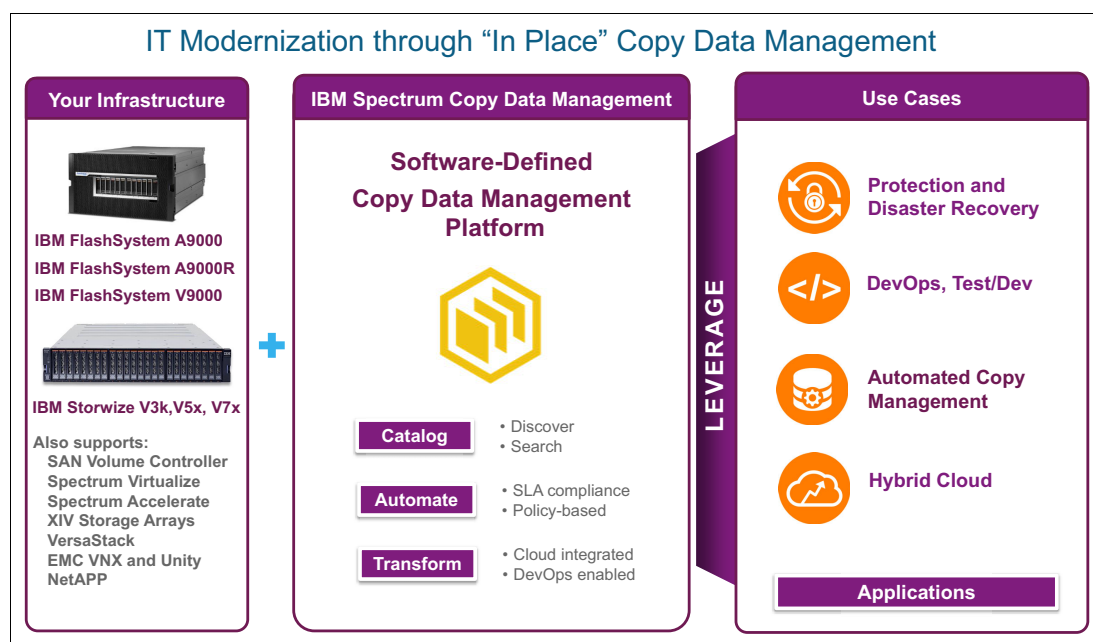


Figure 3-15 Software-Defined IBM Spectrum Copy Data Management Platform

IBM Spectrum Copy Data Management includes support for the following copy data management use cases:

- ▶ Automated Copy management
- ▶ Development and operations
- ▶ Data protection and disaster recovery
- ▶ Test and development
- ▶ Hybrid cloud computing

Automated copy management

IT functions that rely heavily on copies or ‘snapshots’ are typically managed by using a complex mix of scripts, tools, and other products, none of which are optimized for copy management. With IBM Spectrum Copy Data Management, organizations have a holistic, simplified approach that greatly reduces cycle time and frees staff to manage more productive projects.

IT teams can use the core policy engine, catalog, and reporting of IBM Spectrum Copy Data Management to dramatically improve IT operations that rely on copies of data, including disaster recovery, testing and development, business analytics, and local recovery. IBM Spectrum Copy Data Management improves operations by using automated, service-level based copy policies that are consistent, reliable, and easily repeatable. This feature provides huge savings in operating expenses.

Development and operations (DevOps)

Organizations are increasingly moving toward DevOps for faster delivery of new applications to market. IBM Spectrum Copy Data Management enables IT teams to use their existing storage infrastructure to enable DevOps, helping to meet the needs of the development teams for rapid deployment of the infrastructure. IBM Spectrum Copy Data Management templates define the policies for infrastructure deployment. The whole system is accessible through the REST API.

Rather than following legacy processes to requisition IT resources, developers include the infrastructure deployment commands directly within their development systems, such as IBM Cloud (IBM Bluemix®), Chef, or Puppet. Predefined scripts and plug-ins for popular DevOps tools simplify implementation.

Next-generation data protection and disaster recovery

Through its template-based management and orchestration of application-aware copies, IBM Spectrum Copy Data Management can support next-generation data protection and recovery workflows. IBM Spectrum Copy Data Management enables IT to mount and instantly access copies that are already in the production storage environment. IBM Spectrum Copy Data Management catalogs all snapshots and replicas, and alerts you if a snap or replication job was missed or failed.

Disaster recovery can be fully automated and tested nondisruptively. In addition, IBM Spectrum Copy Data Management can coordinate sending data through the AWS Storage Gateway to the AWS storage infrastructure. This feature provides a simplified, low-cost option for longer term or archival storage of protection copies.

Automated test and development

The speed and effectiveness of test and development processes are most often limited by the time it takes to provision IT infrastructure. With IBM Spectrum Copy Data Management, test and development infrastructure can be spun-up in minutes, either on an automated, scheduled basis, or on-demand basis.

Hybrid cloud computing

IBM Spectrum Copy Data Management is a powerful enabler of the hybrid cloud, enabling IT to take advantage of cloud compute resources. IBM Spectrum Copy Data Management not only helps customers move data to the cloud, it also enables IT organizations to create live application environments that can use the less expensive, elastic compute infrastructure in the cloud. Being able to spin up workloads and then spin them back down reliably helps maximize the economic benefit of the cloud by only using and paying for infrastructure as needed.

IBM Spectrum Copy Data Management is a software platform that is designed to use the infrastructure in the IT environment. It works directly with hypervisor and enterprise storage APIs to provide the overall orchestration layer that uses the copy services of the underlying infrastructure resources. IBM Spectrum Copy Data Management also integrates with IBM Cloud (IBM Bluemix is now integrated with IBM Cloud), AWS S3 for cloud-based data retention, Puppet, and others.

Database-specific functionality

IBM Spectrum Copy Data Management allows the IT team to easily create and share copies of all popular database management systems by integrating key database management system (DBMS) tasks within well-defined policies and work-flows. The solution also includes application-aware integration for Oracle and Microsoft SQL Server platforms, providing a deeper level of coordination with the DBMS.

Secure multi-tenancy

Secure multi-tenancy meets the needs of both managed service providers and large organizations that need to delegate resources internally. Individual “tenants” can be created within a single IBM Spectrum Copy Data Management instance, allowing each tenant its own set of resources and the ability to support administration within the tenancy to create users, define jobs, and perform other functions.

Policy templates for automation and self-service

Template-based provisioning and copy management provides easy self-service access for internal customers to request the resources that they need, when they need them. Templates are pre-defined by the IT team, and they are accessible through a self-service portal interface or through API calls.

Compatibility

IBM Spectrum Copy Data Management is a simple-to-deploy software platform that is designed to use the existing IT infrastructure. It works directly with hypervisor and storage APIs to provide the overall orchestration layer that uses the copy services of the underlying infrastructure resources. It also integrates with Amazon Web Services S3 for cloud-based data retention.

IBM Spectrum Copy Data Management delivers the following benefits:

- ▶ Automate the creation and use of copy data on existing storage infrastructure, such as snapshots, vaults, clones, and replicas
- ▶ Reduce time that is spent on infrastructure management while improving reliability
- ▶ Modernize existing IT resources by providing automation, user self-service, and API-based operations without the need for any additional hardware
- ▶ Simplify management of critical IT functions such as data protection and disaster recovery
- ▶ Automate test and development infrastructure provisioning, drastically reducing management time
- ▶ Drive new, high-value use cases, such as using hybrid cloud compute
- ▶ Catalog and track IT objects, including volumes, snapshots, virtual machines, data stores, and files

3.10 Block, file, and object storage

Block, file, and object are different approaches to accessing data. This section provides a high-level overview of each method. Figure 3-16 is a high-level view of these differences.

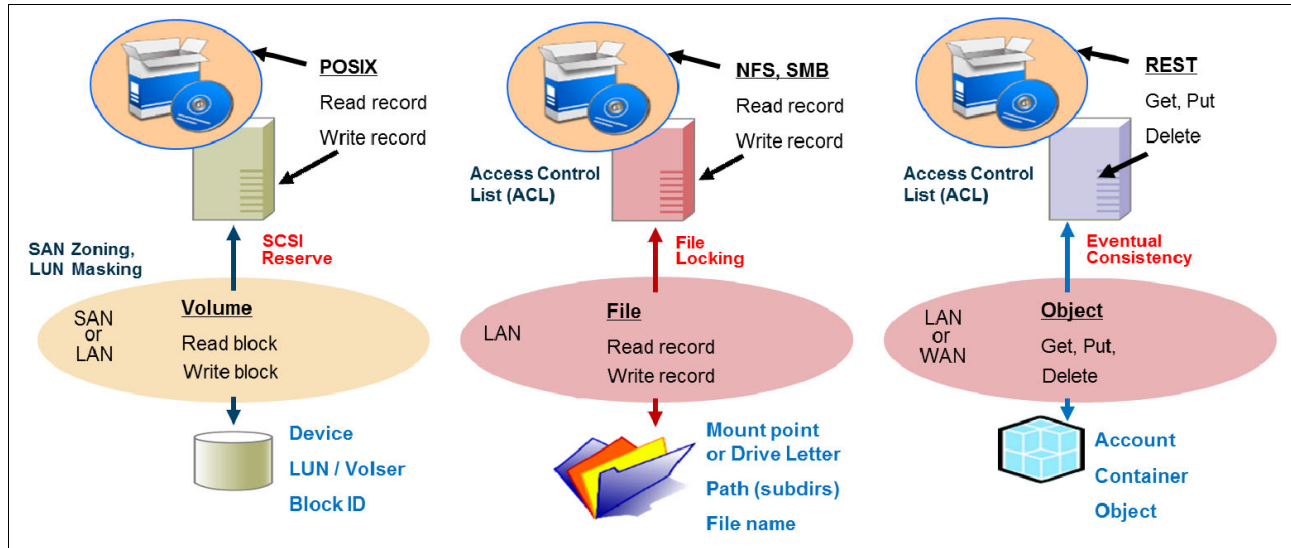


Figure 3-16 High-level view of data access differences between file, block, and object storage

3.10.1 Block storage

Block storage offerings are differentiated by speed/throughput (as measured in IOPS) and segmented by lifecycle of the disk. Data is split into evenly sized chunks or blocks of data, each with its own unique address. How blocks of data are accessed is up to the application. Few applications access the blocks directly. Rather, a Portable Operating System Interface (POSIX) file system is used in a hierarchical way of organizing files so that an individual file can be located by describing the path to that file.

3.10.2 File storage

File storage uses protocols to access individual directories and files over various protocols, including NFS, CIFS/SMB, and so on. Certain file attributes might describe a file and its contents, such as its owner, who can access the file, and its size. This metadata is all stored along with the data or related directory structure.

3.10.3 Object storage

With object storage, data is written into self-contained entities called objects. Unlike file systems, an object storage system gives each object a unique ID, which is managed in a flat index. There are no folders and subfolders. Unlike files, objects are created, retrieved, deleted, or replaced in their entirety, rather than being updated or appended in place.

Object storage also introduces the concept of *eventual consistency*. If one user creates an object, a second user might not see that object listed immediately. Eventually, all users will be able to see the object listed.

When a user or application needs access to an object, the object storage system is provided with a unique ID. This flat index approach provides greater scalability, enabling an object storage system to support faster access to a massively higher quantity of objects or files as compared to traditional file systems.

3.11 IBM Block Storage solutions

This section describes IBM block storage solutions.

3.11.1 IBM Spectrum Virtualize



IBM Spectrum Virtualize software is at the heart of IBM SAN Volume Controller, IBM Storwize family, IBM FlashSystem® V9000, and VersaStack. It enables these systems to deliver better data value, security, and simplicity through industry-leading virtualization. This virtualization transforms existing and new storage and streamlines deployment for a simpler, more responsive, scalable, and cost-efficient IT infrastructure.

IBM Spectrum Virtualize systems provide storage management from entry and midrange up to enterprise disk systems, and enable hosts to attach through SAN, FCoE, or iSCSI to Ethernet networks. IBM Spectrum Virtualize is easy to use, which enables staff to start working with it rapidly.

IBM Spectrum Virtualize uses virtualization, thin provisioning, and compression technologies to improve storage utilization and meet changing needs quickly and easily. In this way, IBM Spectrum Virtualize products are the ideal complement to server virtualization strategies.

Key Capabilities

IBM Spectrum Virtualize software capabilities are offered across various platforms, including SAN Volume Controller, Storwize V7000, Storwize V5000, and FlashSystem V9000. The following IBM Spectrum Virtualize products are designed to deliver the benefits of storage virtualization and advanced storage capabilities in environments from large enterprises to small businesses and midmarket companies:

- ▶ IBM Real-time Compression™ for inline, real-time compression
- ▶ Stretched Cluster and IBM HyperSwap® for a high-availability solution
- ▶ IBM Easy Tier® for automatic and dynamic data tiering
- ▶ Distributed RAID for better availability and faster rebuild times
- ▶ Encryption for internal and external virtualized capacities
- ▶ FlashCopy snapshots
- ▶ Remote data replication

Benefits

The sophisticated virtualization, management, and functions of IBM Spectrum Virtualize provide the following storage benefits:

- ▶ Improves storage utilization up to 2x
- ▶ Supports up to 5x as much data in the same physical space
- ▶ Simplifies management of heterogeneous storage systems
- ▶ Enables rapid deployment of new storage technologies for greater ROI
- ▶ Improves application availability with virtually zero storage-related outages

The SAN Volume Controller combines software and hardware into a comprehensive, modular appliance that uses symmetric virtualization.

Symmetric virtualization is achieved by creating a pool of managed disks from the attached storage systems. Those storage systems are then mapped to a set of volumes for use by the attached host systems. System administrators can view and access a common pool of storage on the SAN. This function helps administrators to use storage resources more efficiently and provides a common base for advanced functions.

The IBM Spectrum Virtualize functions are shown in Figure 3-17.

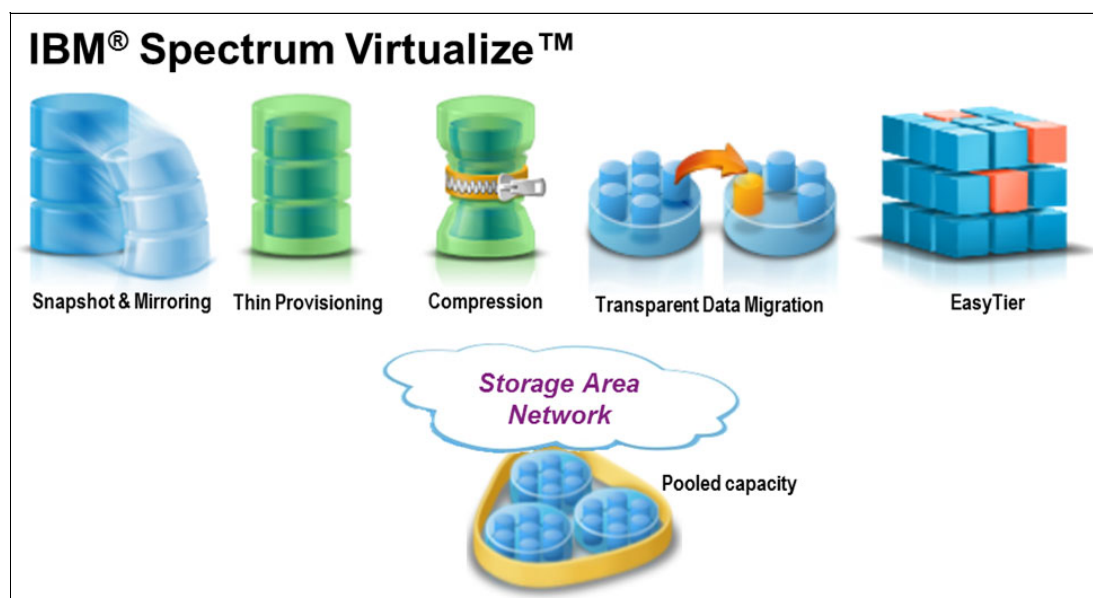


Figure 3-17 IBM Spectrum Virtualize functions

IBM Spectrum Virtualize features and benefits are listed in Table 3-3.

Table 3-3 IBM Spectrum Virtualize features and benefits

Feature	Benefits
Single point of control for storage resources	<ul style="list-style-type: none"> ▶ Designed to increase management efficiency ▶ Designed to help support business application availability
Pools the storage capacity of multiple storage systems on a SAN	<ul style="list-style-type: none"> ▶ Helps you manage storage as a resource to meet business requirements and not just as a set of boxes ▶ Helps administrators better deploy storage as required beyond traditional “SAN islands” ▶ Can help increase use of storage assets ▶ Insulates applications from physical changes to the storage infrastructure
Clustered pairs of IBM SAN Volume Controller data engines	<ul style="list-style-type: none"> ▶ Highly reliable hardware foundation ▶ Designed to avoid single points of hardware failure
IBM Real-time Compression	<ul style="list-style-type: none"> ▶ Increases effective capacity of storage systems up to five times, helping to lower costs, floor-space requirements, and power and cooling needs ▶ Can be used with a wide range of data, including active primary data, for dramatic savings ▶ Hardware compression acceleration helps transform the economics of data storage

Feature	Benefits
Innovative and tightly integrated support for flash storage	<ul style="list-style-type: none"> ▶ Designed to deliver ultra-high performance capability for critical application data ▶ Move data to and from flash storage without disruption; make copies of data onto hard disk drive (HDD)
Support for IBM FlashSystem	Enables high performance for critical applications with IBM MicroLatency®, coupled with sophisticated functions
Easy-to-use IBM Storwize family management interface	<ul style="list-style-type: none"> ▶ Single interface for storage configuration, management, and service tasks regardless of storage vendor ▶ Helps administrators use their existing storage assets more efficiently
IBM Storage Mobile Dashboard	Provides basic monitoring capabilities to securely check system health and performance
Dynamic data migration	<ul style="list-style-type: none"> ▶ Migrate data among devices without taking applications that are using that data offline ▶ Manage and scale storage capacity without disrupting applications
Manage tiered storage	Helps balance performance needs against infrastructure costs in a tiered storage environment
Advanced network-based copy services	<ul style="list-style-type: none"> ▶ Copy data across multiple storage systems with IBM FlashCopy ▶ Copy data across metropolitan and global distances as needed to create high-availability storage solutions
Integrated Bridgeworks SANrockIT technology for IP replication	<ul style="list-style-type: none"> ▶ Optimize use of network bandwidth ▶ Reduce network costs or speed replication cycles, improving the accuracy of remote data
Enhanced stretch cluster configurations	<ul style="list-style-type: none"> ▶ Provide highly available, concurrent access to a single copy of data from data centers up to 300 km apart ▶ Enable nondisruptive storage and virtual machine mobility between data centers
Thin provisioning and “snapshot” replication	<ul style="list-style-type: none"> ▶ Dramatically reduce physical storage requirements by using physical storage only when data changes ▶ Improve storage administrator productivity through automated on-demand storage provisioning
Hardware snapshots integrated with IBM Spectrum Protect Snapshot Manager	<ul style="list-style-type: none"> ▶ Performs near-instant application-aware snapshot backups, with minimal performance impact for IBM DB2®, Oracle, SAP, VMware, Microsoft SQL Server, and Microsoft Exchange ▶ Provides advanced, granular restoration of Microsoft Exchange data

Virtualizing storage with SAN Volume Controller helps make new and existing heterogeneous storage arrays more effective by including many functions that are traditionally deployed within disk array systems. By including these functions in a virtualization system, SAN Volume Controller standardizes functions across virtualized storage for greater flexibility and potentially lower costs.

How SAN Volume Controller stretches virtual volume with heterogeneous storage across data centers is shown in Figure 3-18.

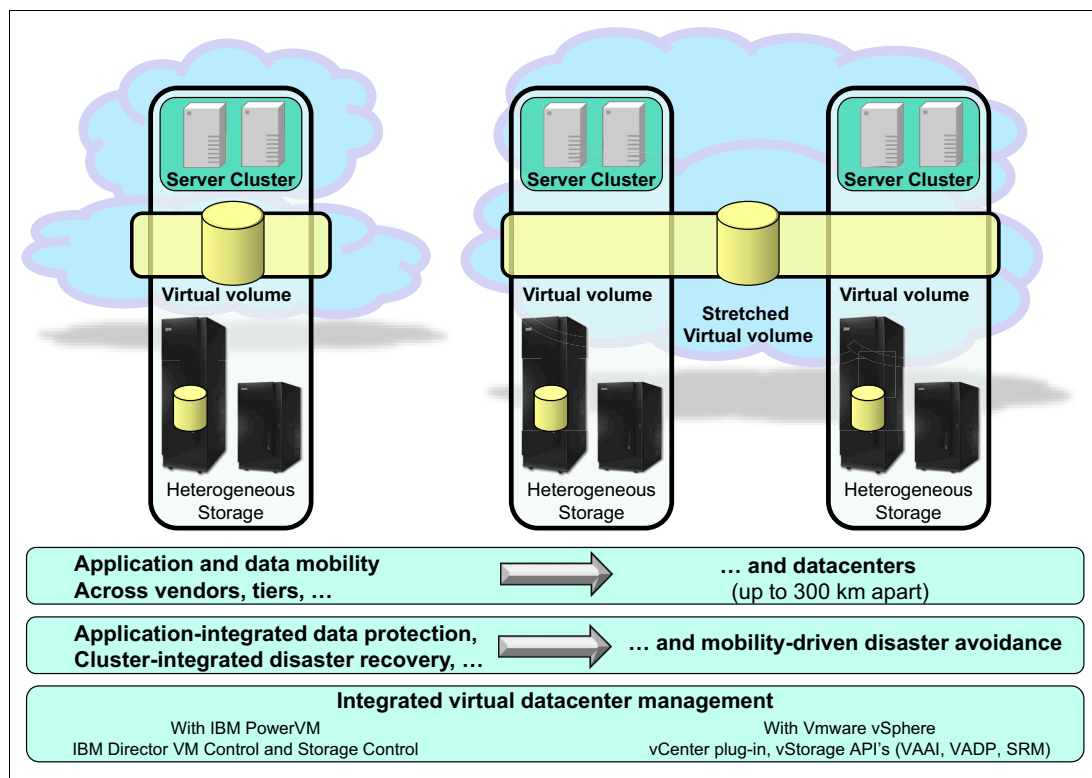


Figure 3-18 Stretching virtual volume across data centers with heterogeneous storage

SAN Volume Controller functions benefit all virtualized storage. For example, IBM Easy Tier optimizes use of flash memory, and Real-time Compression enhances efficiency even further by enabling the storage of up to five times as much active primary data in the same physical disk space¹. Finally, high-performance thin provisioning helps automate provisioning. These benefits can help extend the useful life of existing storage assets, reducing costs.

Integrating these functions into SAN Volume Controller also means that they are designed to operate smoothly together, reducing management effort:

- **Storage virtualization:** Virtualization is a foundational technology for software-defined infrastructures that enables software configuration of the storage infrastructure. Without virtualization, networked storage capacity utilization averages about 50 percent, depending on the operating platform. Virtualized storage enables up to 90 percent utilization by enabling pooling across storage networks with online data migration for capacity load balancing.

Virtual Storage Center supports a virtualization of storage resources from multiple storage systems and vendors (that is, heterogeneous storage). Pooling storage devices enables access to capacity from any networked storage system, which is a significant advantage over the limitations inherent in traditional storage arrays.

- **IBM Easy Tier:** Virtual Storage Center helps optimize flash memory with automated tiering for critical workloads. Easy Tier helps make the best use of available storage resources by automatically moving the most active data to the fastest storage tier, which helps applications and virtual desktop environments run up to three times faster.

¹ Compression data based on IBM measurements. Compression rates vary by data type and content.

- ▶ Thin provisioning: Thin provisioning helps automate provisioning and improve productivity by enabling administrators to focus on overall storage deployment and utilization, and on longer-term strategic requirements, without being distracted by routine storage-provisioning requests.
- ▶ Remote mirroring: IBM Metro Mirror and Global Mirror functions automatically copy data to remote sites as it changes, enabling fast failover and recovery. These capabilities are integrated into the advanced GUI, making them easy to deploy.
- ▶ IBM Real-time Compression: Real-time Compression is patented technology that is designed to reduce space requirements for active primary data. It enables users to store up to five times as much data in the same physical disk space, and can do so without affecting performance.

IBM Spectrum Virtualize for Public Cloud

A flexible declination of the product delivers a powerful solution for the deployment of IBM Spectrum Virtualize software in public clouds, starting with IBM Cloud. This new capability provides a monthly license to deploy and use IBM Spectrum Virtualize in IBM Cloud to enable Hybrid Cloud solutions. It also offers the ability to transfer data between on-premise data centers by using any IBM Spectrum Virtualize based appliances (including SAN Volume Controller, Storwize family, V9000, VersaStack with Storwize family, or SAN Volume Controller appliance), or IBM Spectrum Virtualize Software Only and the IBM Cloud.

Through IP-based replication with Global or Metro Mirror, users can now create secondary copies of their own premises data in the public cloud for Disaster Recovery, workload redistribution, or migration of data from on premises data centers to public clouds.

IBM Spectrum Virtualize Hybrid Cloud opportunities are shown in Figure 3-19.

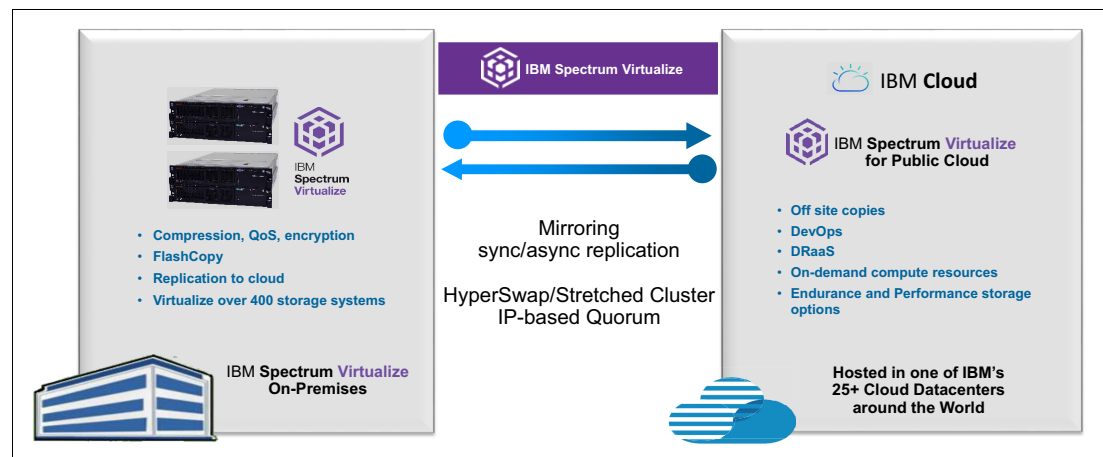


Figure 3-19 IBM Spectrum Virtualize Hybrid Cloud opportunities

3.11.2 IBM Spectrum Access Blueprint



IBM Spectrum Access Blueprint provides enterprise data management and protection for hybrid and multicloud environments with operational control and efficiency. Deployable with VersaStack solutions from IBM and Cisco, IBM Spectrum Access provides a blueprint to deliver the economics and simplicity of the public cloud with the accessibility, virtualization, security and performance of an on-premises implementation.

Ideally, a cloud platform and cloud technologies are essential to implement microservices. Microservices are based on the ability to spin up business services in distinct containers and intelligently route from one container or endpoint to another. This means that having a cloud-service fabric is important to stand up microservices instances and have them connect and reliably offer the quality of service expected.

One of the challenges with microservices compared with classic service-oriented architecture (SOA) lies in the complexity of orchestrating identity and access management in harmony with the dynamic nature of microservices components. With new services being spun up dynamically, the host names for the instances of each service become dynamic as well.

This end-to-end private cloud solution of IBM Spectrum Access Blueprint with VersaStack converged infrastructure and IBM Cloud Private technologies delivers the essential private cloud service fabric for building and managing on-premises, containerized applications and guarantees to provide seamless access to persistent storage for stateful services, such as database applications. This solution also comes with services for data, messaging, Java, blockchain, DevOps, analytics, and many others.

IBM Spectrum Access Blueprint includes the following key features:

- ▶ Enterprises can build new applications delivering infrastructure services easily and efficiently
- ▶ Simple scalability
- ▶ Optimizes cloud deployments
- ▶ Quickly deploy storage classes to comply with business SLAs
- ▶ Provision capacity directly with containerized applications

IBM Spectrum Access currently incorporates the following features:

- ▶ VersaStack to create a compute and storage platform that consists of IBM Spectrum Virtualize and IBM Storwize systems that are paired with the Cisco UCS compute platform
- ▶ IBM Spectrum Connect to provide the ability to connect stateful storage to containerize workloads
- ▶ IBM Cloud Private to provide a management and orchestration platform to deploy Containers

For more information, see this website:

<https://www.ibm.com/us-en/marketplace/ibm-spectrum-access>

IBM Cloud Private 2.1.0

IBM Cloud Private is an application platform for developing and managing on-premises, containerized applications. It is an integrated environment for managing containers that includes the container orchestrator Kubernetes, a private image repository, a management console, and monitoring frameworks.

IBM Cloud Private makes it easy to stand up an elastic runtime that is based on Kubernetes to address each of the following workloads:

- ▶ Deliver packaged workloads for traditional middleware. IBM Cloud Private initially supports IBM Db2, IBM MQ, Redis, and several other open source packages.
- ▶ Support API connectivity from 12-factor apps (software-as-a-service) to manage API endpoints within an on-premises datacenter.
- ▶ Support development and delivery of modern 12-factor apps with Microservice Builder.

Along with great capabilities to run enterprise workloads, IBM Cloud Private is delivering enhanced support to run processor-intensive capabilities, such as machine learning or data analytics quickly by taking advantage of Graphics Processing Unit (GPU) clusters (see Figure 3-20).

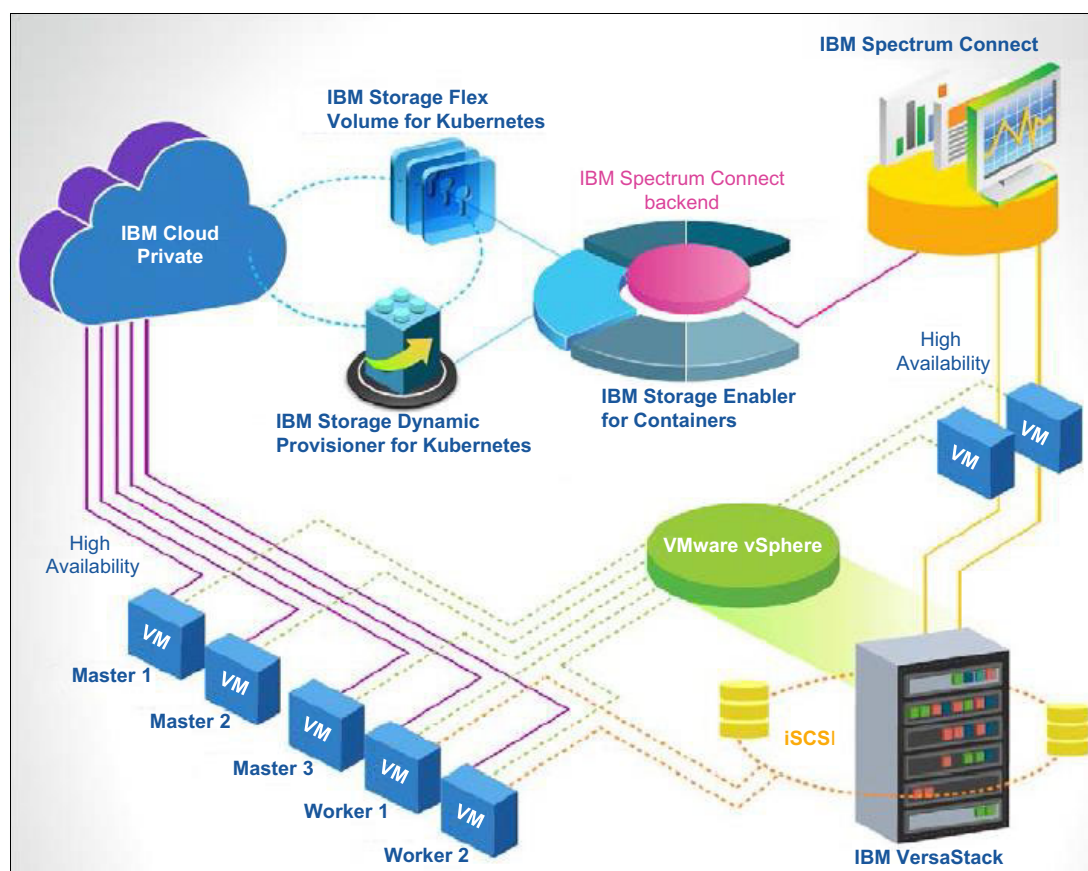


Figure 3-20 End-to-end private cloud solution architecture that uses IBM Spectrum Access

The IBM Spectrum Access blueprint offers a pretested and validated solution to help pool your compute, network, and storage resources for cloud application deployment. It delivers a simplified, standardized, and trusted approach for the deployment, use, and management of your shared infrastructure and cloud environment. It helps you monitor and manage applications while providing resource consumption reports. IBM Spectrum Access is ideal for IBM Cloud Private deployment because it has the essential private cloud service fabric for building and managing on-premises, containerized applications with persistent storage.

3.11.3 IBM Spectrum Accelerate



IBM Spectrum Accelerate is a highly flexible, software-defined storage solution that enables rapid deployment of block data storage services for new and traditional workloads on and off premises. It is a key member of the IBM Spectrum Storage portfolio.

IBM Spectrum Accelerate allows you to run hotspot-free, grid-scale software that runs on the XIV Storage System Gen3 enterprise storage platform in your data center infrastructure or in a cloud provider, such as IBM Cloud. It offers proven grid-scale technology, mature features, and ease of use. It is deployed on over 100,000 servers worldwide.

IBM Spectrum Accelerate delivers predictable, consistent storage performance, management scaling to more than 68 petabytes usable, and a rich feature set that includes remote mirroring and granular multi-tenancy. It deploys on premises on x86 commodity servers and on the optimized XIV Storage System, and off-premises as a public cloud service on IBM Cloud.

You can manage all your IBM Spectrum Accelerate instances, wherever they are deployed, in a single, intuitive interface. Hardware-independent, transferable licensing offers superb operational flexibility and cost benefits.

IBM Spectrum Accelerate delivers a single management experience across software-defined storage infrastructure by using IBM's HyperScale Manager, which can manage IBM Spectrum Accelerate instances, IBM XIV, and the IBM A9000 all flash solution. This combination helps cut costs through reduced administration effort and training, reduces procurement costs, standardizes data center storage hardware operations and services, and provides licensing flexibility that enables cost-efficient cloud building.

How straightforward scaling is by building a storage grid with IBM Spectrum Accelerate is shown in Figure 3-21.

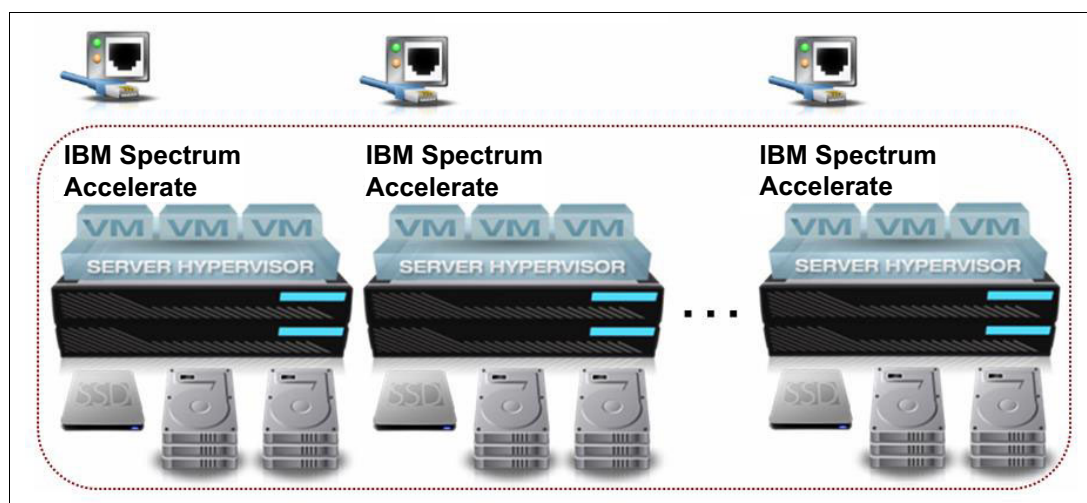


Figure 3-21 IBM Spectrum Accelerate iSCSI storage grid

Key capabilities

IBM Spectrum Accelerate gives organizations these following capabilities:

- ▶ Enterprise cloud storage in minutes, by using commodity hardware
- ▶ Hotspot-free performance and QoS without any manual or background tuning needed
- ▶ Advanced remote replication, role-based security, and multi-tenancy
- ▶ Deploy on-premises or on the cloud (also as a service on IBM Cloud)
- ▶ Hyper-scale management of dozens of petabytes
- ▶ Best in class VMware and OpenStack integration
- ▶ Run IBM Spectrum Accelerate and other application virtual machines on the same server

IBM Spectrum Accelerate runs as a virtual machine on vSphere ESXi hypervisor, which enables you to build a server-based SAN from commodity hardware that includes x86 servers, Ethernet switches, solid-state drives (SSDs), and direct-attached, high-density disks. IBM Spectrum Accelerate essentially acts as an operating system for your self-built SAN storage, grouping virtual nodes and spreading the data across the entire grid.

IBM Spectrum Accelerate release 11.5.3 manages up to 15 nodes in a grid. It provides a single point of management of up to 144 grids connected through Hyper-Scale Manager, up to 2,160 nodes.

IBM Spectrum Accelerate allows you to deploy storage services flexibly across different delivery models, including customer-choice hardware, data center infrastructure, and IBM storage systems.

IBM Spectrum Accelerate includes the following benefits:

- ▶ Cost reduction by delivering hotspot-free storage to different deployment models on- and off-premises, enabling organizations to pay less overall for the same capacity by optimizing utilization, acquiring less hardware, and minimize administrative overhead
- ▶ Increased operational agility through easy cloud building, faster provisioning, small capacity increments, and flexible, transferable licensing
- ▶ Rapid response through enterprise-class storage availability, data protection, and security for the needs of new and traditional workload in the data center and other sites, while flexibly balancing capital and operational expenses
- ▶ Scaleout across 144 virtual systems and seamless management across IBM Spectrum Accelerate instances on- and off-premises and the XIV storage system

The IBM Spectrum Accelerate features with their associated benefits are listed in Table 3-4.

Table 3-4 IBM Spectrum Accelerate features and benefits

Feature	Benefit
Performance	<ul style="list-style-type: none"> ▶ Ensures even data distribution through massive parallelism and automatic load balancing including upon capacity add ▶ Distributed cache
Reliability and Availability	<ul style="list-style-type: none"> ▶ Grid redundancy maintains two copies of each 1-MB data partition with each copy being on a different VM, proactive diagnostics, fast and automatic rebuilds, event externalization ▶ Advanced monitoring; network monitoring; disk performance tracking/reporting; data center monitoring; shared monitoring for some components; data and graphical reports on I/O, usage, and trends ▶ Self-healing, which minimizes the rebuild process by rebuilding only actual data ▶ Automated load balancing across components; minimized risk of disk failure due to rapid return to redundancy
Management	Intuitive GUI: Scales to up to 144 virtual arrays and up to more than 45 PB with IBM Hyper-Scale Manager; extensive CLI; RESTful API; mobile app support with push notifications; multi-tenancy with quality of service by tenant, pool, or host
Cloud automation and Self-service	OpenStack; VMware vRealize Orchestrator through IBM Spectrum Control Base
Snapshot management	Space efficient snapshots: Writable, snapshot of snapshot, restore from snapshot, snapshots for consistency groups, mirroring
Thin provisioning; space reclamation	Thin provisioning per pool, thick-to-thin migration; VMware, Microsoft, Symantec space reclamation support

Feature	Benefit
Mirroring	Synchronous/asynchronous; volumes and consistency groups, recovery point objective (RPO) of seconds; online/offline initialization; failover/failback; mirroring across platforms including with XIV Storage System
Security	Role-based access management, multi-tenancy, iSCSI Challenge Handshake Authentication Protocol (CHAP) and auditing; integrates with Lightweight Directory Access Protocol (LDAP) and Microsoft Active Directory servers

OpenStack device support for IBM XIV

IBM built and contributed the OpenStack Cinder block storage driver for XIV to the OpenStack community. This driver allows IBM Spectrum Accelerate to be the first enterprise class storage system to have OpenStack software support. This feature allows ease of use and fast time to implementation characteristics to be magnified by being able to be automatically managed and provisioned within the OpenStack environment.

The IBM Storage Driver for OpenStack Cinder component added support starting with the Folsom release as shown in Figure 3-22, and then expanded the support for the Grizzly and Havana releases. The driver enables OpenStack clouds to be able to directly access and use IBM Spectrum Accelerate Storage System Gen3.

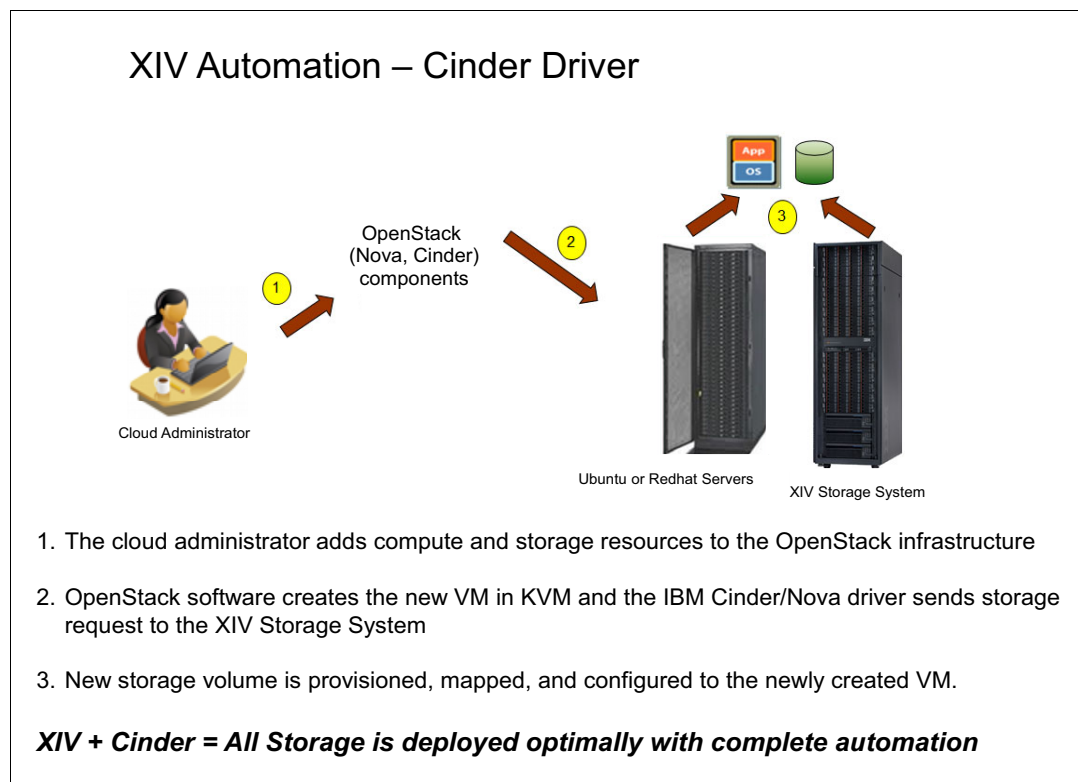


Figure 3-22 OpenStack Cinder support for XIV

Hyperconverged Flexible deployment

IBM Spectrum Accelerate provides customers the capability to create Hyperconverged solutions that run Compute and the Storage services on the same physical x86 servers wherever they are deployed.

By using the VMware ESX hypervisor, extra resources that are not used by the IBM Spectrum Accelerate instances (such as memory and processors) can be provisioned to more guest workloads.

The IBM Spectrum Accelerate instance can be managed and administered from its native GUI or through the IBM HyperScale Manager option. This capability includes tasks, such as creating pools, replication structures, hardware component replacement, and firmware updates (see Figure 3-23).

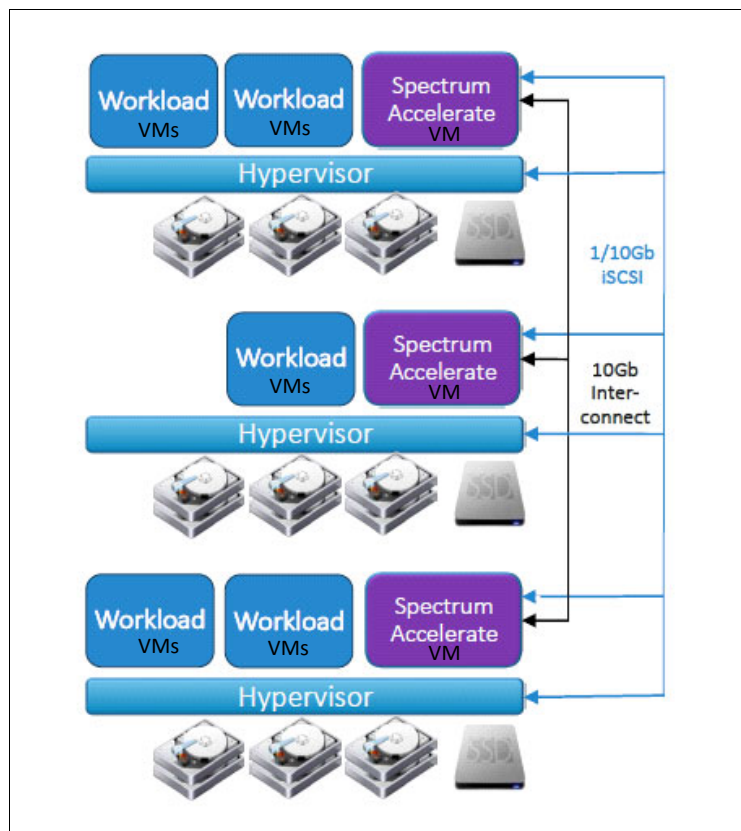


Figure 3-23 IBM Spectrum Accelerate in a Hyperconverged infrastructure

VMware Orchestration

By using IBM Spectrum Connect (see Figure 3-3 on page 25), orchestration of the compute layer, provisioning from predefined IBM Spectrum Accelerate pools, and replication can all be done through the VMware integration points. Therefore, the solution can be controlled through the vRealize suite and can use vCenter and VMware SRM-based APIs.

IBM Spectrum Accelerate as a pre-configured Hyperconverged Solution

Customers looking to benefit from IBM Spectrum Accelerate's Hyperconverged capabilities and its ability to work as a storage system that can replicate to the IBM XIV, but want the ability to deploy it as a pre-integrated solution can order it from Supermicro. IBM and Supermicro have jointly designed an appliance deliverable that combines IBM Spectrum Accelerate software with Supermicro hardware.

This product is delivered as a pre-configured, preinstalled, and pre-tested solution that is ready to be integrated into customer networks. The three basic building blocks are *Small*, *Medium*, and *Large*, and those building blocks can be customized for based on customer-specific requirements.

For more information, see this website:

<https://www.supermicro.com/solutions/spectrum-accelerate.cfm>

3.11.4 IBM XIV Storage System Gen3

IBM Spectrum Accelerate is the common software-defined layer that is inside the IBM XIV Storage System Gen3.

Note: For more information about IBM Spectrum Accelerate, see the following IBM publications:

- ▶ *IBM Spectrum Accelerate Deployment, Usage, and Maintenance*, SG24-8267
- ▶ *Deploying IBM Spectrum Accelerate on Cloud*, REDP-5261
- ▶ *IBM Spectrum Accelerate Reference Architecture*, REDP-5260

3.11.5 IBM FlashSystem A9000 and A9000R

IBM Spectrum Accelerate is the common software-defined layer across the IBM FlashSystem A9000 and A9000R all-flash arrays.

Note: For more information about IBM FlashSystem A9000 and A9000R, see the following IBM publications:

- ▶ *IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation*, SG24-8345
- ▶ *IBM FlashSystem A9000 Product Guide*, REDP-5325

3.12 IBM File Storage solutions

This section describes IBM file storage solutions.

3.12.1 IBM Spectrum Scale



IBM Spectrum Scale is a proven, scalable, high-performance file management solution that is based on IBM's General Parallel File System (GPFS™). IBM Spectrum Scale provides world-class storage management with extreme scalability, flash accelerated performance, and automatic policy-based storage tiering from flash to disk, then to tape. IBM Spectrum Scale reduces storage costs up to 90% while improving security and management efficiency in cloud, big data, and analytics environments.

First introduced in 1998, this mature technology enables a maximum volume size of 8 YB, a maximum file size of 8 EB, and up to 18.4 quintillion (two to the 64th power) files per file system. IBM Spectrum Scale provides simplified data management and integrated information lifecycle tools such as software-defined storage for cloud, big data, and analytics. It introduces enhanced security, flash accelerated performance, and improved usability. It also provides capacity quotas, access control lists (ACLs), and a powerful snapshot function.

Key capabilities

IBM Spectrum Scale adds elasticity with the following capabilities:

- ▶ Global namespace with high-performance access scales from departmental to global
- ▶ Automated tiering, data lifecycle management from flash (6x acceleration) to tape (10x savings)
- ▶ Enterprise ready with data security (encryption), availability, reliability, large scale
- ▶ POSIX compliant
- ▶ Integrated with OpenStack components and Hadoop

Benefits

IBM Spectrum Scale provides the following benefits:

- ▶ Improves performance by removing data-related bottlenecks
- ▶ Automated tiering, data lifecycle management from flash (acceleration) to tape (savings)
- ▶ Enables sharing of data across multiple applications
- ▶ Reduces cost per performance by placing data on most applicable storage (flash to tape or cloud)

IBM Spectrum Scale is part of the IBM market-leading software-defined storage family.

Consider the following points:

- ▶ As a Software-only solution: Runs on virtually any hardware platform and supports almost any block storage device. IBM Spectrum Scale runs on Linux (including Linux on IBM Z® Systems), IBM AIX®, and Windows systems.
- ▶ As an integrated IBM Elastic Storage™ Server solution: A bundled hardware, software, and services offering that includes installation and ease of management with a graphical user interface. Elastic Storage Server provides unsurpassed end-to-end data availability, reliability, and integrity with unique technologies that include IBM Spectrum Scale RAID.
- ▶ As a cloud service: IBM Spectrum Scale delivered as a service provides high performance, scalable storage, and integrated data governance for managing large amounts of data and files in the IBM Cloud.

IBM Spectrum Scale features enhanced security with native encryption and secure erase. It can increase performance by using server-side flash cache to increase I/O performance up to six times. IBM Spectrum Scale provides improved usability through data replication capabilities, data migration capabilities, Active File Management (AFM), transparent cloud tiering (TCT), File Placement Optimizer (FPO), and IBM Spectrum Scale Native RAID.

An example of the IBM Spectrum Scale architecture is shown in Figure 3-24.

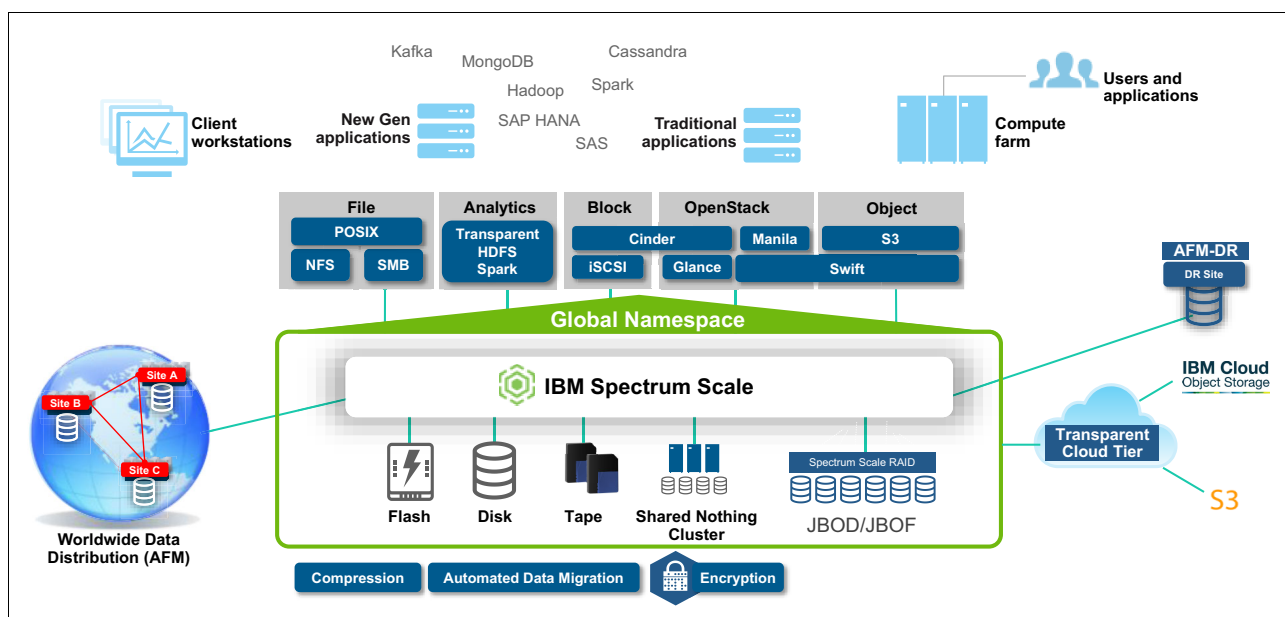


Figure 3-24 IBM Spectrum Scale architecture

IBM Spectrum Scale is based around the following concepts:

- ▶ Storage pools
- ▶ File sets
- ▶ Policy engine
- ▶ Mirroring, replication, and migration capabilities
- ▶ Active File Management
- ▶ File Placement Optimizer
- ▶ Licensing

Storage pools

A storage pool is a collection of disks or arrays with similar attributes. It is an organizational structure that allows the combination of multiple storage locations with identical characteristics. The following types of storage pools are available:

- ▶ System Pool

One system pool is needed per file system. The system pool includes file system metadata and can be used to store data.
- ▶ Data Pool

A data pool is used to store file data. A data pool is optional.
- ▶ External Pool

An external pool is used to attach auxiliary storage, such as tape-to-IBM Spectrum Scale. An external pool is optional.

File sets

IBM Spectrum Scale creates a single name space; therefore, tools are available that provide a fine grained management of the directory structure. A file set acts as a partition of a file system, a subdirectory tree.

File sets can be used for operations, such as quotas or used in management policies. It is a directory tree that behaves as a “file system” within a file system. Consider the following points:

- ▶ It is part of the global namespace.
- ▶ It can be linked and unlinked (such as mount and unmount).
- ▶ Policy scan can be restricted to only scan file sets. This setting can be helpful when the file system has billions of files.
- ▶ A file set can be assigned to a storage pool.

The following type of file sets are available:

- ▶ **Dependent**

A dependent file set allows for a finer granularity of administration. It shares the inode space with another file set.

- ▶ **Independent**

An independent file set features a distinct inode space. An independent file set allows file set level snapshots, independent file scans, and enabled advanced features, such as AFM.

Policy engine

The policy engine uses an SQL style syntax to query or operate on files based on file attributes. Policies can be used to migrate all data that has not been accessed in 6 months (for example) to less expensive storage or used to query the contents of a file system. Management policies support advanced query capabilities, though what makes the policy engine most useful is the performance. The policy engine is capable of scanning billions of objects as shown in Table 3-5.

Table 3-5 Speed comparison for GPFS policy engine

Search through 1,000,000,000 (1 billion) files	
find	~ 47 hours
GPFS policy engine	~ 5 hours

Table 3-5 shows the power of the GPFS policy engine. Although an average **find** across 1 billion files took ~47 hours, the GPFS policy engine can satisfy the request within five hours. The GPFS policy engine can also create a candidate list for backup applications to use to achieve a massive reduction in candidate identification time.

IBM Spectrum Scale has next generation availability with features that include rolling software and hardware upgrades. You can add and remove servers to adapt the performance and capacity of the system to changing needs. Storage can be added or replaced online, and you can control how data is balanced after storage is assessed.

Mirroring, replication, and migration capabilities

In IBM Spectrum Scale, you can replicate a single file, a set of files, or the entire file system. You can also change the replication status of a file at any time by using a policy or command. Using these capabilities, you can achieve a replication factor of two, which equals mirroring, or a replication factor of three.

A replication factor of two in IBM Spectrum Scale means that each block of a replicated file is in at least two failure groups. A failure group is defined by the administrator and contains one or more disks. Each storage pool in a file system contains one or more failure groups. Failure groups are defined by the administrator and can be changed at any time. So when a file system is fully replicated, any single failure group can fail and the data remains online.

For migration, IBM Spectrum Scale provides the capability to add storage to the file system, migrate the existing data to the new storage, and remove the old storage from the file system. All of this can be done online without disruption to your business.

Active File Management

AFM enables the sharing of data across unreliable or high latency networks. With AFM, you can create associations between IBM Spectrum Scale clusters and define the location and flow of file data. AFM allows you to implement a single name space view across clusters, between buildings, and around the world.

AFM operates at the file set level. This configuration means that you can create hundreds of AFM relationships in each file system. AFM is a caching technology though inode. File data in a cache file set is the same as an inode and file data in any IBM Spectrum Scale file system. It is a “real” file that is stored on disk. The job of the cache is to keep the data in the file consistent with the data on the other side of the relationship.

AFM can be implemented in five different modes:

- ▶ Read-Only (ro)
- ▶ Local-Update (lu)
- ▶ Single-Writer (sw)
- ▶ Independent Writer (iw)
- ▶ Asynchronous DR

These modes can be used to collect data at a remote location (single-writer), create a flash cache for heavily read data (read-only), provide a development copy of data (local-update), create a global interactive name space (independent-writer), and create asynchronous copies of file data (asynchronous DR).

Transparent Cloud Tiering

Data in the enterprise is growing at an alarming rate led by growth in unstructured data, leading to a capacity crisis. Cooler and cold data constitutes a large proportion of data in the enterprise. Migrating this data to lower-cost cloud object storage provides cost savings.

Transparent cloud tiering is a new feature of IBM Spectrum Scale 4.2.1 that provides hybrid cloud storage capability. This software-defined capability enables usage of public, private, and on-premises cloud object storage as a secure, reliable, transparent storage tier that is natively integrated with IBM Spectrum Scale without introducing more hardware appliances or new management touch points. It uses the ILM policy language semantics that are available in IBM Spectrum Scale. The semantics allow administrators to define the following policies for tiering cooler and cold data to a cloud object storage:

- ▶ IBM Cloud Object Storage (Cleversafe)
- ▶ Amazon Web Services S3
- ▶ OpenStack Swift

This configuration frees up storage capacity in higher-cost storage tiers that can be used for more active data.

The IBM Spectrum Scale transparent cloud tiering feature is shown in Figure 3-25.

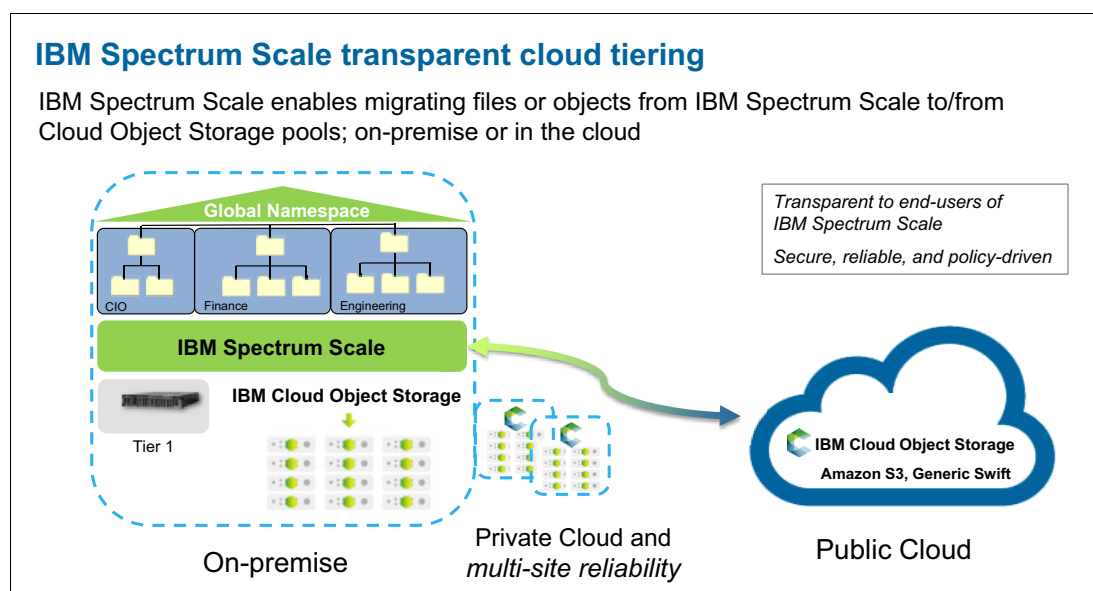


Figure 3-25 IBM Spectrum Scale transparent cloud tiering feature highlights

For more information, see *Enabling Hybrid Cloud Storage for IBM Spectrum Scale Using Transparent Cloud Tiering*, REDP-5411.

IBM Spectrum Scale Management GUI

The IBM Spectrum Scale Management GUI (Graphical User Interface) can be used in conjunction with the existing command line interface. The GUI is meant to support common administrator tasks, such as provisioning more capacity, which can be accomplished faster and without knowledge of the command-line interface.

System health, capacity, and performance displays can be used to identify trends and respond quickly to any issues that arise. The GUI is available to IBM Spectrum Scale Clusters running at or above the 4.2 release for the Standard Edition and Advanced Edition.

For more information, see Figure 3-26.



Figure 3-26 IBM Spectrum Scale management GUI dashboard

The IBM Spectrum Scale management GUI provides an easy way to configure and manage various features that are available with the IBM Spectrum Scale system. You can perform the following important tasks through the IBM Spectrum Scale management GUI:

- ▶ Monitoring the performance of the system based on various aspects
- ▶ Monitoring system health
- ▶ Managing file systems
- ▶ Creating file sets and snapshots
- ▶ Managing Objects, NFS, and SMB data exports
- ▶ Creating administrative users and defining roles for the users
- ▶ Creating object users and defining roles for them
- ▶ Defining default, user, group, and file set quotas
- ▶ Monitoring the capacity details at various levels such as file system, pools, file sets, users, and user groups

File Placement Optimizer

FPO allows IBM Spectrum Scale to use locally attached disks on a cluster of servers that communicate by using the network, rather than the regular case of the use of dedicated servers for shared disk access (such as the use of SAN). IBM Spectrum Scale FPO is suitable for workloads, such as SAP HANA, and IBM Db2 with Database Partitioning Feature. It can be used as an alternative to Hadoop Distributed File System (HDFS) in big data environments.

The use of FPO extends the core IBM Spectrum Scale architecture, which provides greater control and flexibility to use data location, reduces hardware costs, and improves I/O performance. The following benefits are realized when FPO is used:

- ▶ Allows your jobs to be scheduled where the data is located (locality awareness)
- ▶ Metablocks that allow large and small block sizes to coexist in the same file system
- ▶ Write affinity that allows applications to dictate the layout of files on different nodes, maximizing write and read bandwidth
- ▶ Pipelined replication to maximize use of network bandwidth for data replication
- ▶ Distributed recovery to minimize the effect of failures on ongoing computation

For more information about IBM Spectrum Scale FPO, see *GPFS V4.1: Advanced Administration Guide*, SC23-7032.

IBM Spectrum Scale Native RAID

IBM Spectrum Scale Native RAID provides next generation performance and data security. Using IBM Spectrum Scale native RAID, just a bunch of disks (JBOD) are directly attached to the systems running IBM Spectrum Scale software. This technology uses declustered RAID to minimize performance degradation during RAID rebuilds and provides extreme data integrity by using end-to-end checksums and version numbers to detect, locate, and correct silent disk corruption. An advanced disk hospital function automatically addresses storage errors and slow performing drives so that your workload is not affected.

IBM Spectrum Scale native RAID is available with the IBM Power8 architecture in the IBM Elastic Storage Server (ESS) offering.

Licensing

IBM Spectrum Scale V5 offers the following editions so you only pay for the functions that you need:

- ▶ Standard Edition includes the base function plus ILM, AFM, and integrated multiprotocol support, which includes NFS, SMB, and Object and is measured by the number of servers and clients attached to the cluster.
- ▶ Data Management Edition includes encryption of data at rest, secure erase, asynchronous multisite disaster recovery, and all the features of Standard Edition. It is measured by the amount of storage capacity that is supported in the cluster and includes all connected servers and clients. Data Management also includes tiering-to-Object storage (on-prem or cloud), and file audit logging.

For more information, see the following resources:

- ▶ IBM Spectrum Scale:
<http://www.ibm.com/systems/storage/spectrum/scale/index.html>
- ▶ IBM Spectrum Scale (IBM Knowledge Center):
http://www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html
- ▶ IBM Spectrum Scale Wiki:
<https://ibm.biz/BdFPR2>
- ▶ IBM Elastic Storage Server
<http://www.ibm.com/systems/storage/spectrum/ess>

IBM Spectrum Scale for Linux on IBM Z

The IBM Spectrum Scale for Linux on IBM Z implements the IBM Spectrum Scale Software-based delivery model in the Linux on IBM Z environment. The highlights of IBM Spectrum Scale for Linux on IBM Z include the following features:

- Supports extended count key data (IBM ECKD™) DASD disks and Fibre Channel Protocol attached SCSI disks
- Supports IBM HiperSockets™ for communication within one IBM Z System

For more information, see *Getting started with IBM Spectrum Scale for Linux on IBM Z*, which is available at this website:

<http://www.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ZSW03272USEN>

Using IBM Spectrum Scale in an OpenStack cloud deployment

Deploying OpenStack over IBM Spectrum Scale offers benefits that are provided by the many enterprise features in IBM Spectrum Scale. It also provides the ability to consolidate storage for various OpenStack components and applications that are running on top of the OpenStack infrastructure under a single storage management plan (see Figure 3-27).

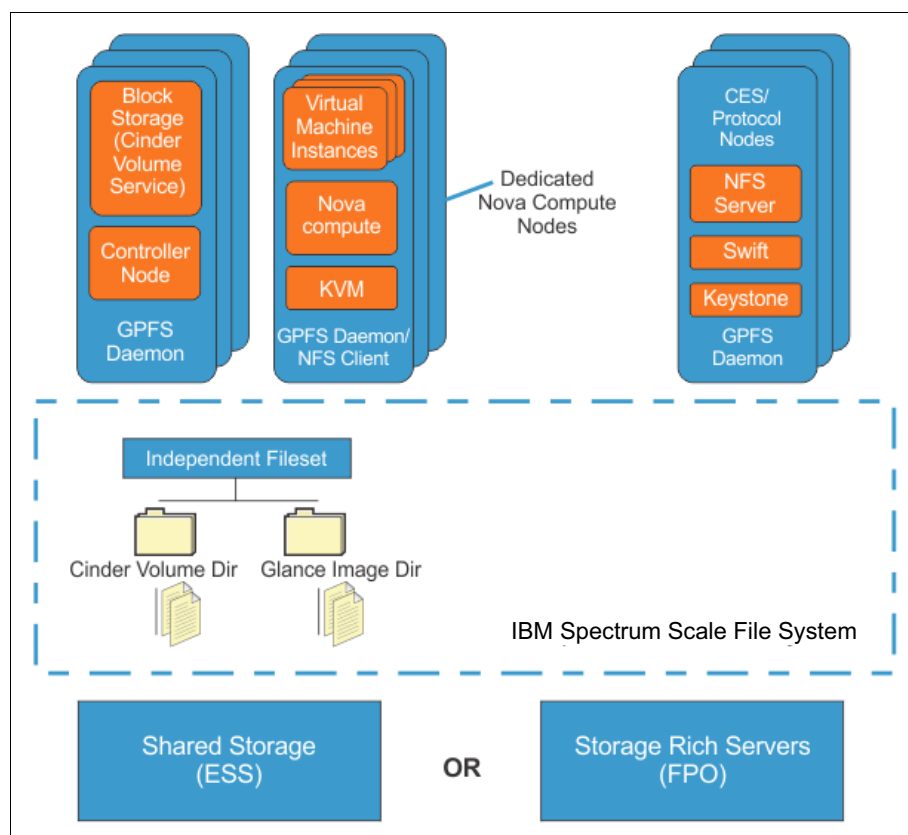


Figure 3-27 IBM Spectrum Scale in an OpenStack cloud deployment

One key benefit of IBM Spectrum Scale is that it provides uniform access to data under a single namespace with integrated analytics.

The following OpenStack components are related to IBM Spectrum Scale:

- ▶ **Cinder:** Provides virtualized block storage for virtual machines. The IBM Spectrum Scale Cinder driver, also known as the GPFS driver, is written to take full advantage of the IBM Spectrum Scale enterprise features.
- ▶ **Glance:** Provides the capability to manage virtual machine images. When Glance is configured to use the same IBM Spectrum Scale fileset that stores Cinder volumes, bootable images can be created almost instantly by using the copy-on-write file clone capability.
- ▶ **Swift:** Provides object storage to any user or application that requires access to data through a RESTful API. The Swift object storage configuration was optimized for the IBM Spectrum Scale environment, which provides high availability and simplified management. Swift object storage also supports native the Swift APIs and Amazon S3 APIs for accessing data. Finally, the Swift object storage also supports access to the same data through object interface or file interface (POSIX, NFS, SMB) without creating a copy.
- ▶ **Manila:** Provides a shared file system access to client, virtual, and physical systems. The IBM Spectrum Scale share driver (GPFS driver) is written to take full advantage of the IBM Spectrum Scale enterprise features.
- ▶ **Keystone:** Although not a storage component, internal keystone with in-built HA is provided by IBM Spectrum Scale as part of the Object protocol. In deployments that already have keystone support, the Object protocol can be configured to use the external keystone server rather than the internal one

IBM Spectrum Scale for Amazon Web Services

Amazon Quick Starts are built by Amazon Web Services (AWS) solutions architects and partners to help you deploy popular solutions on AWS, based on AWS best practices for security and HA. A Quick Start automatically deploys a highly available IBM Spectrum Scale cluster on the AWS Cloud, into a configuration of your choice.

IBM Spectrum Scale is placed into a virtual private cloud (VPC) that spans two Availability Zones in your AWS account. You can build a VPC for IBM Spectrum Scale, or deploy the software into your VPC (see Figure 3-28).

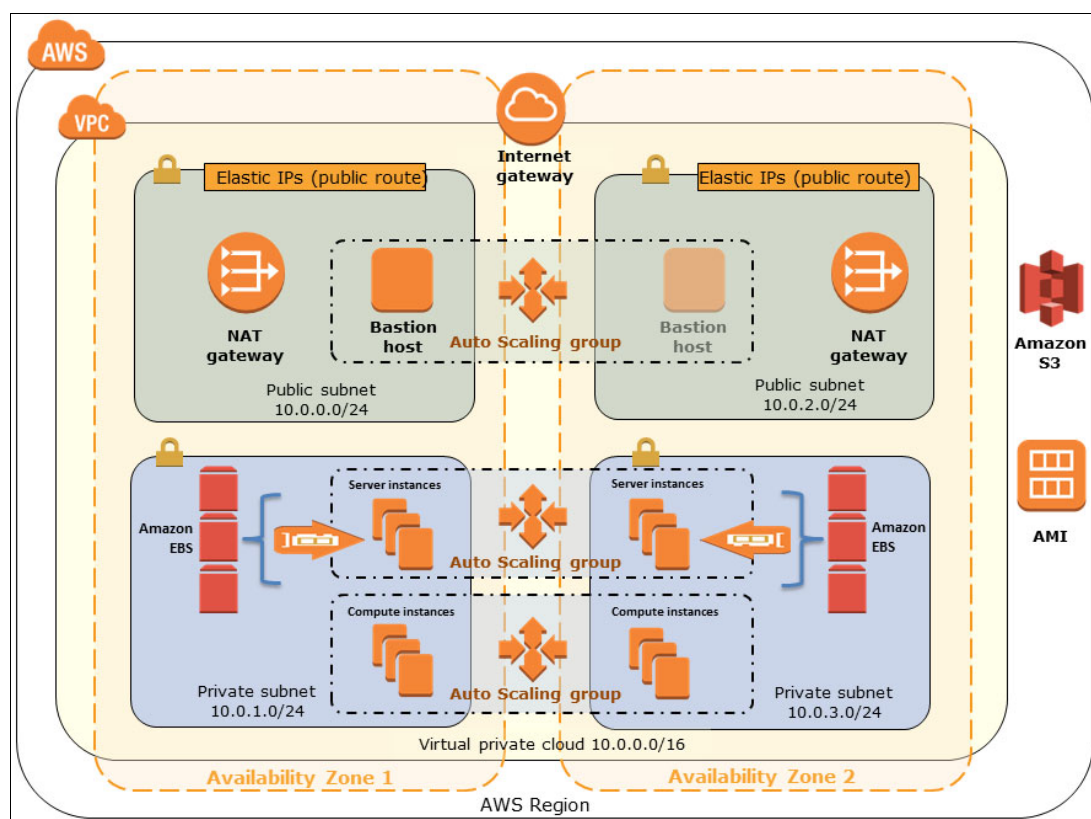


Figure 3-28 IBM Spectrum Scale on AWS

For more information about IBM Spectrum Scale on AWS, see this website:

<https://aws.amazon.com/quickstart/architecture/ibm-spectrum-scale>

3.12.2 IBM Spectrum NAS



IBM Spectrum NAS is a software-defined solution to help customers with enterprise NAS, remote, and departmental file storage needs. It is deployed as a software-defined solution on storage-rich standard x86 servers, with an initial deployment of four instance (physical servers).

IBM Spectrum NAS provides support for SMB and NFS fileshares to users, leveraging internal storage of the servers. IBM Spectrum NAS is designed to support General Purpose Enterprise NAS workloads, such as home directories, Microsoft Applications that require SMB file storage, and Virtual Machines that require file storage.

IBM Spectrum NAS provides the following key capabilities:

- ▶ SMB 1, 2.1 and 3.1.1
- ▶ NFS 3, 4.0 and NFS 4.1

IBM Spectrum NAS is managed from a simple and easy-to-use GUI across the cluster, which removes the need to manage independent file servers or filers. IBM Spectrum NAS data management allows mixing workloads on the same cluster and in multiple tiers in the cluster.

IBM Spectrum NAS includes the following features:

- ▶ Snapshots
- ▶ Quotas
- ▶ Tiering
- ▶ Encryption
- ▶ NENR (Non-Erasable, Non-rewritable) Capabilities
- ▶ Data Retention
- ▶ Synchronous & Asynchronous Replication for DR
- ▶ AntiVirus Integration
- ▶ Authentication by way of LDAP, AD, NIS, Kerberos KDC and local databases

IBM Spectrum NAS Architecture

IBM Spectrum NAS uses a true scale out architecture that allows deployments to add storage nodes as needed. The software architecture features the following tiers (as shown in Figure 3-29):

- ▶ Scale-Out Protocols: Support connections across the file system on all of the nodes in the cluster.
- ▶ Scale-Out Cache Pool: Provides the ability to use NVMe drives to support read and write cache.
- ▶ Scale-Out File Systems: Provides for the file storage across the name space in the cluster.
- ▶ Scale-Out Data Store: Provides data protection across the cluster by using a tunable erasure coding structure that is based on the number of servers in the cluster and the amount of required redundancy.

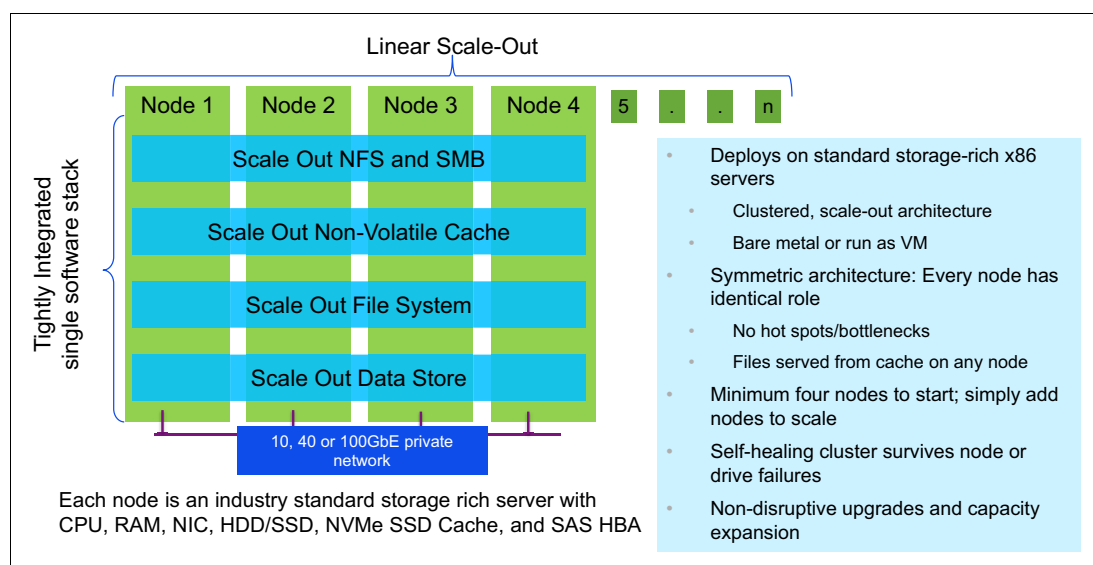


Figure 3-29 IBM Spectrum NAS Software architecture

3.12.3 IBM Spectrum Archive



A member of the IBM Spectrum Storage family, IBM Spectrum Archive enables direct, intuitive, and graphical access to data that is stored in IBM tape drives and libraries by incorporating the IBM Linear Tape File System™ (LTFS) format standard for reading, writing, and exchanging descriptive metadata on formatted tape cartridges. IBM Spectrum Archive eliminates the need for extra tape management and software to access data.

IBM Spectrum Archive offers the following software solutions for managing your digital files with the LTFS format:

- ▶ IBM Spectrum Archive Single Drive Edition (SDE)
- ▶ IBM Spectrum Archive Library Edition (LE)
- ▶ IBM Spectrum Archive Enterprise Edition (EE)

With IBM Spectrum Archive Enterprise Edition and IBM Spectrum Scale, tape can now add savings as a low-cost storage tape tier. The use of a tier of tape for active but “cold” data enables enterprises to look at new ways to cost optimize their unstructured data storage. They can match the value of the data, or the value of the copies of data to the most appropriate storage media.

In addition, the capability to store the data at the cost of tape storage allows customers to build their cloud environments to take advantage of this new cost structure. IBM Spectrum Archive provides enterprises with the ability to store cold data at costs that can be cheaper than some public cloud provider options.

For more information about the potential costs with large-scale cold data storage and retention, see IBM's Tape TCO Calculator, which is available at this website:

<http://www.ibm.com/systems/storage/tape/tco-calculator>

Network attached unstructured data storage with native tape support using LTFS delivers the best mix of performance and lowest cost storage.

Key capabilities

IBM Spectrum Archive options can support small, medium, and enterprise businesses with the following advantages:

- ▶ Seamless virtualization of storage tiers
- ▶ Policy-based placement of data
- ▶ Single universal namespace for all file data
- ▶ Security and protection of assets
- ▶ Open, non-proprietary, cross platform interchange
- ▶ Integrated functionality with IBM Spectrum Scale

Benefits

IBM Spectrum Archive enables direct, intuitive, and graphical access to data that is stored in IBM tape drives and libraries by incorporating the LTFS format standard for reading, writing, and exchanging descriptive metadata on formatted tape cartridges. IBM Spectrum Archive eliminates the need for more tape management and software to access data.

IBM Spectrum Archive takes advantage of the low cost of tape storage while making it easy to use. IBM Spectrum Archive provides the following benefits:

- ▶ Access and manage all data in stand-alone tape environments as easily as though it were on disk
- ▶ Enable easy-as-disk access to single or multiple cartridges in a tape library
- ▶ Improve efficiency and reduce costs for long-term, tiered storage
- ▶ Optimize data placement for cost and performance
- ▶ Enable data file sharing without proprietary software
- ▶ Scalable and low cost

Linear Tape File System

IBM developed LTFS and then contributed it to SNIA as an open standard so that all tape vendors can participate. LTFS is the first file system that works with Linear Tape-Open (LTO) generation 8, 7, 6, and 5 tape technology (or IBM TS1155, TS1150, and TS1140 tape drives) to set a new standard for ease of use and portability for open systems tape storage.

With this application, accessing data that is stored on an IBM tape cartridge is as easy and intuitive as using a USB flash drive. Tapes are self-describing, and you can quickly recall any file from a tape without having to read the whole tape from beginning to end.

Also, any LTFS-capable system can read a tape that is created by any other LTFS-capable system (regardless of the operating system and platform). Any LTFS-capable system can identify and retrieve the files that are stored on it. LTFS-capable systems have the following characteristics:

- ▶ Files and directories are displayed to you as a directory tree listing.
- ▶ More intuitive searches of cartridge and library content are now possible due to the addition of file tagging.
- ▶ Files can be moved to and from LTFS tape by using the familiar drag-and-drop metaphor common to many operating systems.
- ▶ Many applications that were written to use files on disk can now use files on tape without any modification.
- ▶ All standard File Open, Write, Read, Append, Delete, and Close functions are supported.

IBM Spectrum Archive Editions

As shown in Figure 3-30, IBM Spectrum Archive is available in different editions that support small, medium, and enterprise businesses.

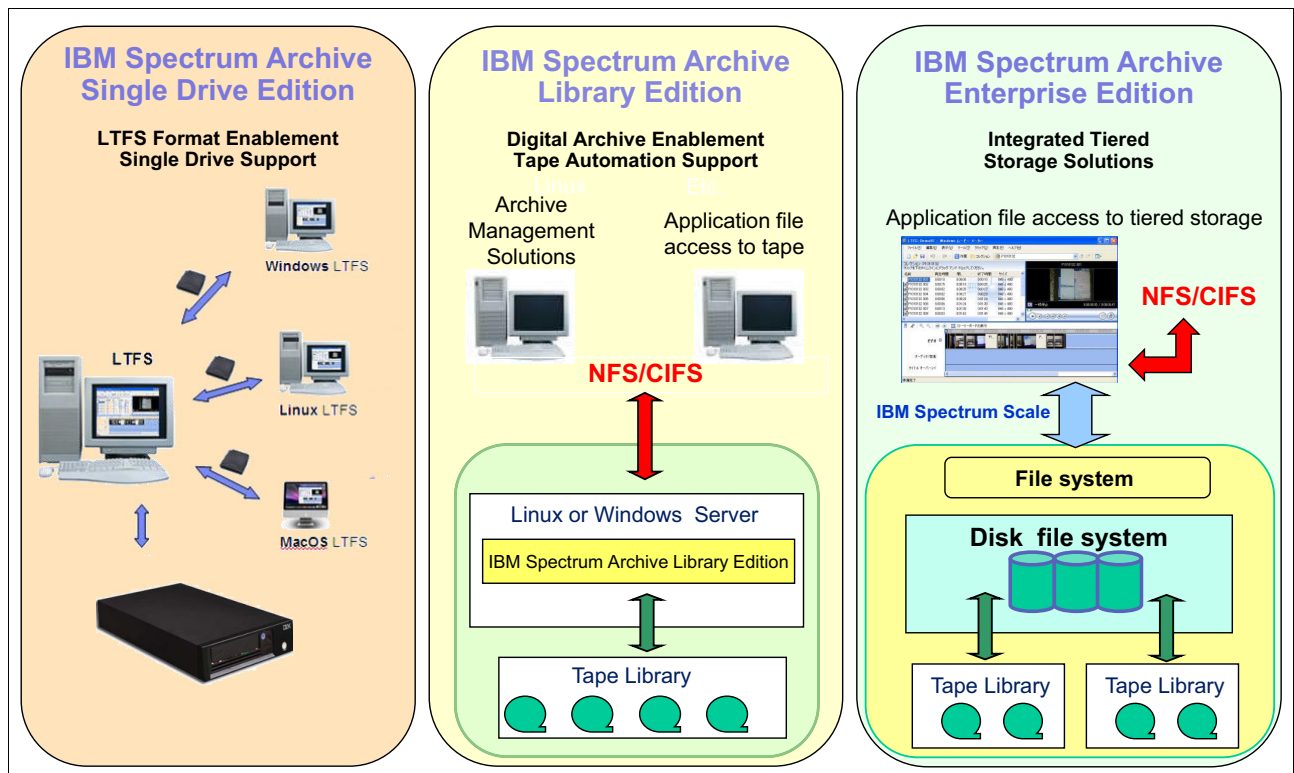


Figure 3-30 IBM Spectrum Archive SDE, LE, and EE implementations

IBM Spectrum Scale Single Drive Edition

The IBM Spectrum Archive Single Drive Edition implements the LTFS Format and allows tapes to be formatted as LTFS Volumes. These LTFS Volumes can then be mounted by using LTFS to allow users and applications direct access to files and directories that are stored on the tape. No integration with tape libraries exists in this edition. You can access and manage all data in stand-alone tape environments as simply as though it were on disk.

IBM Spectrum Archive Library Edition

IBM Spectrum Archive Library Edition extends the file management capability of the IBM Spectrum Archive SDE. IBM Spectrum Archive LE is introduced with Version 2.0 of LTFS. It enables easy-as-disk access to single or multiple cartridges in a tape library.

LTFS is the first file system that works with IBM System Storage tape technology to optimize ease of use and portability for open-systems tape storage. It manages the automation and provides operating system-level access to the contents of the library.

IBM Spectrum Archive LE is based on the LTFS format specification, which enables tape library cartridges to be interchangeable with cartridges that are written with the open source SDE version of IBM Spectrum Archive. IBM Spectrum Archive LE supports most IBM tape libraries, including the following examples:

- ▶ TS2900 tape autoloader
- ▶ TS3100 tape library
- ▶ TS3200 tape library
- ▶ TS3310 tape library
- ▶ TS3500 tape library
- ▶ TS4300 tape library
- ▶ TS4500 tape library

Note: IBM TS1155, TS1150, and IBM TS1140 tape drives are supported on the IBM TS4500 and IBM TS3500 tape libraries only.

IBM Spectrum Archive LE enables the reading, writing, searching, and indexing of user data on tape and access to user metadata. *Metadata* is the descriptive information about user data that is stored on a cartridge. Metadata enables searching and accessing files through the GUI of the operating system. IBM Spectrum Archive LE supports Linux and Windows.

IBM Spectrum Archive LE provides the following product features:

- ▶ Direct access and management of data on tape libraries with LTO Ultrium 8 (LTO-8), LTO Ultrium 7 (LTO-7), LTO Ultrium 6 (LTO-6), LTO Ultrium 5 (LTO-5), and TS1155, TS1150, and TS1140 tape drives
- ▶ Tagging of files with any text, allowing more intuitive searches of cartridge and library content
- ▶ Exploitation of the partitioning of the media in LTO-5 tape format standard
- ▶ One-to-one mapping of tape cartridges in tape libraries to file folders
- ▶ Capability to create a single file system mount point for a logical library that is managed by a single instance of LTFS and runs on a single computer system
- ▶ Capability to cache tape indexes, and to search, query, and display tape content within an IBM tape library without having to mount tape cartridges

The IBM Spectrum Archive LE offers the same basic capabilities as the SDE with additional support of tape libraries. Each LTFS tape cartridge in the library appears as an individual folder within the file space. The user or application can browse to these folders to access the files that are stored on each tape. The IBM Spectrum Archive LE software automatically controls the tape library robotics to load and unload the necessary LTFS Volumes to provide access to the stored files.

IBM Spectrum Archive Enterprise Edition

IBM Spectrum Archive Enterprise Edition (EE) gives organizations an easy way to use cost-effective IBM tape drives and libraries within a tiered storage infrastructure. By using tape libraries instead of disks for Tier 2 and Tier 3 data storage (data that is stored for long-term retention), organizations can improve efficiency and reduce costs.

In addition, IBM Spectrum Archive EE seamlessly integrates with the scalability, manageability, and performance of IBM Spectrum Scale, which is an IBM enterprise file management platform that enables organizations to move from simply adding storage to optimizing data management.

IBM Spectrum Archive EE includes the following highlights:

- ▶ Simplify tape storage with the IBM LTFS format, which is combined with the scalability, manageability, and performance of IBM Spectrum Scale
- ▶ Help reduce IT expenses by replacing tiered disk storage (Tier 2 and Tier 3) with IBM tape libraries
- ▶ Expand archive capacity by simply adding and provisioning media without affecting the availability of data already in the pool
- ▶ Add extensive capacity to IBM Spectrum Scale installations with lower media, floor space, and power costs
- ▶ Support for attaching up to two tape libraries to a single IBM Spectrum Scale cluster

IBM Spectrum Archive EE for the IBM TS4500, IBM TS4300, IBM TS3500, and IBM TS3310 tape libraries provides seamless integration of IBM Spectrum Archive with Spectrum Scale by creating an LTFS tape tier. You can run any application that is designed for disk files on tape by using IBM Spectrum Archive EE.

IBM Spectrum Archive EE can play a major role in reducing the cost of storage for data that does not need the access performance of primary disk. This configuration improves efficiency and reduces costs for long-term, tiered storage.

With IBM Spectrum Archive EE, you can enable the use of LTFS for the policy management of tape as a storage tier in a IBM Spectrum Scale environment and use tape as a critical tier in the storage environment. IBM Spectrum Archive EE supports IBM LTO Ultrium 8, 7, 6, and 5, IBM System Storage TS1155, TS1150, and TS1140 tape drives that are installed in TS4500, TS3500, or LTO Ultrium 8, 7, 6, and 5 tape drives that are installed in the TS4300 and TS3310 tape libraries.

The use of IBM Spectrum Archive EE to replace disks with tape in Tier 2 and Tier 3 storage can improve data access over other storage solutions. It also improves efficiency and streamlines management for files on tape. IBM Spectrum Archive EE simplifies the use of tape by making it transparent to the user and manageable by the administrator under a single infrastructure.

The integration of IBM Spectrum Archive EE archive solution with Spectrum Scale is shown in Figure 3-31.

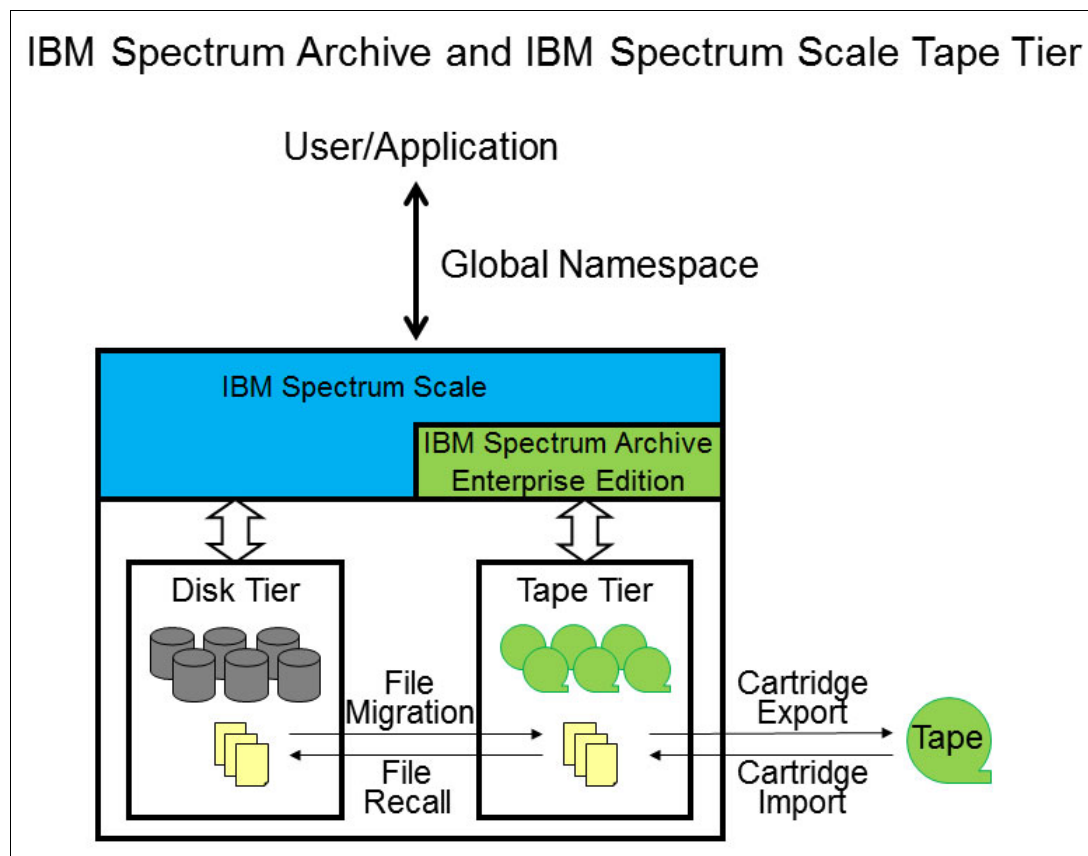


Figure 3-31 Integration of Spectrum Scale and Spectrum Archive EE

The seamless integration offers transparent file access in a continuous name space. It provides file level write and read caching with disk staging area, policy-based movement from disk to tape, creation of multiple data copies on different tapes, load balancing, and HA in multi-node clusters.

It also offers data exchange on LTFS tape by using import and export functions, fast import of file name space from LTFS tapes without reading data, built-in tape reclamation and reconciliation, and simple administration and management.

For more information, see this website:

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

IBM Spectrum Archive in two site mode

Asynchronous Archive Replication is an extension to the stretched cluster configuration, in which users require the data that is created is replicated to a secondary site and can be migrated to tape at both sites by incorporating IBM Spectrum Scale AFM to the stretched cluster. In addition to geolocation capabilities, data that is created on home or cache is asynchronously replicated to the other site.

Asynchronous Archive Replication (see Figure 3-32 on page 76) requires two remote clusters configured: the home cluster and a cache cluster with the independent writer mode. By using the independent writer mode in this configuration, users can create files at either site and the data and metadata is asynchronously replicated to the other site.

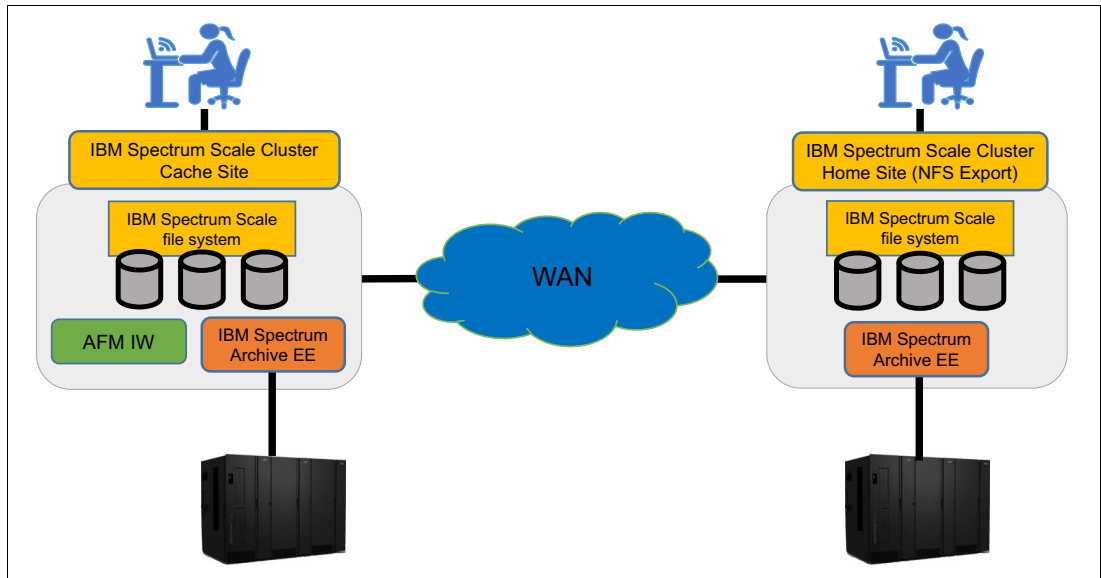


Figure 3-32 Asynchronous Archive Replication

For more information about the latest IBM Spectrum Archive release see *IBM Spectrum Archive Enterprise Edition V1.2.6 Installation and Configuration Guide*, SG24-8333.

Monitoring statistics of IBM Spectrum Archive

IBM Spectrum Archive Enterprise Edition (IBM Spectrum Archive EE) supports a dashboard that helps storage administrators manage and monitor the storage system by using a browser-based graphical interface. By using the dashboard, you can see the following information without the need to log in to a system and enter a command:

- ▶ If a system is running without error. If an error exists, the type of detected error is indicated.
- ▶ Basic tape-related configuration, such as how many pools are available and the amount of space that is available.
- ▶ A time-scaled storage consumption for each tape pool.
- ▶ A throughput of each drive for migration and recall.

The IBM Spectrum Archive EE dashboard is shown in Figure 3-33.

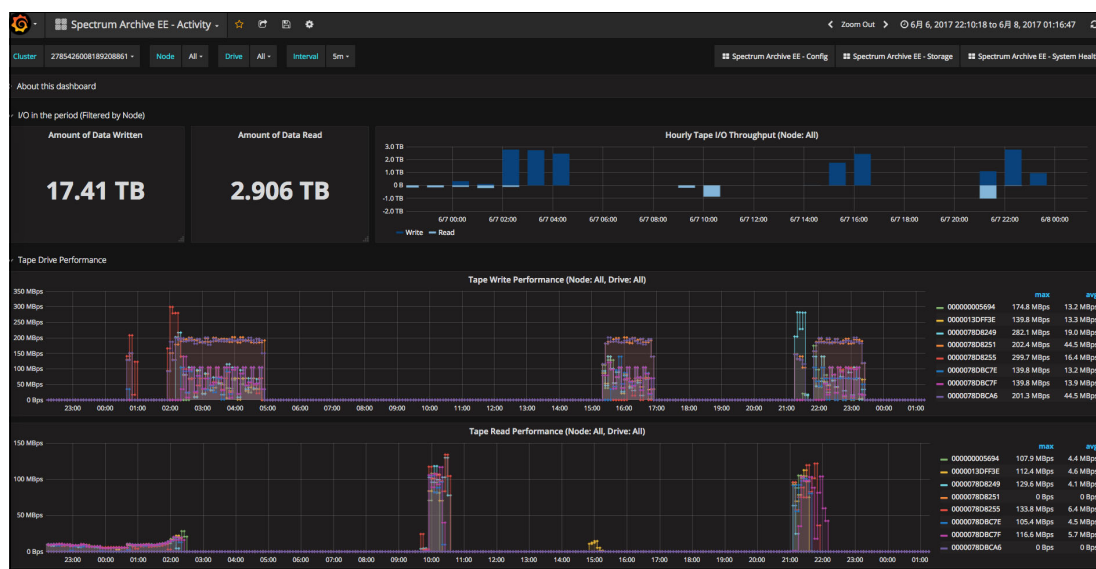


Figure 3-33 IBM Spectrum Archive sample dashboard

For more information, see the [IBM Spectrum Archive Dashboard Deployment Guide](#).

OpenStack and IBM Spectrum Archive

IBM Spectrum Archive Enterprise Edition can also be used to provide object storage by using OpenStack Swift. By using this configuration, objects can be stored in the file system and exist on disk or tape tiers within the enterprise.

For more information about creating an object storage Active Archive with IBM Spectrum Scale and Spectrum Archive, see *Active Archive Implementation Guide with IBM Spectrum Scale Object and IBM Spectrum Archive*, REDP-52377.

3.13 IBM Object Storage solutions

This section describes IBM object storage solutions.

3.13.1 IBM Cloud Object Storage



The IBM Cloud Object Storage (COS) system is a breakthrough cloud platform that helps solve petabyte and beyond storage challenges for companies worldwide. Clients across multiple industries use IBM Cloud Object Storage for large-scale content repository, backup, archive, collaboration, and SaaS.

The Internet of Things (IoT) allows every aspect of life to be instrumented through millions of devices that create, collect, and send data every second. These trends are causing an unprecedented growth in the volume of data being generated. IT organizations are now tasked with finding ways to efficiently preserve, protect, analyze, and maximize the value of their unstructured data as it grows to petabytes and beyond. Object storage is designed to handle unstructured data at web-scale.

The IBM Cloud Object Storage portfolio gives clients strategic data flexibility, simplified management, and consistency with on-premises, cloud, and hybrid cloud deployment options (see Figure 3-34).

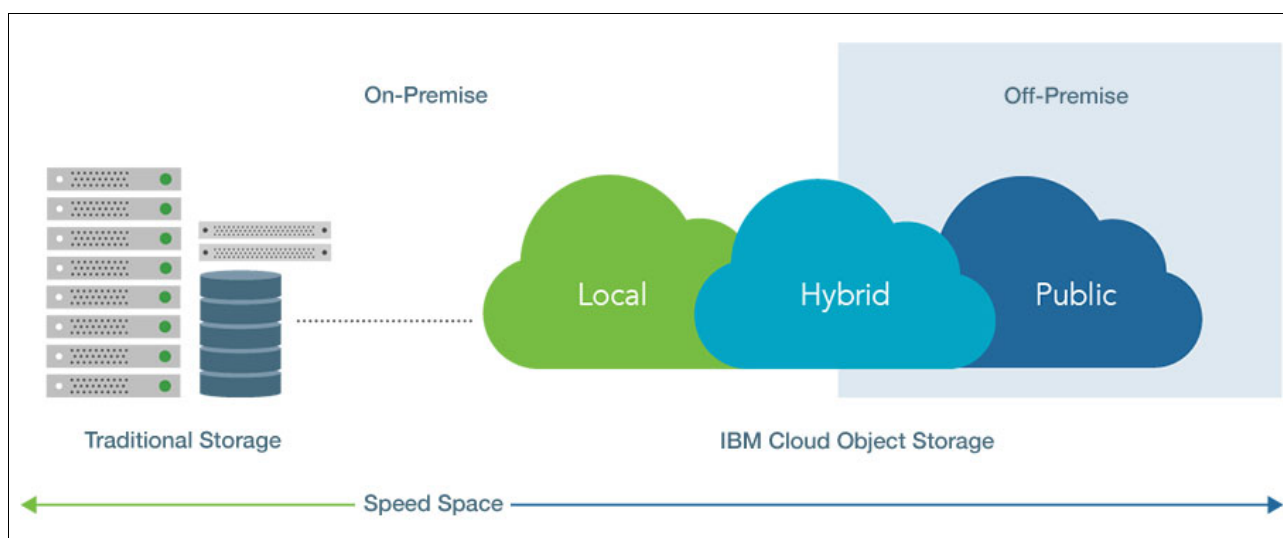


Figure 3-34 IBM Cloud Object Storage offers flexibility for on-premises, cloud, and hybrid cloud deployment options

IBM Cloud Object Storage solutions enhances on-premises storage options for clients and service providers with low-cost, large-scale active archives and unstructured data content stores. The solutions complement the software-defined IBM Spectrum Storage portfolio for data protection and backup, tape archive, and a high-performance file and object solution where the focus is on response time.

IBM Cloud Object Storage can be deployed as an on-premises, public cloud, or hybrid solution, which provides unprecedented choice, control, and efficiency:

- **On-Premise solutions**
Deploy IBM Cloud Object Storage on premises for optimal scalability, reliability, and security. The software runs on industry standard hardware for flexibility and simplified management.
- **Cloud Solutions**
Easily deploy IBM Cloud Object Storage on the IBM Cloud public cloud.
- **Hybrid Solutions**
For optimal flexibility, deploy IBM Cloud object storage as a hybrid solution to support multiple sites across your enterprise (on-premises and in the public cloud) for agility and efficiency.

Access methods

The IBM Cloud Object storage pool can be shared and is jointly accessible by multiple access protocols:

- ▶ Object-based access methods: The Simple Object interface is accessed with a HTTP/REST API. Simple **PUT**, **GET**, **DELETE**, and **LIST** commands enable applications to access digital content, and the resulting object ID is stored directly within the application. The IBM COS Accesser® does not require a dedicated appliance because the application can talk directly to the IBM COS Slicestor® using object IDs (see Figure 3-35).
- ▶ REST API access to storage: REST is a style of software architecture for distributed hypermedia information retrieval systems such as the World Wide Web. REST style architectures consist of clients and servers. Clients send requests to servers. Servers process those requests and return associated responses. Requests and responses are built around the transfer of various representations of the resources. The REST API works in way that is similar to retrieving a Universal Resource Locator (URL). But instead of requesting a web page, the application references an object.
- ▶ File-based access methods: Dispersed storage can also support the traditional NAS protocols (SMB/CIFS and NFS) through integration with third-party gateway appliances. Users and storage administrators are able to easily transfer, access, and preserve data assets over standard file protocol.

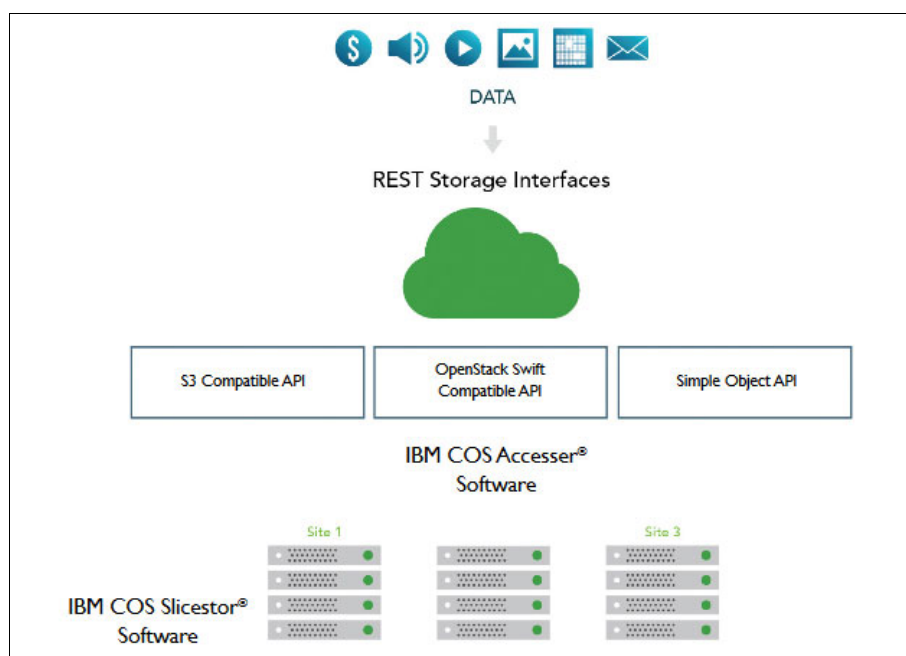


Figure 3-35 REST APIs accessing objects using object IDs with IBM COS Slicestor

The IBM COS System is deployed as a cluster that combines three types of nodes, as shown in Figure 3-36. Each node consists of IBM COS software running on an industry-standard server. IBM COS software is compatible with a wide range of servers from many sources, including a physical or virtual appliance. In addition, IBM conducts certification of specific servers that customers want to use in their environment to help ensure a quick initial installation, long-term reliability, and predictable performance.

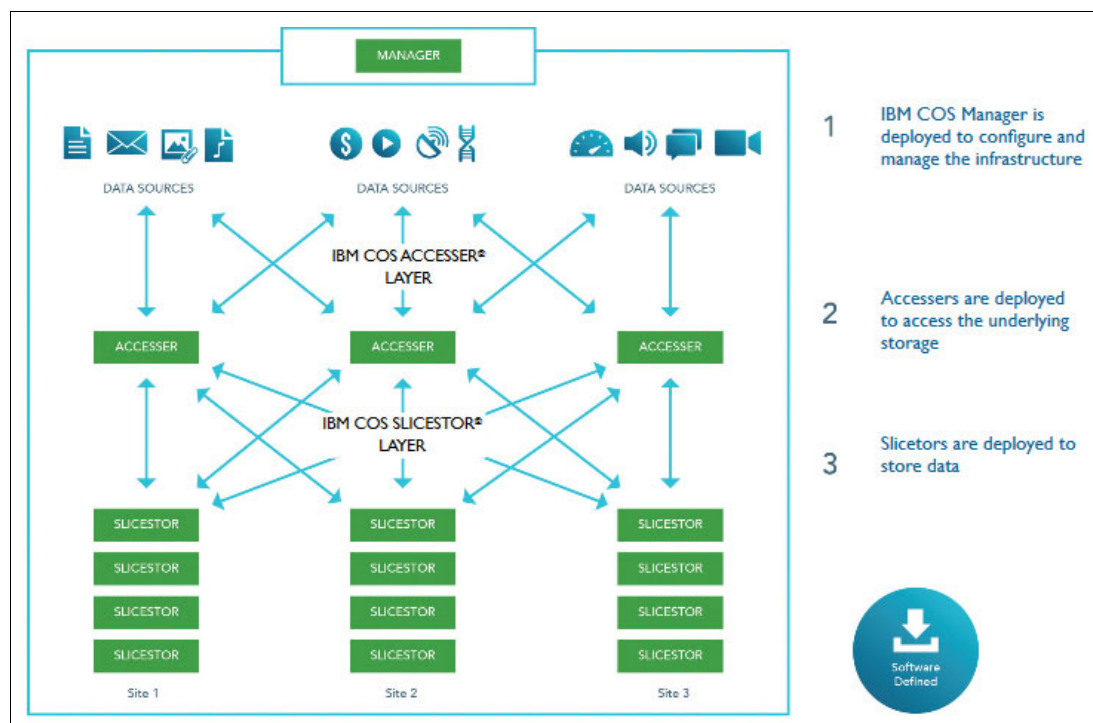


Figure 3-36 IBM COS System deployed as a cluster combining three types of nodes

The following three types of nodes are available:

- ▶ IBM Cloud Object Storage Manager
- ▶ IBM Cloud Object Storage Accesser
- ▶ IBM Cloud Object Storage Slicestor

Each IBM COS System include the following nodes:

- ▶ A single Manager node, which provides out-of-band configuration, administration and monitoring capabilities
- ▶ One or more Accesser nodes, which provide the storage system endpoint for applications to store and retrieve data
- ▶ One or more Slicestor nodes, which provide the data storage capacity for the IBM COS System

The Accesser is a stateless node that presents the storage interface of the IBM COS System to client applications and transforms data using an Information Dispersal Algorithm (IDA). Slicestor nodes receive data to be stored from Accesser nodes on ingest and return data to Accesser nodes as required by reads.

The IDA transforms each object written to the system into a number of slices such that the object can be read bit-perfectly by using a subset of those slices. The number of slices created is called the IDA Width (or Width) and the number required to read the data is called the IDA Read Threshold (or Read Threshold).

The difference between the Width and the Read Threshold is the maximum number of slices that can be lost or temporarily unavailable while still maintaining the ability to read the object. For example, in a system with a width of 12 and threshold of seven, data can be read even if five of the 12 stored slices cannot be read.

Storage capacity is provided by a group of Slicestor nodes, which are referred to as a storage pool. In the diagram in Figure 3-36 on page 80, 12 Slicestor nodes are grouped in a storage pool. A single IBM COS System can have one or multiple storage pools.

A Vault is not part of the physical architecture, but is an important concept in an IBM COS System. A Vault is a logical container or a virtual storage space, upon which reliability, data transformation options (for example, IBM COS SecureSlice and IDA algorithm), and access control policies can be defined. Multiple vaults can be provisioned on the same storage pool.

The Information Dispersal Algorithm combines encryption and erasure-coding techniques that are designed to transform the data in a way that enables highly reliable and available storage without making copies of the data as would be required by traditional storage architectures.

Information Dispersal

At the foundation of the IBM COS System is a technology called *information dispersal*. Information dispersal is the practice of using erasure codes as a means to create redundancy for transferring and storing data. An erasure code is a Forward Error Correction (FEC) code that transforms a message of k symbols into a longer message with n symbols such that the original message can be recovered from a subset of the n symbols (k symbols).

Erasure codes use advanced deterministic math to insert “extra data” in the “original data” that allows a user to need only a subset of the “coded data” to re-create the original data.

An IDA can be made from any Forward Error Correction code. The extra step of the IDA is to split the coded data into multiple segments. These segments can then be stored on different devices or media to attain a high degree of failure independence. For example, the use of forward FEC alone on files on your computer is less likely to help if your hard disk drive fails. However, if you use an IDA to separate pieces across machines, you can now tolerate multiple failures without losing the ability to reassemble that data.

As shown in Figure 3-37 on page 82, five variables (as indicated by a - e in Figure 3-37 on page 82) and eight different equations that use these variables, with each yielding a different output. To understand how information dispersal works, imagine the five variables are bytes.

Following the eight equations, you can compute eight results, each of which is a byte. To solve for the original five bytes, you can use any five of the resulting eight bytes. This process is how information dispersal can support any value for k , $n - k$ is the number of variables, and n is the number of equations.



Figure 3-37 Example of calculations to illustrate how information dispersal works

How the Storage Dispersal and Retrieval works

At a basic level, the IBM COS System uses the following process for slicing, dispersing, and retrieving data (see Figure 3-38 on page 83):

1. Data is virtualized, transformed, sliced, and dispersed by using IDAs. In the example that is shown in Figure 3-38 on page 83, the data is separated into 12 slices. Therefore, the "width" (n) of the system is 12.
2. Slices are distributed to separate disks, storage nodes, geographic locations, or some combination of these three. In this example, the slices are distributed to three different sites.
3. The data is retrieved from a subset of slices. In this example, the number of slices that are needed to retrieve the data is 7. Therefore, the "threshold" (k) of the system is 7. Given a width of 12 and a threshold of 7, this example can be called a "7 of 12" (k of n) configuration.

The configuration of a system is determined by the level of reliability required. In a “7 of 12” configuration, five slices can be lost or unavailable and the data can still be retrieved because the threshold of seven slices has been met. With a “5 of 8” configuration, only three slices can be lost, so the level of reliability is lower. Conversely, with a “20 of 32” configuration, 12 slices can be lost, so the level of reliability is higher.

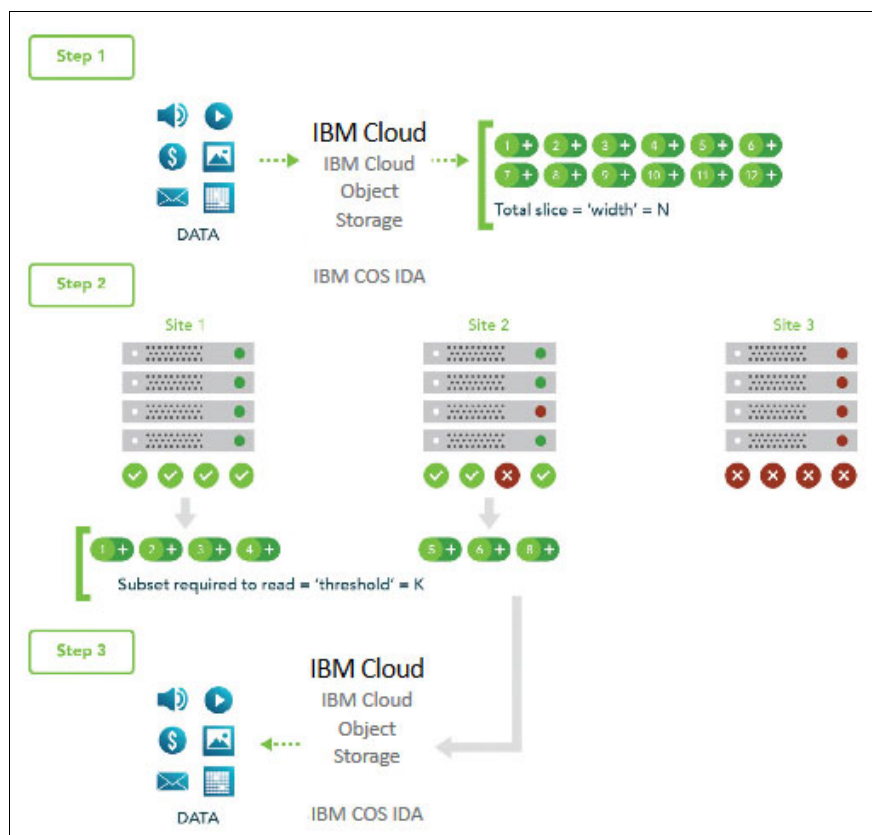


Figure 3-38 COS System's three steps for slicing, dispersing, and retrieving data

Security

IBM COS uses Information Dispersal Algorithm (IDA) to split and disperse data to all SliceStore nodes; therefore, no whole copy of any object is in any single disk, node, or location. IBM COS implements the following security features on top of IDA:

- ▶ Crucial configuration information is digitally signed.
- ▶ Communication between any node is Certificate-based.
- ▶ TLS is supported between IBM COS nodes and on Client to Accesser network connections.
- ▶ SecureSlice algorithm can optionally be applied when storing data. It can implement RC4-128, AES-128, or AES-256 encryption, with MD5-128 or SHA-256 hash algorithm.

For more information about COS security aspects, see *IBM Cloud Object Storage Concepts and Architecture: An Under-the-Hood Guide for IBM Cloud Object Storage*, REDP-5435.

Compliance Enabled Vault support

The Compliance Enabled Vault (CEV) solution provides the user with the ability to create compliance vaults. Objects that are stored in compliance vaults are protected objects that include associated retention periods and legal holds. Protected objects cannot be deleted until the retention period expires and all legal holds on the object are deleted.

Applications can leverage CEV storage and control its retention by way of standard S3 API.

With this feature, IBM COS is natively compliant to the following standard and compliance requirements:

- ▶ Securities and Exchange Commission (SEC) Rule 17a-4(f)
- ▶ Financial Industry Regulatory Authority (FINRA) Rule 4511, which references requirements of SEC Rule 17a-4(f)
- ▶ Commodity Futures Trading Commission (CFTC) Rule 1.31(b)-(c) (July/Aug. '17 Release)

File support

Although IBM COS is primarily an object storage, situations exist in which file support is required. In such cases, the following alternatives are available:

- ▶ Provide a native file support feature
- ▶ Use a solution that is provided by certified partners

Partner-provided file support

The partners that provide a file interface that is certified with IBM COS are listed in Table 3-6.

Table 3-6 IBM COS gateway options

Gateway product	Description
IBM Spectrum Scale	IBM Spectrum Scale is a software-defined parallel file system with a rich HPC heritage. It can make available native IBM GPFS protocol, and CIFS/SMB, NFS, and OpenStack Swift. IBM Spectrum Scale can use IBM COS as an external storage pool and move inactive data to this tier.
IBM Aspera®	IBM Aspera is a software suite that is designed for high-performance data transfer, which is achieved through a protocol stack that replaces TCP with the IBM FASP® protocol. IBM Aspera Direct-to-Cloud provides the capability to make available a file system interface to users and applications through a web interface or through client software that is installed on the server or workstation or mobile device and provide file sync and share capability. Data that is imported through Aspera can be read as an object because one file is uploaded as one object to IBM COS.
Avere FXT Filer	Avere consists of a hardware or virtual offering that makes available CIFS/NFS for users and applications. Because the architecture is caching, active, recent, and hot data is cached locally while the entire data set is kept in IBM COS. Data reduction in the form of compression is available and the customer use cases include I/O intensive workloads, such as rendering and transcoding.
Nasuni Filer	Nasuni filer is a software or hardware solution that provides general purpose NAS capability through SMB/CIFS and NFS. Because the architecture is caching active, recent, and hot data is cached locally while the entire data set is kept in IBM COS. The suite provides file sync and share capability and a global name space. Data reduction in the form of deduplication and compression is included.
Panzura	Panzura is a software or hardware solution that provides general purpose NAS capability through SMB/CIFS and NFS. Because the architecture is caching, active, recent, and hot data is cached locally while the entire data set is kept in IBM COS. Data reduction in the form of deduplication and compression is included.

Gateway product	Description
Ctera	Ctera is a hardware and software solution that is focused on file sync and share capability. The architecture caches the data set onsite with the master copy retained in IBM COS. Client software can be installed on workstations or mobile devices, Data can also be made available through SMB/CIFS and NFS. Data reduction in the form of deduplication and compression is included.
Storage Made Easy (SME)	SME provides File Sync, which is a software-based file sync and share solution with mobile and workstation clients and a focus on inter-cloud compatibility including IBM COS.
CloudBerry Explorer	CloudBerry Explorer is a software-based object storage client that allows users to directly interact with an IBM COS. Data that is imported through CloudBerry Explorer can be read as an object because one file is uploaded as one object to IBM COS.
Seven10 Storfirst	Seven10 Storfirst is a software-based SMB/CIFS and NFS gateway offering that can talk to IBM COS, legacy tape, VTLs.

IBM Cloud Object: Concentrated Dispersal Mode

Concentrated Dispersal Mode enables deployment of a COS System with as few as three SliceStor nodes (six in the case of two-site deployments). Customers can then start with a system as small as 72 TB at reasonable cost, and grow seamlessly to petabytes and beyond. As their object storage capacity needs grow, their cost per TB of capacity decreases significantly.

Concentrated Dispersal Mode is implemented with the following concept:

- ▶ When Standard Dispersal Mode is set, data is sliced; then, each slice is spread on the COS system (one slice per SliceStor node).
- ▶ With Concentrated Dispersal Mode, after data is sliced, multiple slices are stored within the same SliceStor node. This process allows a COS system to be configured with a wider IDA than the number of SliceStor nodes.

An example of an IDA that is 12 slices wide is shown in Figure 3-39. With normal dispersal mode, this IDA requires 12 SliceStor nodes, whereas with concentrated dispersal mode that is configured with four slices per node, IDA requires three SliceStor nodes.

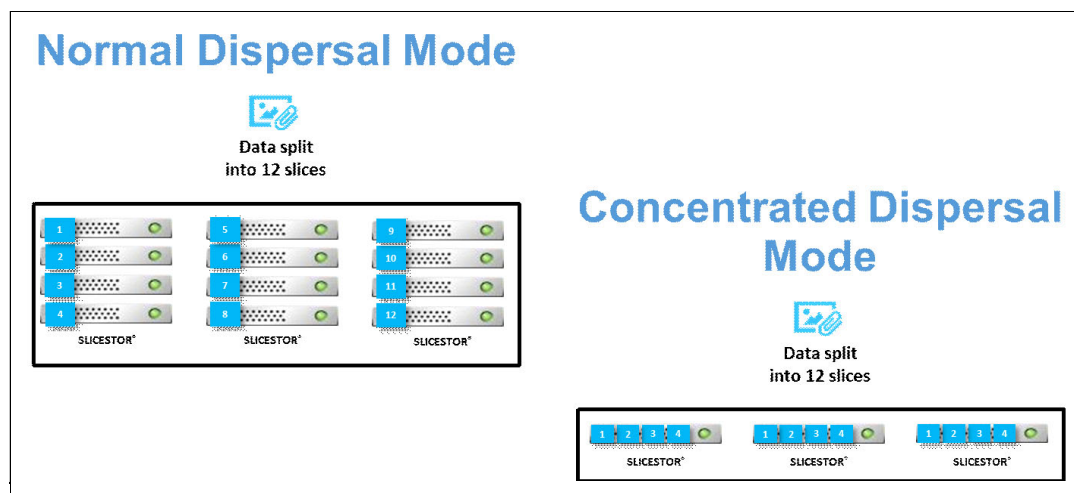


Figure 3-39 Concentrated Dispersal Mode concept

The number of multiple slices that is stored in the same Slicestor node depends on the configured IDA. A wider IDA reduces storage expansion factor, but requires each Slicestor to process more slices, which optimizes storage efficiency, but lowers Ops/sec performance. Although a narrower IDA increases the storage expansion factor, it reduces the number of slices each Slicestor must process, which lowers storage efficiency, but optimizing Ops/sec performance.

After a customer deploys such a low-end system, they can expand it at will to fit their current needs by adding a device set to a new storage pool if they need to change their IDA, or adding a device set to an existing storage pool.

The advantages and drawbacks of both expansion options are listed in Table 3-7.

Table 3-7 IBM COS System storage expansion options

Expansion option	PRO's	CON's
Add a new storage pool over a new device set	Any legal set size and IDA can be added	<ul style="list-style-type: none"> ► Vaults must be created on new storage pool to use added storage ► No system level distribution of writes across storage pools
Add a new device set within the same storage pool	<ul style="list-style-type: none"> ► Existing Vaults can use added storage ► System automatically rebalances storage across sets ► System distributes new writes across sets to balance capacity percentage usage 	Limited set size and IDAs that can be added

VersaStack for IBM Cloud Object Storage: Cisco Validated Design

Cisco servers can now host IBM Cloud Object Storage code. The main purpose of VersaStack for IBM Cloud Object Storage Cisco Validated Design (CVD) is to show which Cisco servers are validated to run IBM Cloud Object Storage code, and how they get integrated into a VersaStack infrastructure.

For more information about VersaStack Cisco Validated Design (CVD), see 4.7, “VersaStack for Hybrid Cloud”.

Cisco UCS C220 M4 are validated to host IBM Cloud Object Storage Manager or Accesser cod. Cisco UCS S3260 are validated to host IBM Cloud Object Storage Slicestor code.

A sample physical layout of an integration of the Cisco servers that host IBM Cloud Object Storage code with VersaStack is shown in Figure 3-40.

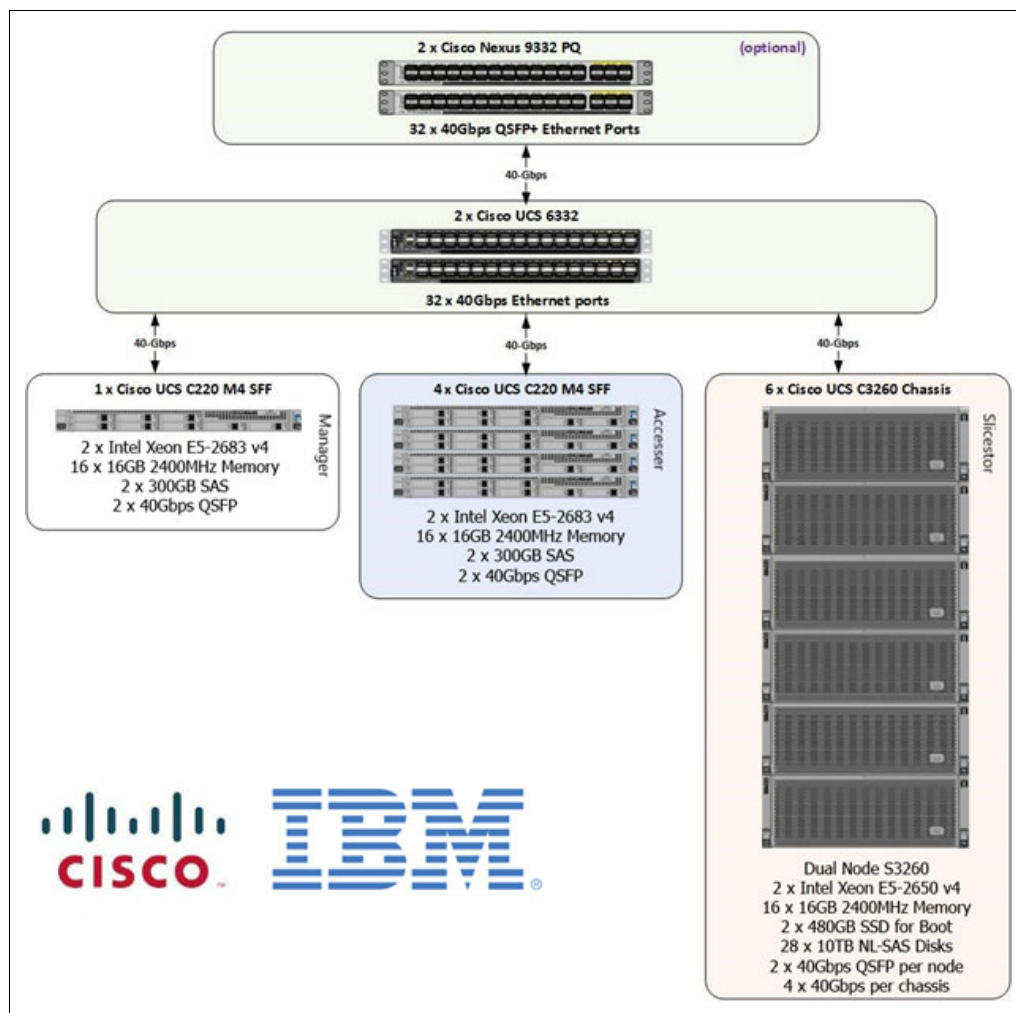


Figure 3-40 IBM COS integration with VersaStack

IBM Spectrum Scale Object support

IBM Spectrum Scale supports file and object solutions. For more information about IBM Spectrum Scale object support, see 3.12.1, “IBM Spectrum Scale” on page 59.

IBM Spectrum Scale includes the ability to provide a single namespace for all data, which means that applications can use the POSIX, NFS, and SMB file access protocols with an HDFS connector plug-in, and the Swift and S3 object protocols, all to a single data set. IBM Spectrum Scale Object Storage combines the benefits of IBM Spectrum Scale with the best pieces of OpenStack Swift, which is the most widely used open source object store today,

Core to the benefit of the use of IBM Spectrum Scale for Object services is the integration of file and object in a single system, which provides applications the ability to store data in one place, and administrators to support multiple protocol services from one storage system.

For storage policies that include enabled file access, the same data can be accessed in place from object and file interfaces so that data does not need to be moved between storage pillars for different kinds of processing. IBM Spectrum Scale Object OpenStack Swift is bundled and managed as part of the deliverable, which hides all of the complexities that otherwise are shown in the raw open source project (see Figure 3-41).

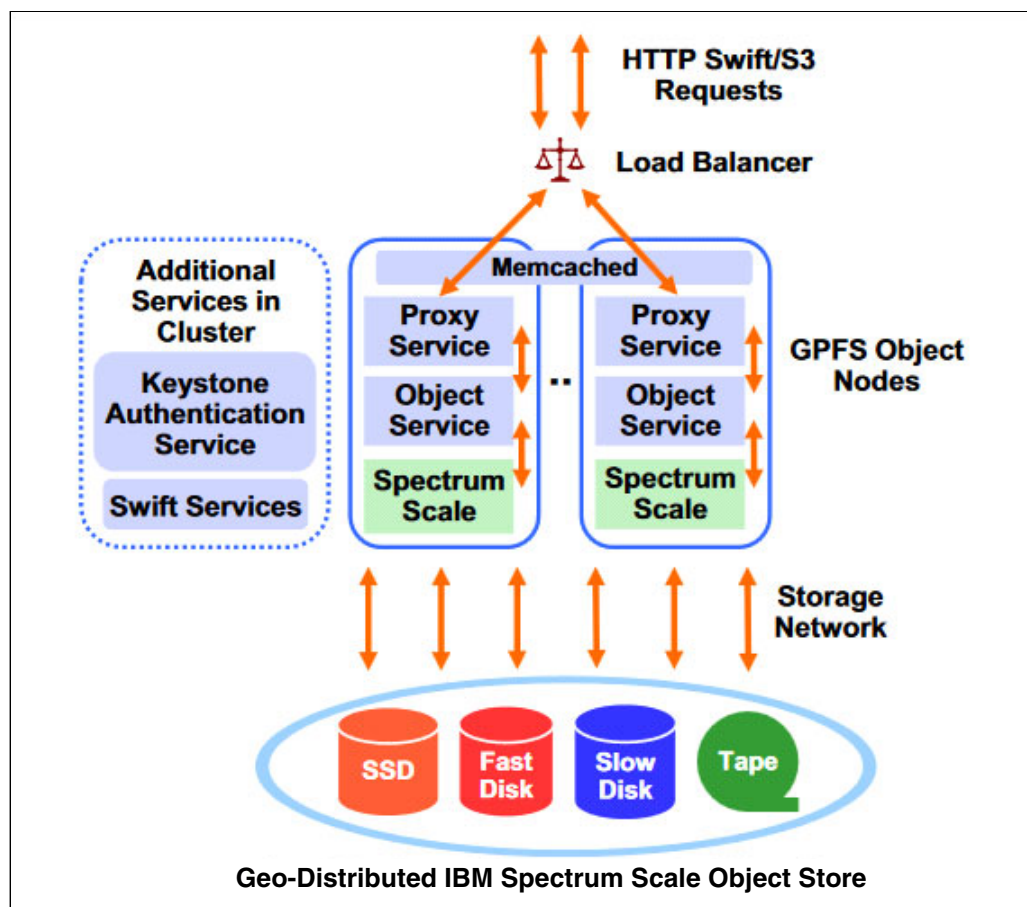


Figure 3-41 IBM Spectrum Scale Object Store architecture

As shown in Figure 3-41, all of the IBM Spectrum Scale protocol nodes are active and provide a front end for the entire object store. The Load Balancer, which distributes HTTP requests across the IBM Spectrum Scale protocol nodes, can be based on software or hardware.

The IBM Spectrum Scale protocol nodes run the IBM Spectrum Scale client and all the Swift services. Clients that use the Swift or S3 API (users or applications) first obtain a token from the Keystone authorization service. The token is included in all requests that are made to the Swift proxy service, which verifies the token by comparing it with cached tokens or by contacting the authorization service.

After applications are authenticated, they perform all object store operations, such as storing and retrieving objects and metadata, or listing account and container information through any of the proxy service daemons (possibly by using an HTTP Load Balancer, as shown in Figure 3-41).

For object requests, the proxy service then contacts the object service for the object, which in turn performs file system-related operations to the IBM Spectrum Scale client. Account and container information requests are handled in a similar manner.

IBM Spectrum Scale Object Storage can provide an efficient storage solution for cost and space, which is built on commodity parts that offer high throughput. The density and performance varies with each storage solution.

IBM Spectrum Scale Object Storage features the following benefits:

- ▶ Use of IBM Spectrum Scale data protection: Delegating the responsibility of protecting data to IBM Spectrum Scale and not using the Swift three-way replication or the relatively slow erasure coding increases the efficiency and the performance of the system in the following ways:
 - With IBM Spectrum Scale RAID as part of the IBM Elastic Storage Server (see “IBM Spectrum NAS” on page 69), storage efficiency rises from 33% to up to 80%.
 - Disk failure recovery does not cause data to flow over the storage network. Recovery is handled transparently and with minimal affect on applications. For more information, see “GPFS-based implementation of a hyper-converged system for a software defined infrastructure” by Azagury, et al.
- ▶ Applications now realize the full bandwidth of the storage network because IBM Spectrum Scale writes only a single copy of each object to the storage servers. Consider the following points
 - Maximum object size is increased up to a configurable value of 5 TB. With IBM Spectrum Scale data striping, large objects do not cause capacity imbalances or server hotspots. They also do not inefficiently use available network bandwidth.
 - No separate replication network is required to replicate data within a single cluster.
 - Capacity growth is seamless because the storage capacity can be increased without requiring rebalancing of the objects across the cluster.
 - IBM Spectrum Scale protocol nodes can be added or removed without failure requiring recovery operations or movement of data between nodes and disks.
- ▶ Integration of file and object in a single system: Applications can store various application data in a single file system. For storage policies that include enabled file access, the same data can be accessed in place from object and file interfaces so that data does not need to be moved between storage pillars for different kinds of processing.
- ▶ Energy saving: High per-server storage density and efficient use of network resources reduces energy costs.
- ▶ Enterprise storage management features: IBM Spectrum Scale Object Storage uses all IBM Spectrum Scale features, such as global namespace, compression, encryption, backup, disaster recovery, ILM (auto-tiering), tape integration, transparent cloud tiering (TCT), and remote caching.

For more information see the following resources:

- ▶ <http://www.redbooks.ibm.com/abstracts/redp5113.html>
- ▶ <https://ibm.biz/BdZgCM>

3.14 IBM storage support of OpenStack components

OpenStack technology is a key enabler of cloud infrastructure as a service (IaaS) capability. OpenStack architecture provides an overall cloud preferred practices workflow solution that is readily installable, and supported by a large ecosystem of worldwide developers in the OpenStack open source community.

Within the overall cloud workflow, specific OpenStack components support storage. The following OpenStack components support storage:

- ▶ IBM Cinder storage drivers
- ▶ Swift (object storage)
- ▶ Manila (file storage)

OpenStack architecture is one implementation of a preferred practices cloud workflow. Regardless of the cloud operating system environment that is used, the following key summary points apply:

- ▶ Cloud operating systems provide the necessary technology workflow to provide truly elastic, pay per use cloud services
- ▶ OpenStack cloud software provides a vibrant open source cloud operating system that is growing quickly
- ▶ OpenStack storage components

3.14.1 Cinder

Cinder is an OpenStack project to provide *block storage as a service* and provides an API to users to interact with different storage backend solutions. Cinder component provides support, provisioning, and control of block storage. The following are standards across all drivers for Cinder services to properly interact with a driver.

Icehouse updates for Cinder are block storage added backend migrations with tiered storage environments, allowing for performance management in heterogeneous environments. Mandatory testing for external drivers now ensures a consistent user experience across storage platforms, and fully distributed services improve scalability.

3.14.2 Swift

The OpenStack Object Store project, which is known as OpenStack Swift, offers cloud storage software so that you can store and retrieve lots of data with a simple API. It is built for scale and optimized for durability, availability, and concurrency across the entire data set. Swift is ideal for storing unstructured data that can grow without bound.

Note: Do not confuse OpenStack Swift with Apple Swift, a programming language. In this paper, the term “Swift” always refers to OpenStack Swift.

3.14.3 Manila

The OpenStack Manila (File) component provides file storage, which allows coordinated access to shared or distributed file systems. Although the primary consumption of shares would be OpenStack compute instances, the service is also intended to be accessed independently, based on the modular design established by OpenStack services.

Manila features the following capabilities:

- ▶ Shared file system services for VMs
- ▶ Vendor-neutral API for NFS/CIFS and other network file systems
- ▶ IBM Spectrum Scale Manila (in Kilo):
 - Extends Spectrum Scale data plane into VM
 - Supports both kNFS and Ganesha 2.0
 - Create/list/delete Shared and Snapshots
 - Allow/deny access to a share based on IP address
 - Multi-tenancy

For more information about OpenStack technology, see the following website:

<http://www.openstack.org>

3.14.4 IBM SDS products that include interfaces to OpenStack component

The following IBM SDS products include interfaces to OpenStack components:

- ▶ The IBM Storage Driver for OpenStack environments: The IBM Storage Driver for OpenStack environments is a software component that integrates with the OpenStack cloud environment. It enables the usage of storage resources that are provided by the following IBM storage systems:
 - DS8880: This storage system can offer a range of capabilities that enable more effective storage automation deployments in private or public clouds. Enabling the OpenStack Cinder storage component with DS8880 allows for storage to be made available whenever it is needed without the traditional associated cost of highly skilled administrators and infrastructure. For more information, see *Using IBM DS8000 in an OpenStack Environment*, REDP-5220.
 - IBM Spectrum Accelerate: Remote cloud users can issue requests for storage resources from the OpenStack cloud. These requests are transparently handled by the IBM Storage Driver. The IBM Storage Driver communicates with the IBM Spectrum Accelerate Storage System and controls the storage volumes on it. With the release of Version 11.5 software, IBM Spectrum Accelerate introduced support for multi-tenancy. *Multi-tenancy* enables cloud providers to divide and isolate the IBM Spectrum Accelerate resources into logical domains, which can then be used by tenants without any knowledge of the rest of the system resources. For more information, see *Using XIV in OpenStack Environments*, REDP-4971.
 - IBM Storwize family/SAN Volume Controller: The volume management driver for the Storwize family and SAN Volume Controller provides OpenStack Compute instances with access to IBM Storwize family or SAN Volume Controller storage systems.
Storwize and SAN Volume Controller support fully transparent live storage migration in OpenStack Havana:
 - No interaction with the host is required: All advanced Storwize features are supported and exposed to the Cinder system.
 - Real-time Compression with EasyTier supports iSCSI + FC attachment.
- IBM FlashSystem (Kilo release): The volume driver for FlashSystem provides OpenStack Block Storage hosts with access to IBM FlashSystems.

- ▶ **IBM Spectrum Scale:** As of OpenStack Juno Release, Spectrum Scale combines the benefits of Spectrum Scale with the most widely used open source object store today, OpenStack Swift. Spectrum Scale provides enterprise ILM features. OpenStack Swift provides a robust object layer with an active community that is continuously adding innovative new features. To ensure compatibility with the Swift packages over time, no code changes are required to either Spectrum Scale or Swift to build the solution. For more information, see *A Deployment Guide for IBM Spectrum Scale Unified File and Object Storage*, REDP-5113.
- ▶ **IBM Spectrum Protect:** IBM data protection and data recovery solutions provide protection for virtual, physical, cloud, and software-defined infrastructures and core applications and remote facilities. These solutions fit nearly any size organization and recovery objective. They deliver the functions of IBM Spectrum Protect.

IBM Spectrum Protect enables software-defined storage environments by delivering automated data protection services at the control plane for file, block, and object backup.

IBM Spectrum Protect enables cloud data protection with OpenStack and VMware integration, cloud portal, and cloud deployment options.

For more information, see the following resources:

Protecting OpenStack with Tivoli Storage Manager for Virtual Environments:

<https://ibm.biz/BdXZmY>

Note: For more information about the IBM storage drivers and functions that are supported in the various OpenStack releases, see the following wiki:

<https://wiki.openstack.org/wiki/CinderSupportMatrix>



IBM Storage Systems for SDS

This chapter is a general overview of IBM Storage Systems that provides broad functions that can be combined to meet the customers' business requirements for their software-defined storage (SDS) environments. Some of the IBM storage systems come with SDS software that can make deployment even easier. Some fit within the SDS because of the functions that are provided to meet specific critical business requirements.

The requirements for a customer's SDS environment can include response times for hot and cold data. A FlashSystem V9000 or FlashSystem A9000/A9000R, with Spectrum Scale, Spectrum Archive, and an IBM TS4500 library might be the answer to the requirement for fast access for hot data. The FlashSystem is integrated with active archive requirements for cold data that can be stored on tape. The storage systems can be seamlessly integrated with Spectrum Scale and Spectrum Archive.

SDS is part of the software-defined infrastructure (SDI), and requirements might need certain types of storage to fulfill those needs. A successful implementation can be traced back to the appropriate planning needed to produce the architecture that meets the requirements.

IBM disk, hybrid disk, and flash memory systems provide storage efficiency solutions such as Real-Time Compression, automated tiering, virtualization, and thin provisioning. These storage solutions increase the data storage optimization opportunities for organizations of all sizes to boost system performance and lower IT costs.

This chapter includes the following sections:

- ▶ Integration with SDS
- ▶ IBM FlashSystem family
- ▶ IBM TS4500 and TS3500 tape libraries
- ▶ IBM DS8880
- ▶ IBM Storwize family
- ▶ IBM Elastic Storage Server
- ▶ VersaStack for Hybrid Cloud

4.1 Integration with SDS

Figure 4-1 shows the IBM SDS vision with all the storage HW/SW products it is built on and many possible configurations supporting current and future business applications.

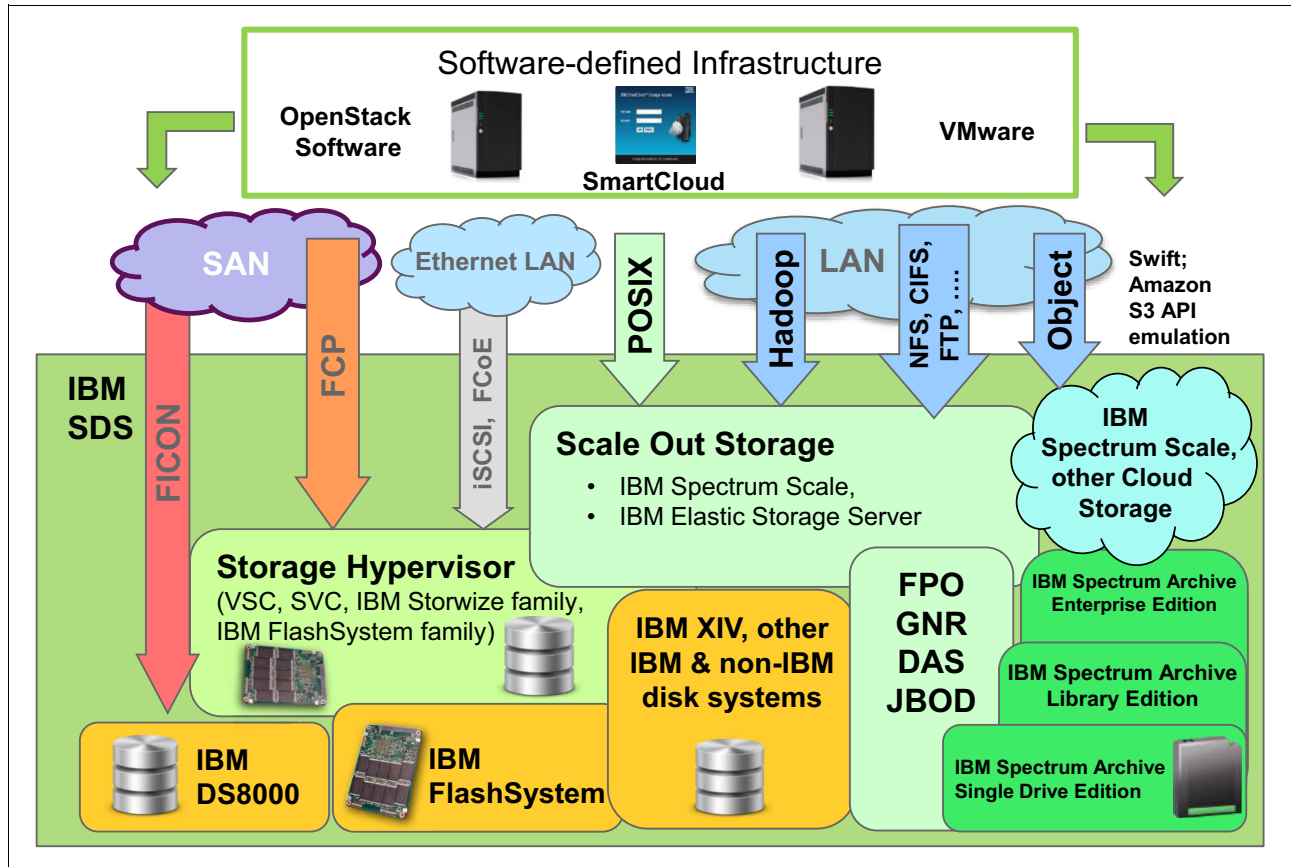


Figure 4-1 IBM SDS Vision with Storage Systems

SDI (some examples are shown at the top of Figure 4-1) can use the storage services provided by the SDS infrastructure as shown with integrated monitoring and control tools.

Storage capacity can be deployed and accessed in different ways according to the particular application and business requirements:

4.2 IBM FlashSystem family

For businesses to act based on insight from data and transform it into competitive advantage, the data-driven applications must operate at HA and provide sufficient performance. IBM FlashSystem delivers extreme performance to provide measurable economic value across the data architecture, including servers, software, applications, and storage. IBM offers a comprehensive flash portfolio with the IBM FlashSystem family, supporting SDS solutions as a concept, and flash-optimized XIV, Storwize V7000, and DS8000 storage.

The IBM FlashSystem family allows you to take advantage of best-in-breed solutions that provide extreme performance, macro efficiency, and microsecond response times. The IBM FlashSystem V9000 Enterprise Performance Solution, the IBM FlashSystem A9000, and the IBM FlashSystem 900 members of the FlashSystem family are described in this section. IBM FlashSystem A9000 is the newest addition to the FlashSystem family of storage systems. You can consider IBM FlashSystem as a major Tier for SDS.

4.2.1 FlashSystem benefits

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

IBM FlashSystem shared flash memory systems offer affordable, high-density, ultra low-latency, high reliability, and scalable performance in a storage device that is both space and power efficient.

IBM Flash products can either augment or replace traditional hard disk drive storage systems in enterprise environments. They empower applications to work faster and scale further. In addition to optimizing performance, the IBM FlashSystem family brings enterprise reliability and macro efficiency to the most demanding data centers, allowing businesses to receive the following benefits:

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

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

- ▶ **Extreme Performance:** Enable business to unleash the power of performance, scale, and insight to drive services and products to market faster.
- ▶ **MicroLatency:** Achieve competitive advantage through applications that enable faster decision making due to microsecond response times.
- ▶ **Macro Efficiency:** Decrease costs by getting more from efficient use of the IT staff, IT applications, and IT equipment due to the efficiencies that flash brings to the data center.
- ▶ **Enterprise Reliability:** Durable and reliable designs that use enterprise class flash and patented data protection technology.

4.2.2 IBM FlashCore technology

IBM FlashCore® technology refers to the IBM innovations that enable FlashSystem storage to deliver extreme performance, IBM MicroLatency, enterprise-grade reliability, and a wide range of operational and cost efficiencies.

These technologies and innovations are represented in the FlashCore hardware-accelerated architecture and IBM MicroLatency modules. They are also in other advanced flash management features and capabilities that are used in the IBM FlashSystem 900, A9000, and V9000:

- ▶ **Hardware Accelerated Architecture:** By using an all-hardware data path, FlashSystem arrays minimize the amount of software interaction during I/O activity, resulting in the highest performance and lowest latency for all-flash storage arrays.
- ▶ **IBM MicroLatency Modules:** By using IBM-designed, purpose-engineered flash memory modules, FlashSystem delivers extreme performance, greater density, unlimited scalability, and mission-critical reliability.
- ▶ **Advanced Flash Management:** Unique, patented IBM hardware and software innovations enable FlashSystem to provide the most reliable, feature-rich, and highly available flash data storage.

4.2.3 FlashSystem 900

Flash memory gives organizations the ability to deliver fast, reliable, and consistent access to critical data. With IBM FlashSystem 900, you can make faster decisions based on real-time insights and unleash the power of the most demanding applications. These applications include online transaction processing and analytics databases, virtual desktop infrastructures, technical computing applications, and cloud environments.

FlashSystem 900 can also lower operating costs and increase the efficiency of IT infrastructure by using much less power and space than traditional hard disk drive (HDD) and SSD solutions. Here are some reasons to consider FlashSystem 900 for implementation:

- ▶ **When speed is critical:** IBM FlashSystem 900 is designed to accelerate the applications that drive business. Powered by IBM FlashCore Technology, FlashSystem 900 delivers high performance at lower cost:
 - 90 μ /155 μ read/write latency.
 - Up to 1.1 million random read 4,000 IOPS.
 - Up to 10 GBps read bandwidth.
- ▶ **High capacity business needs:** IBM FlashSystem 900 has 12 hot-swappable IBM MicroLatency storage modules: 1.2 TB, 2.9 TB, and 5.7 TB.
- ▶ **Provides higher density:** IBM FlashSystem 900 employs 20 nm multi-level cell (MLC) chips with IBM-enhanced Micron MLC technology for higher storage density and improved endurance.
- ▶ **Highly scalable:** FlashSystem 900 is configurable with 2.4 - 57 TB of capacity for increased flexibility.
- ▶ **Easy to integrate into VMware environments:** FlashSystem 900 is easy to integrate with VMware VASA by using Spectrum Connect to use the following features:
 - Greater communication between vSphere and FlashSystem.
 - Ability of vSphere to monitor and directly manage FlashSystem, allowing greater efficiency.
 - Integration of VASA Unmap for greater storage efficiency.

Through SAN Volume Controller, the FlashSystem 900 includes support for OpenStack cloud environments.

For more information, see the following IBM publications:

- ▶ *FlashSystem 900 Product Guide*, TIPS1261
- ▶ *Implementing IBM FlashSystem 900*, SG24-8271

4.2.4 IBM FlashSystem V9000

IBM FlashSystem V9000 is a comprehensive all-flash enterprise storage solution. FlashSystem V9000 delivers the full capabilities of IBM FlashCore technology plus a rich set of storage virtualization features.

FlashSystem V9000 offers the advantages of software-defined storage at the speed of flash. These all-flash storage systems deliver the full capabilities of the hardware-accelerated I/O provided by FlashCore Technology. FlashSystem V9000 also delivers the enterprise reliability of MicroLatency modules and advanced flash management. These features are coupled with a rich set of the features that are found in the most advanced software-defined storage solutions. These features include Real-time Compression, dynamic tiering, thin provisioning, snapshots, cloning, replication, data copy services, and high-availability configurations.

V9000 now also supports physical expansions that receive SSD drives and NearLine SAS drives that offer three tiers of storage.

FlashCore technology plus a rich set of storage virtualization features allow FlashSystem V9000 to deliver industry-leading value to enterprises in scalable performance, enduring economics, and agile integration:

- ▶ **Fast:** Optimize your infrastructure with scale-up, scale-out capabilities of fast FlashSystem performance.
- ▶ **Cost-effective:** Powerful virtualized storage enables you to realize immediate and long-term economic benefits.
- ▶ **Easy:** Unlike conventional storage, FlashSystem is easy to deploy, can virtualize legacy systems, and delivers value in hours.

The FlashSystem V9000 has connectivity to OpenStack cloud environments through the Cinder driver.

For more information, see the following IBM publications:

- ▶ *IBM FlashSystem V9000 Product Guide*, TIPS1281
- ▶ *IBM FlashSystem V9000 and VMware Best Practices Guide*, REDP-5247

4.2.5 IBM FlashSystem A9000/A9000R

IBM FlashSystem A9000 is a new comprehensive all-flash enterprise storage solution. It delivers the full capabilities of IBM FlashCore technology combined with advanced reduction mechanism and all the features of the IBM Spectrum Accelerate software stack.

As a cloud optimized solution, IBM FlashSystem A9000 suits the requirements of public and private cloud providers who require features, such as inline data deduplication, multi-tenancy, and quality of service.

It also uses powerful software-defined storage capabilities from IBM Spectrum Accelerate, such as Hyper-Scale technology and VMware integration:

- ▶ An enhanced management interface simplifies storage administration
- ▶ Data reduction: Pattern removal, data deduplication, and compression
- ▶ VMware vStorage API for Array Integration (VAAI)
- ▶ Multi-tenancy
- ▶ Host Rate Limiting: QoS
- ▶ Fibre Channel and iSCSI support
- ▶ Snapshots
- ▶ Synchronous and asynchronous remote mirroring
- ▶ Data Migration
- ▶ Hyper-Scale Mobility
- ▶ Encryption
- ▶ Authentication by using Lightweight Directory Access Protocol (LDAP)
- ▶ OpenStack and REST support
- ▶ VMware synergy

IBM FlashSystem A9000 is a fixed storage solution to provide up to 300 TB of effective capacity by using 8U of rack space.

IBM FlashSystem A9000R is a scalable storage solution into a single rack from 300 TB to 1800 TB of effective capacity.

FlashSystem A9000 and FlashSystem A9000R use the same firmware, and both offer onsite setup and service that are provided by IBM. They also share a feature set.

For more information, see the following publications:

- ▶ *IBM FlashSystem A9000 Product Guide*, REDP-5325
- ▶ *IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation*, SG24-8345

4.3 IBM TS4500 and TS3500 tape libraries

IBM Spectrum Archive Enterprise Edition, along with IBM Spectrum Scale, allows you to connect tape libraries to the cloud. Tape has always been a cost effective solution. With the growth in data controlled by cloud environments, TS4500 and TS3500 can be used as an efficient tape tier.

How a TS4500 or TS3500 tape library can be configured as the tape tier in the storage cloud through IBM Spectrum Scale and IBM Spectrum Archive Enterprise Edition is shown in Figure 4-2.

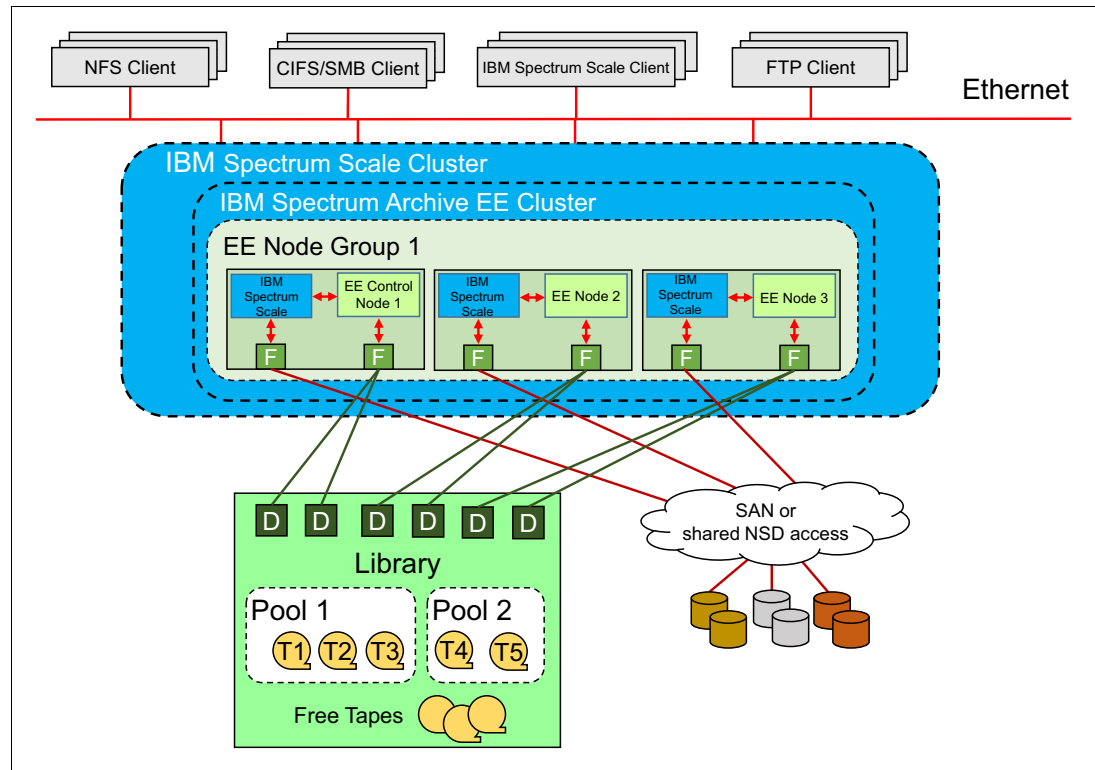


Figure 4-2 TS4500/TS3500 tape library tape tier configuration for cold storage

4.3.1 TS4500 tape library

The IBM TS4500 tape library is a next-generation storage solution that is designed to help midsize and large enterprises respond to storage challenges. Among these challenges are high data volumes and the growth in data centers. These factors in turn increase the cost of data center storage footprints, the difficulty of migrating data across vendor platforms, and increased complexity of IT training and management as staff resources shrink.

In the TS4500, IBM delivers the density that today's and tomorrow's data growth requires, along with the cost efficiency and the manageability to grow with business data needs while preserving existing investments in IBM tape library products.

You can now achieve both a low cost per terabyte (TB) and a high TB density per square foot. The TS4500 can store up to 5.5 PBs of uncompressed data in a single 10-square foot library frame and up to 175.5PBs of uncompressed data in a 17 frame library.

The TS4500 tape library includes the following highlights:

- ▶ Improve storage density with more than two times the expansion frame capacity and support for 33 percent more tape drives
- ▶ Proactively monitor archived data with policy-based automatic media verification
- ▶ Improve business continuity and disaster recovery with automatic control path and data path failover

- ▶ Help ensure security and regulatory compliance with tape-drive encryption and Write Once Read Many (WORM) media
- ▶ Support Linear Tape-Open (LTO) Ultrium 8, LTO Ultrium 7, LTO Ultrium 6, LTO Ultrium 5, and IBM TS1155, TS1150, and TS1140 tape drives
- ▶ Increase mount performance and overall system availability with dual robotic accessors
- ▶ Provide a flexible upgrade path for users who want to expand their tape storage as their needs grow
- ▶ Reduce the storage footprint and simplify cabling with 10U of rack space on top of the library

For more information, see the *IBM TS4500 R4 Tape Library Guide*, SG24-8235.

4.3.2 TS3500 tape library

TS3500 continues to lead the industry in tape drive integration with features such as persistent worldwide name, multipath architecture, drive/media exception reporting, remote drive/media management, and host-based path failover.

The IBM TS3500 tape library is designed to provide a highly scalable, automated tape library for mainframe and open-systems backup and archive. The library can scale from midsize to large enterprise environments. Here are the highlights of the tape library:

- ▶ Support highly scalable, automated data retention on tape by using LTO Ultrium and IBM 3592 tape drive families
- ▶ Deliver extreme scalability and capacity, growing from one to 16 frames per library and from one to 15 libraries per library complex
- ▶ Provide up to 2.25 exabytes (EB) of automated, low-cost storage under a single library image, improving floor space utilization and reducing storage cost per TB with IBM 3592 JD enterprise advanced data cartridges (10 TB native capacity)

Massive scalability (300,000+ LTO tape cartridges) can be achieved with the TS3500 Shuttle Complex, as shown in Figure 4-3.



Figure 4-3 TS3500 Shuttle Complex moves tapes cartridges between physical libraries

For more information, see *IBM Tape Library Guide for Open Systems*, SG24-5946.

4.4 IBM DS8880

The IBM DS8000 series is the flagship block storage system within the IBM System Storage portfolio.

The IBM DS8880 family offers business-critical, all-flash, and hybrid data systems that span a wide range of price points. Consider the following points:

- ▶ The IBM DS8884F, DS8886F, and DS8888F are all-flash offerings. The DS8880 all flash includes up to 2.9PB of raw Flash capacity in a three-rack footprint.
- ▶ The IBM DS8884 and IBM DS8886 are two high-performance hybrid models that scale to more than 5.2 petabytes (PB) of raw drive capacity. A total of 14 types of media can be managed in up to three different tiers (Flash cards and flash drives, SAS, and Nearline SAS drives).
- ▶ IBM enhanced the OpenStack Cinder driver with DS8880 support.
- ▶ Integration of storage systems requires an OpenStack Block Storage driver on the OpenStack Cinder nodes.

How Horizon (Dashboard) and Nova (Compute) interact with Cinder over the Ethernet control-path is shown in Figure 4-4. Also shown is the data path from Nova to the DS8880 storage system.

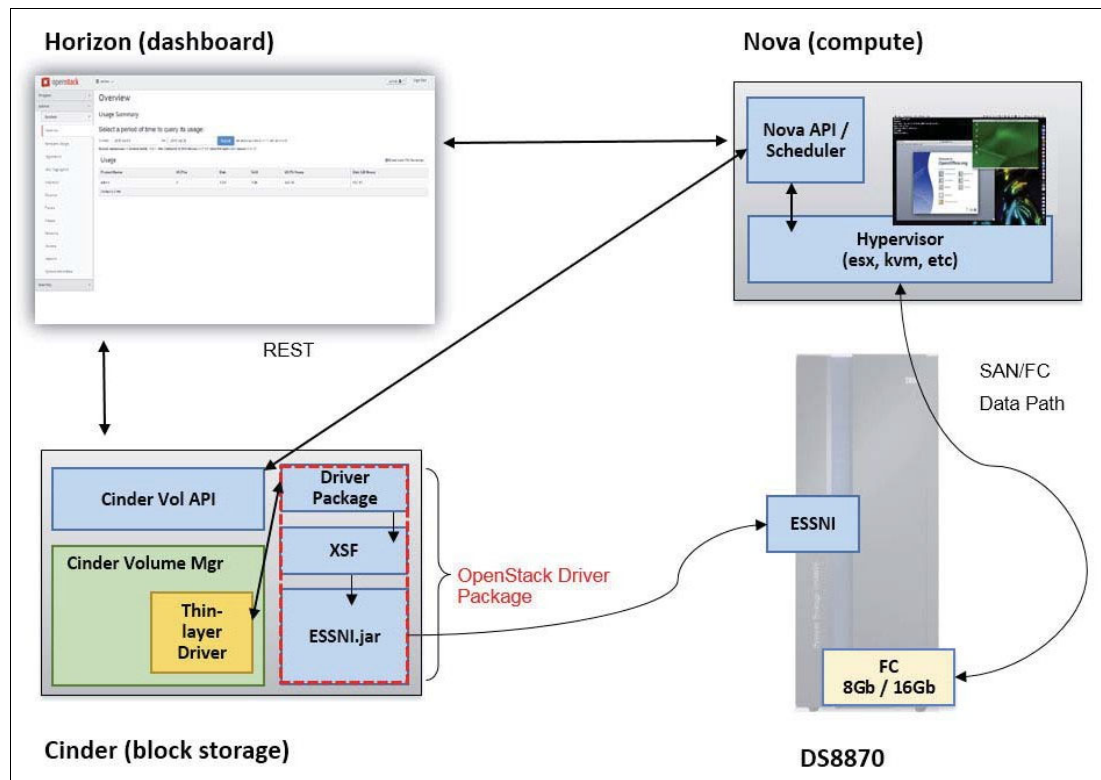


Figure 4-4 OpenStack Cinder driver support for DS8880

With the availability of the IBM Storage Driver for the OpenStack Cinder component, the IBM DS8880 storage system can now extend its benefits to the OpenStack cloud environment.

The IBM Storage Driver for OpenStack Cinder enables OpenStack clouds to access the DS8880 storage system. The IBM Storage Driver for OpenStack is fully supported by Cinder and provides “block storage as a service” through Fibre Channel to VMs.

Cloud users can send requests for storage volumes from the OpenStack cloud. These requests are routed to, and transparently handled by, the IBM Storage Driver. The IBM Storage Driver communicates with the DS8880 storage system and controls the storage volumes on it.

The last version of Cinder Driver 2.1.0 provides the following capabilities:

- ▶ Create/Delete Volume
- ▶ Volume Attach/Detach (by way of Nova Compute)
- ▶ Snapshots, Clones (FlashCopy with background copy)
- ▶ Backups (Copy Volume Images to Object Store)
- ▶ Swift, Ceph and TSM Support
- ▶ Volume Types, Volume Retype
- ▶ Support volume Quality of Service (QoS)
- ▶ Quotas
- ▶ Consistency Groups for FlashCopy and replication
- ▶ Volume Retype: Ability to change the type of a Cinder volume to a new tier, add capabilities, and so on
- ▶ Volume Replication: Ability to do synchronous and asynchronous replication of cinder volumes between two subsystems

For more information about the use of the DS8000 in a OpenStack environment, see *Using IBM DS8000 in an OpenStack Environment*, REDP-5220.

IBM also provides Storage integration between VMware and DS8000 (see Figure 4-5 on page 103). DS8880 supports vStorage API for Array Integration (VAAI) and includes the following features:

- ▶ Integration with vStorage APIs to improve performance.
- ▶ Full copy (also known as XCOPY) primitive offloads work from production virtual servers to storage, which helps reduce host and SAN resource utilization.
- ▶ Hardware-assisted locking (by using Atomic Test & Set) primitive enables a finer-grained level of locking (block-level instead of LUN-level) on VMware Virtual Machine File System (VMFS) metadata, which is more efficient and also scales better in larger VMware clusters.
- ▶ Write Same (zero blocks) primitive allows the process of zeroing the VMDK to be offloaded to the storage subsystem.
- ▶ Unmap (Block Delete) allows the host to request that a range of blocks on a thin-provisioned volume be unmapped, which allows the underlying space to be released.

DS8880 and VMware Integration

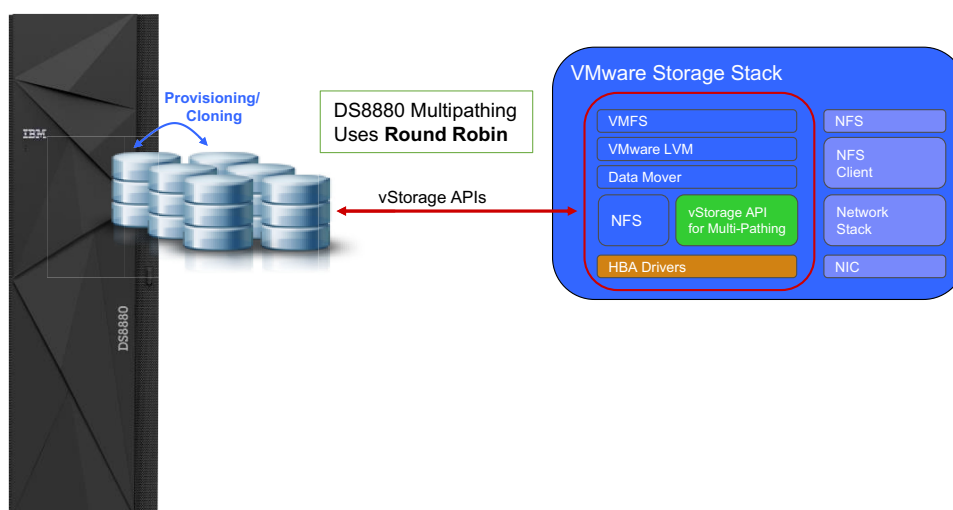


Figure 4-5 Storage integration between VMware and DS8000

The tasks that can benefit from improved performance include the following examples:

- ▶ VM creation, cloning, snapshots, and deletion
- ▶ vMotion and storage vMotion
- ▶ Extending a VMFS Volume
- ▶ Extending the size of a VMDK file

DS8880 also supports the following components:

- ▶ vCenter plug-in
- ▶ VMware vCenter Site Recovery Manager (SRM)
- ▶ VMware Web Client

For more information, see the *IBM DS8880 Product Guide (Release 8.3)*, REDP-5344.

Transparent cloud tiering

Transparent cloud tiering (TCT) provides a native cloud storage tier for IBM Z environments. TCT moves data directly from the DS8880 to cloud object storage, without sending data through the host. TCT provides cloud object storage (public, private, or on-premises) as a secure, reliable, transparent storage tier that is natively integrated with the DS8880.

TCT on the DS8880 is fully integrated with DFSMSHsm, which reduces CPU utilization on the host when you are migrating and recalling data in cloud storage. DFSMSdss can also be used to store data that is generated by using the **DUMP** command.

The DS8880 supports the OpenStack Swift and Amazon S3 APIs. The DS8880 also supports the IBM TS7700 as an object storage target and the following cloud service providers:

- ▶ Amazon S3
- ▶ IBM Cloud
- ▶ OpenStack Swift Based Private Cloud

4.5 IBM Storwize family

Designed for software-defined environments, the IBM Storwize family includes technologies that complement and enhance virtual environments, and built-in functions, such as Real-time Compression and Easy Tier technology that deliver extraordinary levels of efficiency. Available in a wide range of storage systems, the Storwize family delivers sophisticated capabilities that are easy to deploy, and help to control costs for growing businesses.

The IBM Storwize family consists of the IBM SAN Volume Controller, IBM Storwize V7000, IBM Storwize V5000, and the all flash memory systems including V7000F and V5000F. Benefits of the Storwize family include high-performance thin provisioning, Real-time Compression, IP replication, Easy Tier, Tiering to the Cloud encryption, QOS, an new advanced GUI, and storage virtualization.

Data protection is provided with the standard and proven set of features that include FlashCopy, Metro and Global Mirror, and volume mirroring.

The Storwize family uses IBM Spectrum Virtualize software the same proven software as SAN Volume Controller, and provides the same interface and similar capabilities across the product line (see Figure 4-6).

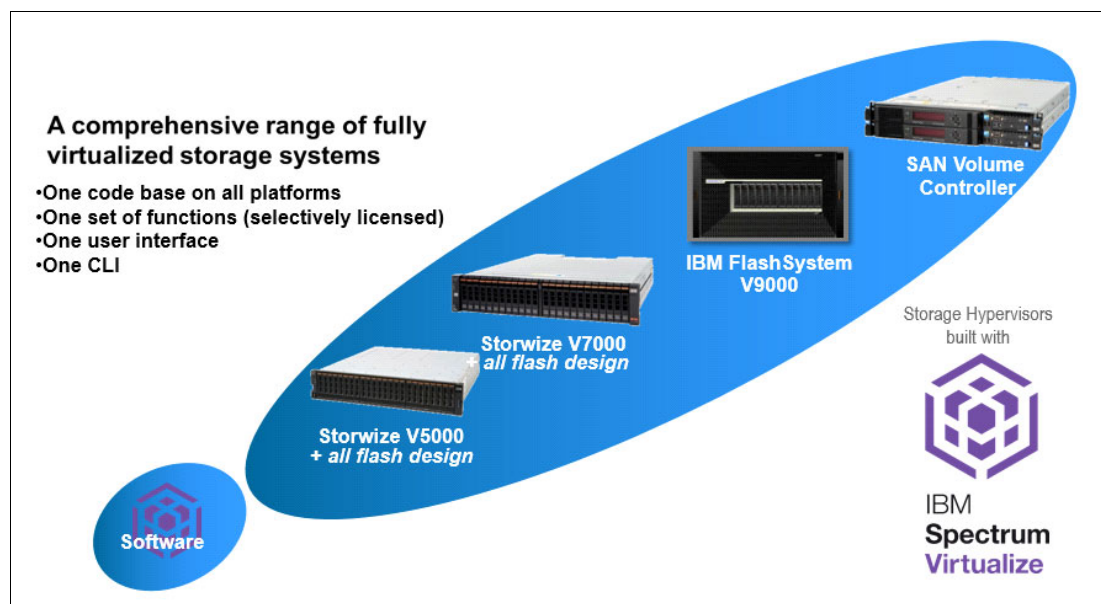


Figure 4-6 IBM's Storage Hypervisors powered by IBM Spectrum Virtualize

4.5.1 Storwize V5000 Gen2 and V5000F Gen2

The Storwize V5000 Gen2 and V5000F Gen2 entry systems provide Hybrid (V5000) and All-Flash (V5000F) block storage capability. They deliver efficient, entry-level configurations that are designed to meet the needs of small and midsize businesses.

Designed to provide organizations with the ability to consolidate and share data at an affordable price, the Storwize V5000 Gen2 offers advanced software capabilities that are found in more expensive systems.

Storwize V5030 system can scale up to two systems clusters that support 1,520 drives (see Figure 4-7).



Figure 4-7 IBM Storwize V5000 family

For more information, see *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030) with IBM Spectrum Virtualize V8.1, SG24-8162*, which is available at the following website:

<http://www.redbooks.ibm.com/abstracts/sg248162.html>

4.5.2 Storwize V7000 Gen2+ and V7000F Gen2+

These midrange systems provide Hybrid (V7000) and All-Flash (V7000F) Block storage capability. They are highly scalable virtualized, enterprise-class, flash-optimized storage systems that are designed to consolidate workloads into a single system for ease of management, reduced costs, superior performance, and HA. Storwize V7000 systems can scale up to four systems cluster supporting 3,040 drives.

IBM Storwize V7000 solution incorporates IBM Spectrum Virtualize software and provides a modular storage system that includes the capability to virtualize its internal and external SAN-attached storage. IBM Storwize V7000 solution is built on IBM Spectrum Virtualize.

The IBM Storwize V7000 is a modular, midrange virtualization RAID storage subsystem that employs the IBM Spectrum Virtualize software engine. It has the following benefits:

- ▶ Brings enterprise technology to midrange storage
- ▶ Specialty administrators are not required
- ▶ Client setup and service are simplified
- ▶ The system can grow incrementally as storage capacity and performance needs change
- ▶ Multiple storage tiers are in a single system with nondisruptive migration between them
- ▶ Simple integration can be done into the server environment

The IBM Storwize V7000 (see Figure 4-8) consists of a set of drive enclosures. Control enclosures contain disk drives and two nodes (an I/O group), which are attached to the SAN fabric or 10 gigabit Ethernet (GbE) fast Ethernet. Expansion enclosures contain drives and are attached to control enclosures.



Figure 4-8 IBM Storwize V7000

For more information, see *Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.1*, SG24-7938, which is available at this website:

<http://www.redbooks.ibm.com/abstracts/sg247938.html>

4.5.3 IBM SAN Volume Controller as an SDS appliance

IBM SAN Volume Controller is a virtualization appliance solution, which implements the IBM Spectrum Virtualize software. It maps virtualized volumes that are visible to hosts and applications to physical volumes on storage devices.

Each server within the storage area network (SAN) features its own set of virtual storage addresses that are mapped to physical addresses. If the physical addresses change, the server continues to run by using the same virtual addresses that it used previously. Therefore, volumes or storage can be added or moved while the server is still running.

The IBM virtualization technology improves the management of information at the “block” level in a network, which enables applications and servers to share storage devices on a network.

IBM Spectrum Virtualize software in SAN Volume Controller helps make new and existing storage more effective. SAN Volume Controller includes many functions that are traditionally deployed separately in disk systems. By including these functions in a virtualization system, SAN Volume Controller standardizes functions across virtualized storage for greater flexibility and potentially lower costs.

The total storage capacity manageable per system is 32 PetaBytes.

IBM SAN Volume Controller highlights

Improving efficiency and delivering a flexible, responsive IT infrastructure are essential requirements for any cloud deployment. Key technologies for delivering this infrastructure include virtualization, consolidation, and automation. SAN Volume Controller provides these technologies to help you build your storage cloud.

SAN Volume Controller's enhanced storage capabilities with sophisticated virtualization, management, and functions have these benefits:

- ▶ Enhance storage functions, economics, and flexibility with sophisticated virtualization
- ▶ Employ hardware-accelerated data compression for efficiency and performance
- ▶ Use encryption to help improve security for data on existing storage systems
- ▶ Move data among virtualized storage systems without disruptions
- ▶ Optimize tiered storage, including flash storage, automatically with IBM Easy Tier
- ▶ Improve network utilization for remote mirroring and help reduce costs
- ▶ Implement multi-site configurations for HA and data mobility

An overview of the IBM SAN Volume Controller architecture is shown in Figure 4-9.

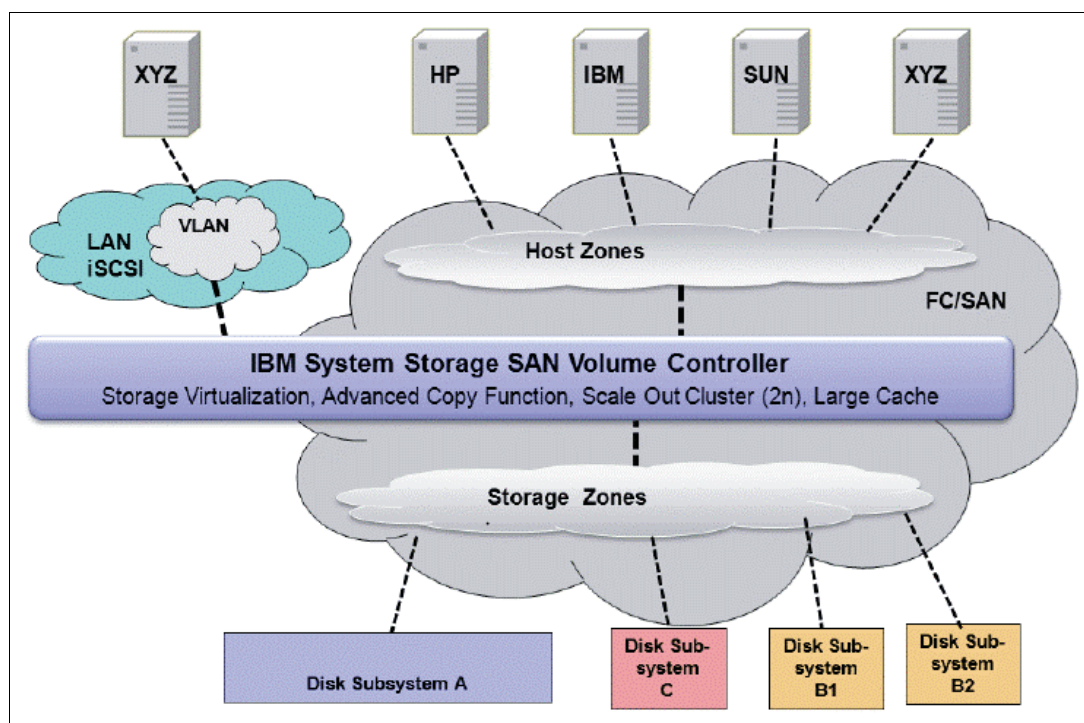


Figure 4-9 SAN Volume Controller overview

4.5.4 Ongoing IBM contributions to OpenStack Cinder for the Storwize family

The Storwize family, which includes SAN Volume Controller, supports these items:

- ▶ Folsom, Grizzly, Havana, Icehouse, Juno
- ▶ iSCSI and Fibre Channel
- ▶ Advanced Storwize features such as Real-time Compression and Easy Tier
- ▶ Software-defined placement by using OpenStack filter scheduler
- ▶ Storage-assisted volume migration (Storwize family is the only storage to support this function in Havana)

With the OpenStack Havana release, a new administrator feature for migrating volumes between Cinder instances was added. Volumes can be migrated with Host Assisted Data Migration or by Storage Assisted Data Migration with the IBM Storwize family.

The common use cases for migrating volumes are shown in Figure 4-10.

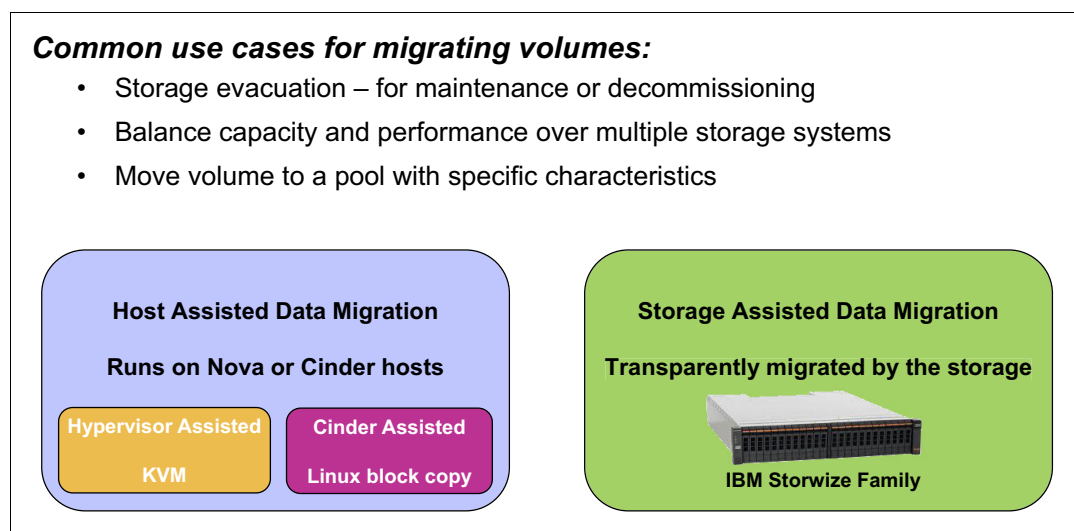


Figure 4-10 Common use cases for volume migration in OpenStack environment

IBM Storwize family is the only storage in the Havana release to support storage assisted migration. Volumes move between two storage pools that are managed by a Storwize family system.

The use of the Storwize family storage assisted migration results in the following key benefits:

- ▶ No interaction with the host
- ▶ No impact on VM and node
- ▶ Instantaneous
- ▶ No effect on VM operations or volume management

4.6 IBM Elastic Storage Server

The IBM Elastic Storage Server (ESS) is a modern implementation of software-defined storage, combining IBM Spectrum Scale software with IBM servers and disk arrays (see Figure 4-11).



Figure 4-11 IBM Elastic Storage Server

This technology combines the CPU and I/O capability of the IBM POWER8® architecture and matches it with 2U and 4U storage enclosures. This architecture allows the IBM Spectrum Scale Native RAID software capability to actively manage all RAID functionality that is formerly accomplished by a hardware disk controller. RAID rebuild times are reduced to a fraction of the time that is needed with hardware RAID.

The IBM building block solution for high-performance storage allows you to accomplish the following tasks:

- ▶ Deploy petascale class high-speed storage quickly with pre-assembled and optimized servers, storage, and software.
- ▶ Host multiple tenants, adjust resource allocation, and scale as your needs evolve.
- ▶ Experience higher performance and scalability with lower cost.
- ▶ Rebuild failed disks faster with IBM developed de-clustered RAID technology.

IBM has implemented Elastic Storage Server configurations for various workloads, from high-velocity import through high-density cloud storage usage models, deploying the latest SSD, serial-attached SCSI (SAS), and Nearline SAS drives. For performance-oriented workloads, the system's affordable building block configurations start at 24 or 48 drives.

For high-capacity storage, IBM offers configurations that can support almost 2 petabytes of usable, deployable storage in a single industry-standard 42U rack. For mixed workloads, the server supports varied configurations of building blocks, with placement rules for the creation and management of all data on the appropriate storage tier.

For more information, see the following websites:

- ▶ <http://www.ibm.com/systems/storage/spectrum/ess>
- ▶ https://www.ibm.com/support/knowledgecenter/SSYSP8_5.2.0/sts52_welcome.html

Newly developed RAID techniques from IBM use this CPU and I/O power to help overcome the limitations of current disk drive technology.

Elastic Storage Server is a building block that provides the following benefits:

- ▶ Deploy petascale class high-speed storage quickly with pre-assembled and optimized servers, storage, and software
- ▶ Host multiple tenants, adjust resource allocation, and scale as your needs evolve
- ▶ Experience higher performance and scalability with lower cost
- ▶ Achieve superior sustained streaming performance
- ▶ Rebuild failed disks faster with IBM developed de-clustered RAID technology

For more information, see the Elastic Storage Server in IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/POWER8/p8ehc/p8ehc_storage_landing.htm

4.7 VersaStack for Hybrid Cloud

Converged infrastructures allow data center administrators to minimize compatibility issues between their devices. They group compute, network, and storage components into a single IT management entity.

IBM's long history and leadership of IBM in storage, combined with experience and expertise of Cisco in data center computing, networking, mobility, collaboration, and analytics technologies made the IBM and Cisco partnership a natural move for building best-of-breed converged infrastructures.

VersaStack is an IBM and Cisco converged infrastructure solution, which combines computing, networking, and storage into a single integrated and managed system. It is built over Cisco Unified Computing System (Cisco UCS) Integrated Infrastructure, including Cisco UCS and Cisco UCS Director management interface, combined with IBM storage that is based on IBM Spectrum Virtualize and IBM Spectrum Accelerate technologies.

VersaStack offers quick deployment and rapid time to value for the implementation of modern infrastructures. It enables enterprises to easily and cost-effectively scale compute, network and storage capacity as needed. This solution allows data center administrators to be provided an Infrastructure-as-a-service private cloud solution.

VersaStack is backed by the so-called "Cisco Validated Design" (CVD) program, which is a Cisco program delivering VersaStack as a tested, validated and documented solution, almost as a single product provides.

The components of VersaStack are shown in Figure 4-12.

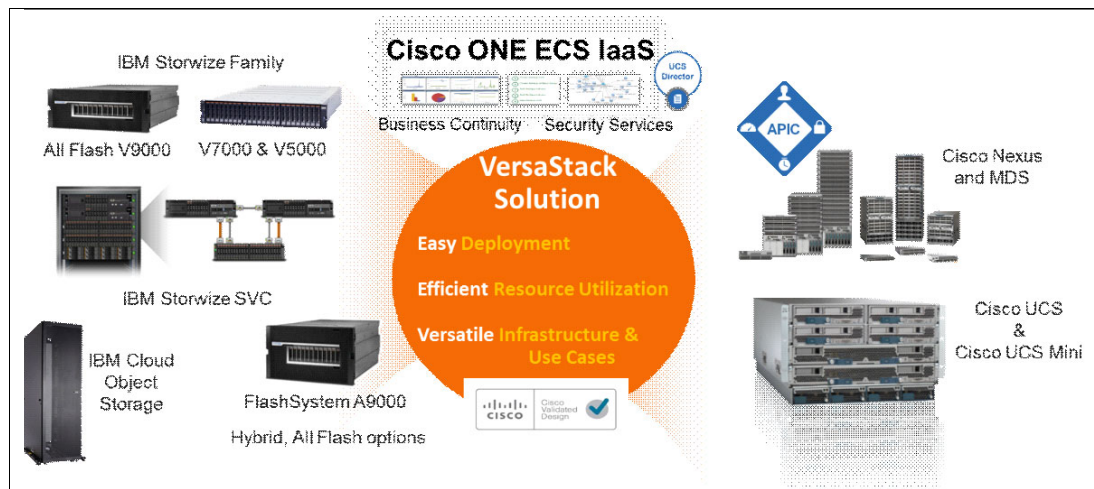


Figure 4-12 VersaStack Components

For more information about the VersaStack CVDs, see this website:

<https://ibm.biz/BdZhdB>

For more information about VersaStack, see this website:

<http://www.versastack.com>

The latest VersaStack for Hybrid Cloud CVD, which combines VersaStack CVDs with IBM Spectrum Copy Data Management and Cisco ONE Enterprise Cloud Suite, allows VersaStack to provide a hybrid cloud solution. In this solution, any workload can be orchestrated, deployed, and managed across data centers, public cloud, and private cloud environments.

IBM Spectrum Copy Data Management allows for the creation, distribution, and lifecycle management of application aware data copies. Cisco ONE Enterprise Cloud Suite delivers Data center automation, cloud management, self-service ordering, and assured application performance over a hybrid cloud.

An overview of VersaStack for Hybrid Cloud capabilities and components is shown in Figure 4-13.

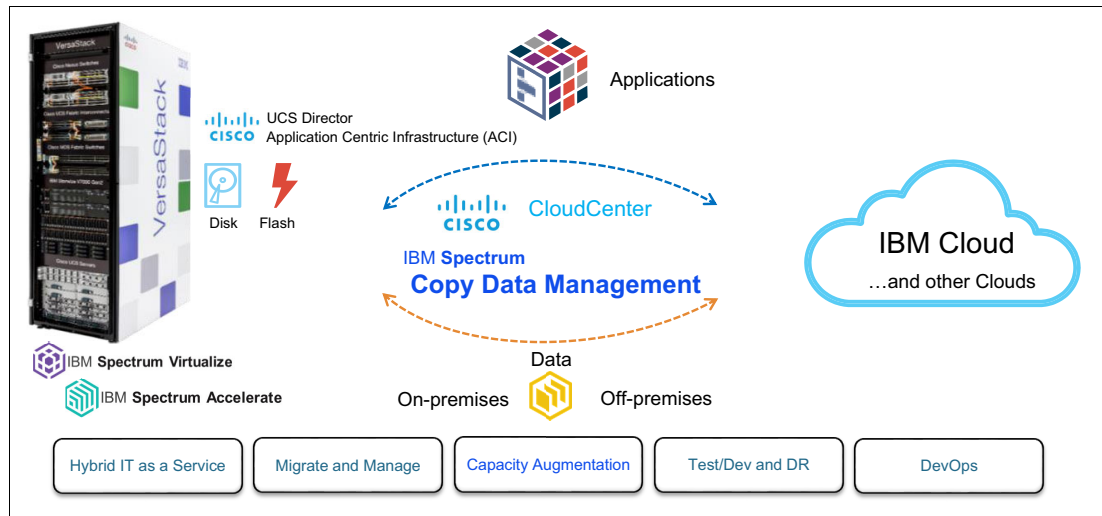


Figure 4-13 VersaStack for Hybrid Cloud in a nutshell

VersaStack for Hybrid Cloud can then apply to many use cases, such as Hybrid IT as a Service, DevOps, Resource Optimization, and Workload Life Cycle Management.



Use cases

This chapter describes software-defined storage use cases across various industries. Each description covers the clients' needs, a proposed solution, and results.

This chapter includes the following sections:

- ▶ 5.1, "Secure file and object storage for backups" on page 114
- ▶ 5.2, "Hybrid Cloud telecommunications storage optimization project" on page 116
- ▶ 5.3, "Government sector transformation to private cloud" on page 118
- ▶ 5.4, "Cloud service provider use case with SAP HANA" on page 119
- ▶ 5.5, "Media and entertainment company Hybrid Cloud" on page 121
- ▶ 5.6, "Life sciences and genomics" on page 124
- ▶ 5.7, "Oil and gas solutions" on page 126
- ▶ 5.8, "Automobile analytics with Hortonworks" on page 127
- ▶ 5.9, "Ensuring an insurance company's SAS Grid performance" on page 129

5.1 Secure file and object storage for backups

An important finance institute engaged IBM to help them standardize their existing backup environment based on different software and hardware technology that included protection against ransomware.

Business objective

The finance company's business goal was to find a cheaper and faster way to implement a backup solution based on unique software for their whole application environment, and to find an infrastructure able to combine backup and disaster recovery capability. IBM was faced with the following requests:

- ▶ Backup software able to provide a unique interface across different server platforms
- ▶ Provide a disaster recovery solution across different sites
- ▶ Provide a central point of management
- ▶ Implement backup storage pool tiers based on both short-term and long-term. In this case, a *short-term* backup has a retention time of no more than one year, and the *long-term* backups (when specifically requested) will be retained for 7 years:
 - Short-term backups will exist on premise at each site
 - Long-term backups will be pushed to a cloud storage environment
 - Tracking and traceability capability of backup data are to be subject to policies defined by the EU's General Data Protection Regulation (GDPR)

Proposed solution

IBM selected IBM Spectrum Protect was the solution chosen to address all of the customer requests to have a unique software solution to manage the backup of all their different server environment.

The Elastic Storage Server (ESS), which runs IBM Spectrum Scale, was selected to host short-term storage pools and enable quick restores for backups. The long-term storage pools were created on an IBM Cloud Object Storage solution that supports dispersed file system at three different data centers.

This solution helped the customer achieve cost savings in addition to reaching their goals for high availability and for the disaster recovery solution for the entire backup environment. The node-to-node replication of IBM Spectrum Protect and the IBM Spectrum Scale replication completed the solution requirements.

Figure 5-1 shows the proposed and implemented solution.

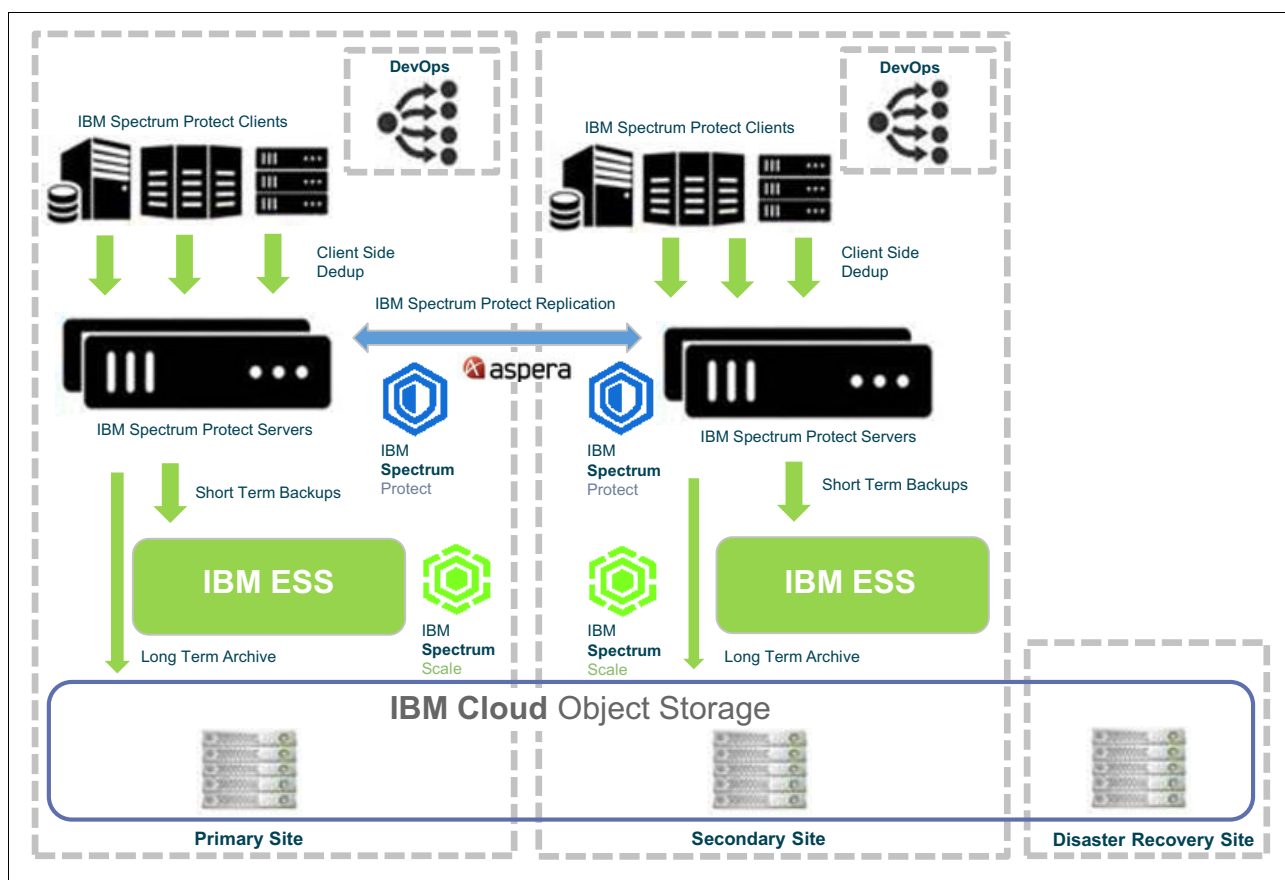


Figure 5-1 Proposed and implemented solution

Benefits of the solution

By implementing the solution, IBM provided the following benefits to the customer:

- ▶ Improved service levels and reduced errors and costs
- ▶ Turned the backup storage environment into a software-defined, cloud-ready solution
- ▶ Improved visibility, control, and automation for the backup environment
- ▶ Improved capacity utilization using IBM Spectrum Protect data deduplication
- ▶ Introduced the use of IBM COS dispersed file system, reducing the necessity of duplicate full capacity in all the sites
- ▶ Introduced high availability for backup servers thanks to node-to-node replication
 - Introduced an air-gapped solution with IBM Spectrum Protect talking to IBM COS
 - Introduced the use of IBM ESS for short-term backup storage pools that allow quick restore capabilities.
- ▶ Allowed application-aware backups and restores with IBM Spectrum Protect Snapshot
- ▶ Reduced the risk of data loss
- ▶ Enabled near-instant restores at any time

5.2 Hybrid Cloud telecommunications storage optimization project

A large telecommunications company engaged IBM to help them consolidate their storage environment and solve IT challenges, which were crucial for further business expansion.

Business objective

The large telecommunications company's business goal was to instantly respond to new market challenges and to expand their market share, which is supported by adaptive, simplified, and responsive IT infrastructure. IBM was faced with the following requests:

- ▶ Provide storage consolidation and optimization, and a unique storage access interface across the enterprise
- ▶ Provide a high availability solution across campus
- ▶ Provide a central point of management
- ▶ Implement storage tiers that are based on performance
- ▶ Non-disruptively migrate online data between storage tiers
- ▶ Provide a central point of monitoring across the enterprise
- ▶ Provide Snapshot functions as a first line of defense for critical applications
- ▶ Offload backup jobs from production LUNs

Their storage environment was four years old and based on a non-IBM solution. The challenge was to design a storage environment that is based on customer requests to support application load and future application requests, and size a robust solution to meet the customer's expectation.

Regarding capacity, they requested 1.5 PB of usable capacity for the first phase, with expansion up to 4.0 PB in the next four years. During these four years, IBM must provide the same service level agreements (SLAs) for all storage tiers, and the solution must be scalable to accept new workloads. The storage environment before the redesign is shown in Figure 5-2 on page 116.

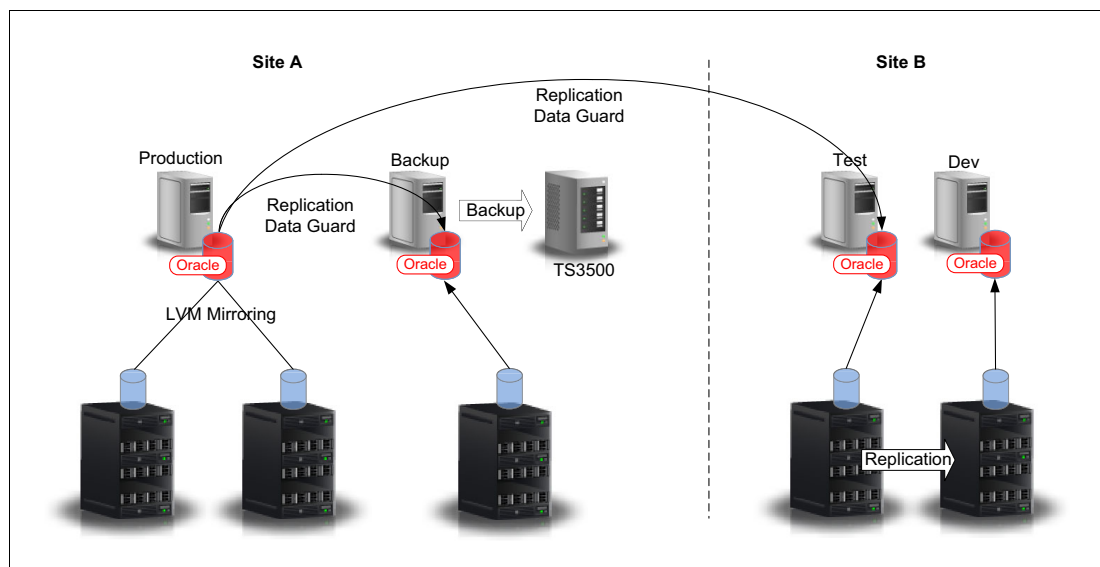


Figure 5-2 Storage environment before the redesign

Proposed solution

IBM selected IBM Spectrum Control and IBM Spectrum Virtualize as a solution to solve all requests from customer side, and to redesign and improve existing environment implementing storage virtualization. By implementing this solution, the storage environment can dynamically adjust to any workload from application side. It can also distribute application workload among storage tiers based on type of workload, performance requests, and data importance.

Implementing SAN Volume Controller back end storage becomes just a storage tier. All storage functionality is transferred from the bottom to the upper layer. With this solution, IBM introduced a storage-defined storage solution on premise. By using the functions that are provided by Transparent Cloud Tiering (TCT), the customer moved some of its testing workload to IBM Cloud. This step was the first step to the Hybrid Cloud storage environment.

The proposed and implemented solution is shown in Figure 5-3.

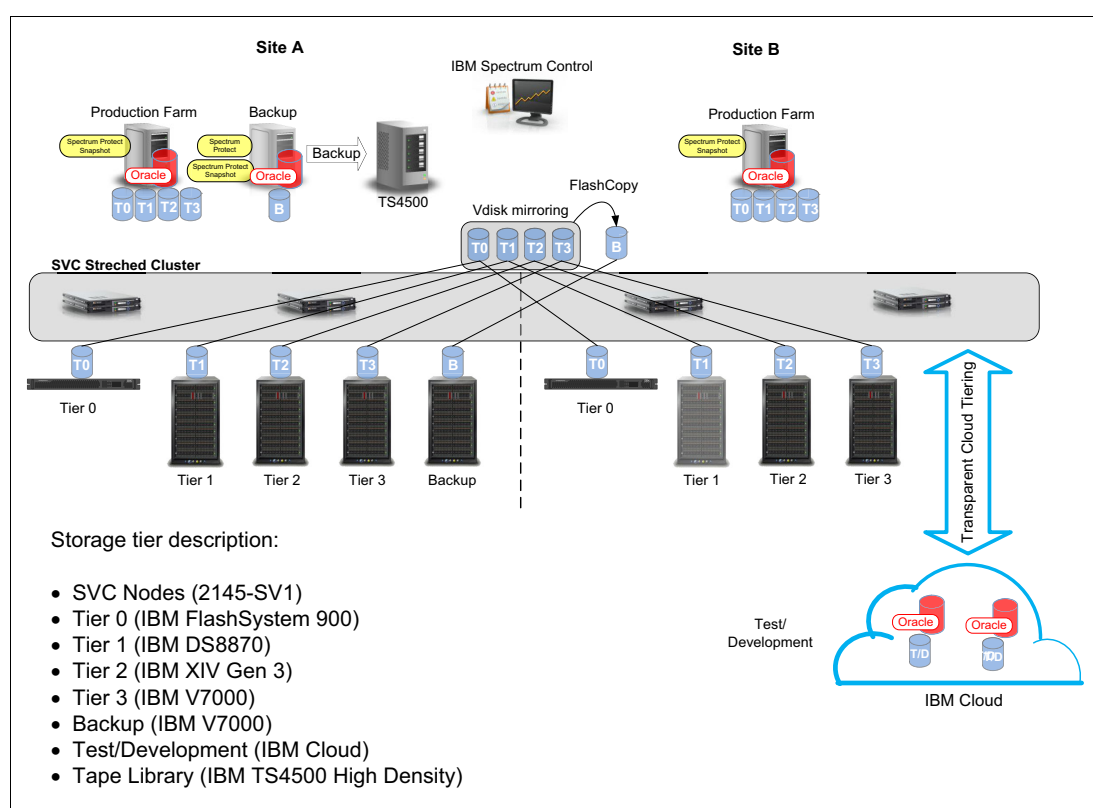


Figure 5-3 Proposed and implemented solution

Benefits of the solution

By implementing the IBM Spectrum Virtualize and IBM Spectrum Control solution, IBM provided the following benefits to the customer:

- ▶ Improved service levels and reduced errors and costs
- ▶ Turned existing storage into software-defined hybrid cloud storage
- ▶ Managed storage performance without requiring a storage expert
- ▶ Improved visibility, control, and automation for virtual and physical storage environments
- ▶ Provided a high-availability solution across the data center
- ▶ Improved capacity utilization
- ▶ Introduced storage tiering
- ▶ Introduced storage service pools

- ▶ Provided a central monitoring approach for the entire storage environment
- ▶ Allowed historical performance reporting and performance analytics
- ▶ Allowed transparent data movement between storage tiers
- ▶ Allowed application-aware backups and restores with IBM Spectrum Protect Snapshot
- ▶ Reduced the risk of data loss
- ▶ Enabled near-instant restores at any time

5.3 Government sector transformation to private cloud

The client is from the government sector, and is looking to transition from a silo-based infrastructure deployment model to a centralized IT provider model.

Business objective

The government sector believed that many agencies had systems that were performing identical workloads and storing duplicate data while having separate staff managing each set of systems. The state recognized that to meet mandated budgetary constraints, they would need to make significant changes in their IT deployment model. They outlined the following objectives:

- ▶ Use economies of scale and the full purchasing power of the state to drive down the cost of hardware and software acquisition
- ▶ Need to standardize IT infrastructure components
- ▶ Control IT operational complexity
- ▶ Increase the efficiency and time-to-market of commodity IT services
- ▶ Increase transparency of IT inventory, utilization, and cost
- ▶ Move IT infrastructure spending from the Capital expenditure format to the Operating expenditure format
- ▶ Form an overall consumption model for IT to allow for greater ingenuity and solution collaboration
- ▶ Decrease the amount of support necessary for maintaining IT services

These goals led them to decide that a software-defined storage infrastructure within a private cloud implementation would be ideal.

Proposed solution

The client recognized that they had aging systems in need of updating to gain efficiencies and capabilities. IBM recommended that they build a completely new storage infrastructure using software-defined infrastructure concepts based on the IBM Spectrum Control and IBM Spectrum Virtualize solution.

The integration with the IBM storage for SDS provided by Storwize, FlashSystem, and the SDS components enabled them to reach all of their objectives. In addition to the storage infrastructure recommendations, IBM also suggested that they implement a cloud stack that is composed of IBM Cloud Orchestrator.

Figure 5-4 shows the proposed architecture.

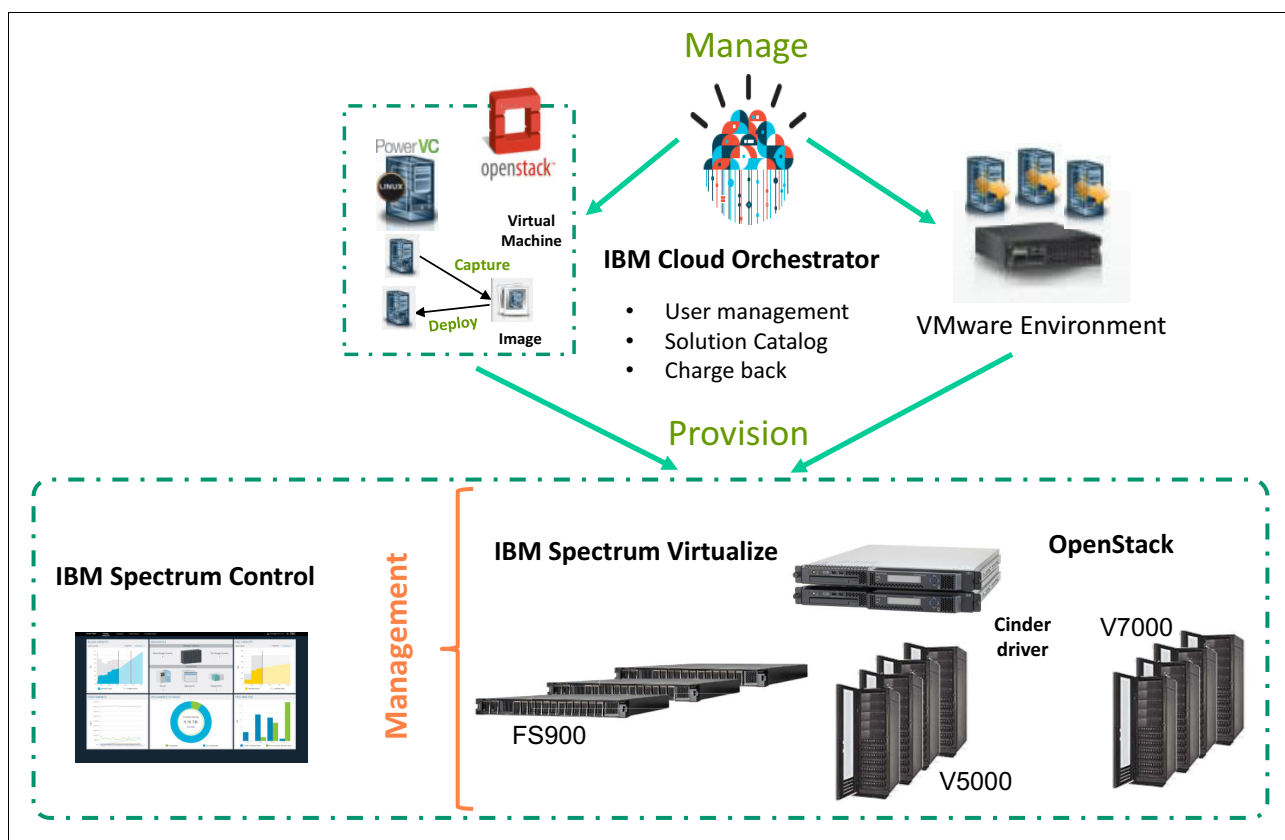


Figure 5-4 Proposed architecture

Benefits of the solution

By using the complete Resource Management Toolset capabilities that are provided in IBM Cloud Orchestrator, IBM Spectrum Control, and IBM Spectrum Virtualize, the state was able to achieve remarkable benefits. They implemented policy-driven service levels that lead to lower required capacities of high-cost tier 1 storage. They also automated monitoring of data not required to run the business, which resulted in drastic reduction of storage use. The chargeback capability resulted in better conservation of resources by agencies.

Overall, they projected that they would recognize a 5-year return on investment of 136% and allow them to reduce their usable storage capacity by 66%. The system was projected to maintain or improve levels of performance and availability.

5.4 Cloud service provider use case with SAP HANA

The client was a private cloud service provider that wanted to enhance analytics support for clients with an in-memory cloud solution that can deliver requisite performance and scalability. The IT environment was a combination of traditional and cloud workloads.

Business requirement

With its core focus on SAP solutions, the client looked for the right configuration for core customer workloads that enable synergies among traditional and new real-time analytics SAP workloads. This configuration operates in a single environment that used its resources as efficiently and flexibly as possible. As a result, they chose to establish and operate a cloud infrastructure that can host the total SAP software stack and the new SAP HANA platform.

Because data is the currency of the new economy, the customer was asking for better ways to protect, recover, access, share, and use it.

Proposed solution

IBM recommended that they build a new infrastructure with IBM Power8 servers and IBM A9000 storage system to address all of the resource-intensive SAP workload requirements. To address the security requirements, IBM Spectrum Copy Data Management was included in the proposal to use existing data in an efficient, automated, scalable manner. The scope was to manage all of those snapshot and IBM FlashCopy images that were made to support DevOps, data protection, disaster recovery, and Hybrid Cloud computing environments.

The proposed architecture is shown in Figure 5-5.

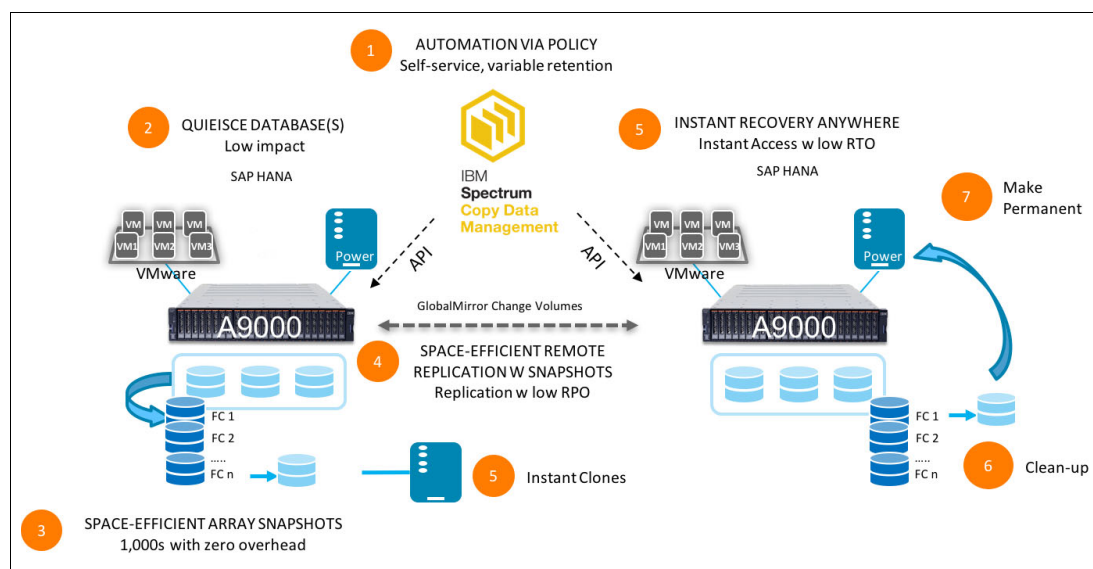


Figure 5-5 Proposed architecture

Benefits of the solution

IBM A9000 that uses IBM Spectrum Accelerate is the correct workload burst for a SAP environment that is running on IBM Power 8 servers and SAP HANA.

IBM Spectrum Copy Data Management (CDM) makes copies available to data consumers when and where they need them, without creating unnecessary copies or leaving unused copies on valuable storage. It catalogs copy data, identifies duplicates, and compares copy requests to existing copies. Data consumers can use the self-service portal to create the copies they need, which enables business agility.

IBM Spectrum CDM includes specific functionality for database environments. It can create application-aware data copies automatically, including snapshots, replication, and clones. Copy creation can be based on SLAs to ensure proper levels of protection.

IBM Spectrum CDM also integrates secure data masking for databases. This integration is vitally important for use cases, such as dev-test and training, where data must remain hidden. Also, one masked copy can be shared with multiple users, which makes it more efficient. Finally, IBM Spectrum CDM allows for integration with DevOps tools.

5.5 Media and entertainment company Hybrid Cloud

The company in this use case is an international multimedia publishing group that operates daily newspapers, magazines, books, radio broadcasting, news media, and digital and satellite TV.

Business needs

The company defined a new strategy that is based on high-quality editorial production, and rethinking products and offers. The client redesigned the business model, mainly regarding new organization of work to develop multimedia and a digital business model.

The media company needs a dynamic storage solution that can provide up to 20 PB of data that is distributed on separate tiers. The system also includes a tape library to ensure a cost-effective solution.

The client was looking for a storage cloud solution that can replicate data in separate sites, where the editors and journalists are based. They needed a system that offered zero downtime while delivering predictable performance in all the digital media information lifecycle phases: Create, manage, distribute, and archive. The storage cloud solution must include a pay-per-use model so that the company's customers can purchase access to old and recent TV programs.

Solution

IBM Spectrum Scale storage systems with a multitier storage pool (including Active File Management and tape management by IBM Spectrum Protect) provide the ability to move information to tape for archiving requests, as shown in Figure 5-6 on page 122.

For the regional sites of the media company, Active File Management (AFM) is used to share the information that is managed by the local site and vice versa. IBM Storage Insights (SaaS running on IBM Cloud) is used to collect and manage storage usage information on the IBM Spectrum Scale environment.

IBM Cost Management is used for reporting for chargeback of the web services customers. IBM Aspera solutions with the patented FASP transfer protocol (SaaS running on IBM Cloud) is used for file transfers to and from external parties for the fastest and the most cost-effective file transfer.

Figure 5-6 shows the proposed architecture.

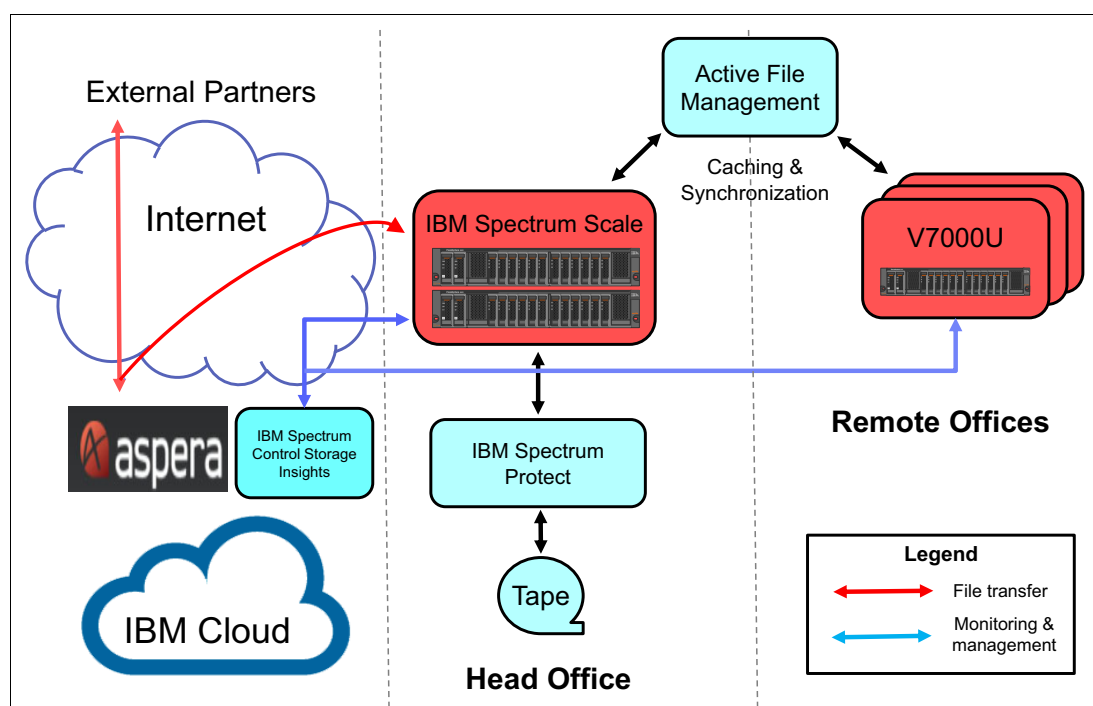


Figure 5-6 Solution based on IBM Spectrum Scale with multitier storage pools

Benefits of the solution

With IBM Spectrum Scale storage, the client can manage more than 100 million files with tens of thousands of users. The IBM Spectrum NAS solution provides concurrent profile logons, of which many have over a thousand small files.

The embedded replication function that is supplied by AFM, combined with IBM Spectrum Protect for backup and the archiving policies of HSM, ensures the continuous availability of the data. By using the ILM approach, the client can move information to the correct storage tier (including tape) to obtain a cost-effective solution.

The IBM cloud storage solution, which is based on IBM Spectrum Scale and its AFM technology, ensures that the remote locations have excellent access response time to the media content.

The IBM Spectrum Scale capability to manage multiple file systems with multiuser file sharing (managed by the HSM policies) provides a secure and cost-effective solution for the requirements. IBM Storage Insights provides intuitive analytics-based monitoring of the usage of storage.

The information that is collected by IBM SmartCloud® Cost Manager provides the customer with a solution to start pay-per-use services. IBM Aspera solutions with the patented FASP transfer protocol send and receive files from and to external partners in the quickest possible time over the internet with the lowest cost in terms of bandwidth used.

The collaborative benefits of the solution for every phase of the digital-media data process are shown in Figure 5-7 on page 123, Figure 5-8 on page 123, and Figure 5-9 on page 124.

Figure 5-7 shows the solution for the Broadcast department.

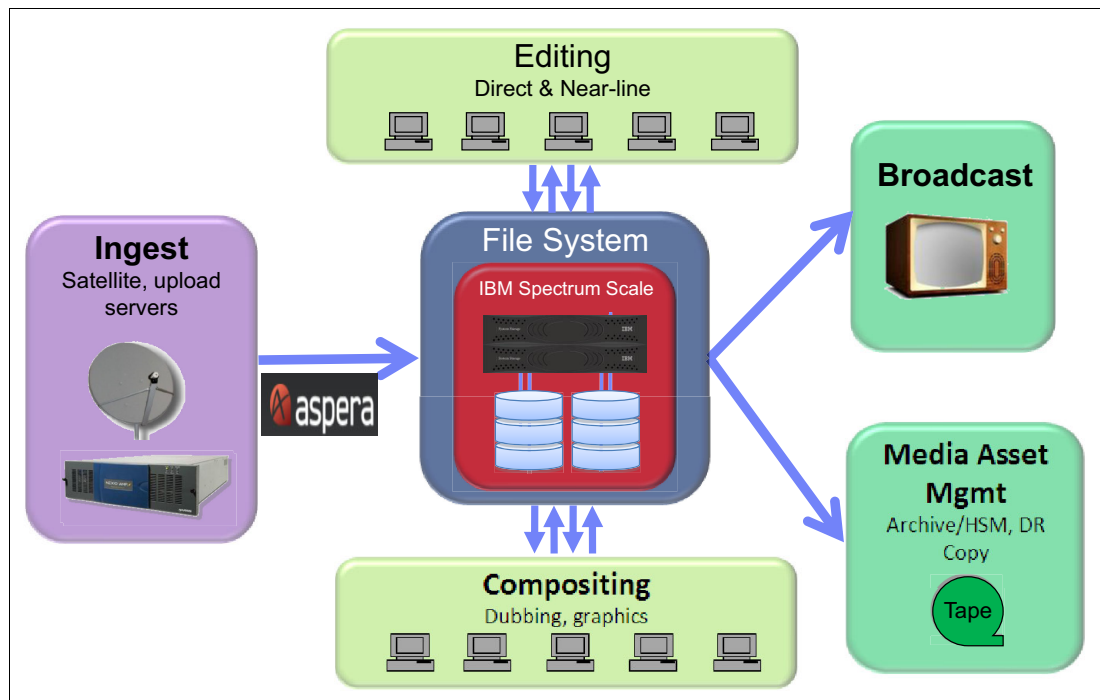


Figure 5-7 Cloud storage value for Broadcast

Figure 5-8 shows the solution for the Post-Production department.

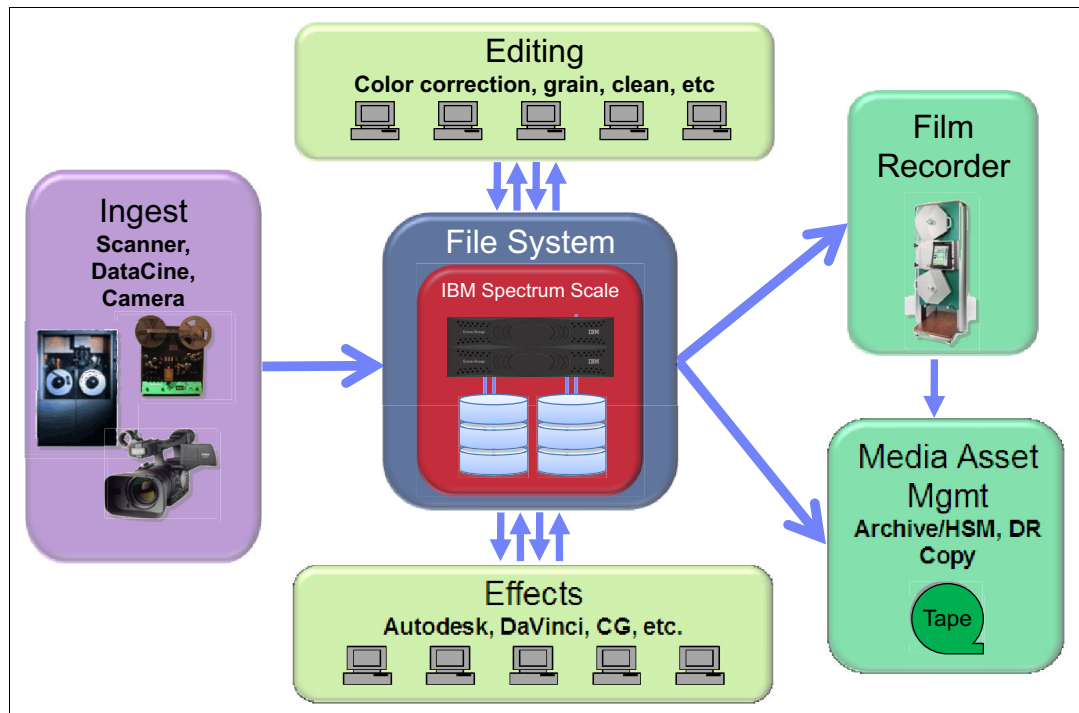


Figure 5-8 Cloud storage value for Post-Production

Figure 5-9 shows the solution for the Content Archiving and Distribution department.

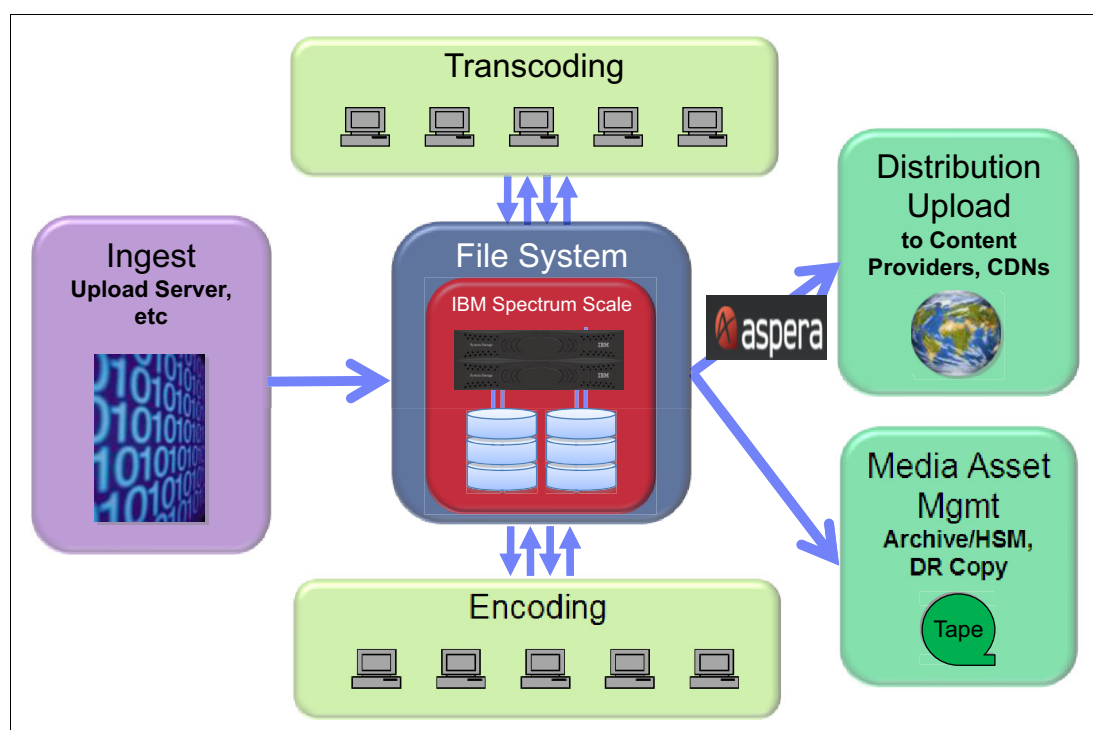


Figure 5-9 Cloud storage value for Content Archiving and Distribution

5.6 Life sciences and genomics

Life sciences research is advancing at a rapid pace, and new techniques, such as next-generation sequencing (NGS), are playing a vital role in growing scientific knowledge, facilitating the development of targeted drugs, and delivering personalized healthcare. By investigating the human genome, and studying it in the context of biological pathways and environmental factors, it's now possible for scientists and clinicians to identify individuals at risk of disease, provide early diagnoses, and recommend effective treatments.

Business needs

Not only are data bandwidth needs exploding to ingest data from newer applications and technologies, but in addition, the need to keep data secure and long term is also playing a major pain point for these solutions. The cost of paying for long term storage increases by greater factors the longer the data is kept. The addition of AI will also play a major role in the data recall of old data to analyze.

Modern biology represents an enormous data management challenge. Full genome sequences from next-generation sequencers can now provide raw data on all three billion base pairs in the human genome, representing terabytes of data that needs to be analyzed and compared to genomic databases.

In cancer research, the whole genome of a tumor and a matching normal tissue sample consumes about one terabyte of data. Large population-level studies can easily involve tens of petabytes of data, with storage requirements frequently doubling every five to twelve months.

In addition to the size of the data, the sheer number of generated files presents unique challenges. In a typical genomics workflow for DNA analysis, the first step involves concatenating and de-multiplexing raw data coming from a sequencer.

A DNA exome sequence may generate 20 to 40 million discrete reads per subject. One IBM customer has published a paper sharing details of their genomics environment. In summary, their genomics core facility (GCF) generates approximately 6 TB of raw data per week, and a further 14 TB of data after analysis and post-processing, for a total of 20 TB of new data per week. At the time of this writing, they manage approximately 11 petabytes of raw data shared between cluster nodes in their HPC center, and store an additional 3 petabytes in their Hadoop-based data science environment.

In clinical environments, storage requirements also include rich data sets, such as MRIs and ultrasound images. Increasingly, storage also must include data from new sources, such as telemetry from remote heart monitors, FITBIT data, and other sensors used together with genomic data for diagnosis. Complicating the data management challenge even more, hospitals need to retain most of this data for future comparative analysis and patient care.

Solution

To address the needs of speed, scale and smarts in genomic medicine, IBM has created an end-to-end reference architecture that defines the most critical components for genomics computing, illustrated in Figure 5-12. The reference architecture is provided for IT practitioners, and to a lesser extent medical personnel interested in the underlying technology. It provides an excellent way to view how application frameworks commonly used in personalized, translational medicine, and sequencing interact with the IT infrastructure. This is just one entry point into the reference architecture, and different workflows and tools apply in disciplines like imaging, cyclometry, and proteomics.

Figure 5-10 shows the IBM Genomics Reference Architecture.

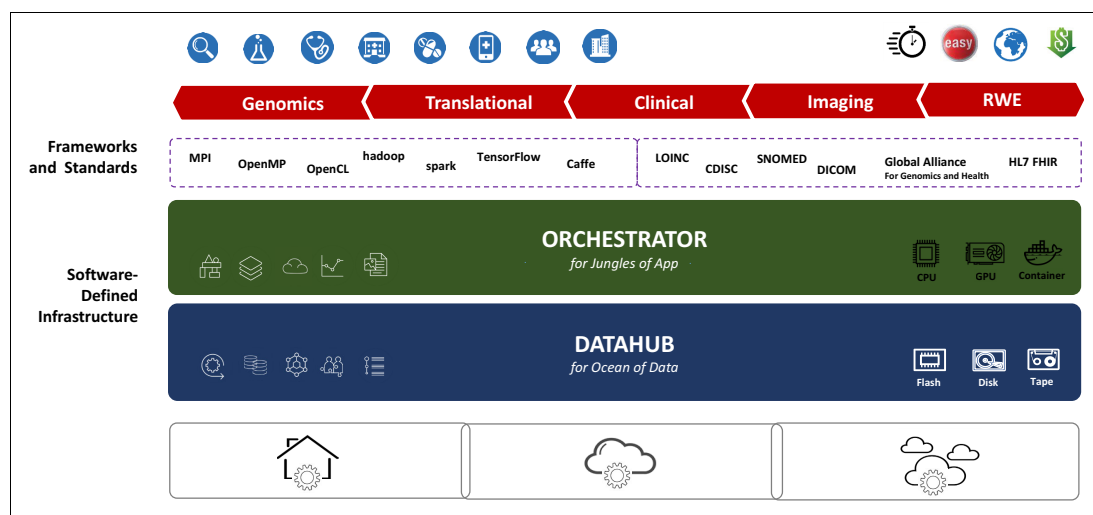


Figure 5-10 Genomics Reference Architecture

Benefits of the solution

Software-defined infrastructures offer opportunities to realize many advantages:

- ▶ Running research and clinical workloads faster, reducing cycle times, and providing better diagnostics and analysis in support of patient care
- ▶ Using resources more efficiently, enabling institutions to do more with less, reduce management costs, and enjoy a better return on assets

- Improved operational agility, enabling institutions to deploy promising new software frameworks more quickly, and better collaborate with others without infrastructure getting in the way

Combining often discrete HPC, data warehouse, data science and analytic environments into a shared environment represents an opportunity to not only increase capacity, but to reduce costs. A more efficient software-defined infrastructure for healthcare can help improve collaboration, reduce cycle times, improve analytic capabilities, and ultimately improve patient care - results that are important to everyone.

5.7 Oil and gas solutions

A government wants to centralize their National Data Repository (NDR) into a central storage where it can resell its valuable seismic data for exploration of resources.

Business needs

All of the upstream companies (both National and independent producers) are struggling to achieve common goals:

- Reduce the price of extracting hydro-carbons
- Reduce service costs
- Reduce transport and storage fees
- Improve recovery ratios

The Ministries want to create a regulation environment like the one that exists for financial companies. For this reason, the government asked IBM for a solution to enable access to, and controlled sharing of, their seismic data.

Solution

The proposed solution was the acquisition of the application that IBM developed with Schlumberger (SLB), *IBM Petroleum Solution for ProSource* (including the IBM software-defined storage infrastructure) to address the following requirements:

- Support the Prosource application from Schlumberger
- 100s of TB of online storage
- PBs of offline storage that is reliable and can be exported to other sites
- Reliable data recalls
- Capability to compress and encrypt offline data

IBM has the software-defined solution that enables ProSource customers to customize based on their existing requirements:

- Secure encryption of data in-flight and at rest, regardless of location (on-premise or in the SLB Cloud)
- Improve operational control of ProSource data management features

The proposed solution included:

- Architected solution to manage pre-stack data
- SLB ProSource Seismic Data Management software
- IBM Spectrum Storage Suite Software tuned for ProSource
- On-premise librarian with ability to connect to SLB Cloud Applications via Hybrid Cloud connector

Figure 5-11 shows the proposed solution.

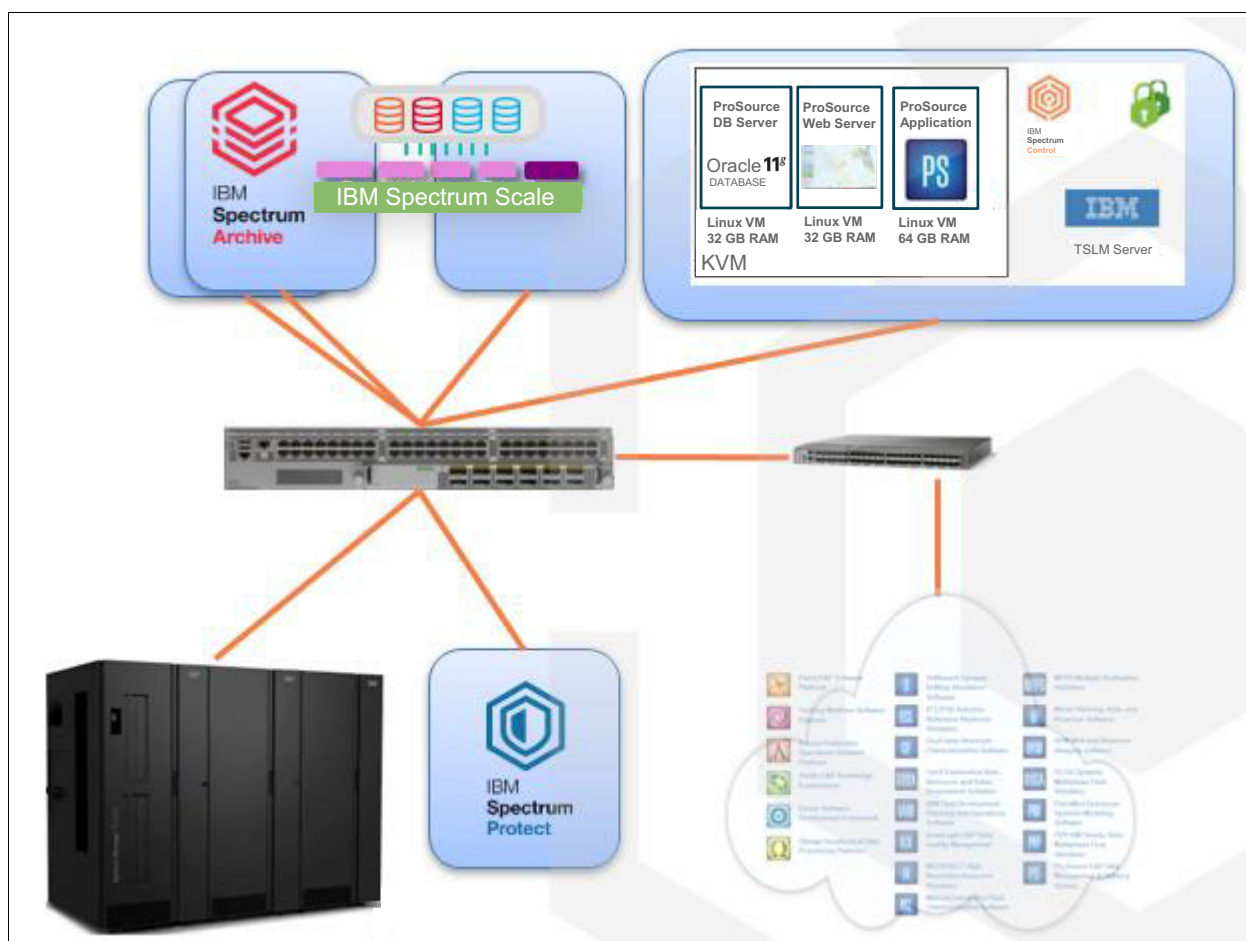


Figure 5-11 Proposed Solution

Benefits of the solution

With IBM Petroleum Solutions for ProSource, the customer started a journey towards a complete transformation of their environment. They built a more cost-effective, faster, and more tightly integrated seismic analytics and data infrastructure that supports high-performance computing (HPC), big data, and new workloads. The IBM and Schlumberger seismic repository enables a digital upstream:

- ▶ Intelligent data lifecycle management proven to handle petabytes of data and billions of files while optimizing storage efficiency
- ▶ Unified open architecture to support HPC, big data and next-generation workloads
- ▶ Scalable tiered storage architecture, from tape to Flash
- ▶ One-stop global support for faster and simpler problem resolution

5.8 Automobile analytics with Hortonworks

An automaker is using Hortonworks Data Platform (HDP) with IBM Spectrum Scale to store Internet of Things (IoT) data.

Business needs

The automaker needed to ingest data from an external application that writes data to storage using POSIX APIs, and then perform in-place analytics to ingest data by using POSIX and perform analytics directly on the data using Hadoop Distributed File System (HDFS). This data will be shared with mobile devices like tablets and smartphones using an object protocol.

Solution

IBM proposed a solution based on HDP with IBM Spectrum Scale. This solution brings business value by analyzing enormous data sources and deriving actionable insights from them.

The first phase of this analytics journey starts with consolidating different data silos to form an *Active Archive*. The Active Archive is then used to get a single view of the customer and perform further predictive analytics on them. With IBM Spectrum Scale, the client can build highly scalable and globally distributed data lakes to form their Active Archives.

IBM Spectrum Scale is the storage layer for the HDP environment as an alternative to native HDFS, because it supports the access to the data by using HDFS APIs. In addition, it is transparent to the applications that use HDP. IBM Spectrum Scale enables more flexible deployment models for storage system optimizing costs.

Figure 5-14 shows the Hortonworks Data Platform Architecture with IBM Spectrum Scale.

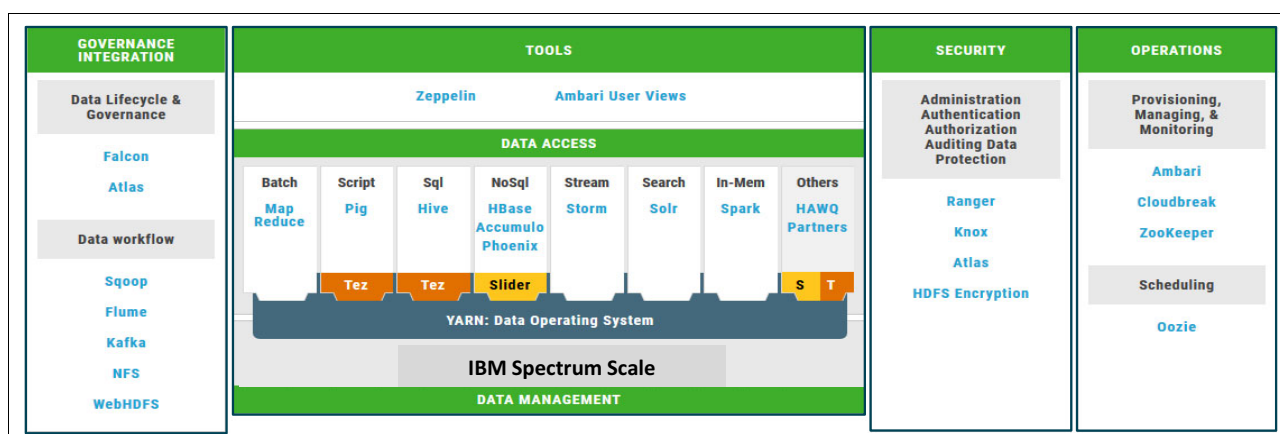


Figure 5-12 Hortonworks Data Platform Architecture

IBM Spectrum Scale V4.2.3 is certified with HDP 2.6/Apache Ambari 2.5 running on IBM Power Systems™ or x86 systems. This certification is for IBM Spectrum Scale software and applies to all deployment models of IBM Spectrum Scale, including IBM Elastic Storage Server. Therefore, the proposed solution uses the ESS appliance.

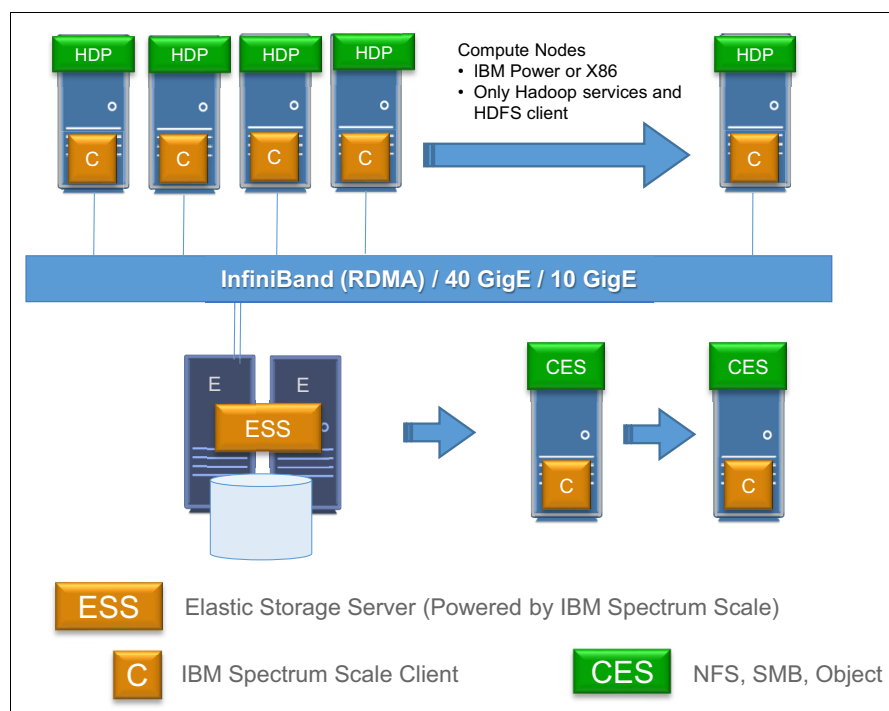
Benefits of the solution

The customer created an extremely scalable software-defined storage solution with a parallel file system architecture and no bottlenecks for data or metadata.

By using an IBM Spectrum Scale global namespace, the customer can span multiple Hadoop clusters and geographies, and clients can create active, remote data copies and enable real-time, global collaboration.

The solution supports data access by using NFS, SMB, Object, POSIX, and the HDFS API. This feature eliminates the need to maintain separate copies of the same data for traditional applications and for analytics.

Figure 5-14 shows the proposed solution.



With IBM Elastic Storage Server running IBM Spectrum Scale, there is no need for three-way replication for data protection, because of its Erasure coding and disk hospital functions. IBM Spectrum Scale includes Federal Information Processing Standards (FIPS)-compliant data encryption for secure data at rest, policy-based tiering/ILM, cold data compression, disaster recovery, snapshots, and backup and secure erase. In this way, the customer was able to address all of its compliance and security requirements.

5.9 Ensuring an insurance company's SAS Grid performance

The client had to run deep analytic scenarios across their entire consumer population, then segment the scenarios by various factors.

Business needs

Having the ability to quickly characterize a factor based on new and current attributes of the client's base of subscribers enables the customer to sift through huge data volumes and apply deep analytics against them.

Solution

The converged infrastructure was created using an IBM ESS GL4 storage system, a multicore IBM Power E880 server, and the IBM Spectrum Scale file system. It also included the IBM 8831-NF2 Ethernet switch made by Mellanox, with Mellanox hi-speed Ethernet adapters used between the server and storage to the switch. The main software building blocks are the IBM AIX operating system and the IBM Spectrum Scale shared file system.

The test bed employed was the SAS Mixed Analytics workload based on the SAS 9.4 M3 Foundation platform. This combination creates a powerful system with enterprise capabilities,

enabling scale-out architecture, affordable storage performance, and storage fabric scalability. See Figure 5-15 for one architecture depicting the software-defined converged infrastructure.

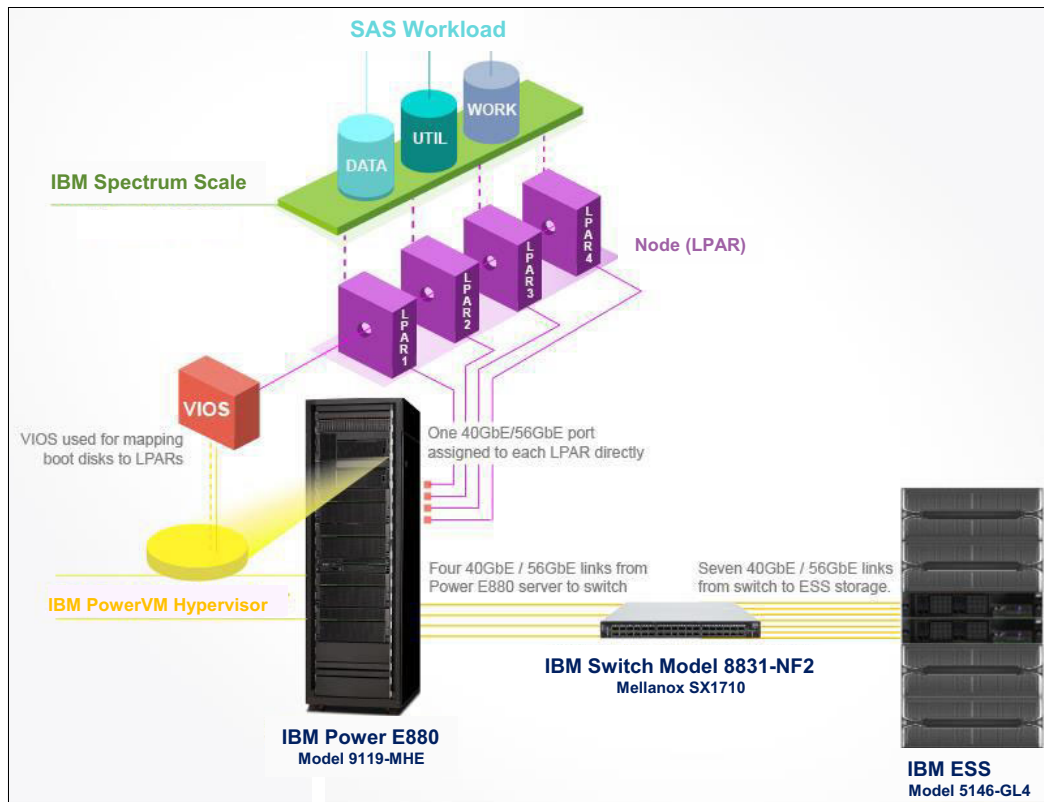


Figure 5-14 Proposed solution

Benefits of the solution

This solution can achieve 200 MBps per core, double the 100 MBps per core requirement for maintaining SAS Grid workloads.

Figure 5-15 shows a summary view of the test results for the different node configurations and GbE configurations.

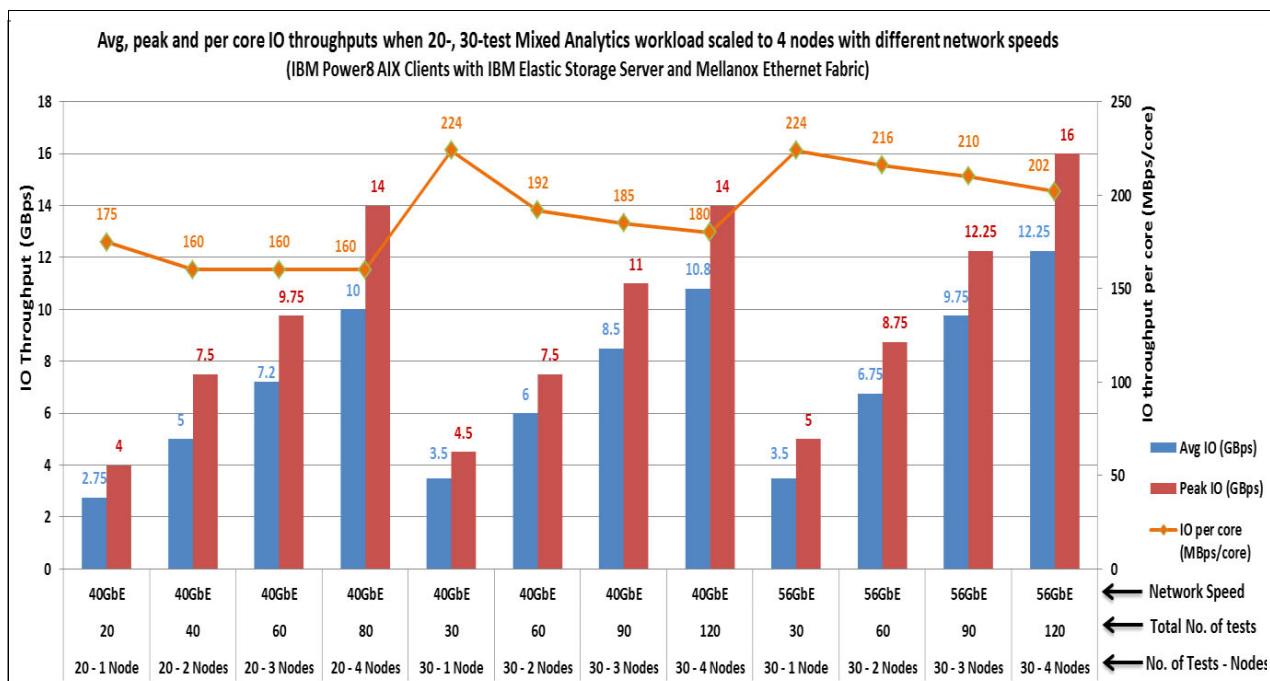


Figure 5-15 Test results for different configuration

The test results demonstrate that the proposed software-defined converged infrastructure is viable and performing well when used with SAS software. The Mellanox Ethernet high-speed storage network was crucial in facilitating the IBM ESS full I/O throughput. The ability to change from 40 GbE to 56 GbE fabric at no additional cost is a key capability that can be used by SAS environments.

The IBM ESS storage server with its JBOD technology proved to be an effective storage solution for SAS workloads. Its performance was comparable to the more expensive mid-tier Fiber Channel-attached Flash storage but without the higher costs, making it more cost effective.



Next steps

Getting started on the journey to a software-defined storage (SDS) solution can be relatively straightforward or fairly complex, depending on the scope of the project under consideration. It is important to understand current organizational capabilities and challenges, and identify the specific business objectives to be achieved by implementing an SDS solution in your enterprise. Ask the following questions:

- ▶ Where is my organization today in the maturity of its cloud, analytics, mobile, social, and traditional computing models?
- ▶ How will SDS meet my company's business needs?
- ▶ What strategy should my organization follow to build SDS infrastructure?
- ▶ Is my SDS plan integrated with my overall software-defined infrastructure (SDI) plan?
- ▶ How does the SDS plan help me create a *System of Insights*?
- ▶ Does my plan clearly define the metrics to be used to prove the value of the SDS implementation?
- ▶ Does IBM have service offerings that can help implement my SDS strategy (Cloud Managed Services)?

IBM personnel can assist you in your journey to smart storage cloud by developing a high-level architecture and implementation plan. This plan includes a supporting business case to justify investment, based on a compelling return on investment (ROI), and on improved service levels and lowered costs.

Business requirements have moved to the forefront of driving IT decisions. The plan for moving to the next step of an SDS implementation needs to be validated with the business plan. As shown in Figure 6-1, the business model becomes the focus, and the plan supports the business model.

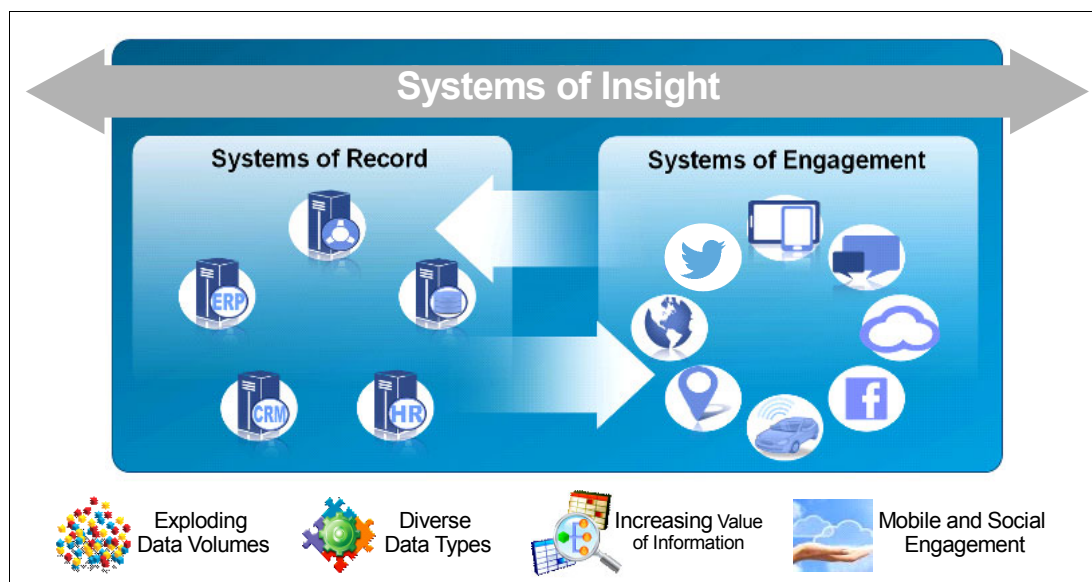


Figure 6-1 Systems of Insight

After the business model is defined, the SDS plan can be mapped to meet the business needs. This applies to both the traditional IT computing models and the Cloud, Analytics, Mobile, Social, and Security (CAMSS) model.

Figure 6-2 on page 135 helps lay out the products that provide the software-defined value as part of the Control Plane or Data Plane. As described earlier, some products have functions that straddle the Control and Data Planes. The key is to map the business requirement to the functions of the products as you define your SDS road map.

In some cases, you should incorporate the storage that you have into the SDS implementation. In other cases, it makes sense to invest in newer-technology storage to meet specific business needs. Optimization of your traditional workload compared to your mobile environment can have different implementations to meet the needs of your business. The business needs are also the driving force of the amount of change you can make to your current IT environment.

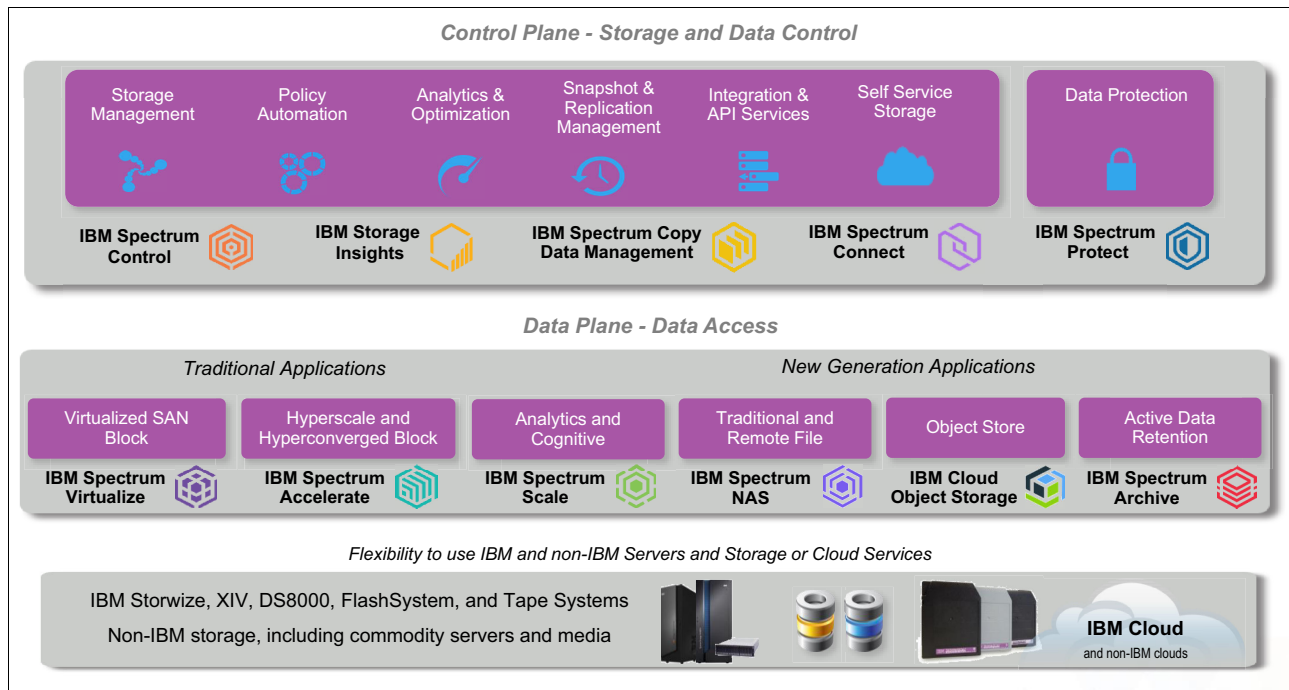


Figure 6-2 IBM SDS Vision

Before embarking on any journey, it is important to understand where you are currently, and what your chosen destination is. Developing your own cloud storage strategy should reflect these important considerations, which help you to define your path. So take the time needed to ensure that you understand how cloud storage can help your business. Justify your move by using ROI, total cost of ownership (TCO), and other business measures that are relevant to your organization. Be sure that you consider technical or compliance concerns, and develop risk-mitigation plans.

Remember that although storage cloud can be a key component of an overall cloud computing approach, you should determine how a storage cloud strategy fits within your broader cloud computing architectural plans. Overall integration of the following system parameters is essential to successful implementations:

- ▶ Optimal workload allocation
- ▶ Transaction integrity
- ▶ Agility and scalability
- ▶ Universal data access
- ▶ Operations
- ▶ Security
- ▶ Compliance

Consider your security needs and how a storage cloud is affected by the confidentiality of the data that you need to store. Data that is highly sensitive, or subject to security-compliance regulations, might not be able to be stored on a public network. Therefore, your storage cloud might need to be located behind an enterprise firewall, which is a private cloud solution requirement. The same might be true for instances where users need to easily access, share, and collaborate on data, without compromising data security, integrity, availability, and control of that data.

Figure 6-3 shows the high-level differences of current organizations versus the implementation of SDI that includes SDS. Planning for the next steps needs to include all aspects of the current business processes and infrastructure, along with the short-term and long-term goals of the business. From there, the SDI infrastructure can be planned. The SDS plan is a subset of the SDI plan.

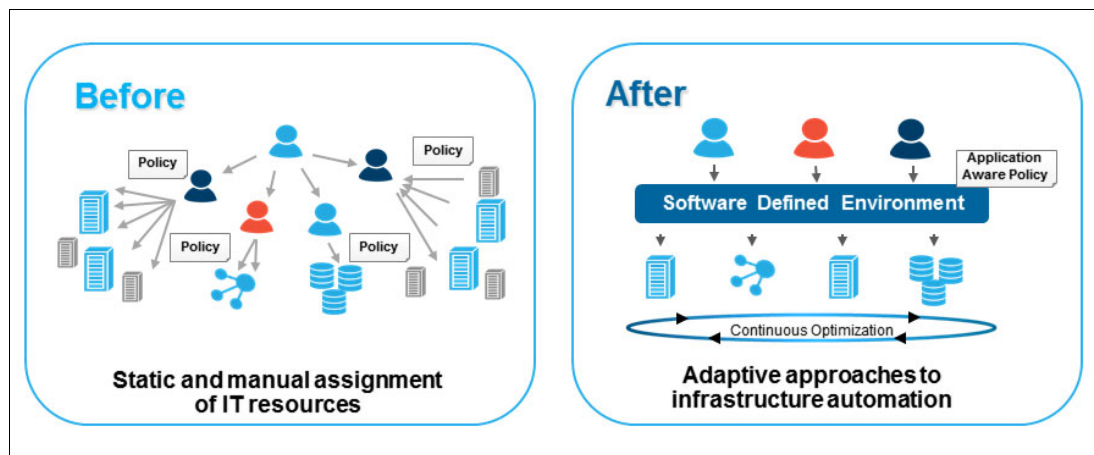


Figure 6-3 Organizational and infrastructure improvements with SDI implementation

The next step is to contact IBM to finalize your plan to meet your business needs.

IBM Systems Client Centers

Located around the world, IBM Client Centers provide access to technical experts and the latest technology to assist you in your purchase decision-making process. The centers include experts who are ready to work with you and your IBM Account Team or Business Partner as you consider the IBM Systems options that are needed to remake enterprise IT and transform your business.

For more information about IBM Client Centers, see this website:

<https://www.ibm.com/it-infrastructure/services/client-centers>

From informational briefings to demonstrations to performance testing, let the Client Centers help you take the next step.

IBM cloud offerings

For more information about IBM cloud offerings, see this website:

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

IBM personnel can assist you by developing a high-level architecture and implementation plan with a supporting business case to justify investment based on a compelling return on investment, with improved service levels and lowered costs for your cloud infrastructure.

IBM consultants use a unique cloud adoption framework, the Cloud Computing Reference Architecture (CCRA), and the IBM Cloud Workload Analysis Tool to help you analyze your existing environment and determine which cloud computing model is best suited for your business. They help you identify the business areas and workloads that, when changed to a cloud computing model, can enable you to reduce costs and improve service delivery that is in line with your business priorities.

The comprehensive structured approach that IBM brings to a cloud implementation engagement is shown in Figure 6-1 on page 134. This approach helps IBM to perform a rigorous analysis of your IT and application infrastructure, and provides recommendations and project planning for streamlining your infrastructure and processes. The IBM methodology incorporates key practices that were learned from engagements with leading businesses around the globe, and partnering with them on their storage cloud journey.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- ▶ *Active Archive Implementation Guide with IBM Spectrum Scale Object and IBM Spectrum Archive*, REDP-5237
- ▶ *IT Modernization using Catalogic ECX Copy Data Management and IBM Spectrum Storage*, SG24-8341
- ▶ *Cloud Computing Patterns of Expertise*, REDP-5040
- ▶ *A Deployment Guide for IBM Spectrum Scale Unified File and Object Storage*, REDP-5113
- ▶ *Enabling Hybrid Cloud Storage for IBM Spectrum Scale Using Transparent Cloud Tiering*, REDP-5411
- ▶ *Harnessing the Power of ProtecTIER and Tivoli Storage Manager*, SG24-8209
- ▶ *IBM DS8880 Architecture and Implementation (Release 8.3)*, SG24-8323
- ▶ *IBM DS8880 Product Guide (Release 8.3)*, REDP-5344
- ▶ *IBM Private, Public, and Hybrid Cloud Storage Solutions*, REDP-4873
- ▶ *IBM SmartCloud: Building a Cloud Enabled Data Center*, REDP-4893
- ▶ *IBM Spectrum Accelerate Deployment, Usage, and Maintenance*, SG24-8267
- ▶ *IBM Spectrum Accelerate Reference Architecture*, REDP-5260
- ▶ *IBM Spectrum Control Base: Enabling VMware Virtual Volumes with IBM XIV Storage System*, REDP-5183
- ▶ *IBM Spectrum Scale (formerly GPFS)*, SG24-8254
- ▶ *IBM Spectrum Scale in an OpenStack Environment*, REDP-5331
- ▶ *IBM Spectrum Scale and ECM FileNet Content Manager Are a Winning Combination: Deployment Variations and Value-added Features*, REDP-5239
- ▶ *IBM Spectrum Scale Security*, REDP-5426
- ▶ *IBM Spectrum Virtualize and IBM Spectrum Scale in an Enhanced Stretched Cluster Implementation*, REDP-5224
- ▶ *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- ▶ *IBM XIV Storage System Architecture and Implementation*, SG24-7659
- ▶ *Implementing IBM FlashSystem 900*, SG24-8271
- ▶ *Implementing the IBM SAN Volume Controller and FlashSystem 820*, SG24-8172
- ▶ *Implementing IBM Storage Data Deduplication Solutions*, SG24-7888

- ▶ *Implementing the IBM Storwize V7000 Unified Disk System*, SG24-8010
- ▶ *Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.1*, SG24-7938
- ▶ *Implementing the IBM Storwize V7000 Gen2*, SG24-8244
- ▶ *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V8.1*, SG24-7933
- ▶ *Introducing and Implementing IBM FlashSystem V9000*, SG24-8273
- ▶ *Regain Control of your Environment with IBM Storage Insights*, REDP-5231
- ▶ *Understanding IBM Spectrum Scale for Linux on z Systems (Express Edition)*, TIPS1211
- ▶ *Using IBM DS8000 in an OpenStack Environment*, REDP-5220
- ▶ *Using XIV in OpenStack Environments*, SG24-4971
- ▶ *VersaStack Solution by Cisco and IBM with SQL, Spectrum Control, and Spectrum Protect*, SG24-8301

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- ▶ IBM Client Demonstration Center

<https://www.ibm.com/systems/clientcenterdemonstrations>

Note: The IBM Client Demonstration Center (for Business Partners, IBMers, and anyone with an IBMid) provides a catalog of remote demonstrations (video or live connection) that consist of self contained material for customer demonstrations of IBM solutions. Most of the demonstrations are provided with predefined scenarios, and some also allow for the development of new scenarios. Demonstrations can also be considered as 'ready to use' material for enablement or training.

- ▶ IBM System Storage Tumblr provides product videos, customer reference videos, case studies, white papers, infographics, and subject matter expert videos around IBM Storage solutions.

<http://www.ibmstorageexperience.tumblr.com>

- ▶ IBM storage news, hints and technical discussions by EMEA storage experts

<https://www.ibm.com/developerworks/community/blogs/storageneers>

- ▶ The Storage Community sponsored by IBM

<https://developer.ibm.com/storage/blog/>

- ▶ IBM Cloud Computing Reference Architecture wiki

https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Wf3cce8ff09b3_49d2_8ee7_4e49c1ef5d22

- ▶ IBM Cloud Computing Redbooks

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

- ▶ IBM SmartCloud Virtual Storage Center Solution, IBM Redbooks Solution Guide
<http://www.redbooks.ibm.com/abstracts/tips0991.html>
- ▶ IBM System Storage for Cloud
<http://www.ibm.com/systems/storage/solutions/cloud>
- ▶ IBM Cloud Offerings
<http://www.ibm.com/cloud-computing/us/en/index.html>
- ▶ IBM Cloud business continuity and resiliency services:
 - IBM Cloud Managed Backup
<http://www.ibm.com/services/us/en/it-services/business-continuity/cloud-managed-backup>
 - IBM Managed Data Vault
<http://www.ibm.com/services/us/en/it-services/managed-data-vault.html>
 - IBM Storage Services
<http://www.ibm.com/services/us/en/it-services/smart-business-storage-cloud.html>
 - IBM Federal and State Contracts
http://www.ibm.com/shop/americas/content/home/en_US/government-contracts.html
 - IBM Resiliency Services
<http://www.ibm.com/services/us/en/it-services/business-continuity/index.html>
- ▶ IBM SAN Volume Controller Knowledge Center
<http://www.ibm.com/support/knowledgecenter/STPVGU/welcome>
- ▶ IBM Spectrum Archive Enterprise Edition Knowledge Center
http://www.ibm.com/support/knowledgecenter/ST9MBR_1.2.0/1tfs_ee_intro.html
- ▶ IBM Spectrum Accelerate Knowledge Center
http://www.ibm.com/support/knowledgecenter/STZSWD/landing/IBM_Spectrum_Accelerate_welcome_page.html
- ▶ IBM Spectrum Control Knowledge Center
<http://www.ibm.com/support/knowledgecenter/SS5R93>
- ▶ IBM Spectrum Scale Knowledge Center
http://www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html
- ▶ IBM Tivoli Storage Manager
http://www.ibm.com/support/knowledgecenter/SSGSG7/landing/welcome_ssgsg7.html
- ▶ IBM Spectrum Virtualize Knowledge Center
<http://www.ibm.com/support/knowledgecenter/STVLF4>
- ▶ Blueprints for IBM Spectrum Protect
<https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli%20Storage%20Manager/page/IBM%20Spectrum%20Protect%20Blueprints>
- ▶ Demonstration of the Operations Center GUI for IBM Spectrum Protect
<https://www.ibm.serviceengage.com/on-premises-solutions>

- ▶ IBM Spectrum Scale
<http://www.ibm.com/systems/storage/spectrum/scale/index.html>
- ▶ IBM Spectrum Scale Wiki
<https://ibm.biz/BdFPR2>
- ▶ IBM Elastic Storage Server
<http://www.ibm.com/systems/storage/spectrum/ess>
- ▶ IBM Storwize V7000 Unified Knowledge Center
http://www.ibm.com/support/knowledgecenter/ST5Q4U/landing/v7000_unified_welcome.htm
- ▶ IBM XIV Storage System Knowledge Center
https://www.ibm.com/support/knowledgecenter/STJTAG/com.ibm.help.xivgen3.doc/xiv_kcwelcomepage.html
- ▶ Thoughts on Cloud: cloud computing conversations led by IBMers
<http://thoughtsoncloud.com>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



REDP-5121-02

ISBN 0738457051

Printed in U.S.A.

Get connected

