Redbooks
ibm.com/redbooks

IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation

Bertrand Dufrasne
Francesco Anderloni
Roger Eriksson
Andrew Greenfield
Lisa Martinez
Stephen Solewin

Storage

IBM

Redbooks

International Technical Support Organization

IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation

January 2019

SG24-8345-03

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

Fourth Edition (January 2019)

This edition applies to FlashSystem A9000 and A9000R, Model 425, with system software Version 12.3.1.

Contents

Notices vii
Trademarks viii

Preface ix
Authors x
Thanks to the authors of the previous edition: xi

Thanks to the following people for their contributions to this project: xi
Now you can become a published author, too xi
Comments welcome xii
Stay connected to IBM Redbooks xii

Chapter 1. IBM FlashSystem A9000 and IBM FlashSystem A9000R 1
1.1 FlashSystem A9000 and FlashSystem A9000R overview 2
1.1.1 FlashSystem A9000 3
1.1.2 FlashSystem A9000R 4
1.2 Feature set 5
1.3 Licensing 8
1.4 Management tools 9
1.5 Utility Offering 10

Chapter 2. Logical architecture and concepts 11
2.1 Overall architecture 12
2.1.1 Parallelism and grid architecture 12
2.1.2 Cache resiliency 13
2.2 IBM FlashCore technology 14
2.3 Data reduction and capacity concepts 15
2.3.1 Data reduction technology 15
2.4 Storage provisioning concepts 25
2.4.1 Storage pools 25
2.4.2 Volumes 26
2.4.3 Snapshots 26
2.4.4 Consistency groups 27
2.5 Multi-tenancy 27
2.5.1 Domains 28
2.6 Data security 30
2.6.1 Key management 31
2.6.2 User access control 31
2.7 Auditing through syslog 32

Chapter 3. Hardware and technology 35
3.1 Systems overview 36
3.1.1 FlashSystem A9000R 36
3.1.2 FlashSystem A9000R entry-level capacity option 38
3.1.3 FlashSystem A9000 39
3.1.4 FlashSystem A9000 and A9000R Storage Utility Offering 39
3.2 System components details 40
3.2.1 Grid controller 40
3.2.2 Flash enclosure 49

3.3 FlashSystem A9000R specifics 54
3.3.1 FlashSystem A9000R switched InfiniBand 55
3.3.2 InfiniBand switch battery backup unit 57
3.3.3 FlashSystem A9000R rack 57
3.4 Scaling FlashSystem A9000R 59
3.5 FlashSystem A9000 specifics 61
3.5.1 FlashSystem A9000 direct InfiniBand 62
3.6 Reliability, availability, and serviceability 65
3.6.1 Reliability 65
3.6.2 Availability 66
3.6.3 Serviceability 67
3.7 IBM FlashCore technology 67
3.7.1 Technology and architectural design overview 70
3.7.2 Hardware-only data path 70
3.7.3 Flash card memory chips 70
3.7.4 Flash module capacities 71
3.7.5 Field Programmable Gate Array 72
3.7.6 Overprovisioning 73
3.7.7 Wear leveling 74
3.8 IBM Variable Stripe RAID and 2D Flash RAID protection overview 74
3.8.1 Variable Stripe RAID 75
3.8.2 Two-dimensional (2D) Flash RAID 77

Chapter 4. Capacity planning and management 79
4.1 Introduction and definitions 80
4.1.1 System capacity specifications 80
4.1.2 Capacity allocation 83
4.2 Capacity planning with data reduction 84
4.2.1 Workload types 84
4.2.2 Data reduction planning 85
4.3 Capacity representation and reporting 90
4.3.1 Data usage collection 90
4.3.2 Capacity representation in the Hyper-Scale Manager GUI 91
4.3.3 Capacity planning report 98
4.4 Intelligent capacity management for deduplication 105

Chapter 5. Installation requirements 109
5.1 Physical space requirements 110
5.1.1 FlashSystem A9000 110
5.1.2 FlashSystem A9000R 110
5.2 Delivery requirements for IBM FlashSystem A9000R 111
5.2.1 Delivery clearance requirements 112
5.2.2 Delivery weight requirements 112
5.2.3 Height and weight reduction features 112
5.3 Site requirements 113
5.3.1 Power supply requirements 113
5.3.2 Power consumption and thermal dissipation 113
5.3.3 Other equipment 115
5.4 Basic configuration 117
5.5 Network connections 118
5.5.1 Remote mirroring connectivity 119
5.5.2 Planning for growth 119
5.5.3 IPv6 addressing and planning 119

5.5.4 IP protocols . . . 120
5.5.5 SMTP server . . . 120
5.5.6 IBM service ports (model 415 only) . . . 120
5.6 Physical installation . . . 121
5.6.1 Initial setup . . . 121
5.6.2 Completing the physical installation . . . 121
5.6.3 System shutdown/power-off and power-on . . . 122
5.7 Host connections . . . 125
5.7.1 Fibre Channel connections . . . 126
5.7.2 iSCSI connections . . . 126
5.7.3 Fibre Channel cabling and configuration . . . 126
5.7.4 iSCSI cabling and configuration . . . 126
5.7.5 iSCSI IP configuration . . . 126

Chapter 6. Performance . . . 129
6.1 Performance considerations . . . 130
6.1.1 Sizing . . . 130
6.1.2 Multipathing considerations . . . 130
6.2 Quality of service . . . 131
6.2.1 Limitation by bandwidth . . . 131
6.2.2 Limitation by input/output operations per second . . . 132
6.3 Performance monitoring . . . 135
6.3.1 Using the Storage Management GUI . . . 135
6.3.2 Using the command-line interface . . . 138

Chapter 7. Monitoring and troubleshooting . . . 141
7.1 Monitoring . . . 142
7.1.1 Monitoring events . . . 142
7.1.2 Monitoring alerts and events . . . 147
7.1.3 Monitoring statistics . . . 148
7.1.4 Monitoring hardware . . . 149
7.1.5 Simple Network Management Protocol monitoring . . . 150
7.1.6 Call Home . . . 150
7.1.7 Encrypting Call Home and heartbeat notifications . . . 151
7.1.8 Subscribe to support notifications . . . 152
7.2 Troubleshooting . . . 152
7.2.1 IBM Support process . . . 152
7.2.2 Collecting logs . . . 155
7.2.3 IBM Remote Support . . . 161

Related publications . . . 169
IBM Redbooks publications . . . 169
Other publications and online resources . . . 169
Help from IBM . . . 170

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.

AIX®
DS8000®
Global Technology Services®
HyperSwap®
IBM®
IBM FlashCore®
IBM FlashSystem®
IBM Spectrum™
IBM Spectrum Accelerate™
IBM Spectrum Control™
IBM Spectrum Storage™
IBM Spectrum Virtualize™
MicroLatency®
PowerPC®
Real-time Compression™
Redbooks®
Redbooks (logo)  ®
System Storage®
Variable Stripe RAID™
XIV®

The following terms are trademarks of other companies:

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

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

Linear Tape-Open, LTO, 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

This IBM® Redbooks® publication presents the architecture, design, concepts, and technology that are used in IBM FlashSystem® A9000 and IBM FlashSystem A9000R. This revised edition applies to software version 12.3.2.

FlashSystem A9000 and FlashSystem A9000R deliver the microsecond latency and high availability of IBM FlashCore® technology with grid architecture, simple scalability, and industry-leading IBM software that is designed to drive your business into the cognitive era.

The Hyper-Scale Manager highly intuitive user interface simplifies management.

Comprehensive data reduction capabilities, including inline deduplication and a powerful compression engine, help lower total cost of ownership. With software version 12.3.1 and Hyper-Scale Manager version 5.5.1 (or later) the system can compute reclaimable and attributed capacity information, without performance impact.

Software version 12.3.2, with Hyper-Scale Manager version 5.6 or later, introduces support for VLAN tagging and port trunking.

From a functional standpoint, FlashSystem A9000 and FlashSystem A9000R take advantage of most of the software-defined storage features that are offered by the IBM Spectrum™ Accelerate software, including multi-tenancy and business continuity functions. FlashSystem A9000 and FlashSystem A9000R supports HyperSwap and multi-site High Availabilty / Disaster Recovery (HA/DR) configurations.

This publication is intended for those individuals who need to plan, install, tailor, and configure FlashSystem A9000 and FlashSystem A9000R.

For detailed information about configuration, management, host attachment, and replication functions and their usage, see the following publications:

- *IBM Spectrum Accelerate Family Storage Configuration and Usage for IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Gen3,* SG24-8376
- *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401
- *IBM HyperSwap and multi-site HA/DR solution for IBM FlashSystem A9000 and A9000R*, REDP-5434
- *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368.

Authors

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

Bert Dufrasne is an IBM Certified Consulting IT Specialist and Project Leader for IBM System Storage® disk products at the International Technical Support Organization (ITSO), San Jose Center. He has worked at IBM in various IT areas. He has authored many IBM Redbooks publications, and he has also developed and taught technical workshops. Before Bert joined the ITSO, he worked for IBM Global Services as an Application Architect. He holds a Master's degree in Electrical Engineering.

Francesco Anderloni is an Infrastructure Specialist working for the IBM Global Technology Services® organization in IBM Italy. Since he joined IBM in 2015, he has been working in the Storage Management area, delivering IBM Storage solutions to clients in Italy and Europe. He is a member of the Young Technical Expert Council in IBM Italy. Francesco holds a Bachelor's Degree in Information Engineering and a Master's Degree in Computer Science, both from the University of Padova, Italy.

Roger Eriksson works at IBM Systems Lab Services Nordic, based in Stockholm, Sweden. He is a Senior Accredited IBM Product Service Professional. Roger has over 25 years of experience working on IBM servers and storage, including Enterprise and Midrange disk, NAS, SAN, IBM x86, and IBM Power. He has done consulting, proof of concepts and education, mainly with the IBM Spectrum Accelerate™ and IBM Spectrum Virtualize™ product line, since December 2008. He also works with the Swedish technical sales team for storage solutions on a daily basis. Working with both clients and various IBM teams worldwide is a normal day's work. He holds a Technical College Graduation in Mechanical Engineering.

Andrew Greenfield is an IBM Global XIV and Flash Solution Engineer who is based in Phoenix, Arizona. He holds numerous technical certifications from Cisco, Microsoft, and IBM. Andrew brings over 24 years of data center experience with Fortune 100 companies to the team. He graduated Magna cum Laude, Honors, from the University of Michigan, Ann Arbor. Andrew has also written other IBM FlashSystem and XIV Gen3 IBM Redbooks publications.

Lisa Martinez has been working in the Washington Systems Center - Storage Team as a storage specialist since January 2012. Her focus has been with pre-sales support for IBM DS8000® and the IBM Spectrum Accelerate family of storage products (FlashSystem A9000/A9000R, XIV, and IBM Spectrum Accelerate). She is also the lead instructor for FlashSystem A9000/A9000R, XIV, and IBM Spectrum Accelerate customer-based workshops. Prior experience includes roles as a storage architect in the Specialty Services Area in IBM Global Technology Services (GTS), and test architect in disk storage focusing on system-level test for XIV for three years, and copy services for DS8000. Lisa holds degrees in Computer Science from New Mexico Highlands University and Electrical Engineering from the University of New Mexico.

Stephen Solewin is an IBM Corporate Solutions Architect based in Tucson, Arizona. He has over 20 years of experience in working on IBM storage, including enterprise and midrange disk, Linear Tape-Open (LTO) drives and libraries, SAN, storage virtualization, and storage software. Steve is a global resource, working with customers, IBM Business Partners, and fellow IBMers worldwide. He has been working on the XIV product line since 2008 and FlashSystem since 2013. He holds a BS degree in electrical engineering from the University of Arizona, where he graduated with honors.

Thanks to the authors of the previous edition:

Mariya Atanasova
Matthew DiVito
John Hyams
Petar Kalachev
David Nierhoff
Markus Oscheka
Jagadeesh Papaiah
Patrick Pollard
Ann Sargent
Axel Westphal
Roey Yeshua

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

Tomer Carmeli
Harry McGregor
Moshe Weiss
Randi Galron
Samuel Krikler
Daniel Lereya
Rivka Matosevich
Sahar Mualem-Yosi
Rivka Pollack
Avi Sabzerou
Yossi Siles
Vadim Steckler
IBM

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

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

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, 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

1

IBM FlashSystem A9000 and IBM FlashSystem A9000R

FlashSystem A9000 and FlashSystem A9000R are functionally built with IBM Spectrum Accelerate software that is optimized to run on a flash storage grid architecture. The optimized systems also include always-on data deduplication and compression.

FlashSystem A9000 and FlashSystem A9000R are designed to deliver performance and ease of use. Data placement is determined by the system, freeing the operators to concentrate on other aspects of their daily tasks.

This chapter provides a high-level overview of FlashSystem A9000 and A9000R. It includes the following sections:

- FlashSystem A9000 and FlashSystem A9000R overview
- FlashSystem A9000R smaller entry-level configuration
- Licensing
- Management tools
- Utility Offering

1.1 FlashSystem A9000 and FlashSystem A9000R overview

FlashSystem A9000 and FlashSystem A9000R bring together the world-class ease of use from IBM Spectrum Accelerate software and the microsecond response times that are provided by IBM FlashCore technology. Designed for enterprise cloud environments, both FlashSystem A9000 and FlashSystem A9000R offer fast and reliable data storage across a wide variety of workloads.

IBM FlashCore technology forms the foundation of FlashSystem A9000. This technology includes many hardware-accelerated I/O features, such as redundant non-blocking crossbar switching, hardware-based encryption, and hardware-based RAID controllers, all designed to deliver consistent microsecond latency response times. IBM FlashCore technology keeps your mission-critical data safe with enterprise-class data protection features, such as IBM Variable Stripe RAID™ (VSR) technology, IBM-engineered error correction codes, overprovisioning capabilities, ultra-fast write, and hardware-based offload data.

The proven IBM Spectrum Accelerate tight integration with cloud automation, local and remote mirroring, advanced data reduction capabilities, and encryption with multi-tenant environments enable FlashSystem A9000 and FlashSystem A9000R to fit the needs of many sizes of business and application workloads with a wide variety of hypervisor and virtualization software, including VMware, OpenStack, and Microsoft solutions. Both FlashSystem A9000 and FlashSystem A9000R use the same software, and both offer onsite setup and service that are provided by IBM. They also share the same feature set, but significant hardware differences exist.

FlashSystem A9000 is machine type 9836 for a 1-year warranty and machine type 9838 for a 3-year warranty. FlashSystem A9000R is a machine type 9835 (1-year warranty) or machine type 9837 (3-year warranty). Concurrent with the 12.2 version software, model number 425 was made generally available in November of 2017. IBM FlashSystem A9000 (9838-U25) and A9000R (9837-U25) storage utility models are available as part of the IBM Storage Utility Offerings. In June of 2018, IBM withdrew all four versions of the Model 415. They can no longer be ordered or purchased, but existing A9000R model 415 systems can continue to be upgraded with additional grid elements.

Version 12.3 introduced a new entry level cost and capacity point for the A9000R. Starting with version 12.3.1, this configuration can be scaled out nondisruptively by adding flash enclosures and grid controllers for higher capacity and performance. Software version 12.3.2, with Hyper-Scale Manager version 5.6 or later, introduces support for VLAN tagging and port trunking (for details, refer to Chapter 2 of *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368).

FlashSystem A9000 and FlashSystem A9000R model 425 incorporate improvements to the grid controllers, as well as using the newest version of the FlashSystem 900, the model AE3. For more information about the AE3, see the IBM Redbooks publication, *Implementing IBM FlashSystem 900 Model AE3*, REDP-8414. New orders of IBM FlashSystem A9000/A9000R Model 425 and U25 systems are shipped with an enhanced grid controller. In this controller, the FC ports are dual-purposed: Using a future software upgrade, customers will be able to connect these ports with servers using FC, or servers using FC-NVMe, or both.

One of the key features in the new FlashSystem 900 is the ability to perform compression. Usually, this capability is unused in the FlashSystem A9000 and FlashSystem A9000R, because they continue to do full data reduction, including compression and deduplication, in the grid controller. The FlashSystem A9000 and A9000R send metadata to the FlashSystem 900, to speed metadata retrieval, and therefore provide a performance increase.

FlashSystem A9000 and A9000R provide extensive business continuity solutions:

- Synchronous and asynchronous mirroring
- HyperSwap functionality
- Multi-site high availability and disaster recovery (HA/DR) solutions
- Transparent online volume migration, known as *Hyper-Scale Mobility*, between A9000 and A9000R systems, and from XIV Gen3 systems to A9000 or A9000R systems.

1.1.1 FlashSystem A9000

FlashSystem A9000 is a fixed configuration, with 6U taken up by three grid controllers that provide compute and interface capabilities, and 2U taken up by a flash enclosure that provides the data storage capacity.

FlashSystem A9000 model number 425

FlashSystem A9000 (9836-425 or 9838-425), which is shown in Figure 1-1, is a dense storage solution that can provide up to 900 TB of effective capacity in 8U. Each of the grid controllers contains two 12-core Intel Xeon E5-2650 v4 processors @ 2.20 GHz and 384 GB dynamic random access memory (DRAM).

Figure 1-1 FlashSystem A9000 (model 425)

Four capacity options exist for FlashSystem A9000 (425), depending on the choice of IBM MicroLatency® module size and the number of modules, as summarized in Table 1-1. Only the smallest MicroLatency modules are offered at a reduced count, to create a lower-cost entry point.

Table 1-1 FlashSystem A9000 (9836/9838-425) capacity options

MicroLatency module usable capacity	Usable capacity[a]	Effective capacity (assumes 5:1 data reduction)
3.6 TB (qty 8)	22 TB	110 TB
3.6 TB (qty 12)	36 TB	180 TB
8.5 TB (qty 12)	85 TB	420 TB

MicroLatency module usable capacity	Usable capacity[a]	Effective capacity (assumes 5:1 data reduction)
18 TB (qty 12)	180 TB	900 TB

a. For the definition of usable capacity or effective capacity, see 4.1.1, “System capacity specifications” on page 80.

1.1.2 FlashSystem A9000R

As shown in Figure 1-2, FlashSystem A9000R (9835-415, 9837-415, 9835-425, or 9837-425) is integrated into an IBM T42 rack, and it can scale up within a single frame. FlashSystem A9000R uses a scale-up building block approach for growth.

Figure 1-2 FlashSystem A9000R

Each building block, which is known as a *grid element*, contains 2x 2U grid controllers, for compute and interface capabilities, with a 2U flash enclosure, for a total of 6U per grid element.

Capacity, caching, and processing power are added at the same time by adding one or more grid elements to scale up, according to the configuration ordered.

FlashSystem A9000R Model number 425

Part of the change that was implemented in the model 425 is both the A9000 and A9000R have the exact same specifications for processors and DRAM. The grid controllers in all versions of the 425 hardware are dual 12 core Intel Xeon E5-2650 v4 processors @ 2.2 GHz and 384 GB DRAM.

The model 425 offers all three capacity points for the MicroLatency modules, however it only scales to four grid elements instead of six for the model 415. With the increased density of the

storage enclosures, the effective capacity of the solution still doubles. See Table 1-2 on page 5 for the usable and effective capacities.

Table 1-2 FlashSystem A9000R (9835/9837-425) capacity options

Number of grid elements	Usable capacity[a] 3.6 TB, 8.5 TB, and 18 TB	Effective capacity (assumes 5:1 data reduction ratio)
2	72 TB 170 TB 360 TB	360 TB 850 TB 1800 TB
3	108 TB 255 TB 540 TB	540 TB 1275 TB 2700 TB
4	144 TB 340 TB 720 TB	720 TB 1700 TB 3600 TB

a. For the definition of usable capacity and effective capacity, see 4.1.1, “System capacity specifications” on page 80.

FlashSystem A9000R smaller entry-level configuration

With the 12.3 code, IBM has introduced a new and smaller entry capacity point commonly called *Grid-Starter* (this is not a formal name). It is essentially an A9000, (3 grid controllers and a Flash enclosure) mounted in the standard A9000R T42 rack. It also includes both infiniband switches. This configuration allows for the same starting capacities as the A9000 (see Table 1-1 on page 3), with the exception of the smallest capacity point, which includes the partially populated FS900.

Starting with software version 12.3.1, this entry-level configuration can be nondisruptively upgraded by adding flash enclosures and grid controllers as required by the target configuration to scale out system capacity and performance. Scale out from the entry point configuration supports target configurations of two, three, and four flash enclosures with four, six, and eight grid controllers, respectively.

Important: It is not possible, nor supported to convert an installed FlashSystem A9000 system to a FlashSystem A9000R entry-level configuration.

1.2 Feature set

FlashSystem A9000 and FlashSystem A9000R provide an all-inclusive software license. All features, including replication, migration, encryption, and data reduction, are included at no additional charge and apply to the entire storage system:

- Data reduction: pattern removal, data deduplication, and compression

  FlashSystem A9000 and FlashSystem A9000R use industry-leading data reduction technology that combines inline, real-time pattern matching and removal, data deduplication, and compression. Compression also uses hardware cards inside each grid controller. Data deduplication applies to data blocks of 8 KB or larger.

  Data reduction is implemented below the global cache to ensure rapid response times, provide a global scope for data reduction services, and allow other data services to be unaffected, including snapshots, replication, and host offload features, such as the VMware vStorage application programming interface (API) for Array Integration (VAAI).

- Multi-tenancy

  FlashSystem A9000 and FlashSystem A9000R enable the secure isolation of logical domains of storage resources among numerous tenants, with the ability to set different quality of service (QoS) levels for each domain. Multi-tenancy enables the division of storage system administration tasks into logical domains, by using role-based permissions. It also enables rapid deployments while it minimizes the need for extensive planning, tuning, or field upgrades.

- VLAN tagging and port trunking

  Starting with software version 12.3.2, support for VLANs in iSCSI environments allows provisionning of private virtual network per tenant. FlashSystem A9000/R VLAN support applies to the data iSCSI ports for hosts and replication, as well as management ports and other Ethernet ports. VLAN support capabilities include VLAN tagging and port trunking as defined by the IEEE 802.1Q specification. Hyper-Scale Manager GUI, version 5.6 or later is required. For more information, refer to Chapter 2 in *IBM FlashSystem A9000/R and IBM XIV Storage System Host Attachment and Interoperability*, SG24-8368.

- Host rate limiting: Quality of service (QoS)

  FlashSystem A9000 and FlashSystem A9000R system resources, such as storage and cache, constitute a virtualized environment that is shared by all hosts and applications. This approach lends itself well to accommodate high-performance requirements for multiple applications with similar performance objectives through fair resource allocation. QoS is available at the domain, pool, host, and volume level.

  In environments with applications with various performance objectives, the QoS feature enables the client to restrict input/output operations per second (IOPS), bandwidth, or both to the correct domain, pool, host group, or volume. QoS can be used to ensure that applications do not use too much of the storage system resources. Therefore, QoS maximizes the resources that are available for applications that require the best performance.

- Intelligent capacity management for deduplication

  Starting with software version 12.3.1, FlashSystem A9000 and A9000R offers patented IBM Research technology that can analyze large amounts of data, without performance impact and provide effective estimates, per volume, for reclaimable capacity, attributed capacity, compression saving, deduplication saving, and total capacity saving. That information is displayed in the Volumes view in the Hyper-Scale Manager GUI (version 5.5.1 minimum is required).

- Fibre Channel (FC) and internet Small Computer System Interface (iSCSI)

  FlashSystem A9000 and FlashSystem A9000R both support the Fibre Channel and iSCSI communications protocols for host attachment and remote mirroring. Newer systems have grid controllers equipped with FC-NVMe adapters. In these new controllers, the FC ports are dual-purposed and NVMe ready: a future software upgrade may enable these ports to connect with servers using FC, or using FC-NVMe, or both.

- Snapshots

  The snapshot capabilities use a redirect on write design that allows snapshots to occur in a subsecond time frame with no performance impact. The system supports multiple differential snapshots of a volume. Any of the snapshots can be made writable. Then, snapshots can be taken of the newly writable snapshots (snapshots of snapshots). Volumes can even be restored from these writable snapshots.

- Synchronous and asynchronous remote mirroring to another FlashSystem A9000 or FlashSystem A9000R. Starting with version 12.2, asynchronous remote mirroring between FlashSystem A9000 or A9000R and XIV Gen3 is enabled.

  Synchronous or asynchronous remote mirroring can be performed over Fibre Channel (FC) or Internet Protocol (IP) iSCSI connections. Both protocols are also supported for two-way mirroring connectivity. Synchronous remote mirroring is used when a zero recovery point objective (RPO) is required. For practical reasons (latency), ensure that the distance is shorter than 100 km (62 miles). For longer distances, asynchronous replication is more appropriate.

  Starting with FlashSystem A9000 Software V12.2.1, up to 3072 sync mirrors and up to 1024 async mirrors are supported.

- HyperSwap

  Starting with version 12.1, FlashSystem A9000 or FlashSystem A9000R can use IBM HyperSwap for high availability solutions. For details, see *IBM HyperSwap and Multi-site HA/DR for IBM FlashSystem A9000 and A9000R*, REDP-5434.

- Multi-site replication

  Using three FlashSystem A9000 and/or A9000Rs, and combining HyperSwap and Asynchronous replication, creates a solution that entails both High Availability (HA) and Disaster Recovery (DR). One side of the HyperSwap pair has an active async link to the third system, and the other side has a standby link. This configuration provides HyperSwap active-active high availability, while keeping data mirrored to a third copy to ensure two levels of business continuity. For more information about Multi-site replication, see *IBM HyperSwap and Multi-site HA/DR for IBM FlashSystem A9000 and A9000R*, REDP-5434.

- VLAN tagging and port trunking

  Starting with software version 12.3.2, support for VLANs in iSCSI environments allows provisionning of private virtual network per tenant. FlashSystem A90000 and A9000R VLAN support applies to the data iSCSI ports for hosts and replication, as well as management ports and other Ethernet ports. VLAN support capabilities include VLAN tagging and port trunking as defined by the IEEE 802.1Q specification. Hyper-Scale Manager GUI, version 5.6 or later is required. For more information, refer to Chapter 2 in *IBM FlashSystem A9000/R and IBM XIV Storage System Host Attachment and Interoperability*, SG24-8368.

- Data migration

  FlashSystem A9000 or FlashSystem A9000R can act as a host, gaining access to volumes on an existing storage system. The system is configured as a proxy to respond to requests between the current hosts and the storage while all existing data is migrated in the background.

- Hyper-Scale Mobility

  IBM Hyper-Scale Mobility allows a volume to be migrated non-disruptively from one FlashSystem A9000 to another over synchronous wide area network (WAN) distances without any host disruption. Starting with version 12.2, Hyper-Scale mobility is enabled for both FlashSystem A9000R and A9000. This capability is in addition to the standard data migration that allows a FlashSystem A9000 system to be a proxy as a host and to migrate volumes from other third-party arrays.

  Starting with Software Version 12.2.1, Hyper-Scale Mobility is also supported from XIV Gen3 (with software level 11.6.2a) to FlashSystem A9000 or A9000R.

  For more information about replication and migration, see the IBM Redbooks publication, *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401.

- Encryption

  FlashSystem A9000 or FlashSystem A9000R helps secure data with industry-standard Advanced Encryption Standard (AES)-256 encryption for data at rest. Encryption is accomplished in hardware to avoid any performance impact. For more information, see the IBM Redbooks publication, *Data-at-rest Encryption for the IBM Spectrum Accelerate Family*, REDP-5402.

- Authentication by using Lightweight Directory Access Protocol (LDAP)

  LDAP can be used to provide user logon authentication, allowing FlashSystem A9000 or FlashSystem A9000R to integrate with Microsoft Active Directory, Open LDAP, or Oracle Java Systems Directory Server. Multiple directory servers can be configured to provide redundancy if one server becomes unavailable.

  For more information, see the IBM Redbooks publication, *Enabling LDAP for IBM FlashSystem A9000 and A9000R with Microsoft Active Directory*, REDP-5387.

- OpenStack and REST support

  FlashSystem A9000 or FlashSystem A9000R can use the well-established IBM code base for OpenStack and Representational State Transfer (REST) API support. For more information, see the IBM Redbooks publication, *Using XIV in Open Stack environment*, REDP-5971.

- VMware synergy

  IBM Spectrum Control™ Base enables a simplified deployment and efficient integration of FlashSystem A9000 and FlashSystem A9000R with the VMware vRealize suite. For more information, see the IBM Redbooks publication, *Using the IBM Spectrum Accelerate Family in VMware Environments: IBM XIV, IBM FlashSystem A9000 and IBM FlashSystem A9000R, and IBM Spectrum Accelerate*, REDP-5425.

- Container support

  IBM Spectrum Connect allows for use of persistent storage for containers in Kubernetes environments. See the IBM Redbooks publication, *IBM Spectrum Connect and IBM Storage Enabler for Containers*, REDP-5470.

1.3 Licensing

As members of the IBM Spectrum Accelerate family, the FlashSystem A9000 and FlashSystem A9000R feature all-inclusive licensing. All the features and functions in the software come with a single license, there is no need for separate licenses for specific features.

IBM provides two licensing options for FlashSystem A9000 and FlashSystem A9000R:

- Enclosure license: The license is ordered with and attached to the specific hardware. When the hardware is no longer being used, the license is also no longer active.
- Customer license: As part of IBM's efforts to build a storage portfolio that supports the movement to software-defined infrastructure, customers can purchase either IBM Spectrum Storage™ Suite software or IBM Spectrum Accelerate software and apply that license to a FlashSystem A9000 or FlashSystem A9000R.

  That software is licensed to the customer and not to the storage, so a customer can use the licensed capacity on multiple storage systems, and move those licenses from old generations to new generations of hardware, removing a concern about stranded licenses.

1.4 Management tools

The graphical user interface (GUI) of FlashSystem A9000 and FlashSystem A9000R redefines storage management by using a simple “connected rings” metaphor to manage many systems. It is installed as part of the Hyper-Scale Manager (HSM) application, which also includes IBM Spectrum Connect, that enables vStorage APIs for Storage Awareness (VASA) functionality, integration with the VMware vRealize suite, and support for containers in a Docker with Kubernetes environment.

HSM is a flexible, consolidated multi-system management application that virtually transforms multiple systems into a single system, allowing customers to centrally manage up to 144 systems. By using Google type-ahead navigation and HTML5, it can be displayed on many customer devices and easily allows control over any FlashSystem A9000 or FlashSystem A9000R functions and reporting, showing not only performance but also compression and data deduplication yields in real time, as effective capacity across all FlashSystem A9000 or FlashSystem A9000R frames.

Starting with Hyper-Scale Manager v5.1, you can manage all members of the IBM Spectrum Accelerate family under the same GUI. Any environment combining FlashSystem A9000, FlashSystem A9000R, XIV Gen3, or IBM Spectrum Accelerate deployments can be managed and configured through a common GUI.

For detailed information, see the IBM Redbooks publication, *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

Hyper-Scale Manager provides the following functions, as illustrated in Figure 1-3:

- Allows integrated management of all FlashSystem A9000, FlashSystem A9000R, XIV Gen3, and other IBM Spectrum Accelerate-deployed systems across the enterprise, including central configuration of user access rights, hosts, and event rules.
- Provides real-time performance monitoring.
- Provides powerful health monitoring by integrating events and alerts across managed systems.

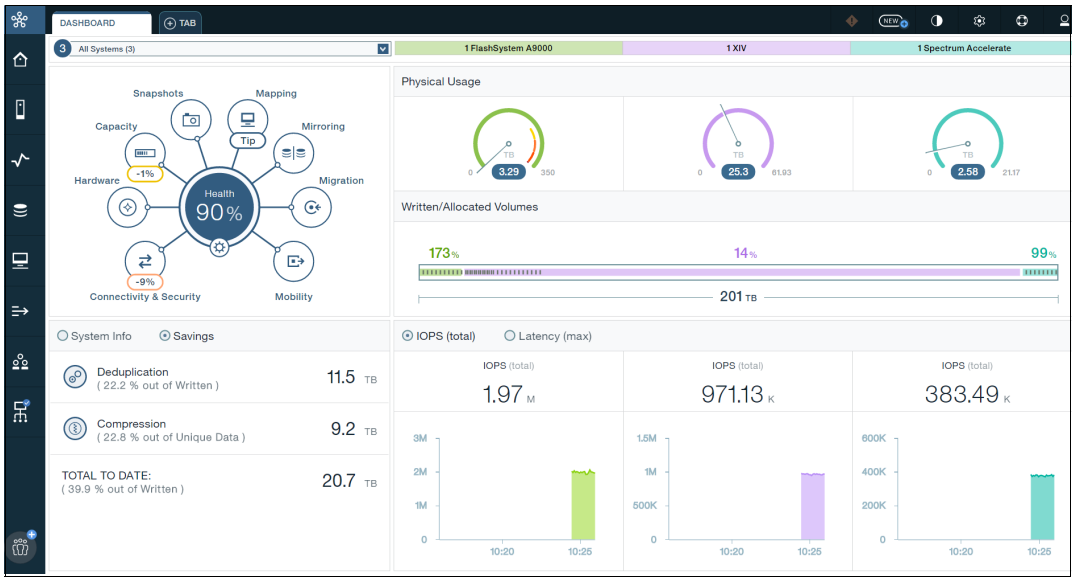


Figure 1-3 Hyper-Scale Manager GUI

- Supports extensive capacity reporting. The system collects capacity-related data over time. Such data can be processed later by analytic applications to reveal insights that can be used to improve capacity planning.
- Automatically open the relevant view and highlight new objects (such as pools, volumes, hosts) after their creation.
- Smart view of object relationships and dependencies in a visual map (for example, a view of all volumes that are mapped to a host) and one-click selection and operation options (map, unmap, delete, and so on)
- Extends management scalability to smartphones and tablets.
- Provides the capability for the Hyper-Scale Manager server to exist as a single instance on a virtual machine server, or on several servers.
- Ability to use wildcards in filter expressions
- Ability to schedule snapshots for volumes and consistency groups on a pool basis
- Saving of user-defined preferences
- Enhanced usability by being able to maximize certain working areas

For detailed information, see the IBM Redbooks publication, *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

1.5 Utility Offering

As of March 2018, along with software version 12.2.1, IBM FlashSystem A9000 (9838-U25) and IBM FlashSystem A9000R (9837-U25) storage utility models are available as part of the IBM Storage Utility Offerings. Although technically identical to Model 425, Model U25 delivers variable capacity usage and billing.

The Utility offering allows for variable billing based on capacity used. A baseline capacity is determined, and the bill remains the same until additional capacity is allocated. Capacity used is calculated on a monthly basis, and billed on a quarterly basis. If capacity used decreases, the additional monthly charge will also decrease.

2

Logical architecture and concepts

This chapter gives an overview of the design concepts, architecture, and technology that make up IBM FlashSystem A9000 systems and IBM FlashSystem A9000R systems.

This chapter includes the following sections:

- Overall architecture
- IBM FlashCore technology
- Data reduction and capacity concepts
- Storage provisioning concepts
- Multi-tenancy
- Data security
- Auditing through syslog

2.1 Overall architecture

The logical architecture of FlashSystem A9000 and FlashSystem A9000R is built on the IBM Spectrum Accelerate software, with many added enhancements to optimize the software stack for use with flash storage.

An important feature and differentiation in this new architecture is the software separation between the computation and storage resources of the product. In terms of the physical architecture, the compute functions are implemented in grid controllers, and the storage function is implemented in flash enclosures.

This separation of the cache, compute, and interface functions from the storage resource separates load balancing across compute and storage resources. It also enables support for a different resiliency scheme in which cache data is triplicated. This cache data protection is unique to FlashSystem A9000 and FlashSystem A9000R.

Another significant enhancement to the underlying IBM Spectrum Accelerate software is the data reduction feature that combines pattern-matching, data deduplication and compression. FlashSystem A9000 and FlashSystem A9000R also offer significant processing power and memory so that they can efficiently address real-time data deduplication and compression.

2.1.1 Parallelism and grid architecture

FlashSystem A9000 and FlashSystem A9000R use a grid architecture, where all grid modules, which are known as *grid controllers*, participate as equal members of the grid. This grid architecture plays an important role in ensuring that all components of FlashSystem A9000 and FlashSystem A9000R have an active role in servicing host requests. Each grid controller performs important roles, which are implemented by specific software functions, which are designated as *nodes*:

- Interface: The interface node processes host I/Os. From the physical standpoint, every grid controller has Fibre Channel (FC) and/or iSCSI cards that can connect to a network for attachment to application hosts.
- Data reduction: The data reduction node implements data deduplication and compression functions. Each grid controller dedicates processing capacity and memory for data reduction.
- Data distribution: The data distribution process ensures that the distribution information is kept up to date for data placement across all flash enclosures by using 16 MB partitions.

  With FlashSystem A9000R, you can order different configurations, by adding grid elements that consist of one flash enclosure and two grid controllers. The data is automatically distributed over all existing flash enclosures to ensure the even usage of the flash enclosure capacity and workload.

Note: FlashSystem A9000 consists of one flash module and three grid controllers, and it cannot be scaled up. However, you can distribute different applications' workloads onto different FlashSystem A9000s, all of which are managed through the same Hyper-Scale Manager instance. Furthermore, the Hyper-Scale Mobility function allows you to do an on-line migration of a volume between A9000 systems transparently to the host system.

- Caching: The cache node implements and manages caching functions. Data reduction is accomplished below the cache in FlashSystem A9000 and FlashSystem A9000R so that each grid module has significant read and write cache to allow for consistent microsecond response times.
- Replication: Grid controllers also share responsibility for replicating data between systems.

As depicted at a high level in Figure 2-1, the various nodes are instantiated and run in parallel on each grid controller. This design enables multiple main threads for higher performance and scalability by adding grid controllers in FlashSystem A9000R. Resiliency and system availability are also reinforced because all functions and data of any failed node on a controller will automatically fail over and be redistributed to another similar node on a different grid controller.

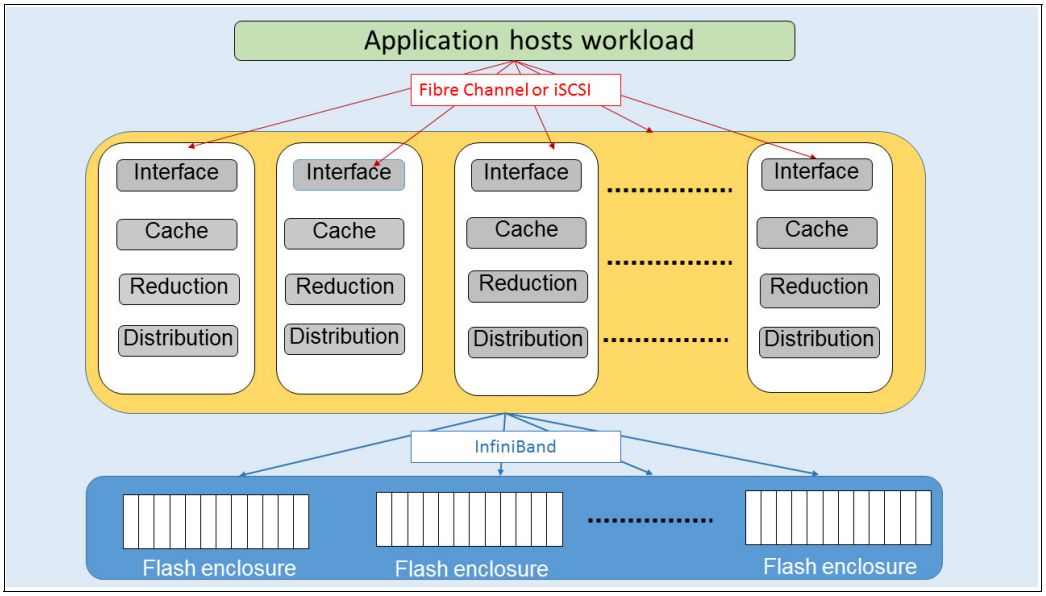


Figure 2-1 Parallelism and grid architecture

2.1.2 Cache resiliency

The cache and cache management design enables decoupling between computation functions (typically, caching and metadata management) and storage resources.

To ensure resiliency, the system is creating multiple copies of the data in cache. Each host data block has a primary cache in a grid controller that is assigned as the primary module, and two additional backup caches in two different grid controllers, as depicted in Figure 2-2 on page 14. (The data reduction nodes are not shown.)

- When information is written to FlashSystem A9000 or FlashSystem A9000R, the primary module ensures that data is also written to both backup modules. The write is acknowledged to the host, only when all three copies are in place.

This design enables FlashSystem A9000R to survive two simultaneous grid controller failures while it ensures data availability.

FlashSystem A9000 with just three grid controllers will remain operational after it loses one grid controller. In the event of dual grid controller loss in the FlashSystem A9000, the system will be offline, but all data is protected and will be available when grid controller function is restored.

Figure 2-2 shows the cache resiliency design.

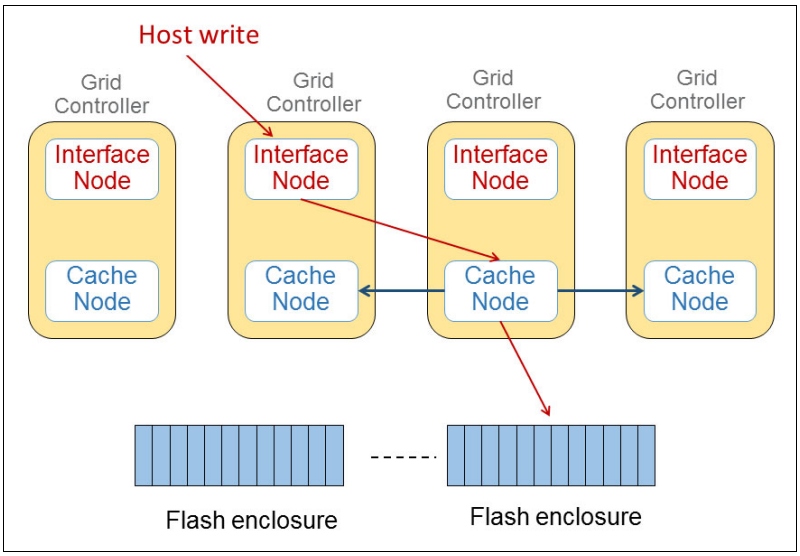


Figure 2-2 Cache resiliency design

In a grid controller failure, the system ensures that another module is selected to take over. The system rebuilds the cache to get back to three copies of each block.

Note: On FlashSystem A9000, with only three grid controllers, the system must wait for the failed controller to be operational again before it can rebuild the third copy of the cache.

2.2 IBM FlashCore technology

The term *IBM FlashCore* refers to the various technologies that are used in IBM FlashSystem A9000 and FlashSystem A9000R flash enclosures to achieve the lowest latency, highest throughput, and longest life span. The data path inside the flash enclosure is hardware that is *accelerated*, which means that no processing or CPU is consumed in the data path. This design enables the lowest latency and highest throughput.

The flash enclosure contains 12 MicroLatency modules, with the exception of the smallest FlashSystem A9000 425, which only has 8. A RAID 5 with a hot spare is built over all MicroLatency modules to protect the flash enclosure from a MicroLatency module failure.

The flash chips inside the MicroLatency module are protected by the IBM Variable Stripe RAID (VSR). VSR protects the MicroLatency module against chip failures or partial chip failures. VSR monitors the health of the flash media, detects and manages flash failures efficiently, and optimizes the usage of all flash resources. With VSR, a flash failure does not result in the need for a maintenance event. No flash module replacement is necessary. The MicroLatency module is not degraded in any way. VSR protects against expected and unexpected failures inside a flash chip.

The combination of the RAID 5 over the MicroLatency module and the VSR inside a MicroLatency module is called two-dimensional RAID protection (2D RAID). Advanced flash management ensures the maximal availability and lifetime of the MicroLatency modules.

Details about the hardware-accelerated I/O, MicroLatency modules, advanced flash management, and 2D RAID are described in 3.7, “IBM FlashCore technology” on page 67.

2.3 Data reduction and capacity concepts

Data reduction is not a new a concept. However, it is now an essential feature of storage systems because data is growing at increasingly fast rates due in part to the growth of mobile platform usage, social media, and big data. Additionally, the analytics that many businesses use to extract business value from these ever-growing sources of data result in yet more data.

FlashSystem A9000 and FlashSystem A9000R perform always-on, inline data reduction that provides application hosts with the capability to provision virtual storage capacity that is much greater than the actual physical capacity. FlashSystem A9000 and FlashSystem A9000R deliver microsecond performance benefits of flash storage at a lower cost per gigabyte.

2.3.1 Data reduction technology

Data reduction in FlashSystem A9000 and FlashSystem A9000R uses the following processes:

- Known pattern matching
- Data deduplication
- Compression

The data reduction runs in a data reduction node in each controller (see Figure 2-1 on page 13), and data reduction occurs after the host write has been acknowledged, or below cache.

Below cache: Placement below the cache means that no latency impact occurs for cached I/Os. The system functions, such as migration, snapshots, and VMware vStorage application programming interface (API) for Array Integration (VAAI) operations are not affected.

Pattern matching and removal

This first layer of data reduction comes from pattern matching.

Pattern matching mechanisms match incoming host writes with a preconfigured set of known patterns that are stored in the system.

When a write is processed, it is split into 8 KB blocks, as shown in Figure 2-3.

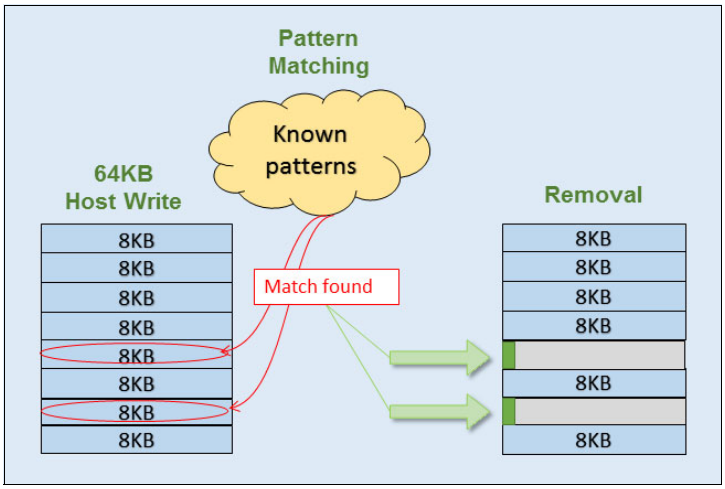


Figure 2-3 Pattern matching and removal

Then, each block is hashed, and the hash value, which is also known as a *fingerprint*, is compared to a table of well-known hashes. If a match is found, the corresponding pattern ID, which is only 2 bytes (little green rectangle in Figure 2-3) is stored.

Data deduplication

Data deduplication is the ability to store data only once, although it can be written many times by various hosts or applications.

The data deduplication mechanism identifies identical blocks of data and stores only one copy of that data in the system. All other identical blocks point to that copy.

In Figure 2-4, each color represents unique data. Every square represents an 8 KB block. The system can detect duplicates, and it stores only one copy of the duplicate 8 KB blocks. For duplicates, Figure 2-4 shows that only the pointers to the data are stored in the system.

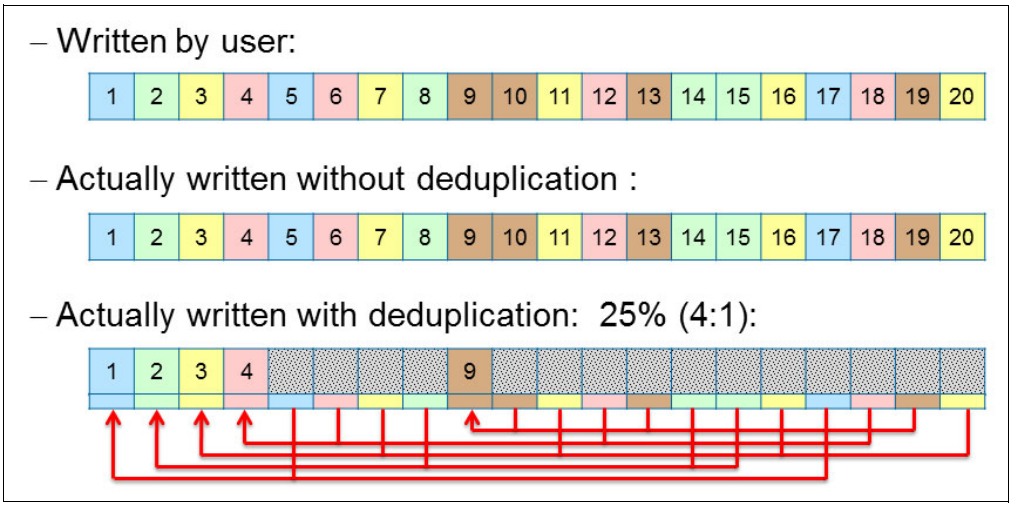


Figure 2-4 Data deduplication principle

In FlashSystem A9000 and FlashSystem A9000R, each 8 KB block for which a duplicate exists is replaced by a pointer to the hash of the duplicate, as shown on the right side of Figure 2-5. Notice that the green block for dedupe is larger than the green block for pattern matching. This is because more metadata has to be stored for deduplication than for pure pattern matching.

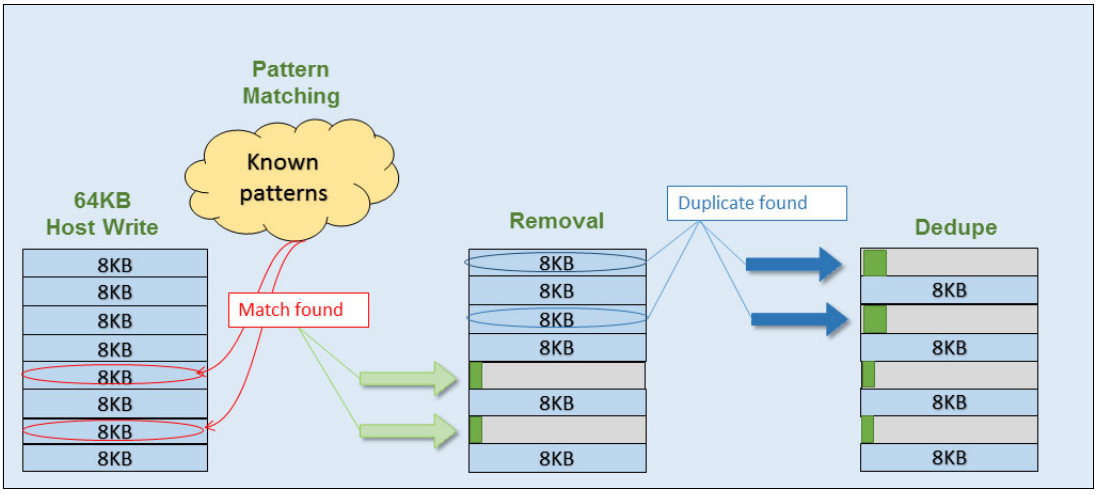


Figure 2-5 Data deduplication in FlashSystem A9000 and FlashSystem A9000R

Whenever a new unique block is found, a new hash is created and stored in a repository. Any future 8 KB writes' hash is checked against the repository for a match.

Important: Deduplication only applies to data blocks of 8 KB or larger.

Data deduplication is performed in sequences, and the system stores hashes in a memory construct, which is known as a *segment*. Each hash (data) has an owning segment, and a certain segment can also contain references to a hash that it owns, or references to a hash in another owning segment. See Figure 2-6. The owning segment of a referenced hash is indicated by the corresponding background color.

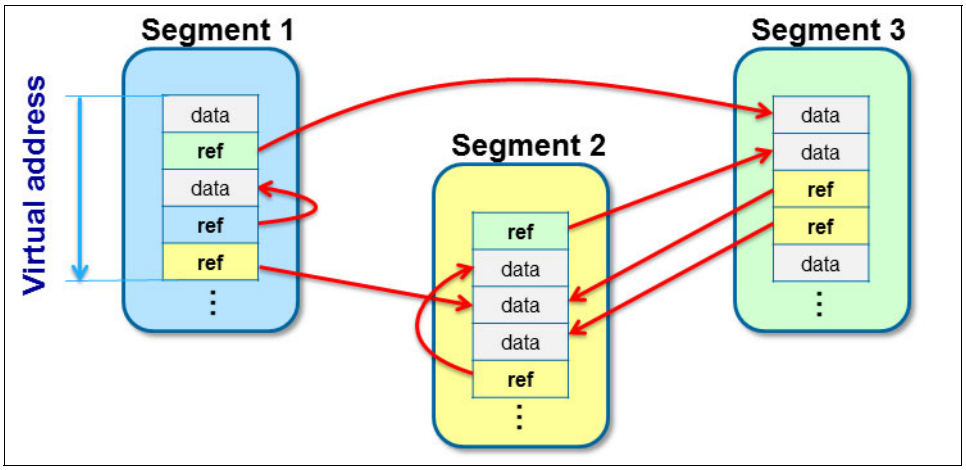


Figure 2-6 Hashes and references in segments

Segments maintain a list of other segments that they created references to recently. Therefore, when the system looks for a match, the recent segments are checked first, which typically speeds up the matching process.

As illustrated in Figure 2-7, the data deduplication of the 8 KB blocks is performed over a 4 KB alignment, which increases the probability of finding a match, resulting in a higher data deduplication percentage.

Note: The 4 KB alignment augments the probability to find a match for deduplication within 8 KB blocks.

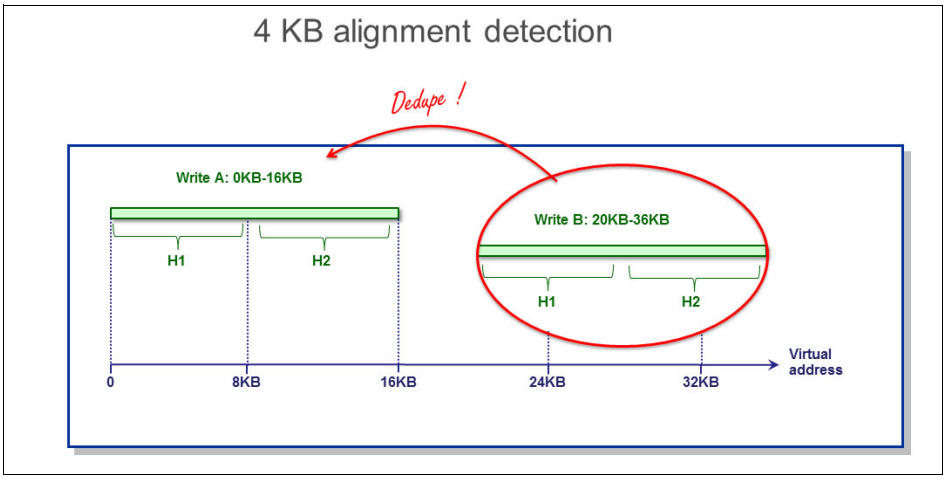


Figure 2-7 Data deduplication with 4 KB alignment detection

Maintaining integrity

Integrity of user data is a major aspect of the system design. There are several data integrity concerns when deduplication is involved, including the following:

- When creating a deduplication reference, the reference must be created to the correct data.
- When reading the data, the read path over the reference must fetch the correct data.
- When deleting data, the data must not be deleted as long as there are any references to this data anywhere in the system.

The first measure of protection in the FlashSystem A9000 and A9000R code design is a dual-layered CRC check. The first CRC is on user data as it enters the system and before compression modifies the data. The second CRC covers what is actually written to the storage after data reduction, and covers both data and metadata.

The A9000 or A9000R uses the industry standard 160 bit SHA1 to fingerprint user data. References are created and defined by their SHA1. The SHA1 is stored in every reference in addition to the SHA1 stored with the data itself. In other referencing methods, such as using ID number or position, there is risk of reading incorrect data if there is a problem in the management of the ID or position. Using the SHA1 as the reference avoids any such potential problems.

Another mechanism protects the reference counters. This mechanism modifies the counter in a transactional manner that is capable of surviving any type of failure, including crashes and communication failures.

As a last resort, the A9000 or A9000R includes a unique offline recovery capability that scrubs the data and reconstructs the metadata, including references and reference counters.

Compression

Finally, data moves on to the compression step (Figure 2-8) for more data reduction.

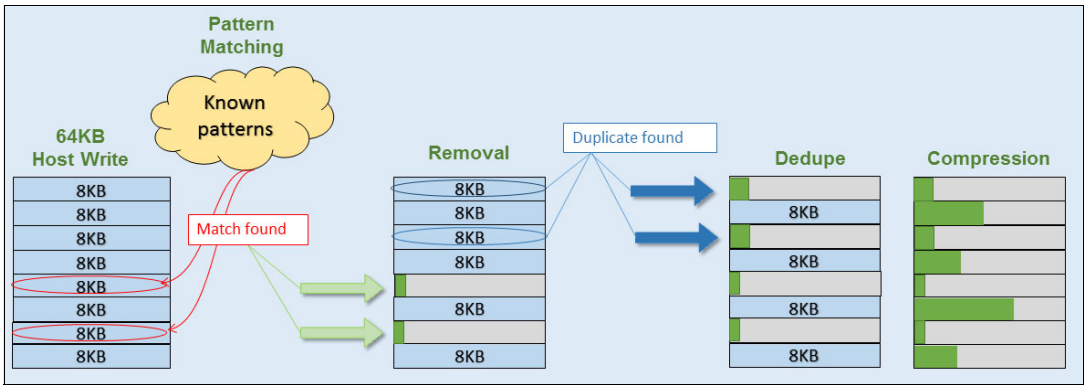


Figure 2-8 Compression

For details about the compression phase, see “Compression technology” on page 21. Figure 2-9 summarizes the data reduction process flow, as implemented in IBM FlashSystem A9000 and FlashSystem A9000R.

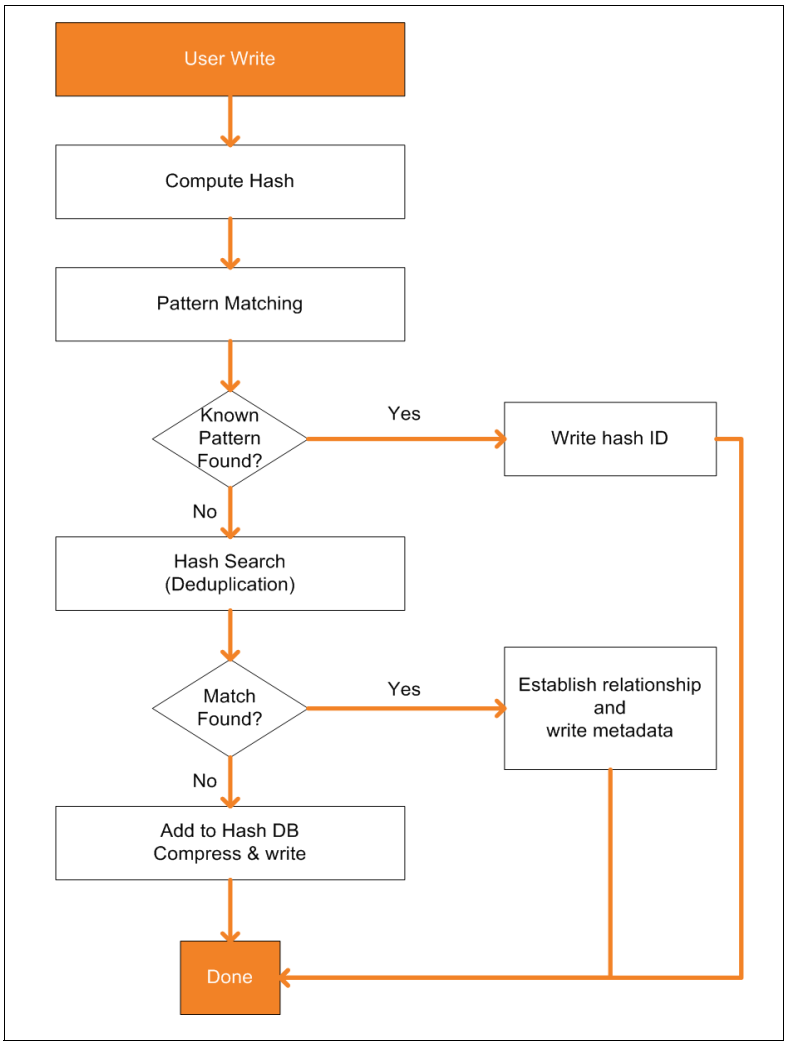


Figure 2-9 Data reduction process flow

Figure 2-10 shows the data reduction data flow. The following list describes the data reduction data flow. (The numbers correspond to the numbers in Figure 2-10.)

1. Write I/O from the host is received by the interface controller. Three copies of data are placed in separate grid controller caches.
2. The write I/O is acknowledged to the host only after the replication of cache is completed.
3. The data in cache is sent to the data reduction engine for data reduction where pattern matching, data deduplication, and compression mechanisms are applied.
4. Data is reduced and then written back to cache.
5. Reduced data is destaged to the flash enclosure.

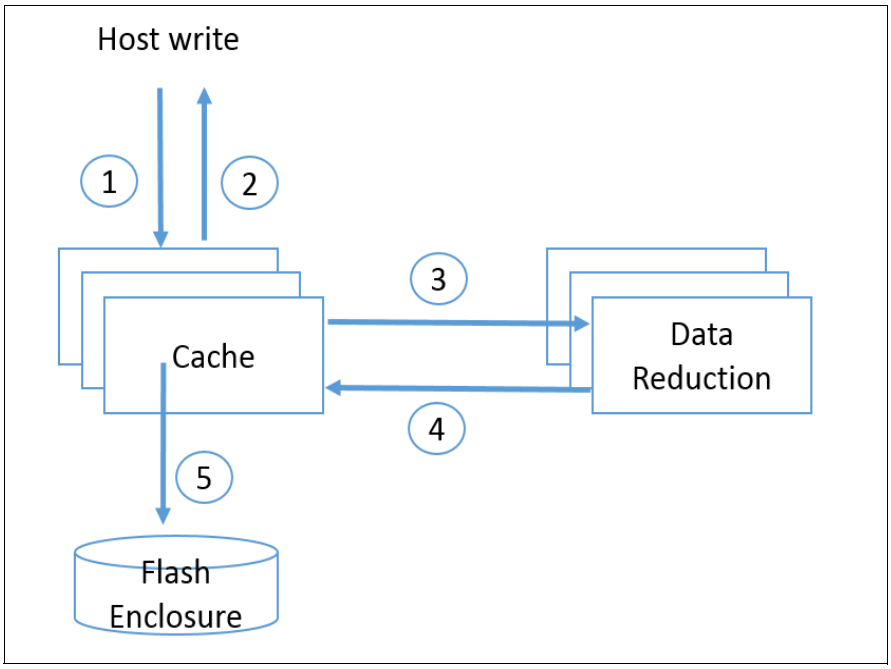


Figure 2-10 Data reduction data flow

Important: The data reduction happens below cache, which means that data is cached before the data reduction, which minimizes latency.

The data reduction is distributed across all of the grid controllers. Most of the system memory is used as read/write cache to absorb and buffer I/O requests and to minimize I/O to the back end.

In addition, compression is assisted by hardware. Two hardware acceleration cards per grid controller are available to process compression activities.

Reduced data is maintained and managed by the system into a fixed number of objects that are called *compression objects*.

Data reduction usage and savings can be monitored by using IBM Hyper-Scale Manager.

Note: For more information about how physical and effective capacity is represented in the system, see Chapter 4, “Capacity planning and management” on page 79.

Compression technology

Compression in FlashSystem A9000 and FlashSystem A9000R is assisted by hardware. Two Intel Coleto cards are included in each grid controller.

The IBM patented compression technology that is used by FlashSystem A9000 and FlashSystem A9000R is based on the *zlib* data compression algorithm that operates in a real-time method.

Traditional versus IBM compression

The traditional approach that is taken to implement data compression in storage systems is an extension of how compression works in compression utilities.

Figure 2-11 shows an example of how the data is broken into fixed size blocks (in the upper-left side of the figure). It also shows how each block is compressed independently into blocks with potentially different sizes (in the upper-right side of the figure). The resulting compressed blocks are stored sequentially in the compressed output.

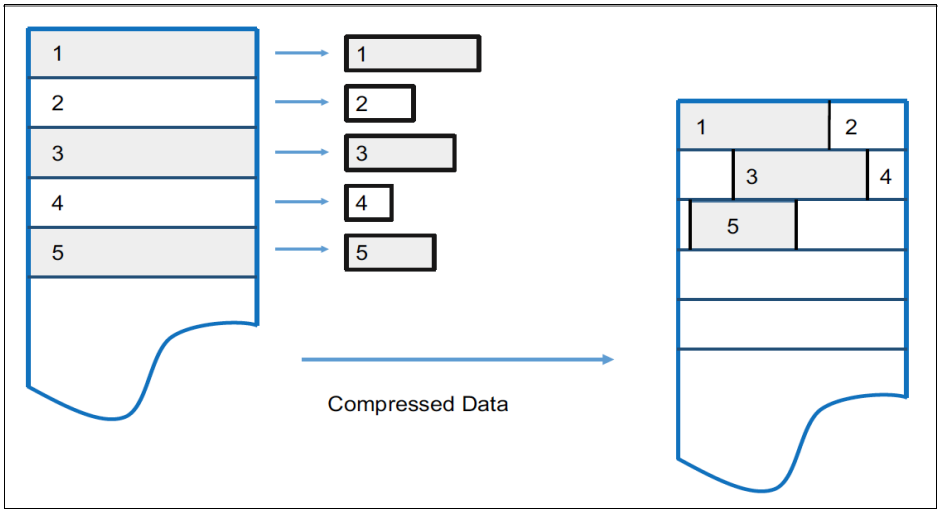


Figure 2-11 Fixed size blocks

Variable size input: The IBM patented compression technology alters the traditional approach to compression. It uses variable-size blocks for the input and fixed-size, 32 KB blocks for the output.

A major difference between traditional compression and the IBM compression technique is in the size of data blocks that are written to the storage device.

The IBM compression technique uses fixed-size writes.

Fixed-size writes: Traditional compression writes variable size blocks to the storage device. IBM compression writes fixed-size output blocks. This difference is the basis of many of its benefits.

This method enables an efficient and consistent method to access the compressed data because it is stored in fixed-size blocks.

Figure 2-12 shows compression that uses a sliding window.

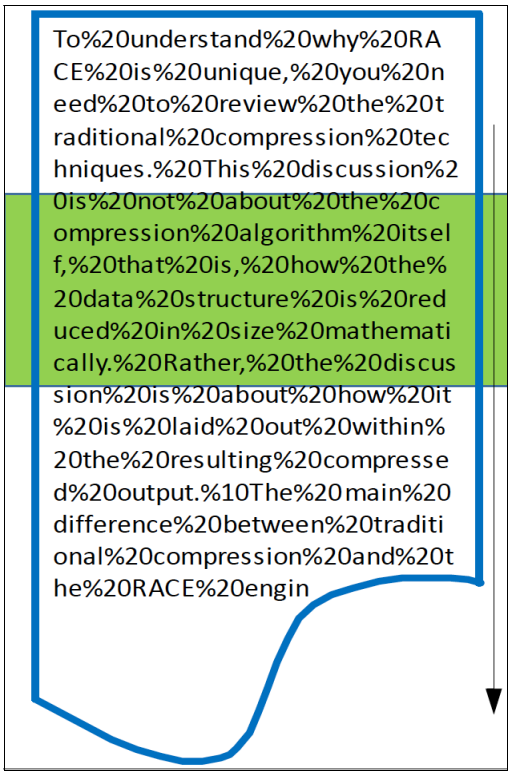


Figure 2-12 Sliding window

Figure 2-13 shows IBM compression with fixed output blocks.

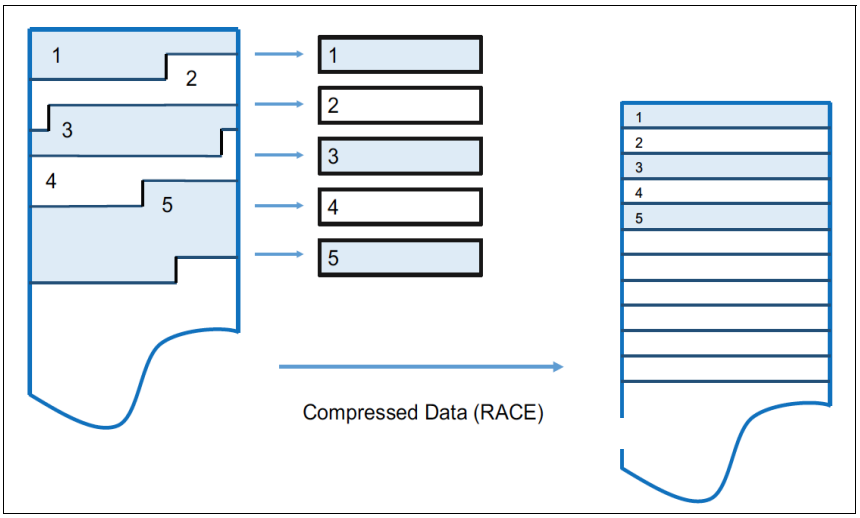


Figure 2-13 IBM Compression with fixed output blocks

Temporal locality

The compression ratio of a data block depends on how many repetitions can be detected within the block. The number of repetitions is affected by how much the bytes that are present in the block relate to each other. The relationship between bytes is driven by the format of the object. For example, an office document might contain textual information and an embedded drawing (such as Figure 2-13 on page 22). Because the division of the file into blocks is arbitrary, it does not know how the data is laid out within the document.

Therefore, a compressed block can be a mixture of the textual information and part of the drawing. This process yields a lower compression ratio because mixing the different data types causes a suboptimal dictionary of repetitions. That is, fewer repetitions can be detected because a repetition of bytes in a text object is unlikely to be found in a drawing.

This traditional approach to data compression can also be called *location-based compression*. The data repetition detection is based on the location of data within the same block. When host writes arrive to IBM Random Access Compression Engine (RACE), they are compressed and fill fixed-size blocks that are called *compressed blocks*. Multiple compressed writes can be aggregated into a single compressed block.

A dictionary of the detected repetitions is stored within the compressed block. When applications write new data or update existing data, it is typically sent from the host to the storage system as a series of writes. Because these writes are likely to originate from the same application and be of the same data type, more repetitions are detected by the compression algorithm. This type of data compression is called *temporal compression* because the data repetition detection is based on the time that the data was written into the same compressed block.

Temporal compression: Temporal compression adds the time dimension that is not available to other compression algorithms. It offers a higher compression ratio because the compressed data in a block represents more homogeneous input data.

Figure 2-14 shows (in the left part of the figure) how three writes (listed in the figure by write number) that are sent one after the other by a host end up in different blocks. They are compressed in different blocks because their location in the volume is not adjacent. This approach yields a lower compression ratio because the same data must be compressed non-natively by using three separate dictionaries. When the same three writes are sent through RACE (in the right part of the figure), the writes are compressed together by using a single dictionary. This approach yields a higher compression ratio than location-based compression.

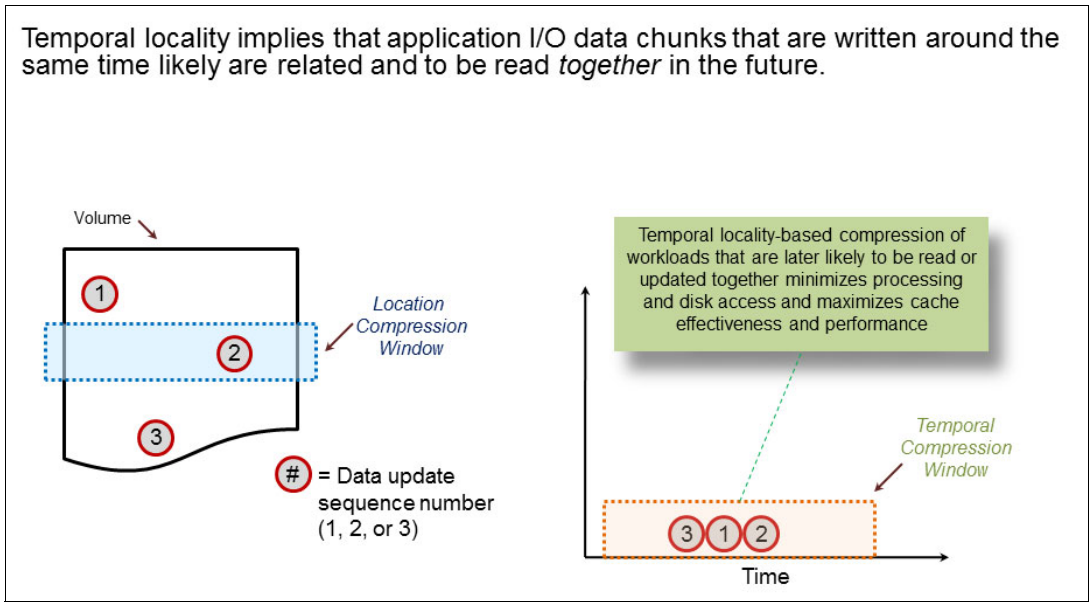


Figure 2-14 Temporal locality

Applications often read those three writes back with temporal correlation, in which case the read benefits from being written together. That is, when this data is read back, the read stream generally follows the same random (non-contiguous volume logical block address) pattern. Therefore, the compression engine reads and extracts the larger block of data, which results in the next few random volume I/O reads by the host.

This data is read from the data that was already extracted by extracting the first large block. Therefore, this process results in what is essentially a cache hit in the compression cache memory.

With real-world applications, in general, truly random I/O does not exist. The reality is that an application reads and writes objects or groups of data. These groups of I/O requests form a repeatable pattern, with the same group of I/O occurring one after another, even if they are written to random locations. IBM invested heavily in understanding these patterns, and IBM Real-time Compression™ uses this understanding to maximize compression ratios and return the best performance.

2.4 Storage provisioning concepts

In this section, we describe how the available physical storage is virtualized and effectively managed in FlashSystem A9000 and FlashSystem A9000R.

FlashSystem A9000 and FlashSystem A9000R implement full storage virtualization, which offers the following benefits:

- Logical volume placement is driven by the distribution algorithms, freeing the storage administrator from planning and maintaining volume layout. The data distribution algorithms manage all of the data in the system collectively without deference to specific logical volume definitions.
- Any interaction, whether it is driven by the host or driven by the system, with a specific logical volume in the system is inherently handled by all resources. The system automatically harnesses all storage capacity, all internal bandwidth, and all processing power that are currently available in the system.
- Consistent performance and scalability. Virtualization algorithms automatically redistribute the logical volumes' data and workload when new hardware is added, maintaining the system balance while it preserves transparency to the attached hosts.

2.4.1 Storage pools

The available capacity in the system can be administratively portioned into separate and independent storage pools. *Storage pools* form the basis for controlling the use of storage space by imposing a capacity quota on specific applications, a group of applications, or departments. This quota enables the isolated management of relationships within the associated group of logical volumes and snapshots. Storage pools can also be uniquely associated with a domain. (See 2.5, "Multi-tenancy" on page 27.)

Tip: In FlashSystem A9000 and FlashSystem A9000R, the minimum pool size is 103 GB.

A storage administrator can create volumes (with the reserved snapshot space) whose total capacity is less than or equal to the size of the pool. In the background, the data reduction methods of deduplication and compression are reducing the amount of data that is written to the MicroLatency module, enabling effective use of the available capacity.

Tip: With Version 12.3, a pool can be system-wide. In other words, a single pool can now utilize the full system effective capacity, up to the system allocation limit.

For FlashSystem A9000 Model 425 and 25U, the system allocation limit is 1200 TB.

For FlashSystem A9000R Model 425 and 25U, the system allocation limit is:

- 1200 TB, for a system with 3 grid controllers and 1 Flash enclosure (Grid Starter)
- 2400 TB, for a system with 4 grid controllers and 2 Flash enclosures
- 3600 TB, for a system with 6 grid controllers and 3 Flash enclosures
- 4800 TB, for a system with 8 grid controllers and 4 Flash enclosures

2.4.2 Volumes

A *volume* is defined within the context of only one storage pool. Because storage pools are logical constructs, a volume and any snapshots that are associated with it can be moved to any other storage pool (within the same domain) if space is sufficient within the target storage pool.

As a benefit of the system virtualization, no limitations exist on the associations between logical volumes and storage pools. In fact, the manipulation of storage pools consists exclusively of metadata transactions and it does not trigger any copying of data. Therefore, changes are completed instantly and without any system performance degradation.

FlashSystem A9000 and FlashSystem A9000R use the grid concept and distribute volume data evenly across hardware storage resources. Volumes are distributed evenly across all flash enclosures by using partitions, and each partition is 16 MB.

The system also uses the concept of an *allocation unit* (AU) size for volumes, which is set at 103 GB.

The minimum volume size that can be created in FlashSystem A9000 or FlashSystem A9000R is 1 GB. However, volumes that are created within 5% of an AU boundary are rounded up to the AU size. This rounding up explains why when you can create a 99 GB volume, 103 GB is used. See Figure 2-15.

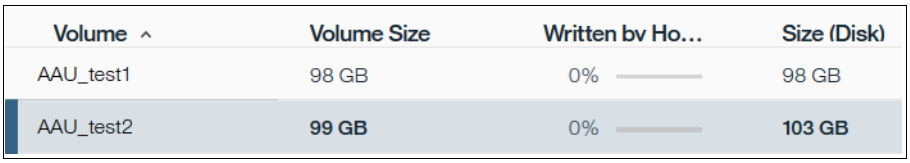


Figure 2-15 Rounding up to Allocation Unit size

2.4.3 Snapshots

A *snapshot* is a point-in-time copy of a volume's data, and it is contained within the same storage pool as the source. As implemented in FlashSystem A9000 and FlashSystem A9000R, snapshots have minimal impact on system performance. When an update is performed to the original data, the update is stored in a new partition and a pointer of the original volume now points to the new partition. However, the snapshot volume still points to the original partition. This method is called *redirect-on-write*.

Because snapshots require capacity because the source and the snapshots differ over time, space for snapshots must be set aside when you define a storage pool. A minimum of 400 GB of snapshot space must be allocated.

FlashSystem A9000 and FlashSystem A9000R implement an automatic snapshot deletion mechanism to protect itself from overutilizing the snapshot space. Each snapshot has a deletion priority property that is set by the user. If the snapshot space is full, any duplicate snapshot is deleted first, even though the original snapshot is older.

2.4.4 Consistency groups

A *consistency group* is a group of volumes of which a snapshot can be made at the same point in time, ensuring a consistent image of all volumes within the group at that time. The concept of a consistency group is common among storage systems in which it is necessary to perform concurrent operations collectively across a set of volumes so that the result of the operation preserves the consistency among volumes.

For example, effective storage management activities for applications that span multiple volumes or the creation of point-in-time backups is not possible without first employing consistency groups.

This consistency among the volumes in the group is paramount to maintaining data integrity from the application's perspective. By grouping the application volumes into a consistency group, you can later capture a consistent state of all volumes within that group at a specified point-in-time by using a special snapshot command for consistency groups (creating a snapshot group).

Issuing this type of a command results in the following process:

1. Complete and destage writes across the constituent volumes.
2. Suspend I/O activity simultaneously across all volumes in the consistency group.
3. Create the snapshots (snapshot group).
4. Resume normal I/O activity across all volumes.

IBM FlashSystem A9000 and FlashSystem A9000R manage these suspend and resume activities for all volumes within a consistency group.

New, with software V12.2.1 and later, you can add a volume to a CG that already contains a snapshot group, and then take a new snapshot of the CG with a refresh option. The new volume will get a new snapshot while the snapshots of the rest of the volumes in the CG are refreshed.

Consistency: Additional mechanisms or techniques, such as the techniques that are provided by the Microsoft Volume Shadow Copy Service (VSS) framework, might still be required to maintain full application consistency from a host system's perspective.

2.5 Multi-tenancy

IBM FlashSystem A9000 and FlashSystem A9000R can easily and securely support the provisioning, administration, and service-level management of storage resources to be used concurrently for multiple tenants or customers. The multi-tenancy of FlashSystem A9000 and FlashSystem A9000R brings flexibility and simplicity to the management of tenant data and storage resources across multiple FlashSystem A9000 and FlashSystem A9000R systems through these methods:

- Secure logical division and isolation of IBM FlashSystem A9000 and FlashSystem A9000R storage pools, volumes, consistency groups, mirrors, data migrations, and quality of service performance classes among numerous tenants
- Simple, quick delegation of administration tasks and role-based permissions
- Simple, rapid deployment without the need for extensive planning and tuning, and the capability to scale out FlashSystem A9000R at the customer site

2.5.1 Domains

IBM FlashSystem A9000 and FlashSystem A9000R multi-tenancy is based on the concept of *domains* where FlashSystem A9000 or FlashSystem A9000R is logically partitioned into one or more independent containers, each with its own assigned administrators. This concept enables secure isolation from the other domains of the logical entities that are contained within a domain.

A user who is associated with a single domain has no knowledge of the other domains that exist on the system or about the pools or volumes that are associated with those domains. Domains can be associated with these entities:

- Users and user groups
- Storage pools (and, inherently, the volumes that they contain)
- Hosts and clusters
- Remote mirror targets

Note: Although a storage pool (and the volumes that it contains) can be associated with only a *single* domain, users (and user groups), hosts (and host clusters), and remote mirror targets can be associated with *multiple* domains.

Conceptually, a user's interactions (for example, the interactions of a storage administrator) with FlashSystem A9000 and FlashSystem A9000R can be viewed as *actions* that are performed upon *objects*. The following examples show these object/action pairs:

- Creating a storage pool
- Resizing a volume
- Mapping a volume to a host
- Viewing the properties of a pool or volume

FlashSystem A9000 and FlashSystem A9000R use a role-based access control (RBAC) model to control *what actions* a specific system user can perform. Multi-tenancy, by using domains, introduces the ability to control *what objects* a specific system user can perform those actions upon.

Note: When you consider multi-tenancy, do not confuse the use of the term *domain* with a fully qualified domain name (FQDN) or directory services domain. In the context of multi-tenancy, a domain is a logical construct that allows partitioning (logically, not physically) of FlashSystem A9000 and FlashSystem A9000R resources.

It is important to understand that the implementation of this domain construct within FlashSystem A9000 and FlashSystem A9000R occurs at the management level. All physical system resources (flash capacity, CPUs, memory, and interfaces) are shared among domains. As such, domain administrators cannot modify the physical system resources.

However, they can be notified of events that relate to physical system attributes because these events might affect the objects within their domain. For example, a domain administrator can be alerted about a hardware failure or the disconnection of an interface link.

Domain policy

An open or closed *domain policy* can be defined on FlashSystem A9000 and FlashSystem A9000R. This policy is set for the system and not for individual domains. This policy affects the behavior of the global administrator user (a user that is not associated with a domain and that has access rights).

If the domain policy is set to *open*, a global administrator can reach into any domain. For example, this global administrator user can manage pools and volumes.

If the domain policy is set to *closed*, everything inside the domain is hidden to the global administrator. For example, the global administrator cannot see or change pools or volumes.

Domain creation

The creation of domains within IBM FlashSystem A9000 and FlashSystem A9000R is performed by a *global storage administrator*, which is a user with the role of storage administrator who is not associated with any domains. By default, the built-in *admin* user account is a global storage administrator.

When the global administrator creates a domain, the global administrator assigns system resources to that domain. Certain resources within FlashSystem A9000 and FlashSystem A9000R have finite limits. Examples include the storage capacity and the number of volumes. When the global administrator creates a domain, the global administrator must determine the quantity of these finite resources to assign to the domain.

For more information about domain creation and the assignment of system resources to a domain, see the IBM Redbooks publication, *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

Domain attributes

Domains include these important characteristics:

- A *domain* is a logical partition of the system's resources. It represents a subset of the system's resources. These resources include, but are not limited to, storage pools, hosts, and mirror targets.
- Users can be assigned to zero or more domains. A user that is assigned to zero domains is considered a global user and has access to all system resources that are not associated exclusively with a domain.
- A domain restricts the resources that a user can manage. A user can manage only the parts of the system that are associated with the domains that the user is associated with, as depicted in Figure 2-16.

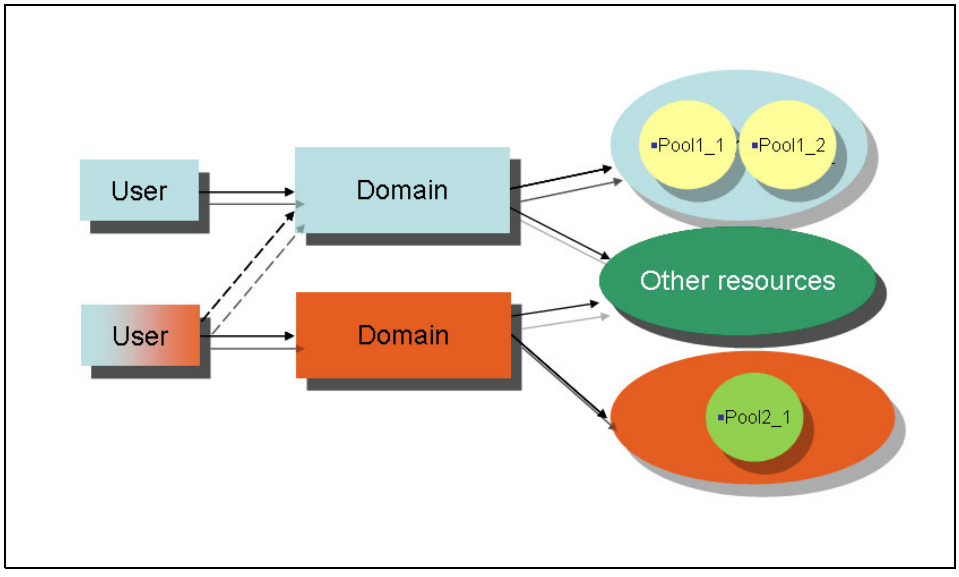


Figure 2-16 User view of domains and resources

- A *domain administrator* is a user who is associated with a domain or domains. The domain administrator is restricted to performing operations on objects that are associated with a specific domain or domains that the domain administrator is associated with.

  A domain administrator can be assigned to manage multiple domains, and a domain can be assigned to be managed by multiple administrators.
- No visibility exists between domains. Domain administrators are not informed of resources outside of their domains. These resources and their related events or alerts are not displayed on lists.
- Within a domain, the storage, security, and application administrators, and read-only users, retain their rights to perform the same operations that they have in an environment that is not based on domains. However, the operations are limited to the resources within the domain.
- Storage that is allocated to a domain can be used to create pools within that domain.
- A pool (and its volumes) can be associated with only one domain.
- Hosts, clusters, and targets can be associated with domains in a non-exclusive (shared) manner.
- Resources that are not associated with any domain are accessible only to global administrators.

Quality of service (QoS) can be set independently for each domain, and it can be used to limit the bandwidth or input/output operations per second (IOPS) that is allowed for a domain or for the storage pools within a domain.

2.6 Data security

FlashSystem A9000 and FlashSystem A9000R provide optional data-at-rest encryption. Data-at-rest encryption is required for regulatory compliance in many environments, such as the healthcare and finance industries.

Data-at-rest encryption protects the data that is stored on the grid controller solid-state drives (SSDs) and flash enclosure MicroLatency modules against unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording, or destruction. Data-at-rest encryption protects the data if the SSDs or flash enclosure MicroLatency modules are stolen or improperly discarded.

Encryption of FlashSystem A9000 and FlashSystem A9000R operating system and metadata is not required, so this data and metadata are not encrypted. The following components are encrypted when encryption is enabled:

- The self-encrypting SSDs, which are located as a pair in each grid controller and serve as a vaulting device for user data. *Vaulted data* is the RAM resident data, such as data reduction information, cache metadata, and cached writes, which will be written to the SSDs when a shutdown of the grid controller is initiated. See “Solid-state drive (SSD) vault devices” on page 48.
- FlashSystem A9000 and FlashSystem A9000R flash enclosure MicroLatency modules. Each MicroLatency module has Field Programmable Gate Array (FPGA) control data-at-rest encryption. Enabling encryption has no performance impact. Data at rest is protected by an Advanced Encryption Standard (XTS-AES) algorithm that uses the 256-bit symmetric option in Xor-encrypt-xor (XEX)-based tweaked-codebook mode with ciphertext stealing (XTS) mode, as defined in the IEEE1619-2007 standard.

Encrypting these two components will encrypt all user data.

An HMAC-SHA256 algorithm is used to create a hash message authentication code (HMAC) for corruption detection, and it is protected by a system-generated cyclic redundancy check (CRC).

2.6.1 Key management

IBM FlashSystem A9000 and FlashSystem A9000R SSDs and flash enclosures have self-encrypting capabilities. To provide centralized and simplified key management and the separation of key storage from data storage, the key management can be accomplished with IBM Security Key Lifecycle Manager (SKLM). SKLM offers production-ready key management and complies with the Key Management Interoperability Protocol (KMIP). Starting with 12.3 code, FlashSystem A9000 and A9000R also support SafeNet KeySecure by Gemalto as external key server.

Starting with 12.1 code, FlashSystem A9000 and FlashSystem A9000R can also manage the keys internally, and support concurrent conversion from external key management to internal key management. However, the reverse operation of changing from local key to an external key server, first erases any data already on disk.

FlashSystem A9000 and FlashSystem A9000R encryption can be enabled during the installation of the system or later. While encryption is not enabled, the system might not meet a customer's standards or legal compliance standards, and the data might not be protected against security issues.

Enabling encryption will encrypt the SSDs and the MicroLatency modules. This non-destructive encryption process applies to the data that is already stored on the system without data rewrite.

FlashSystem A9000 and FlashSystem A9000R encryption can be disabled only when no volumes are defined.

For more information, see the IBM Redbooks publication, *Data-at-rest Encryption for the IBM Spectrum Accelerate Family*, REDP-5402.

Important: You must back up any data that must be kept or migrate it to another system before you deactivate encryption on IBM FlashSystem A9000 or FlashSystem A9000R.

2.6.2 User access control

The security of FlashSystem A9000 and FlashSystem A9000R is achieved on different levels. Data-at-rest encryption ensures the data that is stored on the SSDs and the flash enclosures.

To prevent unauthorized access to the configuration of IBM FlashSystem A9000 or FlashSystem A9000R and to the information that is stored on its volumes, the system uses password-based user authentication. By default, FlashSystem A9000 and FlashSystem A9000R are configured to manage user authentication, which means that all authentication information is stored in the system and the authentication is checked by this information.

FlashSystem A9000 and FlashSystem A9000R can use Lightweight Directory Access Protocol (LDAP) server-based user authentication. When LDAP authentication is enabled, the system accesses a specified LDAP directory to authenticate users.

Every user has a *user role* attribute. The role restricts the user access to the tasks that the user can perform on the system. For example, the role *storageadmin* can manage the whole system, and the role *readonly* can view only system information.

The domain policy setting of *open* or *closed* policy defines whether the global administrator can extend the reach into a domain.

For more information, see the IBM Redbooks publications, *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376 and *Enabling LDAP for IBM FlashSystem A9000 and A9000R with Microsoft Active Directory*, REDP-5387.

2.7 Auditing through syslog

You can configure a syslog server for FlashSystem A9000 and FlashSystem A9000R. All user actions (including all administrative commands and parameters from the CLI or GUI) that are performed on FlashSystem A9000 and FlashSystem A9000R are logged and exported to this server.

Syslog is commonly used for computer system management, security auditing, generalized information, analysis, and debugging messages. FlashSystem A9000 and FlashSystem A9000R can audit all user-entered commands.

Currently, the feature supports the definition of up to two audit servers. The configuration is performed by issuing the **audit_config_set** CLI command, as shown in Example 2-1. This command configures a primary and an optional secondary auditing server for command logging.

Example 2-1 Configuring the audit servers

```
audit_config_set primary_server=Address [ primary_port=port ] [
secondary_server=Address ]
```

The command does not check whether the specified server can be pinged or whether the server is listening on the specified port.

Note: This feature is domain-unaware. It is not possible to define auditing at the domain level.

Next, to effectively enable auditing, you must issue an **audit_enable** command. At a minimum, the `primary_server` must be configured for **audit_enable** to succeed.

Enabling or disabling this feature is restricted to the security administrator.

The following new CLI commands were introduced:

- **audit_enable**

  To effectively enable auditing. For this command to complete successfully, at least one syslog server must be configured.

- **audit_disable**

  A prerequisite for this command is that auditing is enabled (displayed as "`yes`" in **audit_show**).

- **audit_show**

  The results of this command indicate whether auditing is enabled or disabled.

- **audit_config_set**

  This command is used to configure the parameters that are required for enabling audits. Currently, the only supported protocol is SYSLOG over User Datagram Protocol (UDP).

- **audit_config_get**

  This command displays the current audit-related configuration.

3

Hardware and technology

This chapter describes the hardware architecture of IBM FlashSystem A9000 and A9000R.

It provides details about the system components, including the system rack, grid controllers, flash enclosures, power distribution, and interconnect devices.

This chapter includes the following sections:

- Systems overview
- System components details
- FlashSystem A9000R specifics
- Scaling FlashSystem A9000R
- FlashSystem A9000 specifics
- Reliability, availability, and serviceability
- IBM FlashCore technology
- IBM Variable Stripe RAID and 2D Flash RAID protection overview.

3.1 Systems overview

IBM FlashSystem A9000 and A9000R are built on field-proven hardware components, all providing outstanding performance, high resiliency, and redundancy.

Two machine types are associated with each FlashSystem A9000 and A9000R, and they come with two standard warranty periods.

Table 3-1 shows the different machine types and models according to the warranty periods. Note that the Storage Utility Offering Model U25, introduced with software Version 12.2.1 is only available with the 3 year warranty period.

Table 3-1 Machine types and models

	FlashSystem A9000	**FlashSystem A9000R**
One-year warranty	9836-415 or 9836-425	9835-415 or 9835-425
Three-year warranty	9838-415, 9838-425, 9838-U25	9837-415, 9837-425, 9837-U25

Before describing details of hardware components, we start with an overview of each system.

Note: In June of 2018, IBM withdrew all four versions of the Model 415. They can no longer be ordered or purchased, but existing A9000R model 415 systems can continue to be upgraded with additional grid elements. For reference, this chapter still includes relevant information for Model 415.

3.1.1 FlashSystem A9000R

FlashSystem A9000R is delivered in an IBM T42 rack and it consists of two to four grid elements for Model 425 or U25 and of two to six grid elements for Model 415. A *grid element* is always a triplet of two grid controllers and one flash enclosure.

Note: Starting with V12.3, a smaller initial configuration is supported, which essentially is A9000 hardware, mounted in the T42 rack with the Infiniband switches, for more information see 3.1.2, “FlashSystem A9000R entry-level capacity option” on page 38

Figure 3-1 shows the front view of a single grid element. The two grid controllers are at the bottom and the flash enclosure sits on top.

Figure 3-1 A9000R Model 415 (on the left) andA9000R Model 425 or U25 (on the right) Grid element

Grid controllers and flash enclosures are connected over InfiniBand through two redundant 56 Gbps Fourteen Data Rate (FDR) Mellanox InfiniBand switches. For internal management and support, three Ethernet daisy chains are installed between the grid controllers and the flash enclosures.

FlashSystem A9000R ships preconfigured according to the ordered configuration. All grid elements are pre-installed in the rack with all interconnection cables in place. The two InfiniBand switches, two power distribution units (PDUs), and a management patch panel are installed in the middle of the rack. The system requires at least two independent power sources. For the detailed specific requirements that apply to your country, see 5.3, “Site requirements” on page 113.

Figure 3-2 shows a front view of a fully populated FlashSystem A9000R, when the rack door is opened. The grid elements can be seen numbered 1 to 4 for Model 425 and 1 to 6 for Model 415, three below and three above the interconnect and power distribution unit. The front bezel in the center of the rack contains a power switch, which is covered by a transparent lid and a technician port connector.

Figure 3-2 A9000R Model 415 (on the left) and Model 425 /U25 (on the right) full rack view

The system can be ordered and delivered in a weight-reduced or height-reduced option. With the weight-reduced option, the grid controller and flash enclosures are shipped separately, and they will be installed on-site in the rack by an IBM service support representative (SSR).

For more information about the reduced shipping options, see 5.2, “Delivery requirements for IBM FlashSystem A9000R” on page 111. The installation and initial setup of FlashSystem A9000R is always performed by an IBM SSR.

3.1.2 FlashSystem A9000R entry-level capacity option

Beginning with Version 12.3, there is an option to start with a smaller entry-level configuration, which in this publication is called the *Grid-Starter*. This option takes the three capacity points allowed for the A9000R, and removes one grid controller and one flash enclosure. Put another way, it is an A9000 (with the exception of the partially populated FS 900 entry configuration) put into a rack with the infiniband switches.

However, even though the physical hardware configuration resembles the A9000, there is an important difference. The grid controllers and the flash enclosure are all connected via the redundant infiniband switches.

Starting with software version 12.3.1, this entry-level configuration can be nondisruptively upgraded by adding flash enclosures and grid controllers as required by the target configuration to scale out system capacity and performance. Scale out from the entry point configuration supports target configurations of two, three, and four flash enclosures with four, six, and eight grid controllers, respectively. Refer to Figure 3-3.

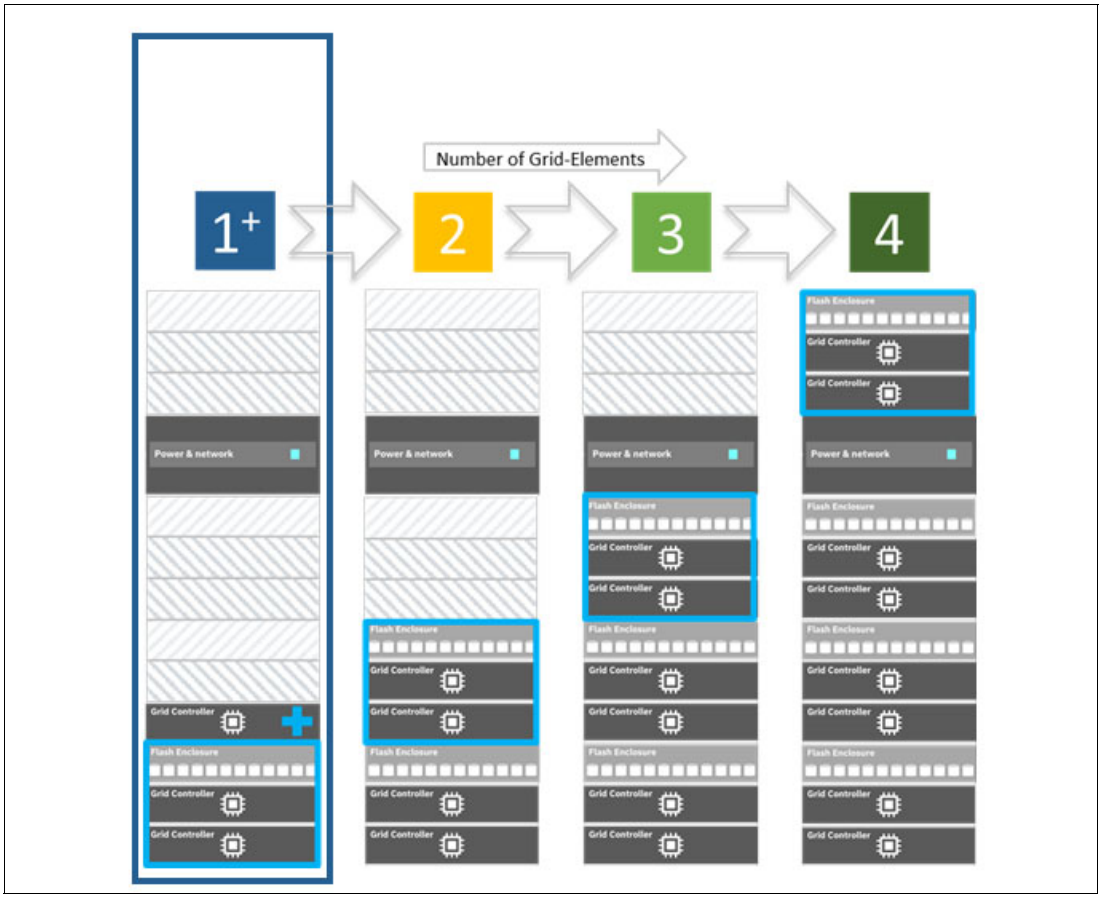


Figure 3-3 Scale out from entry level configuration

3.1.3 FlashSystem A9000

FlashSystem A9000 model 415, 425, or U25 consists of three grid controllers and one flash enclosure. Each component is a 2U unit, for a total of 8U of required rack space.

FlashSystem A9000 must be installed by an IBM SSR, and it can be placed into a customer-provided standard 19-inch rack. Figure 3-4 shows the components' configuration and their positions in the stack, next to the assembled system view.

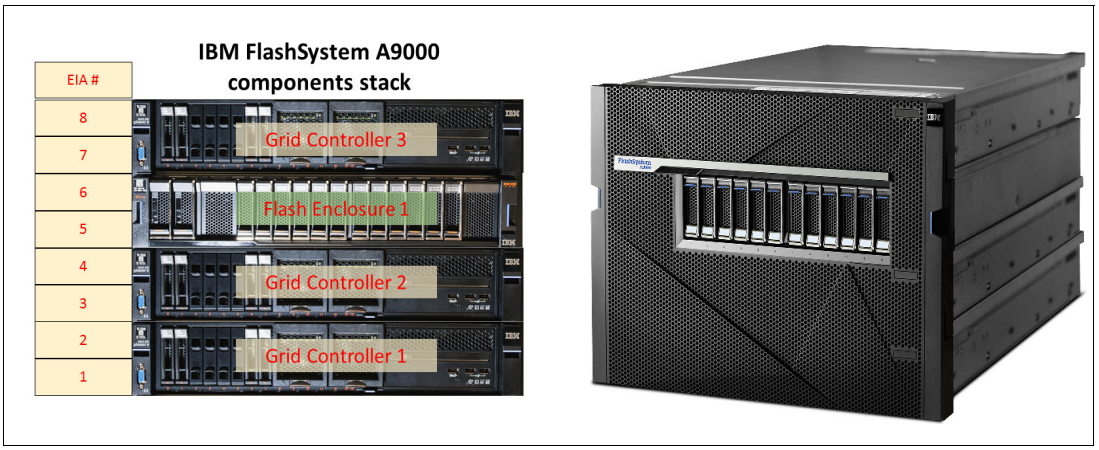


Figure 3-4 FlashSystem A9000 components' stack and system view

Because the system is installed in a customer-provided rack, its position in the rack might vary and other customer servers or storage systems might be installed in the same rack.

Communication between the grid controllers and the flash enclosure is over InfiniBand. All cabling between the grid controllers and the flash enclosure is fully redundant by using two sets of cables. The cables use industry standard plugs.

Each module (the flash enclosure and the three grid controllers) requires two independent power sources.

3.1.4 FlashSystem A9000 and A9000R Storage Utility Offering

Starting with software version 12.2.1, FlashSystem A9000 and A9000R are available through the Storage Utility Offering, as Machine-Type-Model 9837-U25 for FlashSystem A9000R and 9838-U25 for FlashSystem A9000, both with a 3-year maintenance only.

Both models use the same hardware as in the corresponding model 425, except that the A9000 smallest flash enclosure configuration (eight 3.6 TB microlatency modules) is not supported with the Storage Utility Offering. Field upgrades for additional grid elements on a FlashSystem A9000R are also not supported with the utility model.

These models provide a fixed total capacity, with a base and variable usage subscription of that total capacity. The variable capacity usage is billed on a quarterly basis.

IBM Storage Utility Offering allows clients to align their capacity costs to business initiatives. For details, see the IBM Storage Utility Offering home page:

https://www.ibm.com/us-en/marketplace/storage-utility-offering

3.2 System components details

This section describes the system components that are used in FlashSystem A9000 and A9000R. Many of the characteristics apply to both systems.

3.2.1 Grid controller

The grid controller acts as a core component, providing the interface and compute functions. It also provides cache to accelerate both read and write operations. Furthermore, the grid controller is responsible for the inline data reduction through data deduplication and compression. Data compression is assisted by hardware accelerator cards.

Additionally, some grid controllers provide additional functions that include:

- Management ports
- Virtual private network (VPN) ports for remote connection
- Port for the technician's access

The grid controller is based on dual Intel Xeon processors. It offers two CPU sockets, 24 dynamic device reconfiguration 4 (DDR4) error correction code (ECC)-capable memory slots, and high-speed Peripheral Component Interconnect Express (PCIe) 3.0 connectors to attach and serve all I/O ports that are required for FlashSystem A9000 and A9000R. The CPU sockets support Intel Xeon E5-26xx series processors.

Table 3-2 and Table 3-3 list the internal hardware components that are contained in the grid controller for FlashSystem A9000 Model 425 and FlashSystem A9000 Model 415. The differences are indicated in bold. The number of Fibre Channel (FC) or internet Small Computer System Interface (iSCSI) Ethernet adapters depends on the connectivity option that is ordered, as explained in "Grid controller options" on page 44.

NVMe ready adapters: Model 425 and U25, in both A9000 and A9000R included with Software Version 12.2.1 or later contain enhanced grid controllers equipped with FC-NVMe ready adapters.

Table 3-2 Grid controller components in FlashSystem A9000 Model 425 / U25

Component	Hardware element
Processor	2 Intel **E5-2650 v4 @ 2.20 GHz 12 Cores** with Hyper Threading
Memory	**384 GB** DDR4 RAM
InfiniBand Adapter	2 Mellanox IB Connect Dual Port 56 Gbps fourteen data rate (FDR)
16 Gb Fibre Channel Adapter	0 or 2 QLogic QLE2962 Dual Ports (earlier machines may have QLE2662 instead)
10 Gb iSCSI Adapter	1 or 2 Mellanox ConnectX-3 Pro Dual Ports
Compression Accelerator	2 Intel Coleto Creek cards
Boot Media Module	2 hot-swap, RAID1 600GB HDDs for system firmware and logging
Vaulting Module	2 hot-swap, **400 GB** solid-state drives (SSDs) as the vault device
Battery Module	Dual internal redundant battery backup units

Table 3-3 Grid controller components in FlashSystem A9000 Model 415

Component	Hardware element
Processor	2 Intel **E5-2630 v3 @ 2.40 GHz 8 Cores** with Hyper Threading
Memory	**192 GB or 256**[a] **GB** DDR4 RAM
InfiniBand Adapter	2 Mellanox IB Connect Dual Port 56 Gbps fourteen data rate (FDR)
16 Gb Fibre Channel Adapter	0 or 2 QLogic QLE2662 Dual Ports
10 Gb iSCSI Adapter	1 or 2 Mellanox ConnectX-3 Pro Dual Ports
Compression Accelerator	2 Intel Coleto Creek cards
Boot Media Module	2 hot-swap, RAID1 600 GB HDDs for system firmware and logging
Vaulting Module	2 hot-swap, **250 GB** solid-state drives (SSDs) as the vault device
Battery Module	Dual internal redundant battery backup units

a. Depending on the date of manufacturing

Note: In June of 2018, IBM withdrew all four versions of the Model 415. They can no longer be ordered or purchased, but existing A9000R model 415 systems can continue to be upgraded with additional grid elements.

Table 3-4 and Table 3-5 list the internal hardware components that are contained in the grid controller for FlashSystem A9000R Model 425 and FlashSystem A9000R Model 415. The differences are indicated in bold. The number of FC or iSCSI Ethernet adapters depends on the connectivity option that is ordered, as explained in “Grid controller options” on page 44.

Table 3-4 Grid controller components in FlashSystem A9000R Model 425

Component	Hardware element
Processor	2 Intel **E5-2650 v4 @ 2.20 GHz 12 Cores** with Hyper Threading
Memory	384 GB DDR4 RAM
InfiniBand Adapter	1 Mellanox IB Connect Dual Port 56 Gbps FDR (A replacement grid controller may contain 2 adapters.)
16 Gb Fibre Channel Adapter	0 or 2 QLogic QLE2962 Dual Ports (earlier machines may have QLE2662 instead)
10 Gb iSCSI Adapter	1 or 2 Mellanox ConnectX-3 Pro Dual Port
Compression Accelerator	2 Intel Coleto Creek cards
Boot Media Module	2 hot-swap, RAID1 600 GB HDDs for system firmware and logging
Vaulting Module	2 hot-swap, **400 GB** SSDs as vault devices
Battery Module	Dual internal redundant battery backup units

Table 3-5 Grid controller components in FlashSystem A9000R Model 415

Component	Hardware element
Processor	2 Intel **E5-2650 v3 @ 2.30 GHz 10 Cores** with Hyper Threading
Memory	384 GB DDR4 RAM
InfiniBand Adapter	1 Mellanox IB Connect Dual Port 56 Gbps FDR (A replacement grid controller may contain 2 adapters.)
16 Gb Fibre Channel Adapter	0 or 2 QLogic QLE2662 Dual Port
10 Gb iSCSI Adapter	1 or 2 Mellanox ConnectX-3 Pro Dual Port
Compression Accelerator	2 Intel Coleto Creek cards
Boot Media Module	2 hot-swap, RAID1 600 GB HDDs for system firmware and logging
Vaulting Module	2 hot-swap, **250 GB** SSDs as vault devices
Battery Module	Dual internal redundant battery backup units

The grid controller is highly reliable. It has a modular design so that components can be replaced in a failure.

Important: All field-replaceable units (FRUs) can be replaced concurrently from a system perspective. All replacements are performed by an IBM SSR.

By design, in FlashSystem A9000 and A9000R, the grid controller is an isolated failure domain so that any failure or maintenance action that requires you to shut down the grid controller does not affect the overall system.

Figure 3-5 shows the front view of the grid controller:

- On the left, a Video Graphics Array (VGA) connector, but which is not used under normal operations, can be used by the IBM service support representative (SSR) for maintenance.
- The drive cage is next to the VGA connector. It contains two hard disks in a RAID 1 that are the grid controller boot devices, holding FlashSystem A9000 or A9000R microcode and also various system logs and events. In the two right-most slots of the drive cage are two SSDs that are used as vault devices for cache and metadata.
- In the middle, the grid controller contains two battery backup units.
- The front panel is on the right with status LEDs and the power button. Three USB ports are under the front panel. They can be used by an IBM SSR for maintenance.

Important: Do not attempt to power off the grid controller by using the power button.

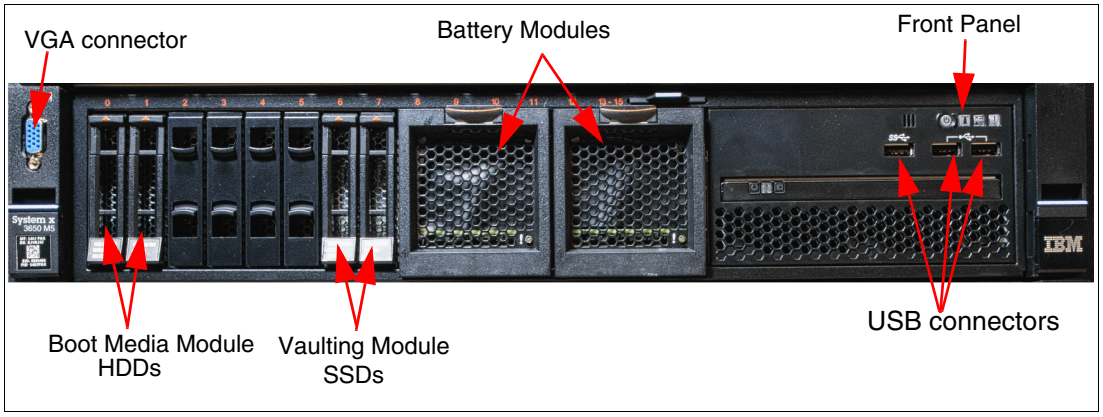


Figure 3-5 Grid controller front view

Figure 3-6 shows the rear view of an FC-enabled A9000R grid controller.

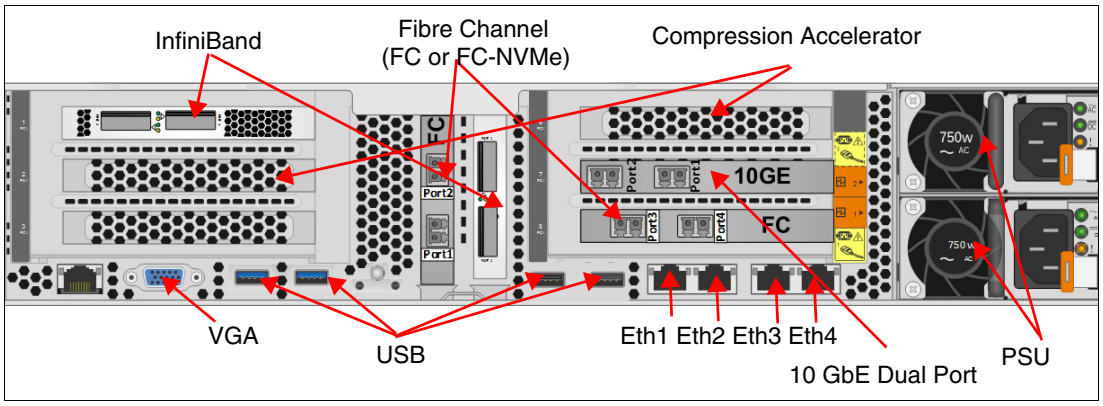


Figure 3-6 Grid controller rear view

The following descriptions refer to the grid controller rear view:

- The PCI slots on the left contain one dual port InfiniBand adapter at the top and one of the compression accelerators in the bottom slot. The middle slot is not populated in any grid controller configuration.
- A VGA connector and four USB ports are next to the Ethernet port. The VGA connector and four USB ports are available for service functions, if needed.
- The middle PCI slots can contain an FC-NVMe connector, an additional InfiniBand card, or both.
- The top PCI slot on the right contains the second compression accelerator card. In the middle is a 10 GbE dual port adapter, and the bottom slot can contain either a FiC-NVMe adapter or another 10 GbE adapter.
- Four 1 GB Ethernet adapters, which are used for the internal communication and as management ports, are under the PCI slots.
- The two power supplies are on the right.

Grid controller battery module

Each grid enclosure contains two hot-swappable batteries. Each battery supplies enough power to allow the grid controller to gracefully shut down in a power failure. The microcode of FlashSystem A9000 and A9000R regularly runs self-tests and calibrations on each battery unit. Those tests check that the batteries are functional (“healthy”) and provide enough energy to sustain a system shutdown.

If any battery in the system shows irregularities, such as low voltage, the system emits an event. When Call Home is configured (see 7.1.6, “Call Home” on page 150) and the unit is under warranty or a valid service contract is in place, a service record with IBM is opened to call for battery replacement.

Figure 3-7 shows a grid controller with one battery pulled out a bit and the second battery inserted next to it.

Figure 3-7 Grid controller with a battery that is pulled out

Grid controller options

The grid controller is available in two versions that provide different I/O connectivity options.

Fibre Channel and iSCSI option

Figure 3-8 shows the Fibre Channel and iSCSI option of the grid controller, which contains the following components:

- One dual-port FC-NVMe adapter in the middle and an additional FC-NVMe adapter at the lower right (symbolized by the orange-yellow rectangles)
- One dual-port 10 GbE controller in the middle right PCI slot (symbolized by the green rectangle) for iSCSI connectivity

Note: When using Fibre Channel, ports 1 and 3 are recommended for host connectivity for load balancing and redundancy as they are located on different adapter cards. Ports 2 and 4 are advised for mirroring and migration connectivity. See Figure 3-8 for the physical port location on the grid controller.

All three adapters can be used to connect hosts, for mirroring or migration connectivity.

The purple rectangles symbolize the InfiniBand adapter cards, which are used for the internal back-end communication only. The InfiniBand cards cannot be used for host or mirroring

access. The InfiniBand adapter card that is shown vertically with a hashing pattern is not present in FlashSystem A9000R. A replacement grid controller FRU may contain two InfiniBand adapters.

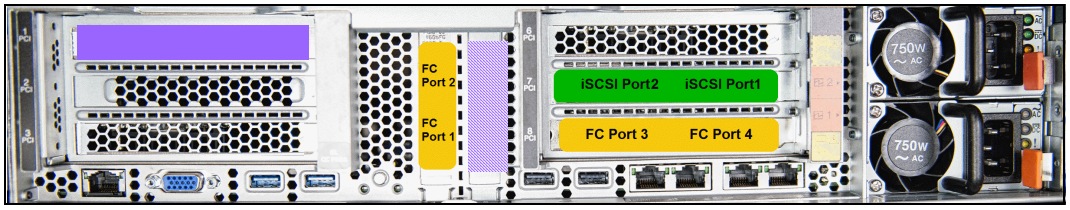


Figure 3-8 Grid controller with FC and iSCSI option

iSCSI only option

Figure 3-9 shows the grid controller with iSCSI option. It contains one 10 GbE dual port adapter in the middle right and a second adapter on the lower right, both for iSCSI connectivity (symbolized by the green rectangles).

The purple rectangles symbolize the InfiniBand adapter cards, which are used for the internal back-end communication only. The InfiniBand cards cannot be used for host or mirroring access. The InfiniBand adapter card that is shown with a hash pattern is not present in FlashSystem A9000R. A replacement grid controller FRU may contain two InfiniBand adapters.

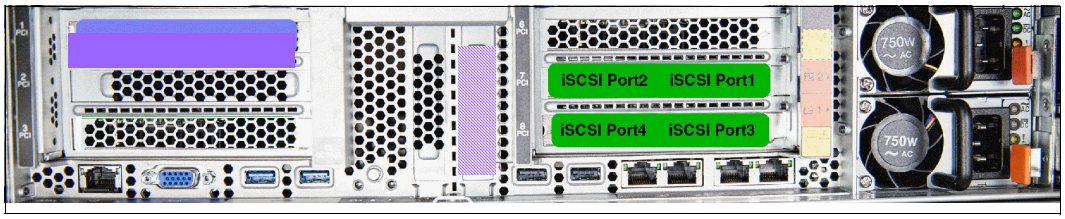


Figure 3-9 Grid controller with iSCSI-only option

NVM express connectivity

Non-Volatile Memory express (NVMe) is a protocol designed for use with flash-based storage. It removes much of the overhead of the existing SCSI command set, improving both bandwidth and response time. NVMe is suitable to be used with many communications protocols, including InfiniBand, Ethernet, and Fibre Channel.

With the 12.2.1 release, the FlashSystem A9000 and A9000R include new Fibre Channel adapters that are capable of utilizing NVMe over Fibre channel (FC-NVMe, or NVMe-f) connectivity between properly configured hosts and the storage system. FC-NVMe is not enabled with release 12.2.1, but IBM plans[1] to enable it in a future release. Until the software upgrade is offered, regular FC activity can be used on the FC-NVMe adapters.

Note: All grid controllers in FlashSystem A9000 or FlashSystem A9000R must be equipped with either the Fibre Channel with iSCSI option or the iSCSI-only option. For FlashSystem A9000 Model 425/U25 and FlashSystem A9000R Model 425/U25, all must be feature code 5003 (4x 16 Gb FC + 2x 10 Gb iSCSI) or all must be feature code 5004 (4x 10 G iSCSI). A mix of those options in the same storage system is not supported. Feature code 5005 is associated with the FC-NVMe adapters mentioned previously.

[1] IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. The development, release, and timing of any future features or functionality described for our products remain at IBM's sole discretion.

Both adapter types (FC 5003 and FC 5005) offer the same 16 Gb bandwidth. The external physical difference between the two adapters is as follows:

- FC adapters are marked with “PCIe x8 16GbFC”
- FC-NVMe adapters are marked with “PCIe FC”

It is not be optional to use feature code 5005 now that it is released, so all new systems model 425/U25 will have the new adapter, and all hardware capacity upgrades (adding a grid element or two) will also include feature code 5005. It is supported to use Fibre Channel connectivity with a mixture of feature codes, for example a 425 purchased with feature code 5003, then upgraded with another grid element.

However, to enable FC-NVMe at a future date, the existing grid controllers with feature code 5003 must be upgraded to feature code 5005. This ability and the field upgrade (MES) is only applicable to model 425; it will not be available on model 415 or U25.

Internal grid controller components

Figure 3-10 shows the inside view of a grid controller as it applies to both FlashSystem A9000 and A9000R:

- The power supply unit cage is at the lower left.
- The PCIe slots that contain the interface, InfiniBand, and compression accelerator cards are on the left.
- In the middle, you see the two CPU sockets (CPU1 at the top and CPU2 at the bottom) and the 24 DDR4 ECC dual inline memory module (DIMM) slots.
- Six fans are to the right of the CPU and memory area. The airflow is from front to back (in the picture, from right to left).
- The storage backplane with the HDD and SSD drives, the power buttons, and status LED panel are on the right.

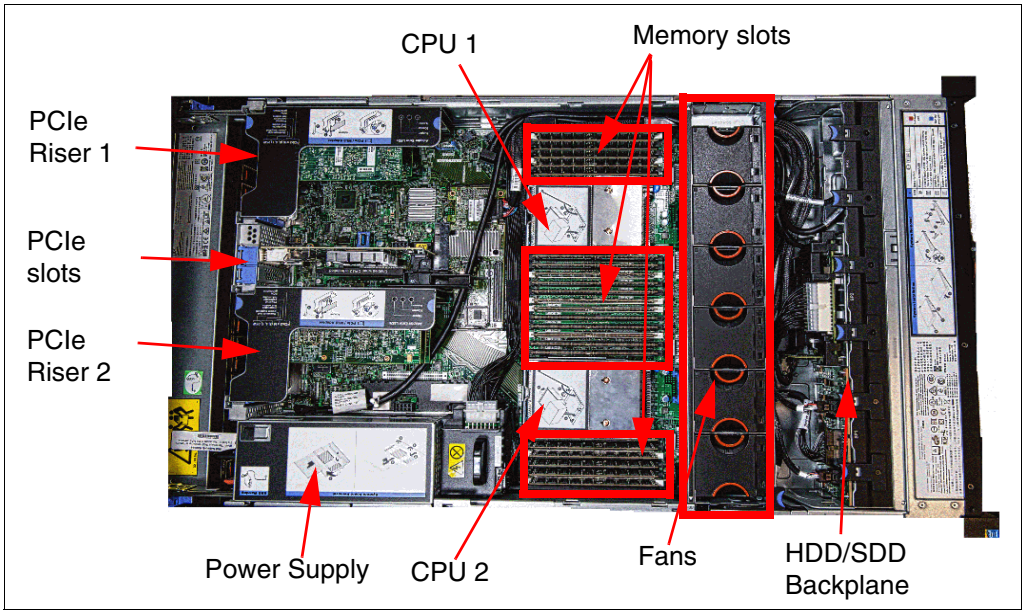


Figure 3-10 Grid controller inside view

Power supplies

The grid controller is powered by a redundant power supply unit (PSU) cage with dual 750 W PSU assemblies.

These power supplies can be replaced individually with no need to power off the grid controller. The power supply is an FRU.

Fans

The grid controller is cooled by six redundant fan units. The cooling design of the grid controller allows it to operate normally, even if two fans have failed completely. Each fan has two independent motors to compensate failures. The fan is an FRU and it is replaced by an IBM SSR in the event of a fan failure.

Figure 3-11 shows a fan that is pulled out of the case. Each fan enclosure contains two motors.

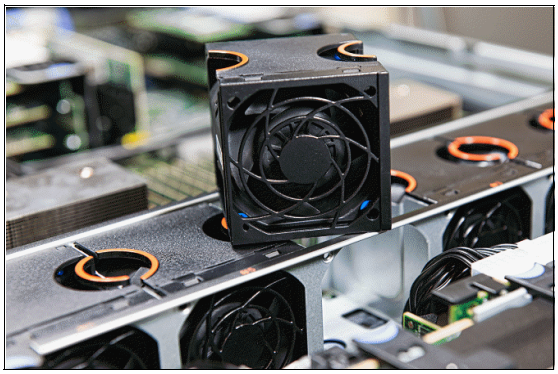


Figure 3-11 Grid controller fan

InfiniBand host channel adapter

The Mellanox Connect-IB Dual Port host channel adapter (HCA) is a 4x 14 data rate (FDR) InfiniBand adapter that features 56 Gbps bandwidth per port at a low latency. The HCA processes protocol, which offloads this workload from the grid controller's processor. InfiniBand performs extensive end-to-end link-level protocol checking to ensure reliability.

The InfiniBand adapter is not a Field Replaceable Unit (FRU) on its own. In case of a failure, the grid controller will need to be replaced.

Fibre Channel adapter

Host, replication, and migration connectivity can be established by using the 16 Gbps Fibre Channel adapters. The grid controller can be equipped with two QLogic QLE2662 Dual Port Fibre Channel adapters. They feature hot-upgradable firmware capability with single firmware images for each port, and they support up to 2,048 concurrent exchanges per port.

The Fibre Channel ports support 4, 8, and 16 Gbps full-duplex data transfer over short wave (SW) fibre links by using 50 micron multi-mode cable. It is not possible to attach this HBA to a 1 or 2 Gbps storage area network (SAN) switch.

The allocation and use of specific ports in each module depend on your environment, your specific requirements in terms of resiliency, the nature of your host I/O traffic, and whether you use mirroring over Fibre Channel.

All Fibre Channel ports can be operated in target mode for host connectivity. They can be configured in either initiator or target mode for mirroring or migration connectivity.

Each of the Fibre Channel adapters is now a separate FRU.

Fibre Channel ports: Always use enough ports to support multipathing without overburdening the host with too many paths to manage. The best practice is to use at least one port from every grid controller for load balancing and redundancy. Using more than 2 ports per grid controller does not necessarily provide more bandwidth and may overburden the host with too many ports to manage. See *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368.

iSCSI adapter

Depending on the configuration that you want, a grid controller can alternatively be equipped with either one or two Mellanox ConnectX-3 Pro dual port 10 GbE adapters. Both ports on the adapter are equipped with optical transceiver modules (SFP+).

The 10 GbE adapter provides enhanced iSCSI host connectivity and adequate infrastructure for potential Fibre Channel over Ethernet (FCoE) functions. Currently, IBM does not offer FCoE with FlashSystem A9000 or FlashSystem A9000R.

You can operate iSCSI connections for various uses:

- As an iSCSI target for host, mirroring, or migration connectivity
- As an iSCSI initiator for remote mirroring or migration

The default and highest possible maximum transmission unit (MTU) that is supported is 9,000 MTU.

The iSCSI adapter is also now a FRU.

Solid-state drive (SSD) vault devices

Each grid controller contains two 400 GB (for Model 425) or 250 GB (for Model 415) enterprise-grade SSDs to save cache data in a power loss. Additionally, the SSDs are used to save metadata and system configuration information regularly (*live vaulting*). During a normal shutdown procedure, the microcode will also write configuration data, metadata, and cache data that is not yet destaged in permanent flash storage to the SSDs. The cache data is mirrored three times across different grid controllers.

Therefore, no mirroring within the grid controller is required. If encryption is configured on FlashSystem A9000 and A9000R, the SSD data will be encrypted, too. The drives are self-encrypting devices. Each SSDs is a FRU on its own.

Hard disk drive

FlashSystem A9000 and A9000R microcode and internal events, logging, and performance data are stored on two mirrored (RAID 1) hard disk drives (HDDs). The mirror RAID configuration ensures that a single disk failure will not result in a grid controller failure. Each HDD is a FRU on its own.

Compression accelerator card

FlashSystem A9000 and A9000R feature a high-performance data deduplication and compression engine. To use the compression efficiency and keep the effect on the performance by the compression minimal, each grid controller contains two hardware compression acceleration cards.

The accelerator card, which is shown in Figure 3-12 on page 49, is based on the Intel Quick Assist Acceleration technology (Coleto Creek Communications Chipset 8950).

The accelerator card is engineered by IBM. It is not a separate FRU. In case of a failure, the grid controller needs to be replaced.

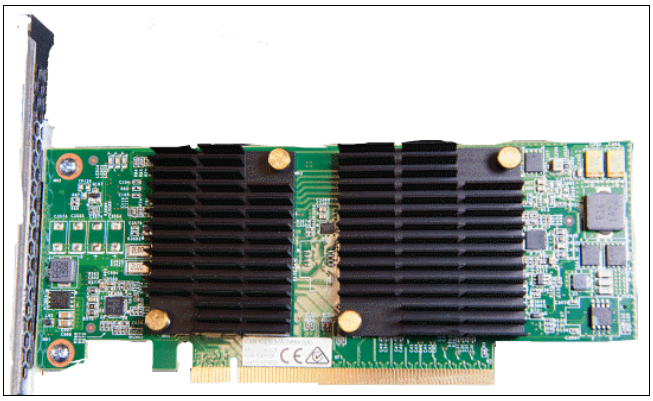


Figure 3-12 Compression accelerator card

3.2.2 Flash enclosure

The flash enclosure is a core component used in FlashSystem A9000 and A9000R providing ultra low-latency, highly reliable, and scalable performance in a space-efficient and power-efficient storage device. FlashSystem A9000 and A9000R Model 425 include an improved flash enclosure, with larger and more flexible capacity points, achieved by combining three-dimensional (3D) chip layout with triple-level cell (TLC) transistors. The new enclosure model also offers inline always-on hardware compression, being done right on the MicroLatency modules without any performance impact.

FlashSystem A9000 and A9000R continue to implement a full data reduction engine, including pattern removal, data deduplication and compression, in the grid controllers, to achieve a very competitive data reduction ratio. However, to reduce CPU workload and improve performance, internal metadata will not be compressed by the grid controllers, allowing the underlying MicroLatency modules to do it instead.

The flash enclosure consists of flash modules (the MicroLatency modules), two fully redundant canister, battery modules, power supplies, fans, a midplane and a power interposer board. The flash enclosure used in FlashSystem A9000 and A9000R Model 425 includes re-designed canisters and midplane. A flash canister used in FlashSystem A9000 and A9000R Model 415 flash enclosure cannot be inserted into a FlashSystem A9000 and A9000R Model 425 flash enclosure due to different internal connectors and alignment.

Figure 3-13 shows a flash enclosure with 12 MicroLatency modules.

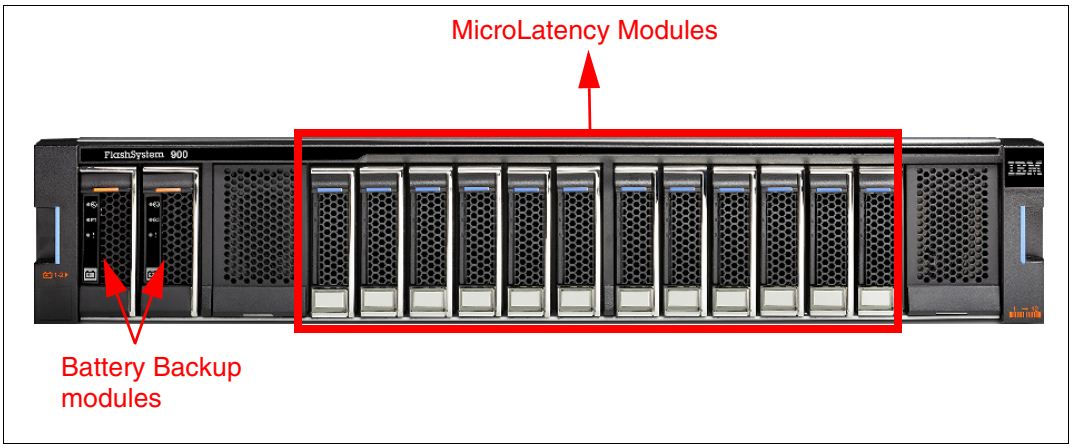


Figure 3-13 FlashSystem A9000 and A9000R flash enclosure

Important: Flash enclosures used in IBM FlashSystem A9000 and A9000R Model 425 are not interchangeable with flash enclosures used in FlashSystem A9000 and A9000R Model 415. Upgrade from one model to the other is not supported as FlashSystem A9000 and A9000R Model 425 delivers a complete hardware refresh.

Figure 3-14 shows the components of the flash enclosure. One of the two canisters was removed. The numbers in brackets show the total numbers of the corresponding component in one flash enclosure. The power supply unit to the right of the fans provides redundant power to the system.

All components are concurrently maintainable except the midplane and the power interposer, which has no active components. All external connections are from the rear of the system.

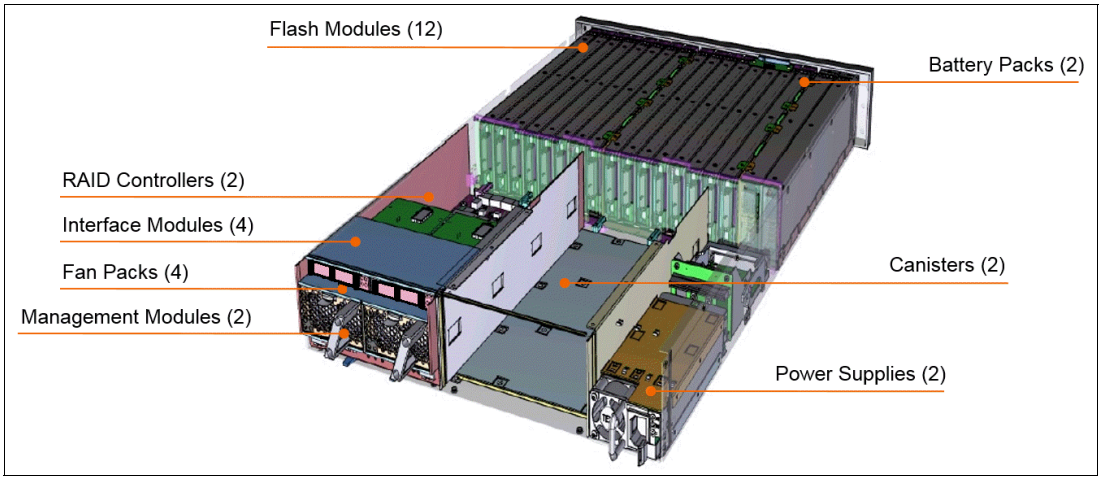


Figure 3-14 Flash enclosure components

Flash Canister

Each flash enclosure canister contains the following components:

- RAID controller
- Two interface modules
- Management module
- Two hot-swappable fan modules
- Two Ethernet ports
- Two USB ports

Figure 3-15 shows one flash enclosure canister with its components. One fan module is pulled out one inch.

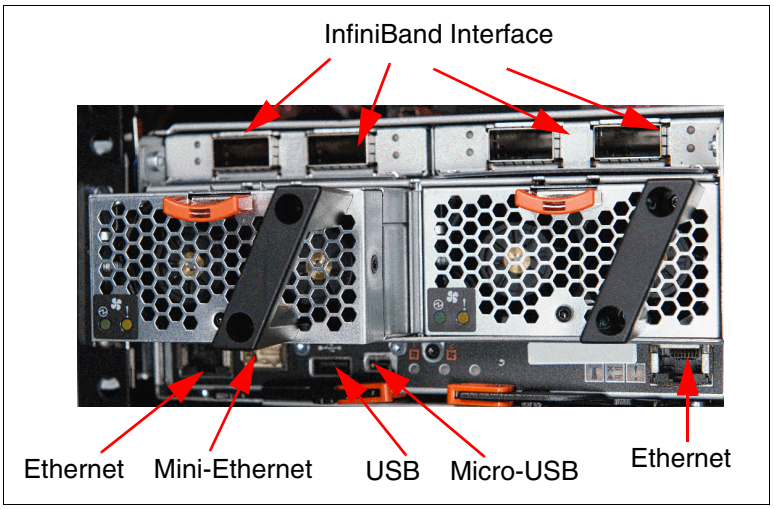


Figure 3-15 Flash enclosure canister

To maintain redundancy, the canisters are hot-swappable.

Interface cards

The flash enclosure supports four 2-port InfiniBand 40 Gbps interface cards. A total of eight ports that offer 40 Gbps InfiniBand connections are supported per flash enclosure:

- FlashSystem A9000R uses four ports (the left port of every InfiniBand interface card).
- FlashSystem A9000 uses six ports: see 3.5.1, “FlashSystem A9000 direct InfiniBand” on page 62.

Figure 3-15 shows one 2-port FlashSystem A9000R flash enclosure canister interface with two cards. For a description of FlashSystem A9000R InfiniBand cabling, see 3.3.1, “FlashSystem A9000R switched InfiniBand” on page 55. 3.5.1, “FlashSystem A9000 direct InfiniBand” on page 62 describes FlashSystem A9000 InfiniBand cabling concept.

InfiniBand support

The flash enclosure InfiniBand interface cards have two 4X quad data rate (QDR) ports each. They are using SCSI Remote Direct Memory Access (RDMA) Protocol (SRP) for establishing the connections.

InfiniBand interface card port LED descriptions

Each InfiniBand interface port in the flash enclosure has a set of LEDs to indicate the status. Table 3-6 lists the InfiniBand LED port descriptions. The two LEDs are next to the InfiniBand ports.

Table 3-6 InfiniBand LED port descriptions

LED name	Color	States
Link state upper LED	Green	► Off: No link is established. ► Solid: Link is established.
Activity lower LED	Amber	► Off: No physical link. ► Solid: Link is established, no activity. ► Flashing: Activity on the link.

Power supply units

The flash enclosure power supply units are accessible from the rear of the unit, and they are fully hot-swappable. Figure 3-16 shows two hot-swappable power supply units with the top one partially pulled out. FlashSystem A9000 and A9000R alerting systems (event logging, Simple Network Management Protocol (SNMP), and so on) will report a power supply fault.

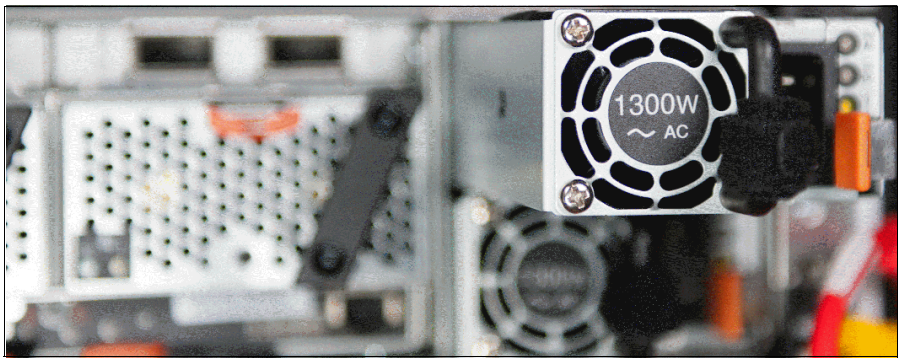


Figure 3-16 Flash enclosure hot-swappable power supply units

Fan modules

Any FlashSystem A9000 and A9000R flash enclosure contains four hot-swappable fan modules. Each flash enclosure canister holds two hot-swappable fan modules. Each fan module contains two fans. The system can remain fully online if one of the fan modules fails. The flash enclosure fan modules are accessible from the rear of the unit (in each canister).

Figure 3-17 shows a flash enclosure hot-swappable fan module.

Figure 3-17 FlashSystem A9000 and A9000R fan module

FlashSystem A9000 and A9000R alerting systems, such as SNMP, will report a fan module fault. It is a FRU, which must be replaced by an IBM SSR.

MicroLatency modules

The flash enclosure supports 12 IBM MicroLatency modules, which are accessible from the enclosure front panel. The MicroLatency modules in Model 425 have usable capacity of either 3.6 TB, 8.5 TB or 18 TB and those for Model 415 have usable capacity of either 1.2 TB, 2.9 TB, or 5.7 TB of flash storage. FlashSystem A9000 Model 425 can be ordered with 8 or 12 MicroLatency modules when using 3.6 TB modules. A miscellaneous equipment specification (MES) upgrade between the two configurable options is not available. All other FlashSystem A9000 and A9000R models are configured with 12 MicroLatency modules only.

Important: All MicroLatency modules in FlashSystem A9000 and A9000R must be the same size. They cannot be mixed.

The usable capacity for MicroLatency modules is described in 2.3, “Data reduction and capacity concepts” on page 15. See 3.7, “IBM FlashCore technology” on page 67 for details about the MicroLatency modules, for example, the hardware-accelerated I/O or the advanced flash management.

Figure 3-18 shows a FlashSystem A9000 and A9000R flash enclosure 2.9 TB MicroLatency module.

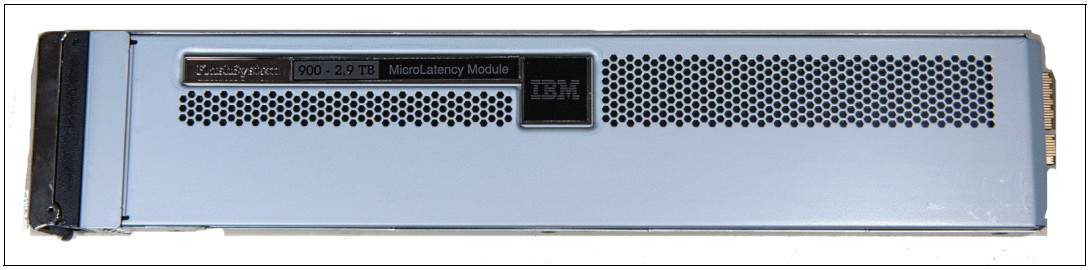


Figure 3-18 FlashSystem A9000 and A9000R flash enclosure 2.9 TB MicroLatency module

Flash enclosure battery modules

The flash enclosure contains two hot-swappable battery modules. The function of the battery modules is to ensure that the system is gracefully shut down (fully flushed and synchronized write cache) when AC power is lost to the unit. Figure 3-19 shows battery module 1, which is in the left-most front of the flash enclosure.

Figure 3-19 Flash enclosure battery module 1 partially pulled out

A battery-reconditioning feature on the flash enclosure calibrates the gauge to report the amount of charge on the batteries. Battery reconditioning is performed automatically.

Flash enclosure high availability

Availability of the flash enclosure is assured by its design points:

- No single point of failure
- FlashCore architecture of the MicroLatency modules with features, such as overprovisioning, wear leveling, and two-dimensional RAID protection, including patented IBM Variable Stripe RAID (VSR)
- Two redundant flash enclosure canisters
- Dual hardware RAID controllers
- Redundant data paths throughout the system
- Active-active Interface cards
- Redundant management controller
- Redundant batteries
- Dual power supplies
- All components are hot-swappable
- Concurrent code loads

3.3 FlashSystem A9000R specifics

This section describes components or other characteristics that are specific to FlashSystem A9000R.

Table 3-7 and Table 3-8 on page 55 show a list of components that are included in IBM FlashSystem A9000R Models 425 and 415, and the available capacity options.

Table 3-7 FlashSystem A9000R Model 425 configurations

FlashSystem A9000R Model 425 configurations				
Grid elements	1.5 Grid-Starter	2	3	4
Grid controllers	3	4	6	8
CPUs (cores)	6 (72)	8 (96)	12 (144)	16 (192)
Memory in GB	1152	1536	2304	3072
iSCSI only ports	12	16	24	32
iSCSI + FC ports	6 + 12	8 + 16	12 + 24	16 + 32
Flash enclosures	1	2	3	4
MicroLatency mod.	12	24	36	48
MicroLatency module in TB	3.6/8.5/18	3.6/8.5/18	3.6/8.5/18	3.6/8.5/18
Usable physical capacity TB	36/85/180	72/170/360	108/255/540	144/340/720
Effective capacity in TB[a]	180/425/900	360/850/1800	540/1275/2700	720/1700/3600
Maximum effective capacity in TB[b]	1800/1800/1800	2400/2400/2400	3600/3600/3600	4800/4800/4800

a. Effective capacity assumes a data reduction calculated at about 5:1.
b. Maximum effective capacity is the up-most provisioning limit that effective capacity can be stretched to, by IBM.

Table 3-8 FlashSystem A9000R Model 415 configurations

FlashSystem A9000R Model 415 configurations					
Grid elements	2	3	4	5	6
Grid controllers	4	6	8	10	12
CPUs (cores)	8 (80)	12 (120)	18 (180)	20 (200)	24 (240)
Memory in GB	1536	2304	3072	3840	4608
iSCSI only ports	16	24	32	40	48
iSCSI + FC ports	8 + 16	12 + 24	16 + 32	20 + 40	24 + 48
Flash enclosures	2	3	4	5	6
MicroLatency modules	24	36	48	60	72
MicroLatency module in TB	2.9/5.7	2.9/5.7	2.9/5.7	2.9/5.7	2.9/5.7
Usable physical capacity TB	58/114	87/171	116/228	145/285	174/342
Effective capacity in TB[a]	300/600	450/900	800/1200	750/1500	900/1800
Maximum effective capacity in TB[b]	1400	2000	2600	3000	3000

a. Effective capacity assumes a data reduction calculated at about 5:1.
b. Maximum effective capacity is the up-most provisioning limit that effective capacity can be stretched to, by IBM.

3.3.1 FlashSystem A9000R switched InfiniBand

InfiniBand is the high-performance network that is used by FlashSystem A9000 and A9000R to connect the components internally. It is the redundant, reliable connection backbone between the components. InfiniBand assures high throughput and low latency, and it is optimal for connecting grid controllers and flash enclosures.

An InfiniBand connection between the grid controllers is used for internal management and data transfer. The management task uses IP over InfiniBand (IPoIB) for concurrent code upgrade, connectivity status checks, and so on. The InfiniBand connection between grid controllers and flash enclosures is only used for data transfer. Management of the flash enclosures is performed from the grid controllers to the flash enclosures by using two daisy chained Ethernet connections.

In FlashSystem A9000R, the internal network is based on two redundant 36-port InfiniBand switches.

Each grid controller has an InfiniBand HCA with two ports as shown in Figure 3-6 on page 43 that is cabled to each of the two InfiniBand switches. Each flash enclosure has four interface cards with two InfiniBand ports. The first port of every flash enclosure interface is attached to a switch so that every flash controller canister is attached to both switches.

The switches are also linked to each other.

Figure 3-20 shows the cabling for FlashSystem A9000R with four grid controllers and two flash enclosures. For clarity, the IB Switch #1 InfiniBand network is yellow, the IB Switch #2 InfiniBand network is red, and the switch interconnect is black.

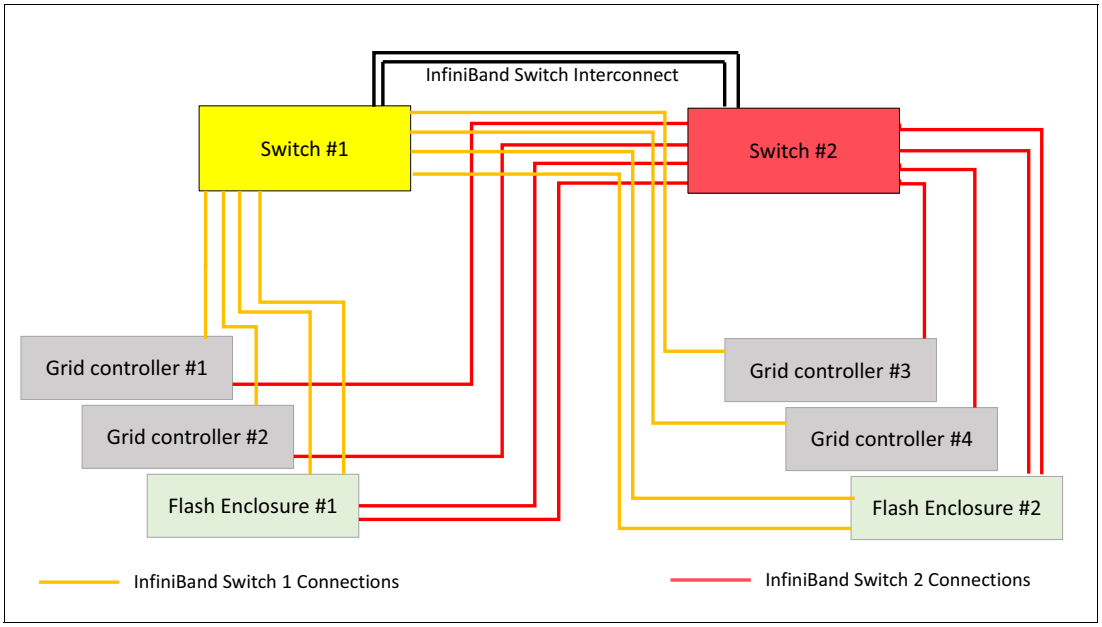


Figure 3-20 Example of InfiniBand cabling

Every added grid element, the triplet of two grid controllers, and one flash enclosure will add one InfiniBand connection of every grid controller to both switches and one InfiniBand connection of every flash enclosure canister to both switches.

This network topology enables the maximum bandwidth use and lowest latency because the switches are used in an active-active configuration. The InfiniBand switches are also tolerant to any failure of the following individual network components:

- Ports
- Links
- Switches

Figure 3-21 shows the two InfiniBand switches that are installed in FlashSystem A9000R.

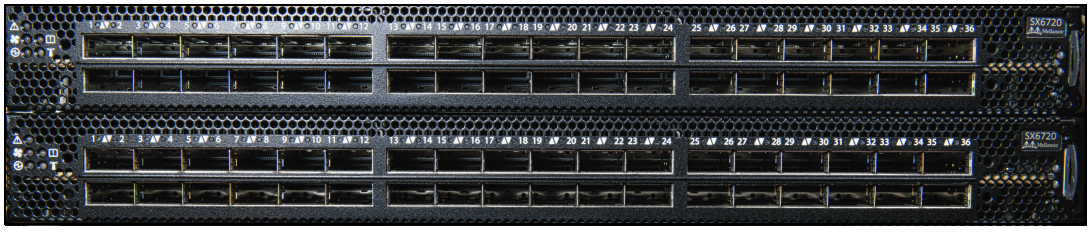


Figure 3-21 FlashSystem A9000R InfiniBand switches

Each InfiniBand switch contains 36 ports, which support InfiniBand Fourteen Data Rate (FDR). A port that supports FDR has four lanes that run a bit rate of 14.0625 Gbps, resulting in an effective bandwidth of more than 56 Gbps full bidirectional bandwidth per port. Port-to-port latency is less than 200 nanoseconds. Each switch has 4 Tbps switching throughput. The switches are powered by two redundant power supplies with an included battery backup unit, and fan modules to eliminate any single point of failure. Additionally, each switch has three RJ-45 management ports that are used by FlashSystem A9000R.

Figure 3-22 shows the InfiniBand cabling of FlashSystem A9000R flash enclosure. The yellow and red cabling reflects the two InfiniBand networks as shown in Figure 3-20 on page 56.

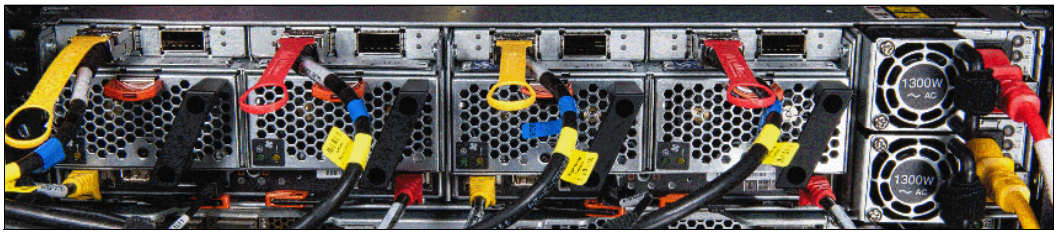


Figure 3-22 FlashSystem A9000R flash enclosure with InfiniBand cabling

3.3.2 InfiniBand switch battery backup unit

The InfiniBand switches build a core component in the back end of FlashSystem A9000R. Therefore, it is essential in a power loss of both power sources that the IB switches remain online while the grid controller and the flash enclosures power down. To ensure the availability of both switches, each power supply contains a battery backup unit (BBU). The BBUs can be replaced, concurrently, without the need to take any switch offline. The batteries in the InfiniBand switch are constantly monitored, as are the batteries in the grid controller and the flash enclosure, and are regularly calibrated, too.

Figure 3-23 shows the battery unit of an IB switch. The left picture shows that the battery is pulled out from the power supply. The right picture shows the battery (black) when it is installed in the power supply.

Figure 3-23 IB switch battery and power supply

3.3.3 FlashSystem A9000R rack

FlashSystem A9000R is delivered in its own rack. The solid rack is a proven standard rack, IBM T42. It provides space for up to 42 standard units so that you can order FlashSystem A9000R to up to 12 grid controllers and six flash enclosures, available with Model 415.

All cables are in place and connected in the rack by IBM manufacturing before the system is delivered. Even in a partially populated system, all required cables for potential future growth (IB cables, network cables, and power cables) are preinstalled.

At positions 19 - 24, the rack holds the back-end infrastructure of FlashSystem A9000R. The components are the power distribution units 1 and 2, management patch panel, and IB switches 1 and 2. For a schematic view of the rack and the installed components, see Figure 3-26 on page 62.

For more information about the rack options and site requirements, see 5.1, “Physical space requirements” on page 110.

Network connections for management

On the back of the rack, a patch panel is installed to connect the management, VPN, and service ports. Figure 3-24 shows the management port patch panel. It contains the following ports (from left to right):

- Management 1 - 3 to connect to the customer internal management network.
- VPN 1 and 2 to provide a connection to a second network, which allows outbound internet connection for remote support.
- Tech so that an IBM SSR can connect a notebook computer for maintenance reasons. A second tech port is in the front bezel at the front of the rack.
- Maint 1 and 2 so that an IBM SSR can connect to the maintenance module (may not be present on newer systems). The maintenance module is discontinued with FlashSystem A9000 and A9000R Model 425 and the ports are no longer present. It is also discontinued for Model 415 units shipped after March 2017.
- Modem to connect a phone line - not used. Port is removed from the new patch panel.
- Power utility to provide AC power for an IBM SSR’s notebook computer or a monitor, which might be required for service by an SSR.

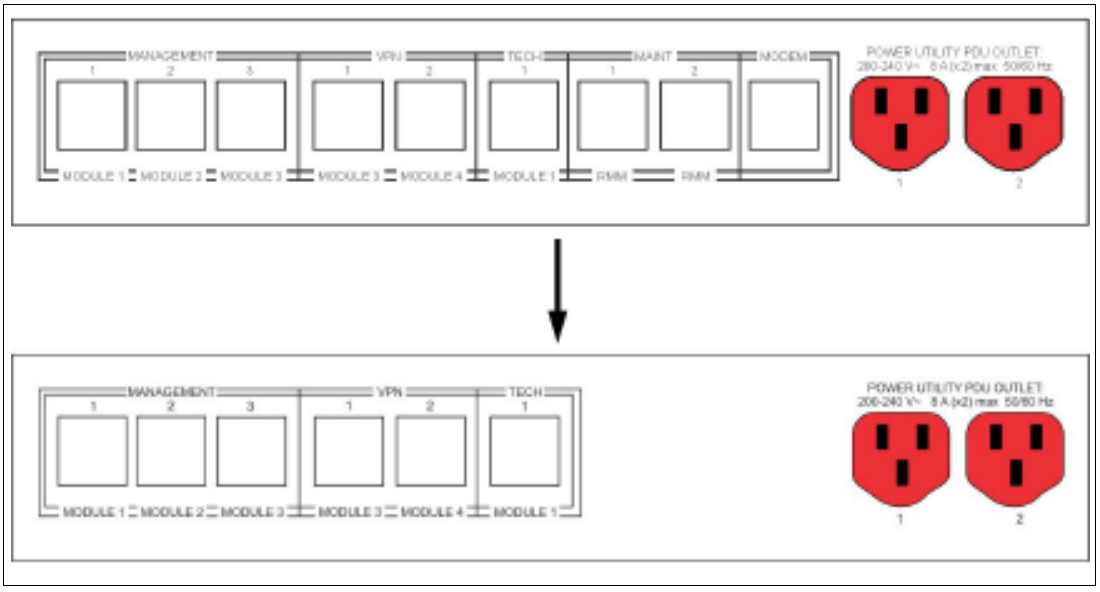


Figure 3-24 FlashSystem A9000R Model 415 (top) and A9000R Model 425 (bottom) utility patch panel

3.4 Scaling FlashSystem A9000R

Any flash enclosure in FlashSystem A9000R is populated with 12 MicroLatency modules. Different capacity options are available for the MicroLatency modules - either 3.6 TB, 8.5 TB, or 18 TB for Model 425 and either 2.9 TB or 5.7 TB for Model 415.

Important: The size of all MicroLatency modules in FlashSystem A9000R must be equal, and they cannot be changed later.

FlashSystem A9000R consists of two to four grid elements for Model 425 or of two to six grid elements for Model 415, depending on the configuration ordered.

All data that is stored on the system will be evenly redistributed across all flash enclosures during the installation process so that all grid modules and flash enclosures are evenly used.

Figure 3-25 on page 60 shows possible scaling options for FlashSystem A9000R. The minimum configuration, which is shown on the left, is the Grid-starter configuration. The Grid-Starter configuration is only valid for a Model 425. A fully configured rack contains four grid elements for Model 425 and was six grid elements for the discontinued Model 415.

Depending on the number of grid elements in your order, they are stacked in the rack in the sequence that is shown. The next grid element is always installed in the next free location in the rack, starting from the bottom.

Grid elements installed must all be the same type (interface options and MicroLatency module size). An intermix of interface options or MicroLatency module capacities is not supported.

With the new entry level configuration, the scaling options planned[2] for the Model 425 are shown in Figure 3-25.

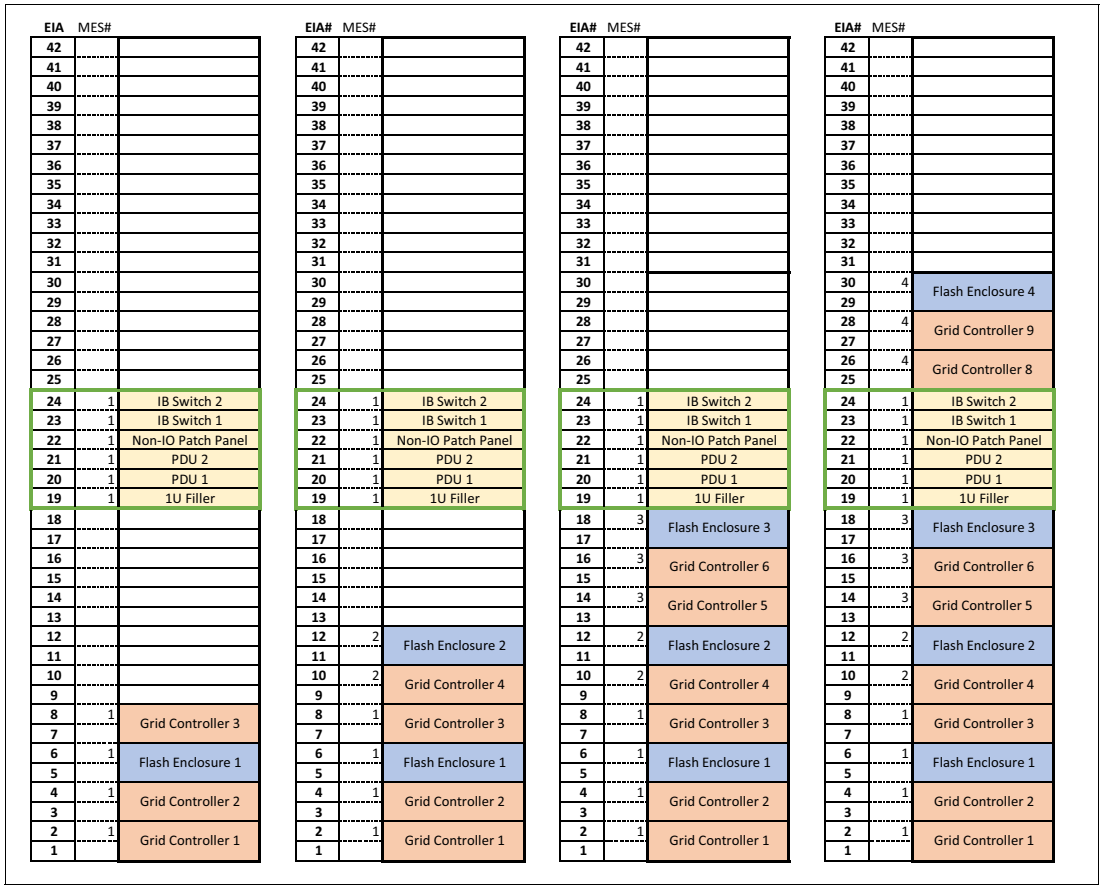


Figure 3-25 FlashSystem A9000R model 425 scaling options

[2] IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remain at IBM's sole discretion.

3.5 FlashSystem A9000 specifics

The following tables show a list of the components in FlashSystem A9000 with capacity details. Table 3-9 shows the details for Model 425 and Table 3-10 shows the details for 415.

Table 3-9 FlashSystem A9000 Model 425 configurations

FlashSystem A9000 Model 425 configurations	
Number of grid controllers	3
CPUs (cores)	6 (72)
Memory in GBs	1152
iSCSI (10 Gb) only ports	12
iSCSI (10 Gb) + FC (16 Gb) ports	6 + 12
Number of flash enclosures	1
MicroLatency modules	8[a] or 12
MicroLatency module capacity in TBs (three options)	3.6/8.5/18
Usable physical capacity in TBs	21.6[b]/36[c]/85/180
Effective capacity in TBs[d]	110[b]/180[c]/425/900
Maximum effective capacity in TBs[e]	1200/1200/1200/1200

a. The option is available only when using the 3.6 TB MicroLatency Modules.
b. Achieved in configuration with 8 MicroLatency Modules of 3.6 TB capacity.
c. Achieved in configuration with 12 MicroLatency Modules of 3.6 TB capacity.
d. Effective capacity assumes a data reduction that is calculated at about 5:1.
e. Maximum effective capacity is the up-most provisioning limit that effective capacity can be stretched to, by IBM.

Table 3-10 FlashSystem A9000 Model 415 configurations

FlashSystem A9000 Model 415 configurations	
Number of grid controllers	3
CPUs (cores)	6 (48)
Memory in GBs	576
iSCSI (10 Gb) only ports	12
iSCSI (10 Gb) + FC (16 Gb) ports	6 + 12
Number of flash enclosures	1
MicroLatency modules	12
MicroLatency module capacity in TBs (three options)	1.2/2.9/5.7
Usable physical capacity in TBs	12/29/57
Effective capacity in TBs[a]	60/150/300
Maximum effective capacity in TBs[b]	1200/1200/1200

a. Effective capacity assumes a data reduction that is calculated at about 5:1.
b. Maximum effective capacity is the up-most provisioning limit that effective capacity can be stretched to, by IBM.

3.5.1 FlashSystem A9000 direct InfiniBand

InfiniBand is the high-performance network that is used by FlashSystem A9000 and A9000R to connect the components internally. It is the redundant, reliable connection backbone between the components. InfiniBand assures high throughput and low latency, and it is optimal for connecting grid controllers and flash enclosures.

An InfiniBand connection between the grid controllers is used for internal management and data transfer. The management task uses IP over InfiniBand (IPoIB) for concurrent code upgrade, connectivity status checks, and so on. The InfiniBand connection between grid controllers and flash enclosures is used for data transfer only. Management of the flash enclosures is performed from the grid controllers to the flash enclosures by using two daisy-chained Ethernet connections.

FlashSystem A9000 uses direct InfiniBand connections to connect the three grid controllers among each other and to connect the flash controller to the three grid controllers. The interconnect between the components is direct without the requirement to use an InfiniBand switch. The cabling is performed by an IBM SSR during the system installation.FlashSystem A9000 back-end communication is based on a high-performance redundant InfiniBand back end. To ensure that no single point of failure exists in the back end, each grid controller has one connection to every other grid controller, plus two InfiniBand connections to the flash enclosure.

See the following figures for a scheme of the FlashSystem A9000 InfiniBand cabling. Figure 3-26 shows the cabling scheme for Model 425, and Figure 3-27 on page 63 shows the cabling scheme for Model 415.

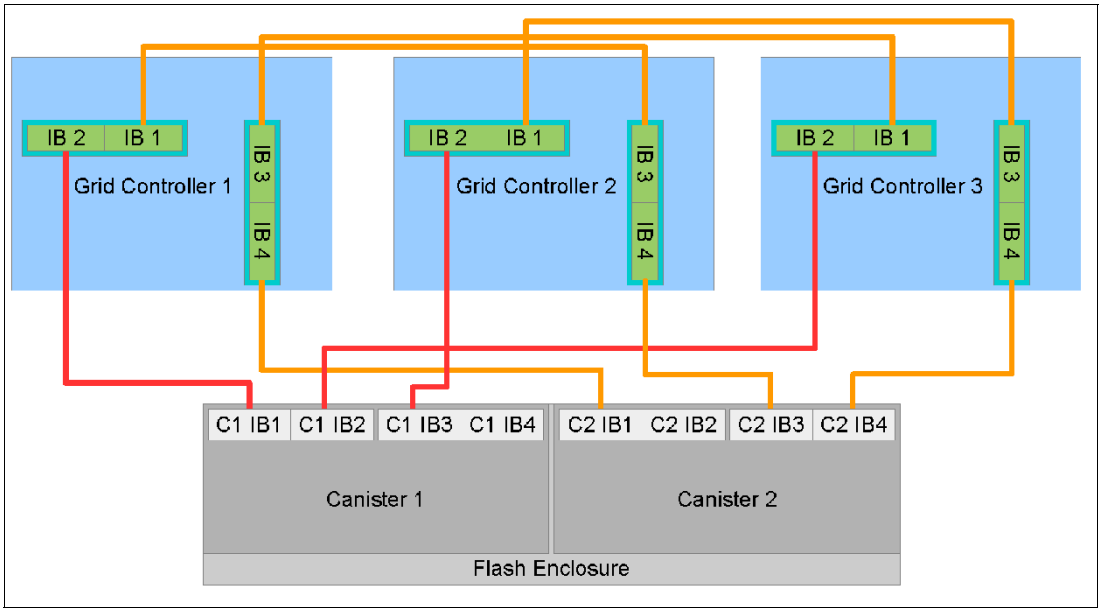


Figure 3-26 FlashSystem A9000 Model 425 InfiniBand cabling scheme

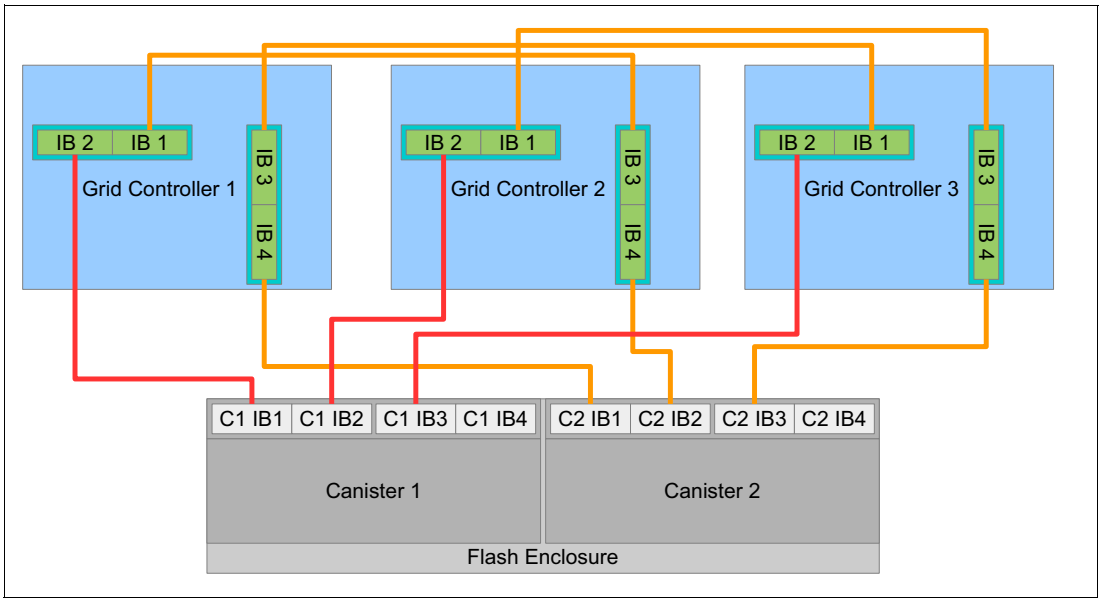


Figure 3-27 IBM FlashSystem A9000 Model 415 InfiniBand cabling scheme

Ethernet and USB cabling in FlashSystem A9000

The grid controller contains four 1 Gb Ethernet ports and four USB ports.

Ports Eth1, Eth2, and Eth4 are used for a daisy chain connection between the grid controller and the flash enclosure. This daisy chain is required for internal management traffic and debugging that are used by IBM Support only. Port Eth3 on grid controllers 1 and 2 on Model 425 and on grid controllers 1 and 3 on Model 415 is used for remote access and management.

IBM technicians use port USB2 on grid controllers 1 and 3 on Model 425 or port Eth3 on grid controller 2 and port Eth4 on grid controller 3 on Model 415 for servicing actions on the system.

For VPN connections, use port USB4 on grid controllers 1 and 2 on Model 425 or port USB4 on grid controllers 2 and 3 on Model 415.

See the figures below for the Ethernet and USB cabling scheme. Figure 3-28 on page 64 shows the cabling on Model 425 and Figure 3-29 on page 64 shows the cabling for Model 415.

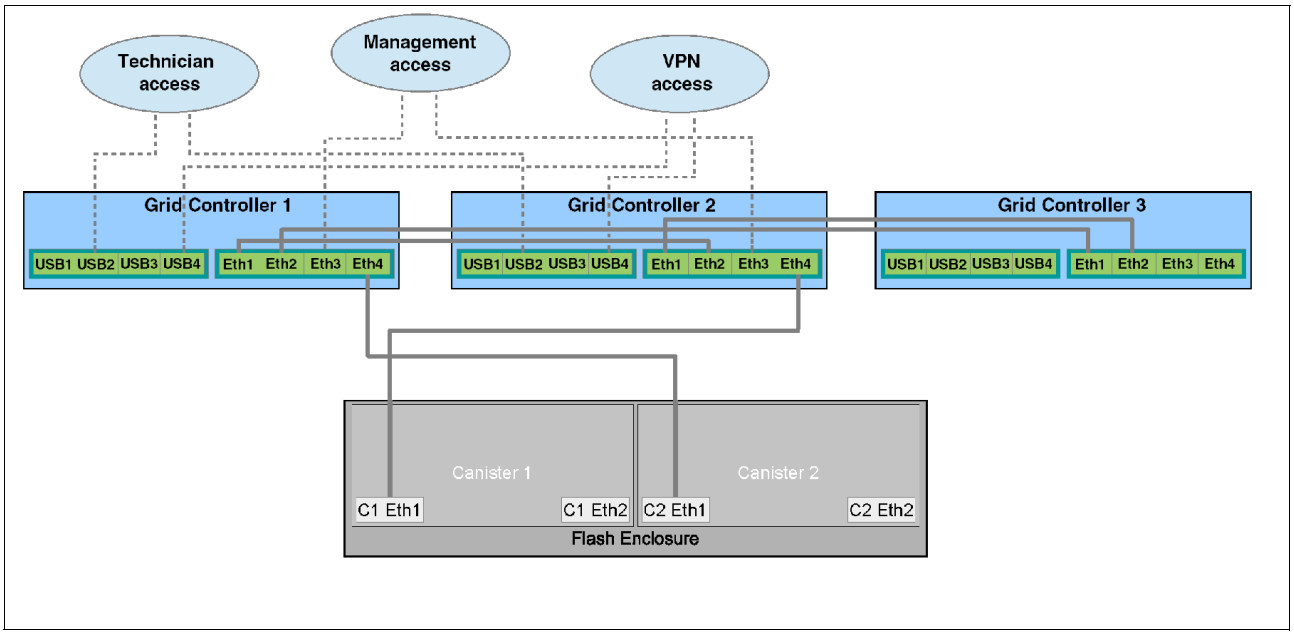


Figure 3-28 FlashSystem A9000 Model 425 Ethernet and USB cabling

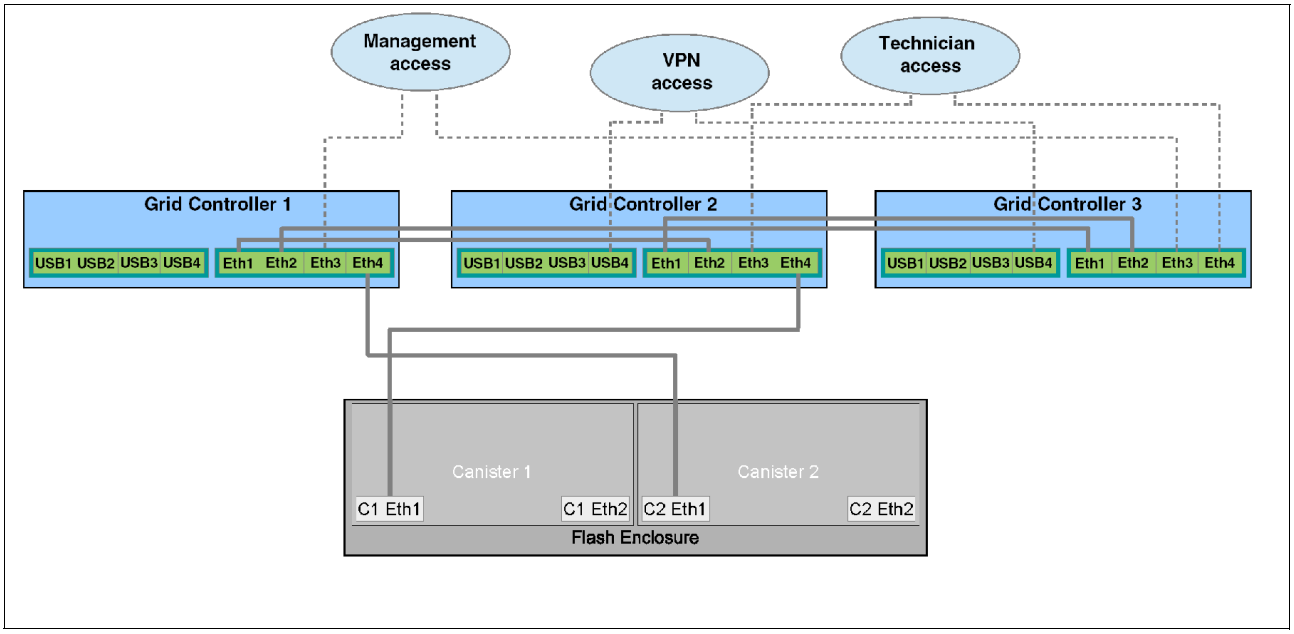


Figure 3-29 FlashSystem A9000 Model 415 Ethernet and USB cabling

I/O connectivity

Depending on the ordered configuration (the number of iSCSI and FC ports), FlashSystem A9000 can have twelve 10 GbE ports or six 10 GbE ports and 12 FC ports. All ports can be used for host, mirroring, or migration connectivity. Every device must be connected through an Ethernet or Fibre Channel switch. Direct connection to FlashSystem A9000 is not supported.

For more information about host connectivity, see *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368.

3.6 Reliability, availability, and serviceability

The unique modular design and logical topology of FlashSystem A9000 and A9000R fundamentally differentiate them from traditional, monolithic systems. This architectural divergence extends to the exceptional reliability, availability, and serviceability (RAS) aspects of the system.

FlashSystem A9000 and A9000R also incorporate autonomic and proactive monitoring and self-healing features. These features enable preventive measures to preserve data redundancy before a component malfunction occurs.

The system is automatically restored to full redundancy within minutes of a hardware failure. When a grid controller fails, its workload is directly taken over by another grid controller.

The reliability features of a flash enclosure are described in 3.7, “IBM FlashCore technology” on page 67.

3.6.1 Reliability

Reliability is engineered at many different levels:

- The separation of grid controllers and the flash enclosure protects against the impacts of a failed grid controller. A grid controller can fail without any effect to the flash enclosure. The tasks of a failed grid controller are taken over by the other grid controllers.

  Note: Correct host multipathing must be used to prevent host disconnection.

- Redundant hot-swappable components:
  - Each flash enclosure has two clustered, hot-swappable canisters, each of which contains two hot-swappable fan modules, two management controllers, four management Ethernet ports, and a USB port for service connectivity. The batteries, fans, and power supplies are all redundant and hot-swappable.
  - If a MicroLatency module failure occurs, critical customer applications can remain online while the defective module is replaced. Each flash enclosure is configured with a hot spare module for use in case of a MicroLatency module failure.
  - Each grid controller contains two redundant, hot-swappable power supplies. The cooling system is resilient to keep the grid controller running, even with up to two fans in the failed state. The microcode is stored on two RAID1 secured HDDs. The data and functions of each grid controller are mirrored in a three-way manner.
- The two-dimensional RAID protection (2D-RAID) of the flash enclosure protects against the MicroLatency module, flash chip, and other flash system-related failures. The 2D-RAID consists of IBM Variable Stripe RAID (VSR) and flash enclosure-wide RAID 5. IBM Variable Stripe RAID is a patented IBM technology that provides an intra-module RAID stripe within each flash module, and it is described in 3.8.1, “Variable Stripe RAID” on page 75. Variable Stripe RAID technology helps reduce downtime and maintain performance and capacity during partial or full flash chip failures.
- A MicroLatency module can fail without any effect to the flash enclosure. When a MicroLatency module fails, the RAID 5 configuration over the MicroLatency modules allows a rebuild by using the spare MicroLatency module in the flash enclosure. Flash enclosure-wide RAID 5 protection against a MicroLatency module failure is described in 3.8.2, “Two-dimensional (2D) Flash RAID” on page 77.

- The flash enclosure runs an internal scrubbing process to verify the integrity of the data. It is a low-priority process to minimize performance impact. If an error is detected during the scrubbing process, the flash enclosure will attempt to correct it. If the attempt was unsuccessful (unrecoverable error), an event will be sent but the flash enclosure will stay online.
- InfiniBand cabling is redundant by using two InfiniBand switches (FlashSystem A9000R), or redundant direct cabling (FlashSystem A9000).
- All active components (grid controllers, flash enclosures, and in FlashSystem A9000R, the IB switches) contain a redundant BBU. In an accidental power loss on both main power sources, a loss in BBU redundancy, or an overheating situation, FlashSystem A9000 and A9000R will automatically perform a shutdown to protect the data (cache and metadata). This process is known as *shutdown vaulting*. During normal system operation, the cache and metadata are also saved to the vaulting devices at regular intervals, approximately every 5 minutes (*live vaulting*). Shutdown vaulting does not overwrite live vaulting.

  Vaulting: The role of the vaulting procedure is to write memory-resident data, which is volatile, onto non-volatile storage during a system shutdown.

  FlashSystem A9000 and A9000R use the SSDs in the grid controller as vault devices. The vaulting process design point is to save three copies of the data, preferably on SSDs in three different grid controllers. If hardware fails, the system still saves at least two copies of the data. The system will not remain online with fewer than two grid controllers.

  The grid controller destages the following data during the shutdown:

  – Data in write cache
  – Metadata that changed since the last vault run

  The flash enclosure also has redundant BBUs to allow a graceful shutdown of the enclosure. During shutdown, it writes the following data:

  – Data in the flash enclosure write cache
  – Recent data distribution table

  After all data is secured, the components power off. If the utility power recovers, after the automatic shutdown, FlashSystem A9000 and A9000R power on automatically.

  When the system reboots, it starts a *devaulting* procedure. The devaulting procedure retrieves the data from the vault devices back to memory upon system boot, as part of the system power-up process. Each vaulting device (SSD) will report to the system the piece of data that it holds. The system uses an internal vault distribution table to assign each piece of data to cache nodes.

  If vaulting or devaulting fails during shutdown or during startup because of a catastrophic hardware or software failure, the system moves to maintenance until IBM intervenes.

  The different shutdown scenarios, such as manual or emergency (automatic) shutdown, are described in 5.6.2, “Completing the physical installation” on page 121.

3.6.2 Availability

Grid controllers have automatic and seamless failover capabilities. The cache of a grid controller is triplicated: one copy is local and two other copies are on two other grid controllers. This design is also described in 2.1.2, “Cache resiliency” on page 13. The system will remain operational if it can still maintain at least one secondary cache.

System-wide performance monitoring assures the availability of the needed software functions. For example, if a software process, such as caching or the data reduction process, is unresponsive, the software process is expelled from the degraded grid controller, and the workload is taken over by other grid controllers. This behavior implies that FlashSystem A9000, which is equipped with three grid controllers, can stand one grid controller outage, and FlashSystem A9000R, which has a minimum of four grid controllers, can stand two simultaneous grid controller failures. The loss of one additional grid controller (an unlikely event) will trigger a graceful system shutdown.

FlashSystem A9000 and A9000R include an enhanced Call Home functionality. When the system is configured for call home, each system reports system events regularly and sends heartbeats with several types of system health information to IBM. Based on the events and heartbeats, IBM Support can automatically open support tickets and initiate service actions, without the need of any manual customer action.

Statistic data, which is gathered by IBM based on the events and heartbeat information, helps IBM identify possible upcoming situations that might need proactive assistance.

The power is monitored, and the battery units are conditioned for longer lifespan. The system temperature is monitored at the flash enclosure level and at the system level. When the temperature gets out of range, a graceful shutdown is triggered.

Concurrent code load enables customer applications to remain online during firmware upgrades to all components, including the flash enclosures and MicroLatency modules.

3.6.3 Serviceability

High levels of serviceability are achieved by providing the following features and functions:

- Thorough testing of all components during the manufacturing process.
- Enhanced Call Home - see 7.1.6, “Call Home” on page 150.
- Logs and statistics are collected and recorded by each interface and port that relate to the data path (Fibre Channel, iSCSI, InfiniBand, and Ethernet).
- Debug utilities: XRAY data collection on FlashSystem A9000 and A9000R and Host Attachment Kit (HAK) diagnostic collection on the host side.
- IBM SSR has access to service tools for guided repairs, pre-upgrade checks, concurrent upgrade, and so on.
- Remote support capabilities - see 7.2.3, “IBM Remote Support” on page 161.

3.7 IBM FlashCore technology

The flash enclosures that are used in FlashSystem A9000 and A9000R offer ultra low-latency, highly reliable, and scalable performance in a space-efficient and power-efficient storage device. In addition to optimizing performance, FlashSystem helps bring enterprise reliability and macro efficiency to the most demanding data centers so that businesses can see the following benefits, based on customer experience:

- Servicing more users with less hardware
- Reducing I/O wait and improving response times of critical applications
- Simplifying solutions
- Reducing power and floor space requirements
- Speeding up applications, enhancing the pace of business

- Improving the utilization of the existing infrastructure
- Extending the existing infrastructure
- Mitigating risk

From the customer business perspective, FlashSystem provides benefits and value in four essential areas:

- Extreme performance enables businesses to unleash the power of performance, scale, and insight to drive services and products to market faster.
- IBM MicroLatency delivers microsecond response time to accelerate critical applications to achieve competitive advantages.
- Macro efficiency decreases costs by getting more from the efficient use of the IT applications and IT equipment because of the efficiencies that flash brings to the data center.
- Enterprise reliability enhances the customer experience through durable and reliable designs that use enterprise-class flash and patented data protection technology.

IBM FlashCore technology, which is used in FlashSystem A9000 and A9000R, is based on patented mechanisms to deliver extreme performance, MicroLatency, macro efficiency, enterprise-grade reliability, and a wide range of operational and cost efficiencies. The flash enclosure of FlashSystem A9000 and A9000R is based on FlashCore technology.

Figure 3-30 shows the three major areas within IBM FlashCore technology and the unique IBM attributes of each area.

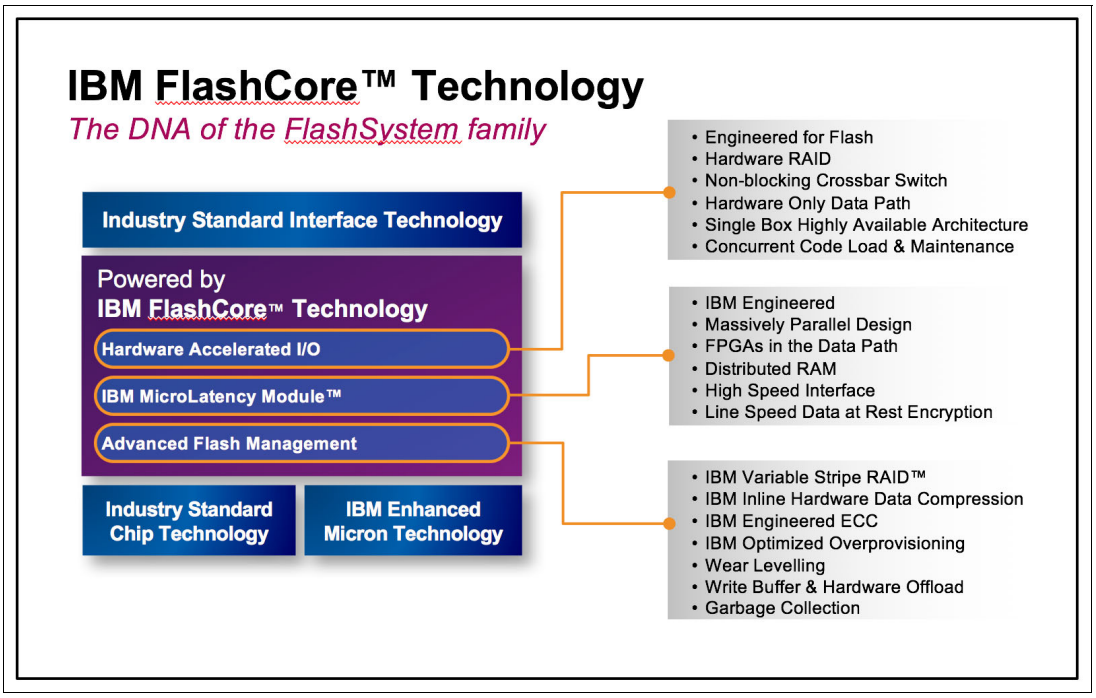


Figure 3-30 IBM FlashCore technology

The three major areas are described:

- Hardware Accelerated I/O

  The flash enclosure hardware design of FlashSystem A9000 and A9000R offers several unique IBM components, including hardware RAID, non-blocking crossbar switch, hardware-only data path, highly available architecture, concurrent code load, and concurrent maintenance. The use of an all-hardware data path design helps to ensure the highest performance and lowest latency of FlashSystem A9000 and A9000R flash enclosures.

- IBM MicroLatency Module

  The flash enclosure of FlashSystem A9000 and A9000R Model 425 uses the new IBM enhanced 3D triple-level cell (3D TLC) flash card memory chips in either 3.6 TB, 8.5 TB, or 18 TB capacity IBM MicroLatency modules.

  The flash enclosure of FlashSystem A9000 and A9000R Model 415 uses the 20 nm IBM enhanced multi-level cell (MLC) flash card memory chips in either 1.2 TB (FlashSystem A9000 only), 2.9 TB, or 5.7 TB capacity IBM MicroLatency modules.

  The design of FlashSystem A9000 and A9000R also employs the use of IBM engineered massively parallel design, field programmable-gate array (FPGA) modules in the data path, distributed RAM, and high-speed interfaces, plus hardware-based data-at-rest encryption.

- Advanced Flash Management

  FlashSystem A9000 and A9000R offer unique patented designs to ensure maximum availability, including IBM Variable Stripe RAID, IBM Inline Hardware Data Compression, IBM engineered error correction code (ECC), IBM optimized over-provisioning, advanced wear-leveling on IBM MicroLatency modules, write buffer and hardware offload, and IBM garbage collection. The wear-leveling algorithm assures the even usage of all blocks. The garbage collection process collects the blocks, which are not used anymore, so that they can be reused for writing.

  The IBM Inline Hardware Data Compression is available only on FlashSystem A9000 and FlashSystem A9000R Model 425. Given that the system makes use of software data reduction, the data being sent to the flash enclosure is already compressed and therefore is bypassing the inline hardware data compression. See 2.3.1, “Data reduction technology” on page 15.

All of this function is possible because of the following IBM patented and world-class innovations:

- ECC algorithms that correct high bit-error rates
- Variable voltage and read-level shifting that help to maximize flash endurance
- Health *binning* and heat segregation, which continually monitor the health of flash blocks and perform asymmetrical wear leveling and sub-chip tiering

These innovations result in up to a 57% improvement in endurance with a potential 45% reduction in write amplification compared to a design without FlashCore.

Next, we explain the technologies, such as hardware-only data path, FPGA, wear-leveling, and MLC/TLC in detail.

3.7.1 Technology and architectural design overview

The flash enclosure of FlashSystem A9000 and A9000R, with an all-hardware data path that uses field programmable-gate array (FPGA) modules, is engineered to deliver the lowest possible latency.

The modules incorporate proprietary flash controllers and use numerous patented technologies. FlashSystem A9000 and A9000R flash enclosure controllers have proprietary logic design, firmware, and system software.

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

The flash enclosure of FlashSystem A9000 and A9000R offers the following notable architectural concepts:

- Hardware-only data path.
- Use of FPGAs extensively.
- Field-upgradable hardware logic.
- Less expensive design cycle.
- High degree of parallelism.
- Intelligent flash modules.
- Distributed computing model.
- Low-power IBM PowerPC® processors (PPCs).
- Interface and flash processors run thin real-time operating systems.
- With minimal management communication, the management processor communicates with the interface and flash processors through an internal network.

3.7.2 Hardware-only data path

The hardware-only data path design of FlashSystem A9000 and A9000R flash enclosure eliminates software layer latency. In the flash enclosure of FlashSystem A9000 and A9000R, data traverses the flash enclosure through FPGAs. No cycles are wasted on *interface* translation, protocol control, or tiering. FPGAs are in the interface cards, on the RAID controllers, and in the MicroLatency modules.

The flash enclosure of FlashSystem A9000 and A9000R with an all-hardware data path design has an internal architecture that differs from other hybrid (SSD and HDD) or SSD-only-based disk systems.

3.7.3 Flash card memory chips

The *flash chip* is the basic storage component of the IBM MicroLatency module. To maintain consistent wear and reliability, all MicroLatency modules in a flash enclosure use flash chips of the same flash technology.

The MicroLatency modules in FlashSystem A9000 and A9000R Model 415 use 20nm IBM enhanced multi-level cell (MLC) chips, which means that the chip lifetime is increased by IBM technology. MLC cells can store 2 bits in one cell, and these 2 bits represent the four values 0, 1, 2, and 3 of the cell. This results in high density flash memory per module while it still retains high performance.

The flash enclosure of FlashSystem A9000 and A9000R Model 425 employs the new 3D-TLC chips, which are of a higher density (storing 3 bits in one cell) than the MLC chips used in Model 415 enclosures. This new design of chips allows the FlashSystem A9000 and A9000R Model 425 to package greater densities of flash memory per card whilst retaining the same if not better performance and wear.

IBM patented ECC checking and correction algorithms ensure the same or greater performance from both the MLC or the 3D-TLC based chips, with a greater capacity for the same footprint and at a lower cost per TB.

Flash chips have a logical structure. A chip is divided into *dies*, then into *planes*, then into *blocks*, and then into *pages*. Figure 3-31 shows the logical structure of a chip. A chip is divided into eight dies, and a die is divided into two planes, resulting in 16 planes inside a chip. The 16 planes are the basis for the variable RAID that is explained in 3.8, “IBM Variable Stripe RAID and 2D Flash RAID protection overview” on page 74.

Figure 3-31 shows the logical flash chip layout.

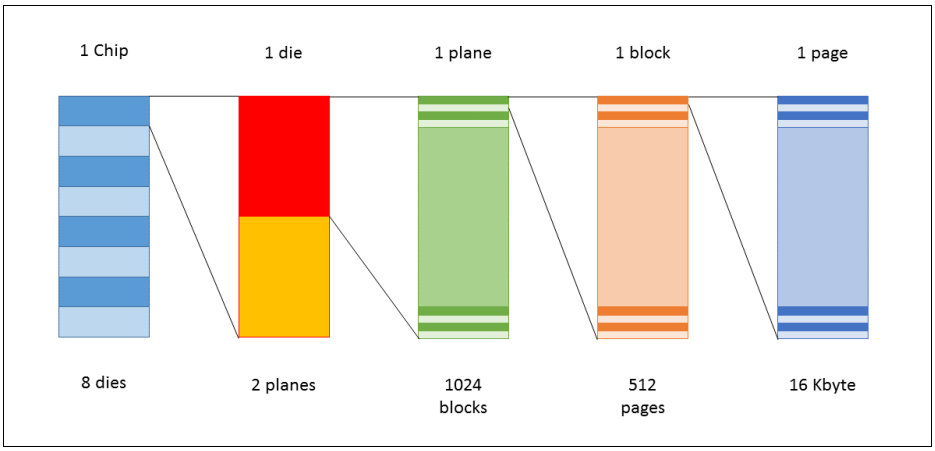


Figure 3-31 Logical flash chip layout

3.7.4 Flash module capacities

The flash enclosure of FlashSystem A9000 and A9000R Model 415 uses 1.2 TB (FlashSystem A9000 only), 2.9 TB, or 5.7 TB IBM MicroLatency modules.

The flash enclosure of FlashSystem A9000 and A9000R Model 425 uses either 3.6 TB, 8.5 TB, or 18 TB IBM MicroLatency modules. This is a 3x increase in capacity per module over Model 415 modules.

The modules must be the same capacity throughout FlashSystem A9000 and A9000R. The flash enclosure is equipped with 12 MicroLatency modules (FlashSystem A9000 Model 425 can be equipped with 8 MicroLatency modules). It uses RAID 5 over n+1 MicroLatency modules and one spare MicroLatency module for a total usable capacity of (6 or) 10 times the capacity of the used MicroLatency modules’ capacity.

Empty bays: If fewer than 12 MicroLatency modules are installed, flash module fillers must be installed in the empty bays to maintain cooling airflow in the system enclosure. Figure 3-32 on page 72 shows the flash module fillers to the left hand side of the installed flash modules.

IBM MicroLatency modules capacity is shown on Model 415 flash enclosures but not shown on the labels of Model 425 flash modules. They use coding for the module type and size as follows. Figure 3-32 shows the coding as “T S03 B”:

- **T** = 3D **T**LC flash technology
- **S** = **S**mall, **M**edium, or **L**arge capacity IBM MicroLatency module
- **03** = **3.6 TB** NAND chips used on the flash module cards
- **B** = Manufacturing use denoting the revision of the IBM MicroLatency module

Figure 3-32 IBM MicroLatency modules capacity labels and flash module filler cards

3.7.5 Field Programmable Gate Array

The flash enclosure of FlashSystem A9000 and A9000R uses FPGAs in many locations.

The two most important locations for lowest latency are inside the IBM MicroLatency module:

- Gateway interface FPGA

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

- Flash controller FPGA

  The flash controller FGPA of the flash module provides access to the flash chips, and it is responsible for the following functions:

  - Provides data path and hardware I/O logic
  - Uses lookup tables and a write buffer
  - Operates independently of other controllers
  - Maintains write ordering and layout
  - Provides write setup
  - Maintains garbage collection
  - Provides error handling

Figure 3-33 shows the flash controller design details.

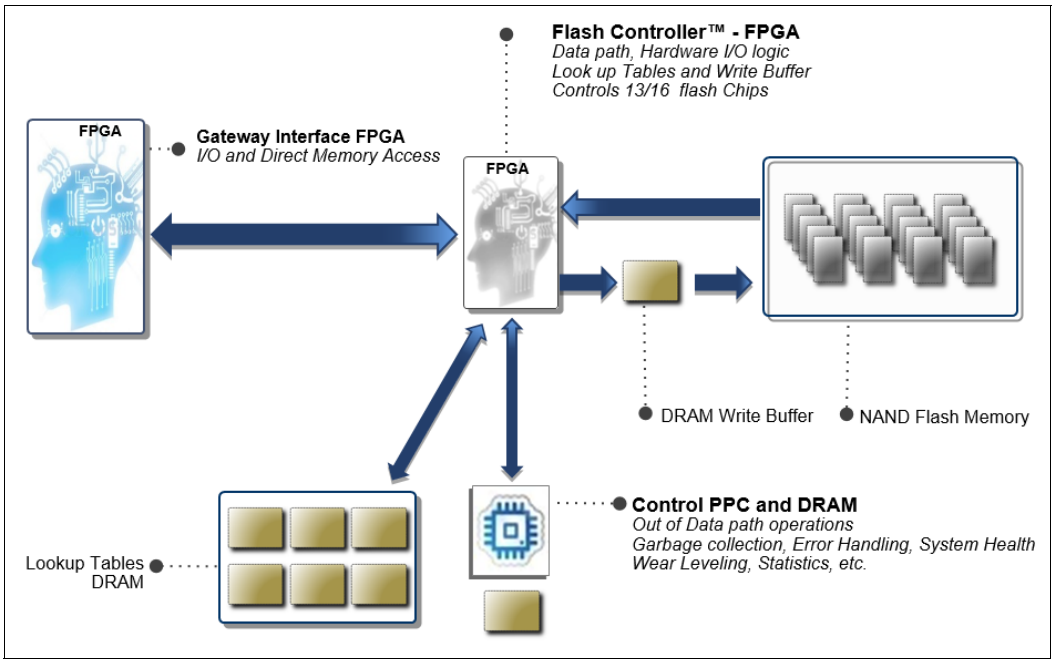


Figure 3-33 FlashSystem MicroLatency module controller details

The concurrent operations that are performed on the flash chips include moving data in and out of the chip through DMA, and by internally moving data and performing erasures. While the system actively transfers user data from the grid controllers, it can simultaneously run garbage collection activities without affecting the I/O.

The number of flash controller FPGAs varies, depending on the size of the MicroLatency module.

3.7.6 Overprovisioning

The MicroLatency modules are *overprovisioned*, which means that the internal flash capacity is much larger than the capacity that is shown to the user. This additional internal flash capacity enhances the lifetime of the modules and enables features, such as IBM variable stripe size, as described in 3.8, “IBM Variable Stripe RAID and 2D Flash RAID protection overview” on page 74.

Flash cells must be erased before they can be rewritten. Overprovisioning allows new data to be written to the MicroLatency modules while a background process erases no longer used cells. This approach prevents performance impacts because writable cells are always available.

The overprovisioning inside the MicroLatency modules is used for FlashCore tasks inside the MicroLatency modules. In addition, spare capacity exists inside a flash enclosure. This spare capacity is the spare MicroLatency module that is used in the RAID 5 context, as described in 3.8.2, “Two-dimensional (2D) Flash RAID” on page 77.

3.7.7 Wear leveling

Writing to flash cells results in wear on those flash cells. To prevent failures from cells wearing out, FlashCore ensures that all cells have the same level of erase and write quality. For example, data that is written only once and then only read will be reallocated so that those cells are the same quality as all of the other cells on the flash enclosure.

3.8 IBM Variable Stripe RAID and 2D Flash RAID protection overview

Storage systems of any kind are typically designed to perform two main functions: store data and protect data. FlashSystem A9000 and A9000R use cache triplication, as explained in 2.1.2, “Cache resiliency” on page 13. The flash enclosure of FlashSystem A9000 and A9000R includes the following data protection:

- IBM enhanced error correction code (ECC)

  IBM developed a proprietary, innovative, and strong ECC algorithm that corrects high bit error rates.

- MicroLatency module RAID data protection

  Each MicroLatency module uses RAID 5 protection for the flash chips inside the controller. The patented IBM RAID protection inside the MicroLatency module helps to ensure a long lifetime for the MicroLatency module. The size of the RAID is variable. For more information, see 3.8.1, “Variable Stripe RAID” on page 75.

- FlashSystem A9000 and A9000R flash enclosure RAID protection

  The flash enclosure is equipped with 12 MicroLatency modules (FlashSystem A9000 Model 425 can be equipped with 8 MicroLatency modules). It uses RAID 5 over n+1 MicroLatency modules and one spare MicroLatency module for a total usable capacity of (6 or) 10 times the capacity of the used MicroLatency modules’ capacity. For more information, see 3.8.2, “Two-dimensional (2D) Flash RAID” on page 77.

The two times RAID 5 protection is called IBM Two-Dimensional (2D) RAID protection.

Table 3-11 lists the various methods of protection.

Table 3-11 Various types of FlashSystem A9000 and A9000R flash enclosures protection

Layer	Managed by	Protection
Flash enclosure-level RAID 5	Centralized RAID controllers	MicroLatency module failure
MicroLatency module-level RAID 5	Each MicroLatency module across the chips by using a dedicated FPGA	Multi-chip, chip, and sub-chip failures
Chip-level error correction code (ECC)	Each MicroLatency module by using the chips	Block, page, and bit errors

Note: The proprietary 2D Flash RAID data protection scheme of the FlashSystem A9000 and A9000R flash storage enclosure combines system-level RAID 5 and module-level Variable Stripe RAID (not only module-level RAID).

3.8.1 Variable Stripe RAID

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

VSR provides high redundancy across chips within a flash module. RAID is implemented at multiple addressable segments within chips, in an *n*+1 RAID 5 manner, and it is controlled at the flash controller level. Because of the massive parallelism of DMA operations that are controlled by each FPGA and the parallel access to chip sets, dies, planes, blocks, and pages, the implementation of VSR has no impact on performance.

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

- VSR is managed and controlled by each of the flash controllers within a single module.
- Each flash controller is in charge of a set of flash chips, depending on the IBM MicroLatency module capacity size.
- Data is written on flash pages of 16 KB and erased in 8 MB flash blocks.
- VSR is implemented and managed at flash chip *plane* levels.
- Sixteen planes are in each chip, and 1,024 blocks are in each plane.
- Before a plane fails, at least 256 flash blocks within a plane must be deemed *failed*.
- A plane can also fail in its entirety.
- Up to 64 planes can fail before a whole MicroLatency module is considered *failed*.
- Up to four chips can fail before a whole MicroLatency module is considered *failed*.
- When a flash module is considered *failed*, 2D Flash RAID takes control of data protection and recovery.
- When a plane or a chip fails, VSR activates to protect data while it maintains system-level performance and capacity.

How Variable Stripe RAID works

VSR is a technology that is patented by IBM. It includes but is more advanced than a simple RAID of flash chips. VSR introduces several key concepts:

- The RAID stripe is not solely across chips. It spans across the planes. Sixteen stripe layers exist for each flash chip controller FPGA.
- The RAID stripe can automatically vary based on observed flash plane failures within a flash module. For example, stripes are not fixed at *n*+1 RAID 5 stripe members, but they can go down to *n*-1, *n*-2, or even *n*-3 based on plane failures.
- VSR reduces maintenance intervals that are caused by flash failures.
- While flash is written to, parity is written across different flash chips.
- Data is written to flash chip pages with ECC. ECC errors are resolved with the stripe parity.
- In a failure, the patented VSR allows RAID stripes to vary.
- In a failure, a bad block is remapped to another location and ignored for any future reuse.
- If the plane of a chip fails, the corresponding VSR is degraded and its data will be moved to other planes. After it is copied, the remaining block stripe members are erased and moved to the “Read To Use” table with one fewer member in the VSR stripe.

This ability to protect the data at *variable* stripes maximizes flash capacity even after flash component failures. Figure 3-34 shows an overview of IBM FlashSystem Variable Stripe RAID.

Figure 3-34 shows a plane failure on the third plane of the second chip. The VSR over the third plane is now a 14+1 RAID and the other planes have a 15+1 VSR.

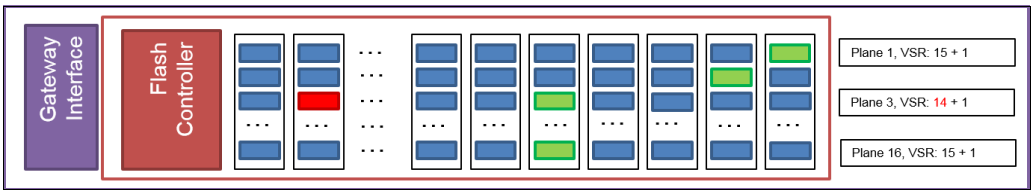


Figure 3-34 FlashSystem Variable Stripe RAID (VSR)

Figure 3-35 shows the benefits of IBM Variable Stripe RAID.

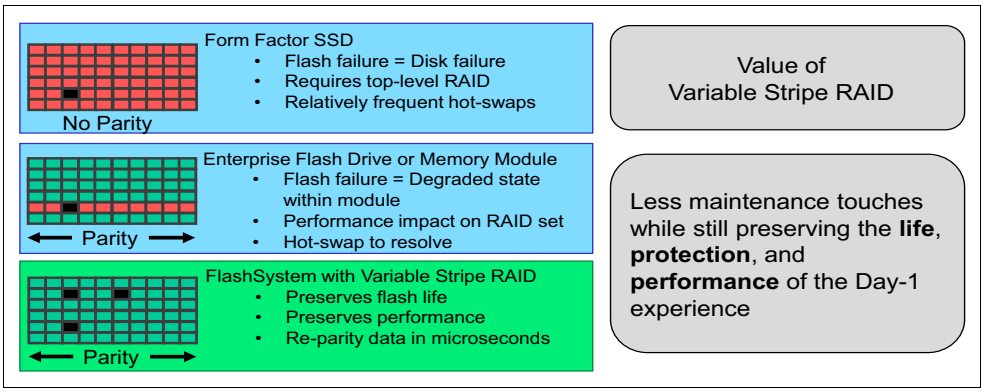


Figure 3-35 The value of FlashSystem Variable Stripe RAID

Important: Variable Stripe RAID affects the *plane* level only. Therefore, *only* the affected planes within a plane failure are converted to a new RAID stripe with one fewer plane. VSR maintains the current stripe member count (*n*+1) layout through the rest of the areas of all other planes and chips that are not involved in the plane failure.

To illustrate how VSR functions, assume that a plane fails within a flash chip and it is no longer available to store data. This failure might occur as a result of a physical failure within the chip or if the address or power lines to the chip are damaged. The plane failure is detected, and the data of that stripe is moved to another location that is protected by RAID 5. Then, the data is protected by RAID 5 again.

The RAID stripe with the failed plane is erased and protected with RAID 5 without the failed plane. If, previously, 16 planes (from 16 different chips) were used for RAID 5, now 15 planes (from 15 different chips) will be used to build a new RAID 5 stripe. Therefore, no data that is stored in the system was lost, and the MicroLatency module self-adapted to the failure and continues to perform and operate by processing read and write requests from host devices.

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

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

Reliability, availability, and serviceability of the flash enclosure

The previous explanation points out an increase in reliability, availability, and serviceability (RAS) levels and FlashSystem A9000 and A9000R flash enclosure RAS levels over other technologies.

In summary, Variable Stripe RAID has these capabilities:

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

3.8.2 Two-dimensional (2D) Flash RAID

Two-dimensional (2D) Flash RAID refers to the combination of Variable Stripe RAID (at the flash module level) and FlashSystem A9000 and A9000R flash enclosure-level RAID 5.

The second dimension of data protection is implemented across MicroLatency modules of RAID 5 protection. This flash enclosure-level RAID 5 is striped across all MicroLatency modules, depending on the selected configuration:

- 8 MicroLatency modules (6 data modules + 1 parity module + 1 hot spare module)
- 12 MicroLatency modules (10 data modules + 1 parity module + 1 hot spare module)

Figure 3-36 shows FlashSystem 2D RAID.

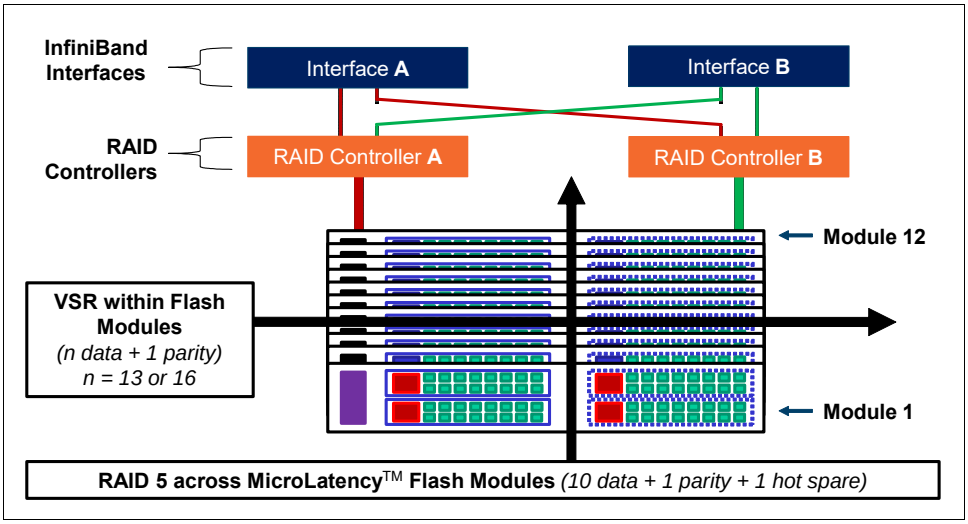


Figure 3-36 IBM FlashSystem 2D Flash RAID protection

FlashSystem A9000 and A9000R flash enclosure-level RAID 5 complements the VSR technology that is implemented within each MicroLatency module, and it provides protection against data loss and data unavailability that result from MicroLatency module failures. It also allows data to be rebuilt onto a hot spare flash module so that a MicroLatency module can be replaced without data disruption.

In addition to 2D Flash RAID protection and VSR data protection, the flash enclosure of FlashSystem A9000 and A9000R incorporates other reliability features:

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

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

4

Capacity planning and management

Capacity planning is an essential consideration when you acquire or provision a new storage system. This chapter helps you determine capacity requirements when using IBM FlashSystem A9000 and A9000R and assumes that you have a good understanding of your environment and applications.

You must also understand how the storage system manages and accounts for capacity. This aspect is relevant for systems, such as IBM FlashSystem A9000 and A9000R, that feature sophisticated and efficient data-reduction functions. The key is to understand how much effective data space your system can provide.

The chapter also explains how capacity is represented and reported by FlashSystem A9000 and A9000R.

It includes the following sections:

- Introduction and definitions
- Capacity planning with data reduction
- Capacity representation and reporting
- Intelligent capacity management for deduplication

4.1 Introduction and definitions

Capacity can be looked at from many sides. This section clarifies the terms in the context of IBM FlashSystem A9000 and A9000R.

4.1.1 System capacity specifications

When you look at the system specifications, the following capacities are presented, as shown in Figure 4-1.

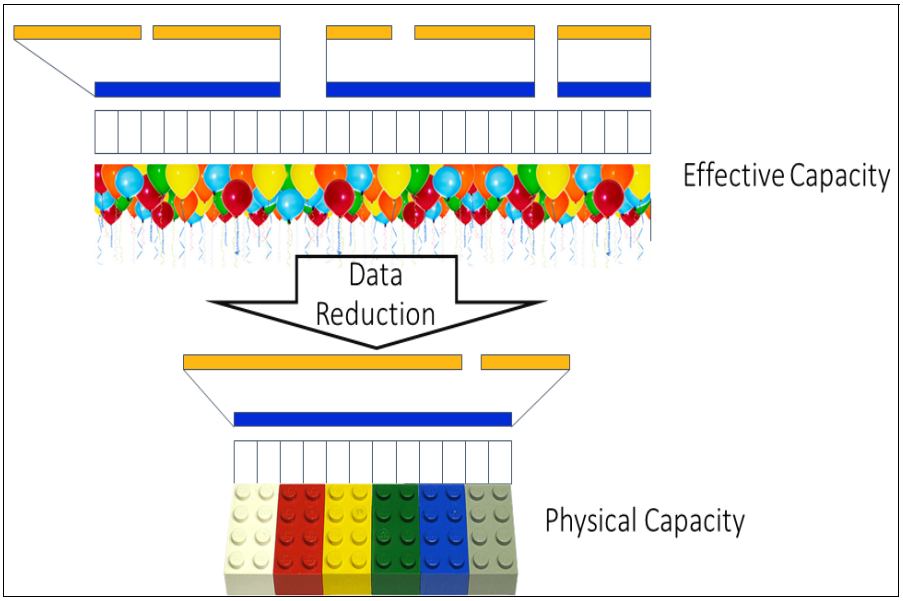


Figure 4-1 Effective capacity and physical capacity

- **Physical capacity**

  The *physical capacity* (or physical usable capacity) represents the amount of space that is physically available in the system to store data. It is the sum of the usable capacity of all flash enclosures.

  Flash enclosure capacity can be categorized into raw and physical usable capacity.

  The Flash MicroLatency modules contain flash chips that are not included in the calculation of usable capacity. The MicroLatency modules have more flash capacity than their stated size. This flash *overprovisioned* capacity is space that is used for several reasons that are inherent to the flash technology, including garbage collection and flash wear leveling.

  When you refer to 1.2 TB, 2.9 TB, or 5.7 TB MicroLatency modules in FlashSystem A9000 and A9000R Model 415, and 3.6 TB, 8.5 TB, or 18 TB in FlashSystem A9000 and A9000R Model 425, the overprovisioned capacity is already discounted.

  Both FlashSystem A9000 and A9000R use a RAID5 implementation across the MicroLatency modules of each flash enclosure. Therefore, MicroLatency modules that are used for hot spare and parity are not counted as part of the usable capacity.

  Out of 8 or 12 MicroLatency modules in a flash enclosure, two modules, one as a hot spare and one as parity, are not counted. For example, a flash enclosure that is equipped with 2.9 TB Microlatency modules has a usable capacity of 2.9 TB x 10 = 29 TB of physical, usable capacity.

- **Effective capacity**

  With data reduction, less data is effectively written and more space becomes available. This is referred to as *effective capacity*. It represents the capacity written by hosts without data reduction.

  The effective capacity varies depending on the type of data that is stored, and the data reduction ratio that can be achieved. It is in fact the maximum capacity available if the reduction ratio remains at its current value.

  By default, a data reduction of about 5:1 is assumed, which means that an effective capacity of 60 TB (when a flash enclosure is equipped with 1.2 TB MicroLatency modules), 150 TB (when a flash enclosure is equipped with 2.9 TB MicroLatency modules), or 300 TB (when a flash enclosure is equipped with 5.7 TB MicroLatency modules) for IBM FlashSystem A9000 and A9000R Model 415.

  For FlashSystem A9000 and A9000R Model 425, the effective capacity is 110 TB (when a flash enclosure is equipped with 8 x 3.6 TB MicroLatency modules), 180 TB (when a flash enclosure is equipped with 12 x 3.6 TB MicroLatency modules), 425 TB (when a flash enclosure is equipped with 8.5 TB MicroLatency modules), or 900 TB (when a flash enclosure is equipped with 18 TB MicroLatency modules). With a higher data reduction factor, you can obtain higher effective capacity.

  Maximum effective capacity or **allocation limit:** By design, the system can support a maximum effective capacity, but no guarantee is implied that you can reach that maximum.

See also 4.3.2, “Capacity representation in the Hyper-Scale Manager GUI” on page 91 for additional capacity representation terminology.

Different system capacities are summarized in Table 4-1 for FlashSystem A9000 Model 425, in Table 4-2 on page 82 for FlashSystem A9000 Model 415, in Table 4-3 on page 82 for FlashSystem A9000R Model 425, and in Table 4-4 on page 83 for FlashSystem A9000R Model 415.

Table 4-1 FlashSystem A9000 Model 425 capacity

FlashSystem A9000 Model 425 capacity	
Number of grid controllers	3
Number of flash enclosures	1
MicroLatency modules	8[a] or 12
MicroLatency module capacity in TB (three options)	3.6/8.5/18
Usable physical capacity in TB	21.6[b]/36[c]/85/180
Effective capacity in TB[d]	110[b]/180[c]/425/900
Maximum effective capacity in TB[e]	1200/1200/1200/1200

a. The option is available only when using the 3.6 TB MicroLatency Modules.
b. Achieved in configuration with 8 MicroLatency Modules of 3.6 TB capacity.
c. Achieved in configuration with 12 MicroLatency Modules of 3.6 TB capacity.
d. Effective capacity assumes a data reduction that is calculated at about 5:1.
e. Maximum effective capacity is the up-most provisioning limit that effective capacity can be stretched to, by IBM.

Note: The entry-level (Grid-starter) capacity points are the same as the A9000 Model 425, with the exception of the smallest capacity point. After the first MES, a grid starter will be a 2 grid element A9000R Model 425.

Table 4-2 FlashSystem A9000 Model 415 capacity

FlashSystem A9000 Model 415 capacity	
Number of grid controllers	3
Number of flash enclosures	1
MicroLatency modules	12
MicroLatency module capacity in TB (three options)	1.2/2.9/5.7
Usable physical capacity in TB	12/29/57
Effective capacity in TB[a]	60/150/300
Maximum effective capacity in TB[b]	1200

a. Effective capacity assumes a data reduction that is calculated at about 5:1.
b. Maximum effective capacity that is supported (design limitation, not expectation).

Table 4-3 FlashSystem A9000R Model 425 capacity

FlashSystem A9000R Model 425 capacity			
Grid controllers	4	6	8
Flash enclosures	2	3	4
MicroLatency modules	24	36	48
MicroLatency module capacity in TB	3.6/8.5/18	3.6/8.5/18	3.6/8.5/18
Usable physical capacity TB	72/170/360	108/255/540	144/340/720
Effective capacity in TB[a]	360/850/1800	540/1275/2700	720/1700/3600
Maximum effective capacity in TB[b]	2400/2400/2400	3600/3600/3600	4800/4800/4800

a. Effective capacity assumes a data reduction calculated at about 5:1.
b. Maximum effective capacity is the up-most provisioning limit that effective capacity can be stretched to, by IBM.

An IBM FlashSystem A9000R entry-level configuration (Grid-Starter) with 3 grid controllers and 1 flash enclosure has a maximum effective capacity of 1200 TB.

Table 4-4 FlashSystem A9000R Model 415 capacity

FlashSystem A9000R Model 415 capacity					
Grid controllers	4	6	8	10	12
Flash enclosures	2	3	4	5	6
MicroLatency modules	24	36	48	60	72
MicroLatency module capacity in TB (two options)	2.9/5.7	2.9/5.7	2.9/5.7	2.9/5.7	2.9/5.7
Usable physical capacity in TB	58/114	87/171	116/228	145/285	174/342
Effective capacity in TB[a]	300/600	450/900	800/1200	750/1500	900/1800
Maximum effective capacity in TB[b]	1400	2000	2600	3000	3000

a. Effective capacity assumes a data reduction that is calculated at about 5:1.
b. Maximum effective capacity that is supported (design limitation, not expectation).

4.1.2 Capacity allocation

FlashSystem A9000 and A9000R have a very efficient capacity allocation, enabled by thin provisioning and data reduction.

Thin Provisioning

With the thin provisioning inherent to FlashSystem A9000 and A9000R, physical capacity is only committed to the logical volume when the associated applications execute writes, not when the logical volume is initially allocated.

Because the total system capacity is designed as a globally available pool, thin-provisioned resources share the buffer of free space. This approach results in highly efficient aggregate capacity use without pockets of inaccessible unused space.

Capacity that is associated with specific applications or departments can be dynamically increased or decreased per the demand imposed at a specified point in time, without necessitating an accurate prediction of future needs.

Storage pools

The available capacity in the system can be administratively portioned into separate andindependent storage pools. The pools are created as *regular pools*. A regular pool means that the storage administrator can create volumes (with the reserved snapshot space) whose total capacity is less than or equal to the size of the pool. Running in the background, the data reduction methods of data deduplication and compression reduce the amount of data that is written to the MicroLatency module, using the available capacity effectively.

Tip: With Version 12.3, a pool can be system-wide. In other words, a single pool can utilize the full system effective capacity.

Volumes

As described in 2.4.2, “Volumes” on page 26, volumes are distributed evenly across flash enclosures by using partitions, and each partition is 16 MB.

The system also uses the concept of allocation unit (AU) size for volumes. The AU size is set at 103 GB.

The minimum volume size that can be created in FlashSystem A9000 or A9000R is 1 GB. However, volumes that are created with a specified size of about 5% or less smaller than the AU size will be rounded to the AU size. For example, when you create a volume, specifying a 98 GB size creates a volume of 103 GB on the system.

4.2 Capacity planning with data reduction

Data reduction in IBM FlashSystem A9000 and A9000R is always in effect. See 2.3.1, “Data reduction technology” on page 15 for a description of the process.

Considering the data reduction capabilities of FlashSystem A9000 and A9000R, it is helpful for capacity planning to start with an overall understanding of the type of data that the system will store. For a better estimate, use the IBM Data Reduction Estimator Tool.

4.2.1 Workload types

Certain workloads will compress better, and other workloads can be deduplicated better. The following types of data are the best candidates for compression and data deduplication:

- Best candidates for compression:
  - Media (video)
  - Data warehousing
  - Online transaction processing
  - Analytics
- Best candidates for data deduplication:
  - Emails
  - Virtual servers
  - Virtual desktops

Understanding data reduction rates, ratios, and savings: To clarify the meaning of the terms data reduction ratio, savings rate, and savings, consider a use case where the original data physical capacity before data reduction was 100 TB, and the physical data capacity after data reduction is 40 TB. The following values help to clarify these terms:

- Data reduction savings rate = 60%
- Data reduction savings = 60 TB
- Data reduction ratio = original size (100 TB) *divided by* the size on disk after data reduction (40 TB) = 2.5:1

When you consider savings, it is easiest to use the data reduction rate. The data reduction ratio helps you understand how much effective data you can store on your system. So, when you have a 5:1 data reduction ratio, you will be able to store 500 TB of data on 100 TB of physical capacity.

4.2.2 Data reduction planning

When you consider how data reduction will affect your space requirements, consider several factors. First, think about the type of data that will be stored and how well that type of data typically can be reduced by using compression and data deduplication.

The best approach is to use a data reduction assessment tool whenever possible. The tool will give you the best estimate.

For information, here are results of observations from the FlashSystem A9000/R already deployed in typical customer environments:

- The overall data reduction of A9000/R population is averaging 3.5:1 (give or take)
- SQL solutions typically have a ratio of between 2:1 and 2.5:1
- Oracle and Db2 typically see higher levels around 3.5:1
- Microsoft Exchange is typically at 2.5:1
- SAP depending on the underlying database:
- SAP HANA is compressed already so very little reduction can be added, the exception would be if the SAP HANA environment is small in capacity compared to a much larger data capacity from other applications.

Web-based Data Reduction Estimator

The Evaluator Group created a tool to assist in this estimating the potential savings through data reduction features. It provides a general idea of typical space savings. This tool is available online for free and public use:

http://www.evaluatorgroup.com/data-reduction-estimator/

This web-based tool allows you to set the space that you are currently using for different types of data. The tool then displays projected space requirements based on the inputs and projected deduplication and compression.

IBM data reduction estimator tool

The IBM data reduction estimator tool is a command-line, host-based utility to estimate the expected compression ratio and data deduplication for a block device.

The IBM data reduction estimation tool supports both data deduplication and compression and the capability to merge all scan results to provide an integrated system-level data reduction estimate (and not just per volume).

The tool also supports “partial scan” for compression only, which is equivalent to the IBM Compresstimator tool as used with the XIV Storage System.

The tool usage flow is depicted in Figure 4-2.

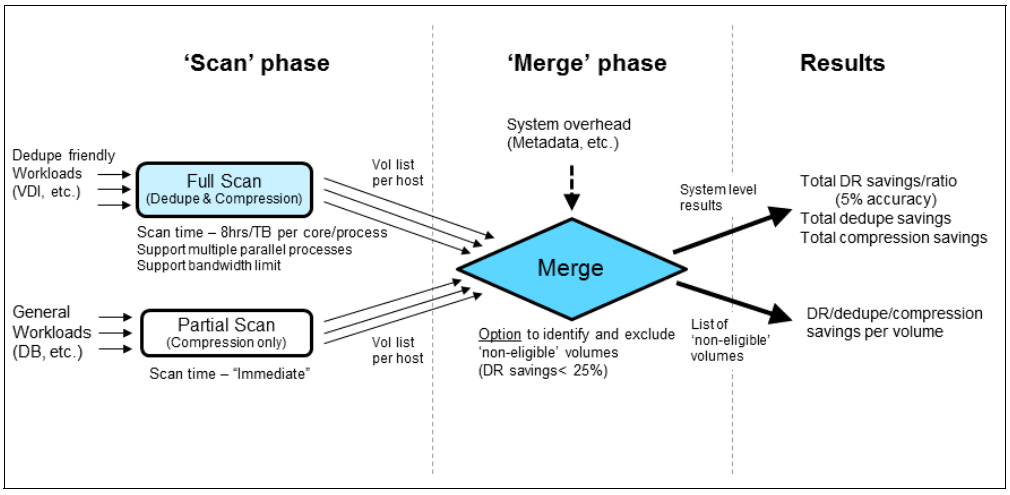


Figure 4-2 Data reduction estimation flow

Installation

Following are the minimum hardware requirements for installing Data Reduction Estimator:

- For HP-UX and ESX, a minimum of 100 MB of free RAM is required in order to achieve the performance goals for the tool.
- For Windows, Red Hat Linux, Ubuntu, IBM AIX®, Solaris, a minimum of 500 MB of free RAM is required.

Proceed as follows:

1. Download the Data Reduction Estimator tool from the Fix central.

   When asked for a Machine Serial Number during the download, enter any combination of seven digits in the XXX-XXXX format to continue.

2. Data Reduction Estimator tool can be installed only on supported Windows operating systems (see list below).
3. After installation, the binary files for other supported operating systems become available in the Windows installation folder.

   By default, the files are copied to:

   – In Windows 64-bit:

   ```
   C:\Program Files (x86)\IBM\Data Reduction Estimation Tool
   ```

   – In Windows 32-bit:

   ```
   C:\Program Files\IBM\Data Reduction Estimation Tool
   ```

   Data Reduction Estimator tool can be used on the following client operating systems:

   – Windows 2008 Server, Windows 2012
   – Red Hat Enterprise Linux Version 5.x, 6.x, 7.x (64-bit)
   – UBUNTU 12.04
   – ESX 5.0, 5.5, 6.0
   – AIX 6.1, 7.1
   – Solaris 10
   – HP-UX 11.31

To install Data Reduction Estimator on a Linux, AIX, Solaris or HP-UX host:

1. Log into the host using the root account.
2. Copy the Data Reduction Estimator binary file from the Windows installation folder to any folder on the host using an SCP utility such as WinSCP or PuTTY.
3. Login to the UNIX host, and confirm that the file has execute permissions, to add execute permissions to the binary file, type **`chmod +x Data-Reduction-Estimator`**.

To install Data Reduction Estimator on IBM i-series through VIOS host:

1. Copy the AIX/Data-Reduction-Estimator tool and ftp it to VIOS in binary mode.
2. Log into the VIOS host using the padmin account.
3. Enter "**`oem_setup_env`**" to exit the padmin shell to a **`# prompt`**.
4. Confirm that the file has executable permissions. To add execute permissions to the binary file, type **`chmod +x Data-Reduction-Estimator`**.

To install Data Reduction Estimator on ESXi:

1. Enable SSH on the ESXi server. Check the following URL for instructions or skip this step if SSH is already enabled:

   http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1017910
2. Copy the ESX/Data-Reduction-Estimator binary file to any folder on the ESXi server. You can copy the file to the ESXi server by using the secure copy protocol (SCP), for example **`scp`** on Linux or **`WinSCP`** on Microsoft Windows.

To install Data Reduction Estimator on another Windows host:

1. Log into the host.
2. Copy the 32/64-bit binary file from the corresponding folder to any folder on the host.

Usage

Follow these steps to use the data reduction estimator tool on a Linux or ESXi server:

1. Log in to the server by using the root account.
2. Obtain the list of device names:
   - In Linux: Use the **`fdisk -l`** command.
   - In ESXi 4.0: Use the **`esxcli corestorage device list | grep Dev`** command.
   - In ESXi 5.0: Use the **`esxcli storage core device list | grep Dev`** command.
3. Run the data reduction estimator tool with the **`–d <device>`** to analyze the device according to the device output list that was created by step 2.

Follow these steps to use the data reduction estimator tool on a Windows server:

1. Log in to the server to use an account with administrator privileges.
2. Open an elevated command prompt with administrator rights. (Run as administrator.)
3. Run the **`wmic DISKDRIVE list brief`** command:

```
D:\>wmic DISKDRIVE list brief
Caption DeviceID Model Partitions Size
HGST HTS725032A7E630 \\.\PHYSICALDRIVE0 HGST HTS725032A7E630 2 320070320640
```

4. Run the data reduction estimator tool with the **`-d <`***device***`>`** to analyze the device according to the DeviceID output.

Syntax

The syntax of the command is shown:

- If you use Linux, ESXi, AIX/Virtual I/O Server (VIOS), or HP-UX, use this syntax:

```
Data-Reduction-Estimator –d <device> [-x Max MBps] [-o result data filename]
[-s Update interval] [--command scan|merge|load|partialscan]     [--mergefiles
Files to merge] [--loglevel Log Level] [--batchfile batch file to process] [-h]
```

- For Windows, use this syntax

```
Data-Reduction-Estimator.exe –d <device> [-x Max MBps] [-o result data
filename] [-s Update interval] [--command scan|merge|load|partialscan]
[--mergefiles Files to merge] [--loglevel Log Level] [--batchfile batch file to
process] [-h]
```

The following descriptions refer to the command:

-d	Specifies the device name, that is, the device path of the device to analyze, for example, `/dev/sda` in Linux or DeviceID in Windows that you can obtain using the **wmic** Windows utility.
-x	Throughput limit up to X MBps. The default is `0 – No limit`.
-o	The name of the output file, the data file which contains the information on the analyzed device. Later it can be used for the “merge” option. If no name is provided, the output file is created with a default name.
-s	Update Interval. The default is 10 seconds.
-p	Specifies the number of processes. The default is 10 seconds.
--command	The operation mode: **scan**: Full scan (default). Estimates total data reduction saving. **merge**: Can be used after the scan of all the devices is completed in order to get the statistics average for all scanned devices. Minimum two files (volumes) are required. **load**: Can be used after the scan of a device is completed in order to load the device statistics from the .dat sketches. **partialscan**: Compression saving estimation. It is used for a quick scan sample.
--mergefiles	Total data saving for more than one scanned device. File list is separated by commas “,”. By default, devices with data saving lower than 90% are ignored. Every such instance is reported.
--mergeall	Override the 90% data saving threshold.
--loglevel	Log level to run. Values are 3 - 7 (default is 3).
--batchfile	Batch file to process. The batch file can contain several devices, with each line referring to a different device.

Note: The progress status will be saved in the log file only when both **-r** and **-t** reach the threshold.

Example

Example 4-1 is an illustration of the data reduction estimator tool command and output.

Example 4-1 Data Reduction Estimator tool illustration

```
root@swfc120:/tmp# ./Data-Reduction-Estimator -d /dev/dm-10
Result data filename not given, auto-generating: file_C8F50050.dat
200.00 GB | 55.60 MBps: 0% [####################] 100%%
Estimated Dedupe Savings:         11.659%
Estimated Compression Savings:    65.797%
Data Reduction Savings:           69.784%
---------------------------------------
Zeroes Detected Savings:          0.227%
Total Data efficiency Savings:    69.853%
Time Consumed:                    00:11:01
```

To get the total data on disk after reduction, take the data reduction saving of the disk size:

`(100% - 69.784%) of 200 GB = 60.432 GB`

The Zeros Detected Savings value refers to large sequences of zeros that were detected on the device. It is not an inherent part of the data reduction saving, as some systems consider this as thin provisioning. The Total Data Efficiency Savings are the total savings, including deduplication, compression and the large zero sequences combined.

Data is first deduped, and then it is compressed. The dedupe saving is `0.11659 * 200 GB = 23.318 GB` and the subsequent compression saving is `0.65797 * (200 - 23.318 GB) = 116.210 GB`. The total saving: `23.318 GB + 116.210 GB = 139.528 GB` (~70% of 200 GB).

Example 4-2 illustrates the total data saving by merging estimates for two volumes: RHEL7 and win2008.

Example 4-2 Saving for two volumes

```
Data-Reduction-Estimator --command merge --mergefiles scan_Linux_RHEL7,scan_win2008_1024
Result data filename not given, auto-generating: merge_out
Estimated Dedup Savings:         97.8%
Estimated Compression Savings:    16.3%
Data Reduction Savings:           98.2%
---------------------------------------
Zeroes Detected Savings:          4.11%
Total Data Efficiency Savings:    98.2%
Time Consumed:                    00:00:00
```

Special considerations

The utility has been written so that it should not be disruptive to a client's environment. It is advised that when the utility is first run on a system that performance is monitored to ensure that performance is not degraded by the utility.

- In general, on ESX there might not be enough RAM to support the data reduction estimation tool performance goals. In this case, expect lower performance on an ESX platform.
- For Windows, Red Hat Linux, Ubuntu, AIX, and Solaris platforms, the default number of concurrent threads performing the scan operation is 10.
- In HP-UX and ESX platforms, the default number of threads is 1. If in your case more than 100 MB of RAM are available for the tool, you can try increasing the number of threads in the same manner described above.

If the scanning task is taking too long, the number of threads can be reduced. You may try to reduce the number of threads to 5, or even to 1 thread if needed.

An example of how to reduce the number of threads:

```
DEDUPEL3=1 ./Data-Reduction-Estimator -command scan --batchfile batchfile -s2 -p3
```

For Windows change from README for Windows. The variable must be set separately:

```
set DEDUPEL3=1
Data-Reduction-Estimator -command scan -batchfile batchfile -s2 -p3
```

In some cases, especially with ESX, changing the number of threads may not result in a significant performance improvement. The partial scan (which estimates compression savings only) requires only a very small amount of RAM, and therefore can be used in appropriate cases.

Note also that compared to the Comprestimator tool, the Data Reduction Estimator tool takes much longer to complete its estimate. Unlike Comprestimator, the Data Reduction Estimator tool has to read an entire disk. This is necessary in order to properly calculate the dedupe ratio.

4.3 Capacity representation and reporting

The Hyper-Scale Manager UI collects and shows different representations of the system capacity, including current usage and an estimate of future needs. In addition, you can also generate reports.

4.3.1 Data usage collection

When connected to the storage system, IBM Hyper-Scale Manager collects information about the actual data usage of the system, to allow forecasting and planning future capacity use. The Hyper-Scale inventory is used in this process.

Data usage collection for a storage system included in the inventory

The capacity usage for systems included in the inventory is once a day by the Management Server and stored in a file on the server.

The capacity usage must meet several criteria for the forecast to be calculated:

- User must have sufficient access rights (role must be Storage Administrator or Read-only) for all of the systems included in the inventory.
- To present the trends, a sampling of 30 days is needed.
- A sufficient number of samples must be available. The forecast trend is not calculated if the number of samples is less than 30. If the system was not sampled for 14 days, then the system gathers 30 new samples, before the trend can again be presented.
- If the sample fails, the Management Server samples the system every 15 minutes for a period of 12 hours. If the sample still fails, the operation is repeated the following day.
- System or storage pool utilization must be above 10%. If the system, or storage pool utilization, is less than 10%, no forecast is calculated.
- Trends cannot be calculated on pools that have no available space for volume allocation.
- If capacity is fluctuating or flat, or space utilization is decreasing, there is no trend.

Data usage collection for a system that is removed from the inventory

The Management Server collects capacity data for systems that are listed on the inventory. Removing a system from the inventory implies stopping the data collection. However, to overcome situations in which the system was mistakenly removed from the inventory, or removed from the inventory for a short period of time, the Management Server applies the following rules on collecting capacity data for systems that are removed from the inventory:

- As long as the system is listed in the inventory, the Management Server collects and keeps its capacity data.
- Whenever the system is removed from the inventory, its capacity data is not immediately deleted. It is kept until the next time slot for which the data is collected from the machine.
- If the system is returned to the inventory prior to the next collection time slot, the capacity usage and its continuity are kept.
- If the system is removed from the inventory, it is impossible to reset its capacity data. To reset the capacity data, the system has to be listed in the inventory.
- If you choose to reset capacity data for all systems, non-monitored systems capacity information is reset as well.

4.3.2 Capacity representation in the Hyper-Scale Manager GUI

Data reduction usage and savings can be monitored by using the Hyper-Scale Manager graphical user interface (GUI). Data reduction savings across individual domains, pools, and volumes can also be monitored. These data reduction values can also be used to see the volumes that achieved the highest reduction savings.

System-level capacity information

The default Dashboard view shows the provisioned capacity for all selected systems in your environment. The following information is shown (see Figure 4-9 on page 94):

- *How much have Hosts written?*

  Represents the total data that is written by hosts to the systems volumes and snapshots before data reduction (Figure 4-3).

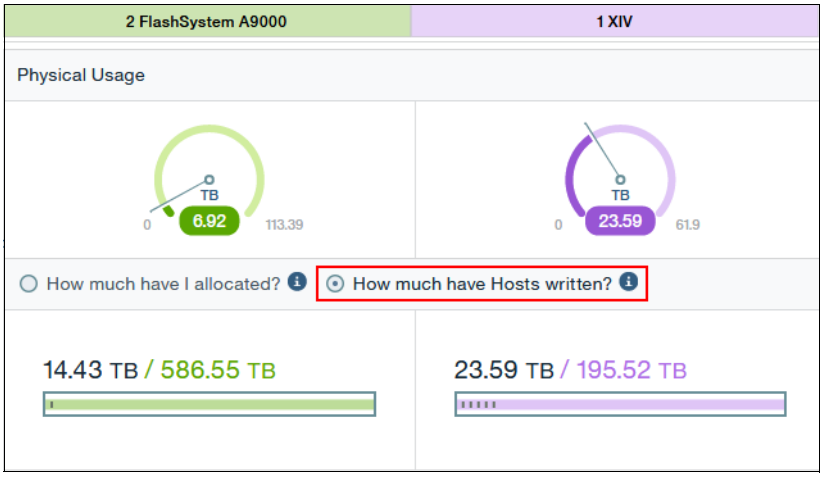


Figure 4-3 How much have Hosts written?

- *How much have I allocated?*

  Allocated capacity is the total capacity that is allocated to hosts from the systems capacity and includes all volumes and snapshots (Figure 4-4).

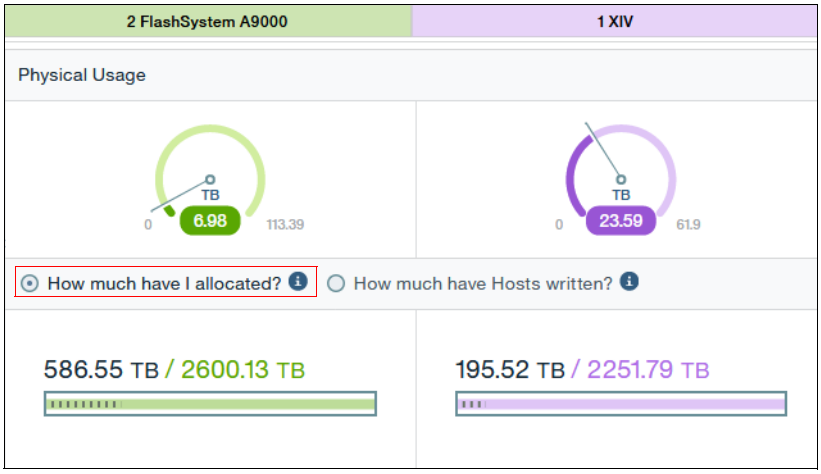


Figure 4-4 How much have I allocated?

- *Projected System Size*

  Projected system size indicates the calculated maximum capacity of the system given the (fixed) physical size and the estimated reduction ratio. The estimation for reduction ratio is a user-configured setting that can be set by specifying any combination of a user estimated reduction ratio and a system actual reduction rate, as follows:

  – User set estimated reduction ratio. See Figure 4-5.

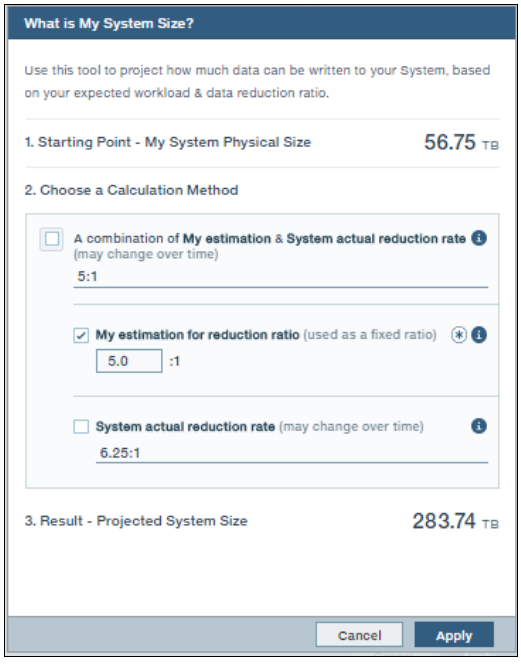


Figure 4-5 User estimated reduction rate

– System actual reduction rate. See Figure 4-6.

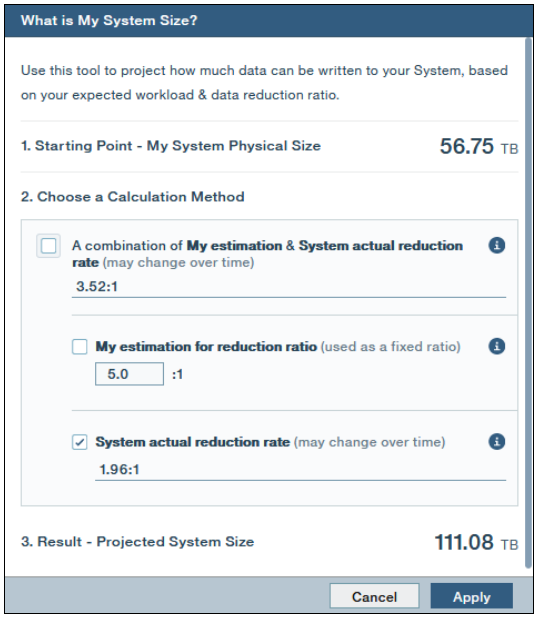


Figure 4-6 Using actual reduction ratio

– Combination of user set and system actual reduction rate. See Figure 4-7.

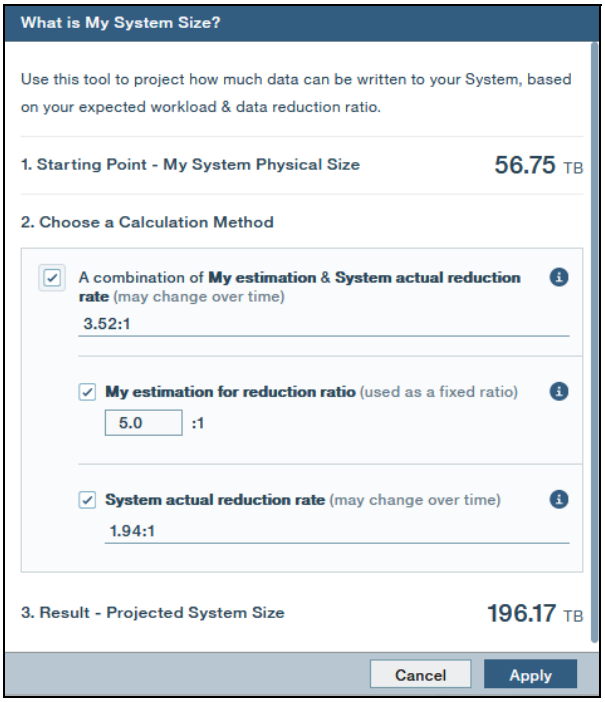


Figure 4-7 Combining user estimated and actual reduction ratios

- *Reclaimable Capacity*

  Total capacity of all volumes that are not used in the environment (not mapped to any hosts and not a mirror target). Therefore, this capacity can be reclaimed and allocated to other hosts. In parentheses, the system displays the percentage of that reclaimable capacity against the total volume allocation (Figure 4-8).

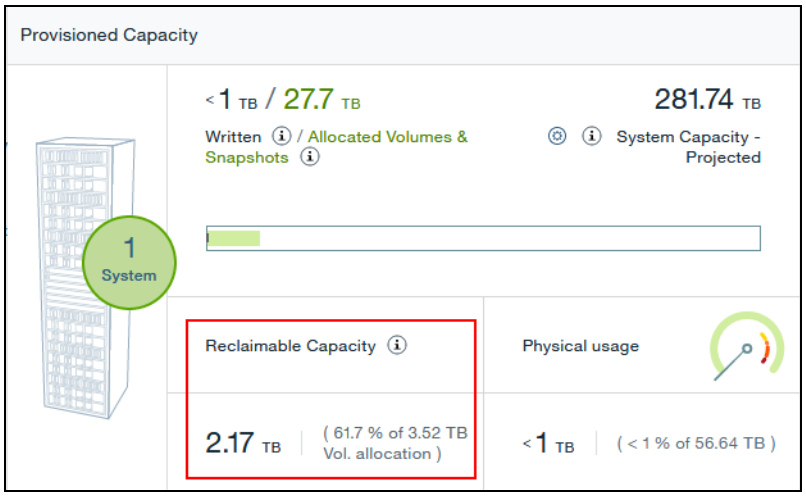


Figure 4-8 Reclaimable capacity

- *Physical usage*

  The actual total physical capacity that is used after data reduction. In parentheses, the system displays the percentage of that physical usage against the total physical space available (Figure 4-9).

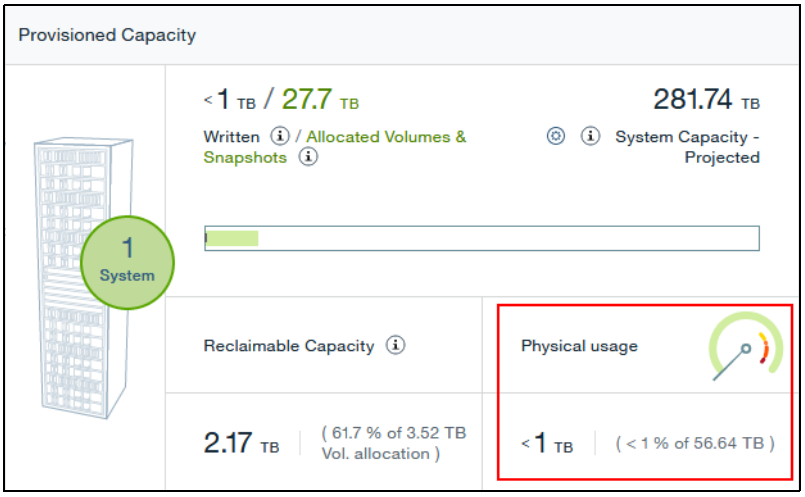


Figure 4-9 Overall provisioned capacity (version 5.2)

In addition, Hyper-Scale Manager continuously estimates and displays data reduction savings on the lower left side of the Dashboard view window, as illustrated in Figure 4-10.

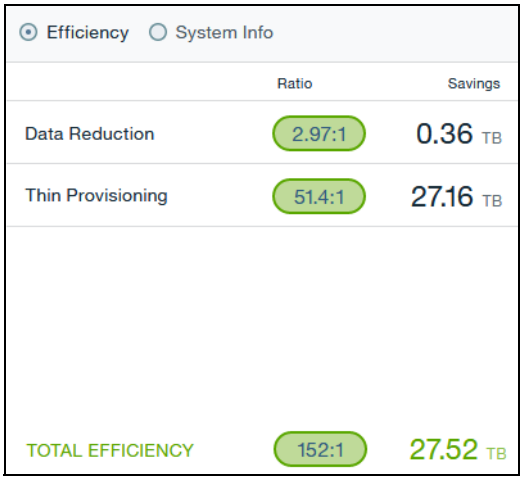


Figure 4-10 Data reduction savings display

To see more details about the data reduction, hover the cursor over the data reduction ratio and the pop-up box shows more details. The additional details include the current deduplication rate as well as the current compression rate, as shown in Figure 4-11.

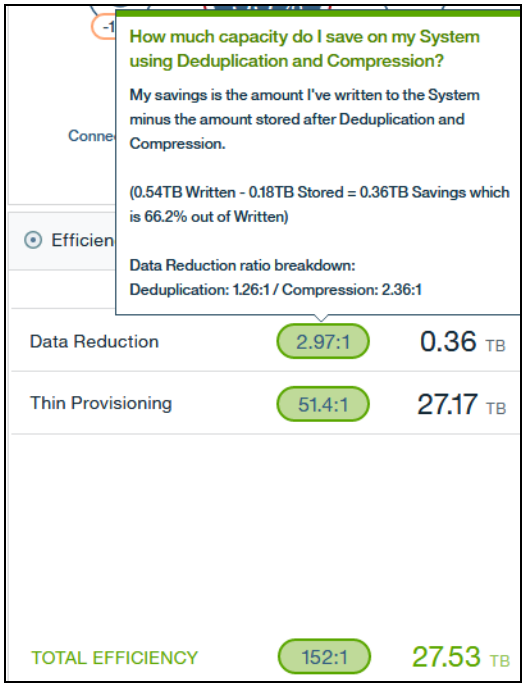


Figure 4-11 Data reduction breakdown

- *Unique Stored Data*

  The deduplication process means that the system avoids writing non-unique data more than once. The **unique stored data** field shows the amount of data actually written. In Figure 4-12 the highlighted row shows out of 835 GB written, the unique stored data is only 212 GB.

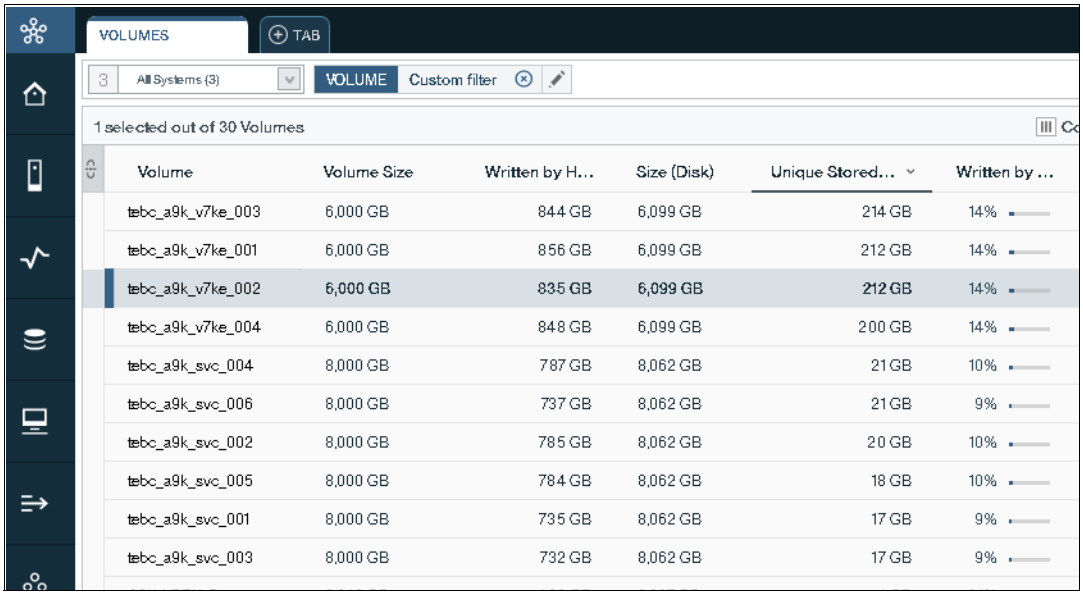


Figure 4-12 Unique Stored Data is often much less than the amount written by the hosts

Storage pool and volume-level capacity information

From the **Pools** view, choose from various filters to view the following volume usage measurements:

- Allocation (soft)

  Shows how much total space has been allocated by Volumes out of the available space on the Pool. It also shows the space reserved for snapshots, and the hard and soft size of the selected pools. See Figure 4-13.

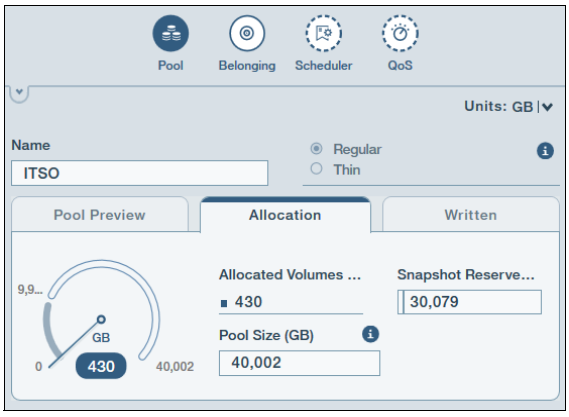


Figure 4-13 Pool's Volume Allocation

- Written (hard)

  Shows how much total data has been written after data reduction to the volumes in these pools. It also shows the space reserved for snapshots. See Figure 4-14.

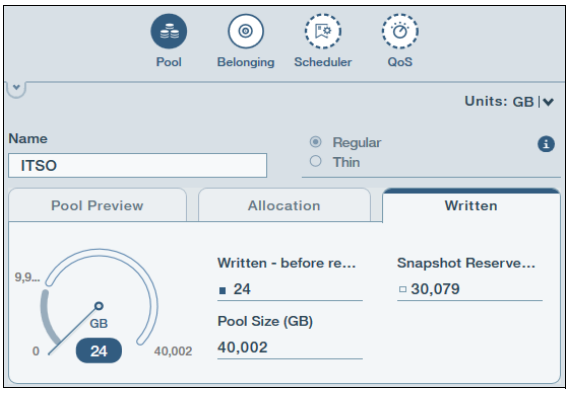


Figure 4-14 Pool's Volume Written

Host-level capacity information

From the **Hosts** view, choose from various filters to view the following host measurements:

- Written by Host

  Shows how much data has been written from the Host to your System. See Figure 4-15.

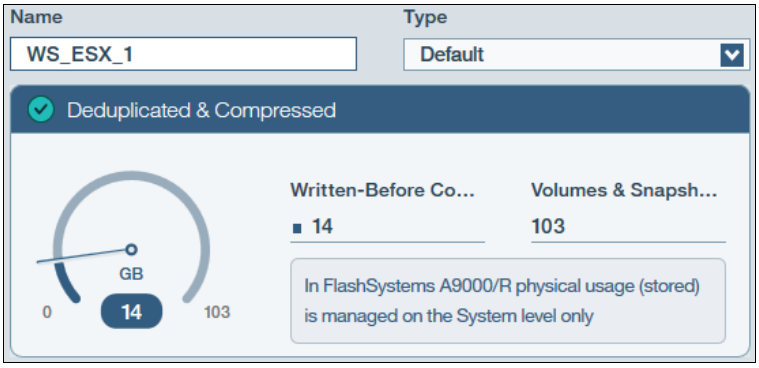


Figure 4-15 Written by host

4.3.3 Capacity planning report

Hyper-Scale Manager includes an infographic-type reporting tool to ease capacity planning across multiple systems. This report provides information about capacity growth, utilization, and trends for the systems, pools, and domains. Capacity planning reports can help administrators forecast space usage and easily view trends in the systems. The Capacity Planning Report is shown in Figure 4-16.

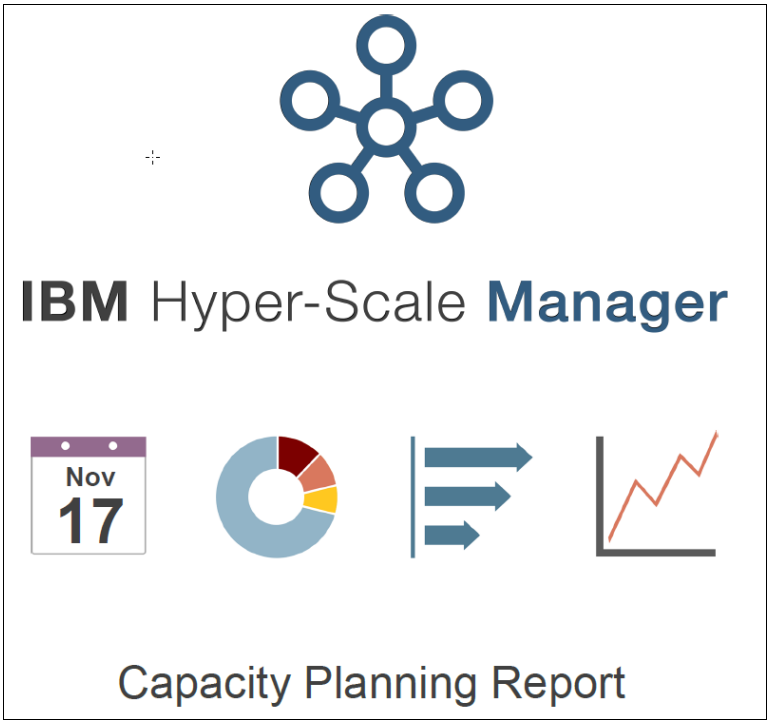


Figure 4-16 The Capacity Planning Report title page

Hyper-Scale Manager collects usage statistics to calculate a forecast of the future of IBM FlashSystem A9000 and A9000R systems, pools, and domains for all managed systems. The data is maintained on the Hyper-Scale Manager server, and this tool requires at least 30 days of data to generate a useful report. This statistical data is used to build the graphical capacity planning report that is shown in this section. Because the capacity data is maintained on the Hyper-Scale Manager server, no interaction is necessary with FlashSystem A9000 or A9000R when the storage administrator requests that a report is generated.

The capacity data must meet several criteria to calculate the forecast. The capacity forecast is not calculated for the following reasons:

- Insufficient number of samples (less than 30 days worth of data).
- Utilization is too low.
- Capacity fluctuates too much, is flat, or is decreasing and will result in the lack of a trend.

The capacity report is generated from the GUI. Instructions about how to generate the report are reviewed later in this section.

Generating a capacity planning report

Whenever needed, you can generate a capacity planning report from IBM Hyper-Scale Manager GUI. The Capacity Planning report is generated from either the System, Domain, or Pool views.

The structure of the file name is:

`CAPACITY_PLANNING_REPORTS_yyyy-mm-dd_hh-mm-ss.zip`

The `.zip` file contains the Capacity Planning PDF with the following name:

`CAPACITY_PLANNING_REPORTS_yyyy-mm-dd_hh-mm-ss.pdf`

Although there are multiple ways to generate a Capacity Planning Report, the following task describes how to generate a report from the Systems view.

If you want to generate a Capacity report for multiple Systems, select the **Capacity Trend/Forecast Report** option from the **Actions** menu.

Tip: You must first select an item in the view to see the Actions menu.

To generate the capacity report:

1. Select the system which you want to generate a capacity report for.
2. From the **Actions** menu, choose **Capacity** →**Capacity Trend/Forecast Report**. A download bar appears at the bottom of your screen with a link to the folder that contains the `.zip` file.
3. Open the `zip` folder and click to open the PDF.

Figure 4-17 shows the generation of a capacity report from the **Actions** menu.

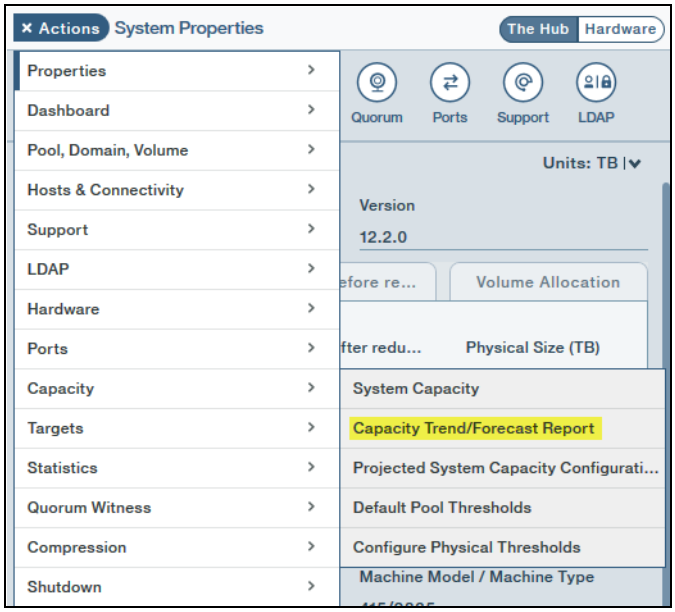


Figure 4-17 Generate a capacity report from the System view

Other paths can lead you to the option to generate the capacity report if those paths are focused on the system, pools, or domains.

Important: Remember to select one or several items in the tabular view to enable the Actions menu.

IBM Hyper-Scale Manager Capacity Planning Report structure

Figure 4-18 shows the table of contents of a sample PDF file with five main categories:

- Part 1 - All System Type Usage (Physical)
- Part 2 - IBM FlashSystem A9000/R Domain Usage (Written)
- Part 3 - IBM XIV Gen3 / SA Domain Usage (Physical)
- Part 4 - IBM FlashSystem A9000/R Pool Usage (Written)
- Part 5 - IBM XIV Gen3 / SA Pool Usage (Physical)

This report lists usage over time. Additional sections, such as usage growth rate and detailed graphs, will be added as Hyper-Scale Manager collects data over time.

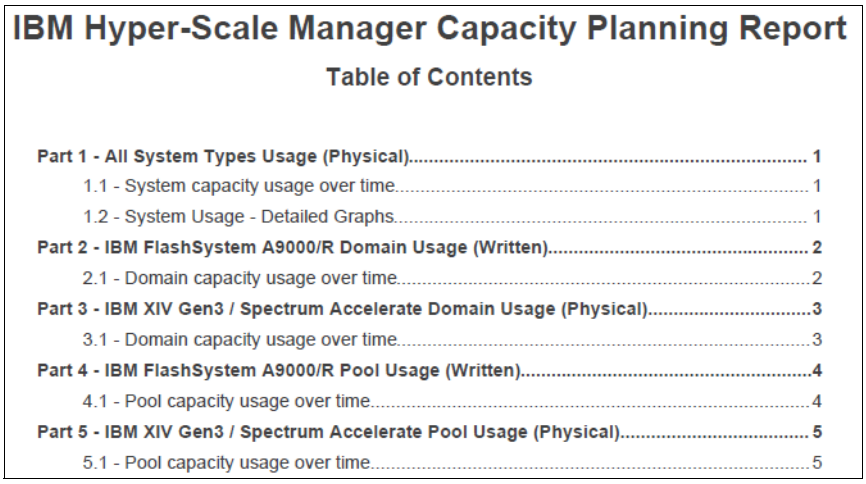

IBM Hyper-Scale Manager Capacity Planning Report

Table of Contents

Part 1 - All System Types Usage (Physical)..... 1
1.1 - System capacity usage over time..... 1
1.2 - System Usage - Detailed Graphs..... 1
Part 2 - IBM FlashSystem A9000/R Domain Usage (Written)..... 2
2.1 - Domain capacity usage over time..... 2
Part 3 - IBM XIV Gen3 / Spectrum Accelerate Domain Usage (Physical)..... 3
3.1 - Domain capacity usage over time..... 3
Part 4 - IBM FlashSystem A9000/R Pool Usage (Written)..... 4
4.1 - Pool capacity usage over time..... 4
Part 5 - IBM XIV Gen3 / Spectrum Accelerate Pool Usage (Physical)..... 5
5.1 - Pool capacity usage over time..... 5

Figure 4-18 IBM Hyper-Scale Manager Capacity Planning table of contents

System usage

This part of the report shows the system capacity allocation over time. Figure 4-19 shows an overall view for a FlashSystem A9000 or a FlashSystem A9000R in the inventory and a capacity forecast that covers the next 6 and 12 months.In this example, there are four systems in the inventory.

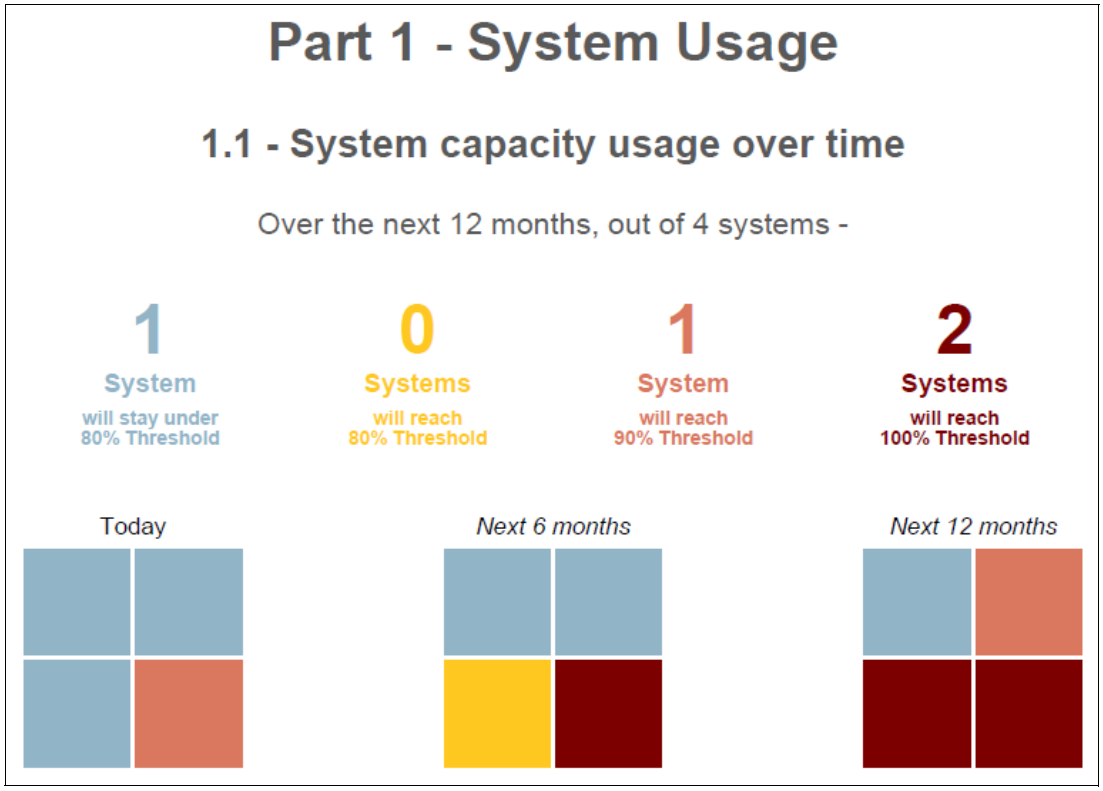


Figure 4-19 Example of system usage in the Capacity Planning report

The next graph, as shown in Figure 4-20, ranks the systems based on their usage growth rate. In this example, four systems are in the Hyper-Scale Manager inventory.

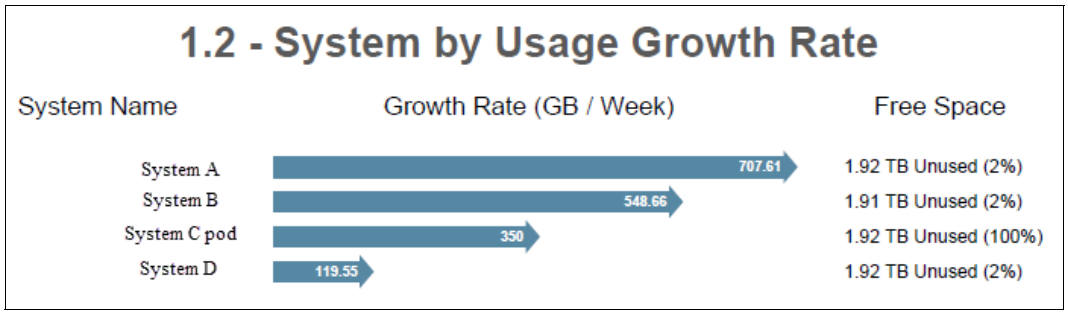


Figure 4-20 System usage growth rate

The final piece for this part of the report is a graph that shows the progression of the systems' capacity allocation growth. An example is shown in Figure 4-21.

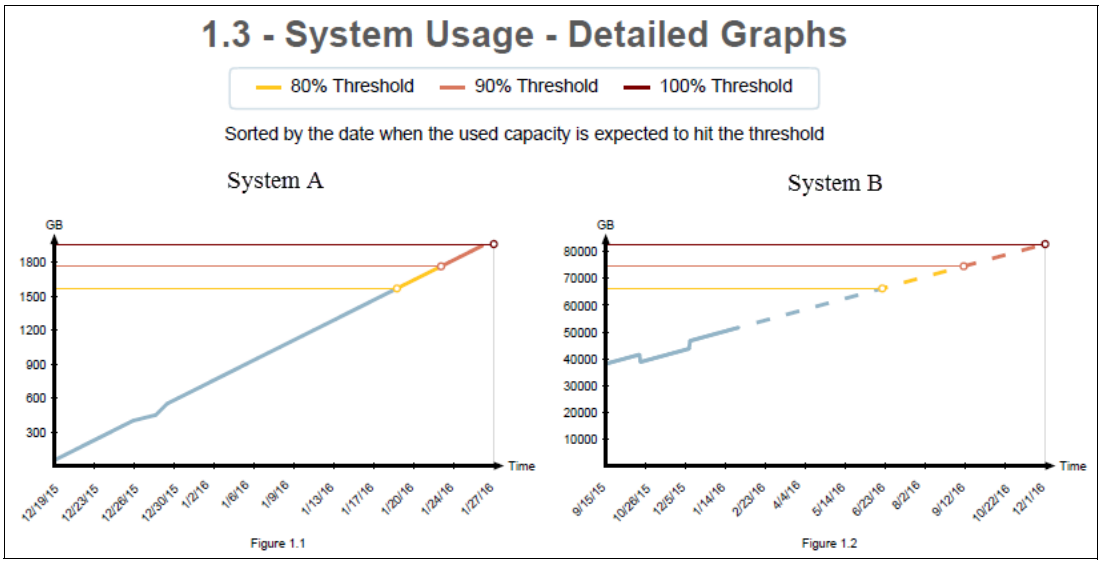


Figure 4-21 Detailed graphs for system usage

Domain usage

This section provides a summary for each of the domains, according to the user domain association. The global administrator sees all of the system domains, regardless of the access policy (which is set by the security administrator). A domain administrator sees only the specific domains for which they are authorized.

Figure 4-22 shows the Summary view. This view does not contain a full 30 days of statistics so the capacity planning report will not generate a growth rate view or the detailed graphical views that are used for forecasting. The forecast cannot be generated with less than 30 days worth of data.

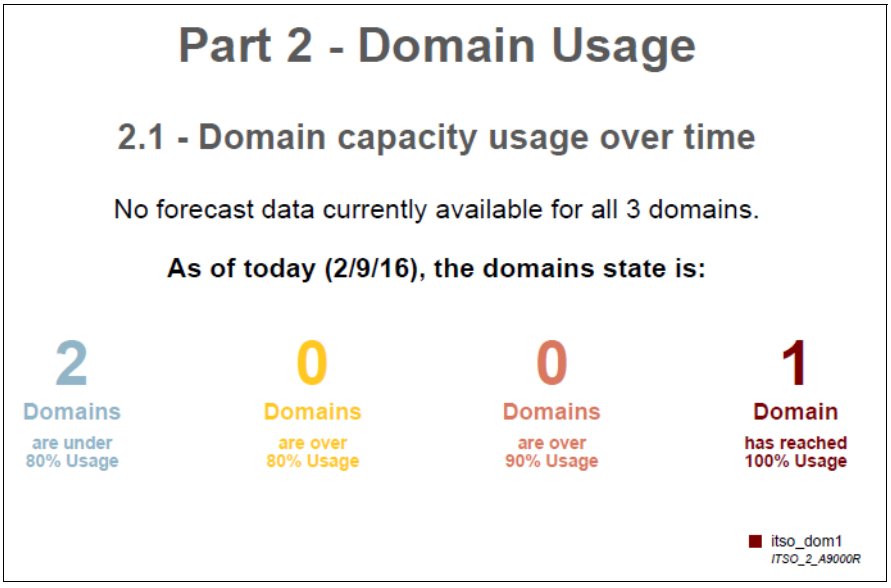


Figure 4-22 Example of domain usage in the Capacity Planning report

The domain growth rate can be useful for cloud service providers where the growth rate is shown and calculated as a forecast for each domain. In addition, for each domain, a graph is provided (with enough data points) that shows the capacity growth progression for allocated and used capacity. You will be able to see similar graphs for pools in the next section.

Pool usage

This section provides a summary for each of the storage pools' capacity information that was collected for each monitored FlashSystem A9000 system or FlashSystem A9000R system. Figure 4-23 shows the Pool Usage view and its predicted usage trend over the next 6 months and 12 months.

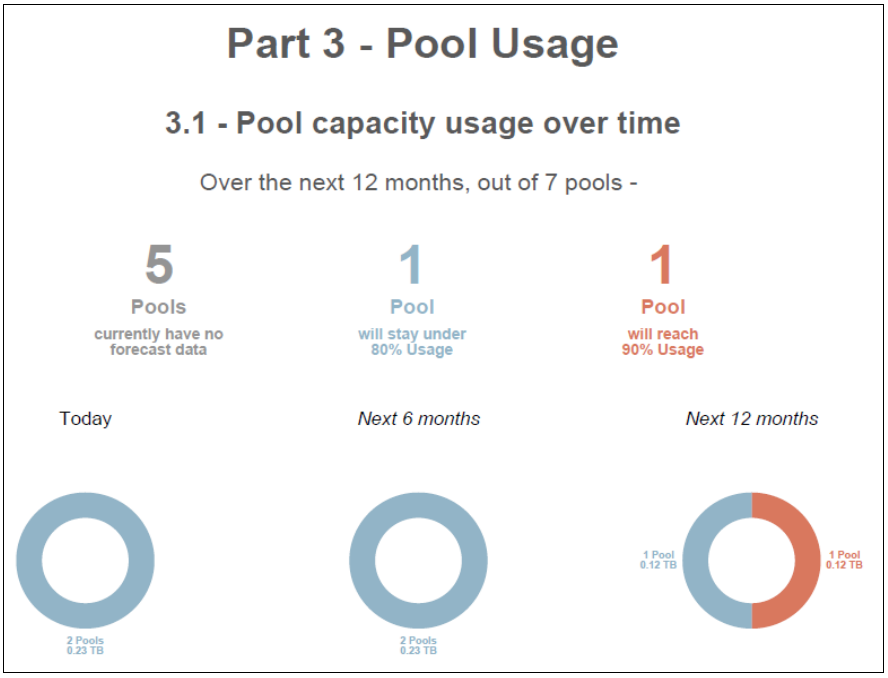


Figure 4-23 Example of pool usage in the Capacity Planning report

This information provides further granularity to help you identify data growth areas in your FlashSystem A9000 or FlashSystem A9000R environment. Detailed bar charts and graphs are also provided for each pool. Figure 4-24 shows the pool by usage growth rate.

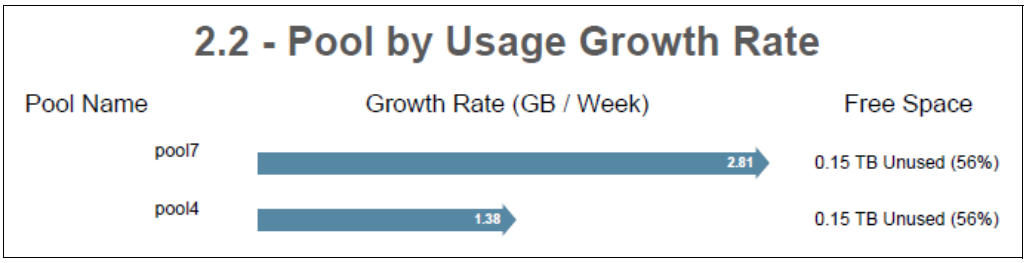


Figure 4-24 Pool growth rate

Figure 4-25 shows the detailed graphs for pool usage.

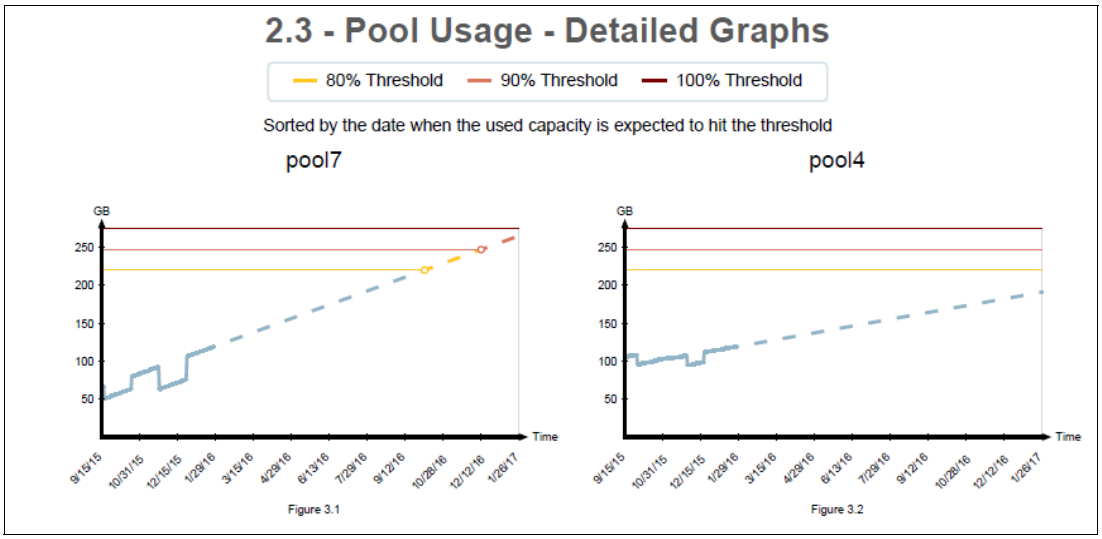


Figure 4-25 Example of detailed graphs for pool usage

Other paths can lead you to the option to generate the capacity report if those paths are focused on the system, pools, or domains.

Important: Remember to select one or several items in the tabular view to enable the Actions menu.

4.4 Intelligent capacity management for deduplication

With software version 12,3,1, IBM has enhanced FlashSystem A9000 and A9000R with patented IBM Research technology that:

- Analyzes vast amounts of data, without performance impact
- Provides effective estimates, per volume, for reclaimable capacity, attributed capacity, compression saving, de-duplication saving, and total capacity saving
- Requires no management, configuration, or training.

This technology is especially relevant for storage systems that apply deduplication. Remember that deduplication is always enforced on FlashSystem A9000 and A9000R; It can not be disabled.

With deduplication, data is shared across multiple volumes, as illustrated in Figure 4-26

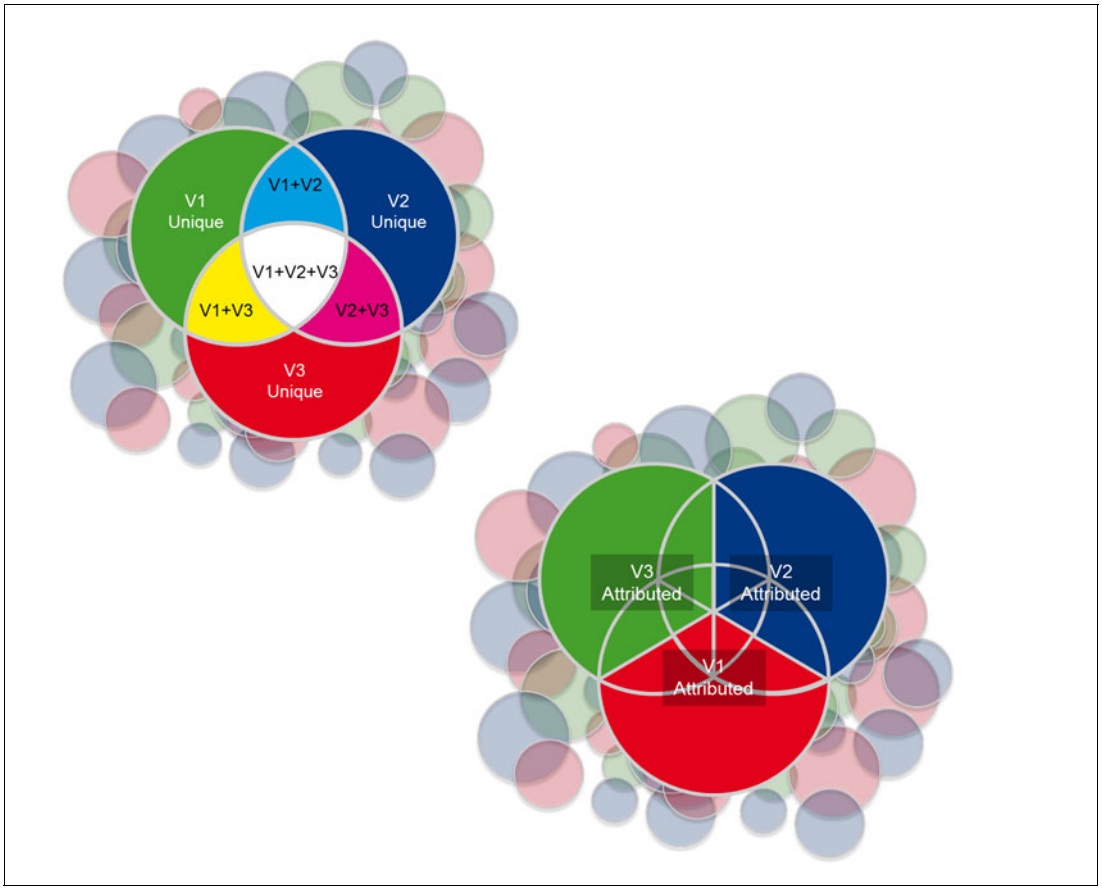


Figure 4-26 Data shared across volumes

Without specific information, it is not easy to know how much of the physical size is actually used by any particular volume. Therefore, it is very difficult for the storage administrator to get a clear understanding of how much physical space could be reclaimed when deleting a volume.

The function introduced in software version 12,3,1, presents the user with information to help answer such questions as:

- How much capacity could be reclaimed by volume deletion
- How much capacity was saved for an application with either compression alone or deduplication alone
- How much effective capacity is really used by that volume (since a volume can share capacity with other volumes through deduplication).

Figure 4-27 provides an example.

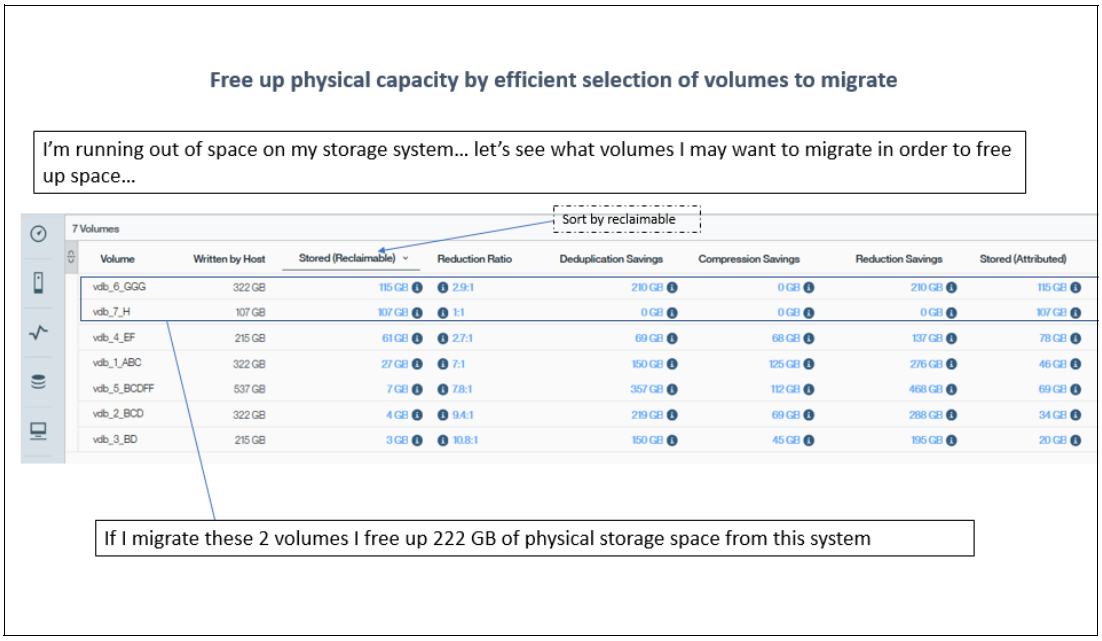


Figure 4-27 Analyze data reduction and options for capacity savings

From a high level standpoint, the technology developed and patented by IBM Research is referred to as Volume Sketches. It is a content-aware extraction technique that is based on a sampling of volumes metadata. Being based on a metadata sampling, significantly reduces the amount of information to analyze. The sample, or sketch data, holds enough information to estimate data reduction attributes with specified error margins.

The function is implemented in the Hyper-Scale Manager server and results are presented in the HSM GUI . Hyper-Scale Manager version 5.5.1 or later is required. During installation, you are prompted as to whether you want o activate the deduplication-aware management feature.

Note that this feature requires the following Hyper-Scale Manager server resources to handle the deduplication-aware management calculations:

- At least 4 physical cores
- At least 6 GB of RAM

As depicted in Figure 4-28, sketch data is harvested from each A900-/R system connected to a single HSM server. The sketches are pulled onto the Hyper-Scale Manager server by a collector and then analyzed by the Analyzer. The results are presented in the Volume view of the HSM GUI and represent the state of Volumes at the time of collection.

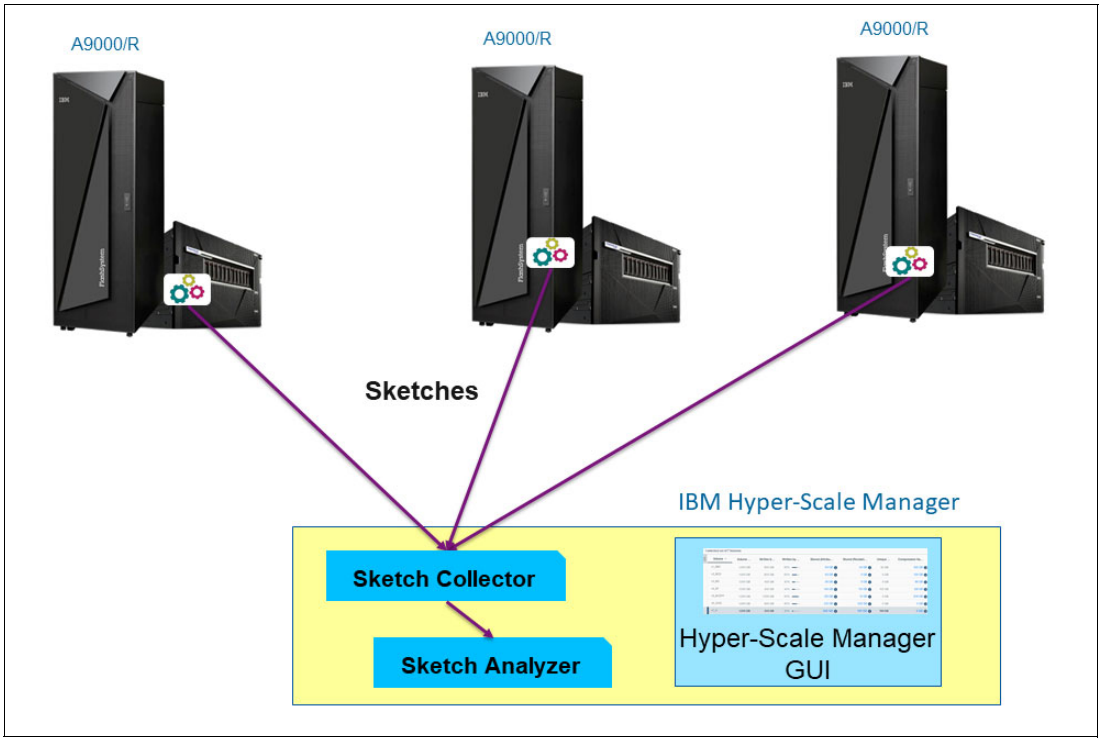
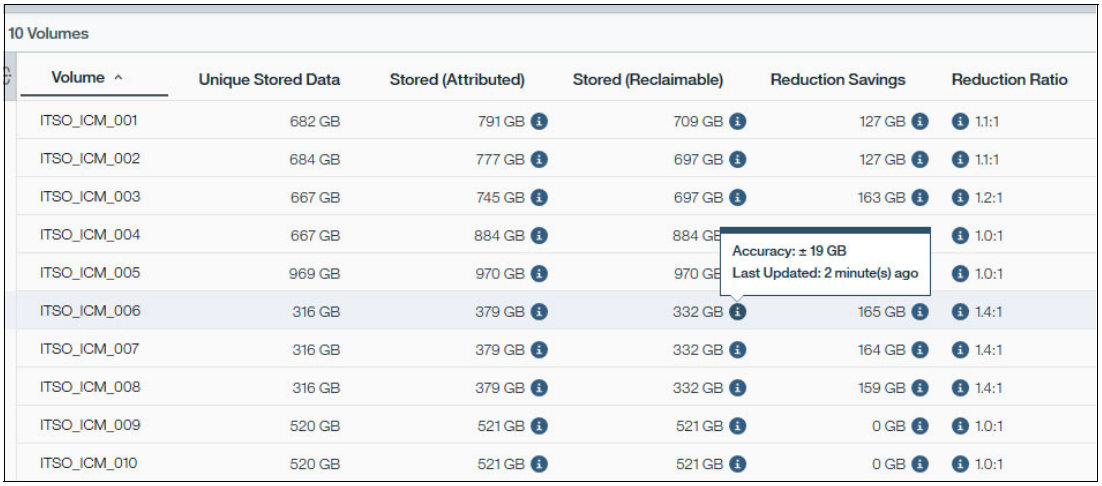


Figure 4-28 IBM Volume Sketches technology

Figure 4-29 illustrates how the information is displayed in the Volumes view in the Hyper-Scale Manager GUI.



10 Volumes

Volume	Unique Stored Data	Stored (Attributed)	Stored (Reclaimable)	Reduction Savings	Reduction Ratio
ITSO_ICM_001	682 GB	791 GB	709 GB	127 GB	1.1:1
ITSO_ICM_002	684 GB	777 GB	697 GB	127 GB	1.1:1
ITSO_ICM_003	667 GB	745 GB	697 GB	163 GB	1.2:1
ITSO_ICM_004	667 GB	884 GB	884 GB		1.0:1
ITSO_ICM_005	969 GB	970 GB	970 GB		1.0:1
ITSO_ICM_006	316 GB	379 GB	332 GB	165 GB	1.4:1
ITSO_ICM_007	316 GB	379 GB	332 GB	164 GB	1.4:1
ITSO_ICM_008	316 GB	379 GB	332 GB	159 GB	1.4:1
ITSO_ICM_009	520 GB	521 GB	521 GB	0 GB	1.0:1
ITSO_ICM_010	520 GB	521 GB	521 GB	0 GB	1.0:1

Figure 4-29 Volumes view in HSM GUI

As shown in Figure 4-29, the relevant information is displayed under additional columns, respectively labeled:

- Stored (Attributed): represents the amount of the volume's unique capacity plus a proportional share of the volume's physical capacity that is common (deduplicated) to additional volumes in the system.
- Stored (Reclaimable): represents the amount of physical capacity that will be freed when a volume is deleted or migrated to another storage system.

- Reduction Savings: represents the amount of data that was written to the system, minus the actual physical capacity that is needed to store that data past compression and deduplication. This is the actual saving.
- Reduction Ratio: is the system-wide data reduction savings ratio. This value is used to project how much more data can be written to the system based on the estimated data reduction savings.

  Remember that both compression and deduplication are always applied.

Note also that the information icon displays when the last update took place and the margin of accuracy, expressed in GB. Accuracy is mainly a function of the volume size: the bigger the volume, the more accurate the estimate will be. Typically, a good accuracy id achieved for volumes greater than 500 GB and with a low data change rate during a calculation cycle. For small volumes, the data will not be presented if the accuracy cannot be guaranteed. Refer to Figure 4-30.

Figure 4-30 Insufficient accuracy

It should also be noted that there is no separation of volumes and their snapshots. In other words, all volume capacities, stored and reclaimable, refer to the volume and all of its snapshots. Reclaimable information is also not available for volumes involved in a volume copy operation.

5

Installation requirements

This chapter gives an overview of the installation requirements, planning, and other required tasks for the deployment of IBM FlashSystem A9000 and IBM FlashSystem A9000R.

For a smooth and efficient installation, planning and preparation tasks must take place before the system is scheduled for delivery and installation in the data center. A sales representative will arrange a Technical Delivery Assessment (TDA) meeting to go over site-specific details and to ensure that the correct information is gathered before the delivery of the system. The IBM installation planning representative, IBM service support representative (SSR), and IBM technical advisor (TA) must participate in this assessment.

The TDA pre-installation checklist and worksheets document must be completed for each FlashSystem A9000 and FlashSystem A9000R and given to the IBM installation planning representative or IBM SSR. The information that is recorded in the worksheets is used for installation and configuration. Before you start an installation, ensure that you have these completed worksheets.

Installation planning involves the following major areas, which are covered in this chapter:

- Physical space requirements
- Delivery requirements
- Site requirements
- Basic configuration
- Network connections
- Physical installation
- Host connections

This chapter presents a summarized version of the preparation tasks and requirements. For detailed and up-to-date information, see the following publications:

- *IBM FlashSystem A9000 Models 9836-415, 9838-415, 9836-425, 9838-425, and 9838-U25 Deployment Guide,* GC27-8564
- *IBM FlashSystem A9000R Models 9835-415, 9837-415, 9835-425, 9837-425, and 9837-U25 Deployment Guide,* GC27-8565

5.1 Physical space requirements

The physical requirements for the location where the equipment will be installed must be checked before system delivery.

5.1.1 FlashSystem A9000

FlashSystem A9000 is intended to be installed in a customer-supplied rack with at least 8U of free space. The physical dimensions are 356 mm (14.02 in.) 8U x 483 mm x 930 mm (8U x 19 in. x 36.6 in.) and the total weight is approximately 125 kg (275.5 lbs).

5.1.2 FlashSystem A9000R

FlashSystem A9000R hardware components are installed in a supplied IBM T42 rack. Adequate space is provided within the rack to house all components and to route all cables correctly.

Note: Partially populated racks are identical to fully populated racks, except that they have fewer grid elements and less storage capacity. FlashSystem A9000R with fewer grid element controllers also has fewer usable Fibre Channel (FC) and internet Small Computer System Interface (iSCSI) ports, and less processing power. Otherwise, partially populated configurations have the same power, floor planning, and software configurations as fully populated systems.

Unused rack space: FlashSystem A9000R rack must be considered a dedicated unit. Spare space within the rack cannot be used for other equipment.

FlashSystem A9000R has the following site requirements:

- The rack dimension and clearance requirements are listed in Table 5-1. Also see Figure 5-1 on page 111.

Table 5-1 Rack dimensions and clearance

Dimension and clearance	Measurement
Height	201.5 cm (79.3 in.)
Depth	129.7 cm (51.1 in.)
Width	64.4 cm (25.4 in.)
Front clearance	120 cm (47.2 in.)
Rear clearance	100 cm (39.4 in.)
Door clearance	10 cm (3.9 in.)

- The following measurements in Figure 5-1 are provided for your convenience.

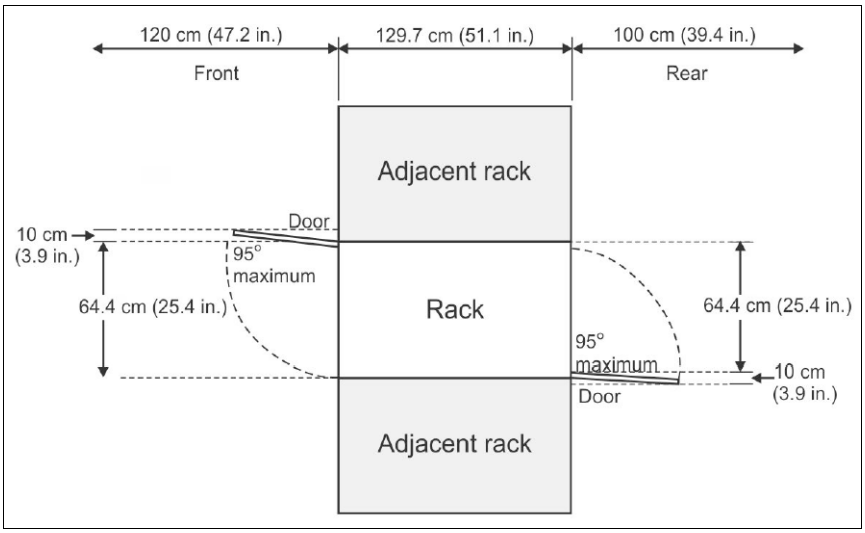


Figure 5-1 FlashSystem A9000R floor space requirements

- The floor must be able to withstand the weight of up to a fully configured IBM FlashSystem A9000R, which is 935 kg (2,061 lbs.) for model 415, without the Management Module, or 774 kg (1,706 lbs) for model 425. For racks which are not fully equipped, subtract 80 kg (176 lbs) for each grid element to get the approximate weight requirement for your rack.
- Enough clearance around the system must be left for cooling and service. Airflow is from front to back.

Building features, such as any ramps, elevators, and floor characteristics, must also be considered.

Rear-door heat exchanger feature for IBM FlashSystem A9000R

The rear-door heat exchanger (Feature Code AFR1) is an optional water-cooled device that is mounted on the rear of IBM FlashSystem A9000R. It cools the air that is heated and exhausted by devices inside the rack. The rack on which you install the heat exchanger can be on a raised floor or a non-raised floor.

5.2 Delivery requirements for IBM FlashSystem A9000R

It is the responsibility of the customer or moving contractor to unpack and move IBM FlashSystem A9000R components as close as possible to the system's final destination before an IBM SSR can start the physical installation. Carefully check and inspect the delivered packages and hardware for any visible damage. If no damage is visible and the tilt and shock indicators show no problem, sign for the delivery.

Important: Be careful not to tilt FlashSystem A9000R storage system rack more than 10 degrees.

5.2.1 Delivery clearance requirements

The clearance measurements (height x width x depth) that are required for delivery through all doors and elevators are listed in Table 5-2. These measurements are for the typical height of the pallet. The measurements do not include more clearance that is needed to raise the pallet on a pallet jack for movement.

Table 5-2 Typical delivery clearance requirements

Dimension	Clearance requirement
Height	216 cm (85.0 in.)
Depth	144 cm (56.7 in.)
Width	94 cm (37.0 in.)

5.2.2 Delivery weight requirements

The path from the truck and to the server room must support the weight of a fully configured rack, which is up to 935 kg (2,061 lb.) for model 415 or 774 kg (1,706 lb.) for model 425. For racks which are not fully equipped, subtract 80 kg (176 lbs.) for each building block to get the approximate weight requirement for your rack.

5.2.3 Height and weight reduction features

IBM FlashSystem A9000R arrives fully assembled with all components in place, unless you order the height or weight reduction shipping options.

If the site does not meet the delivery clearances or more clearance is needed to move IBM FlashSystem A9000R rack through low clearance obstacles, the height reduction shipping feature (Feature Code AFR2) must be ordered.

When the height reduction feature is ordered, the doors, side panels, and top cover of the rack are removed by an IBM SSR before FlashSystem A9000R is moved to its final location by professional movers. The components must be reinstalled by an IBM SSR during FlashSystem A9000R installation.

The weight-reduction feature (Feature code AFR3) removes components from the rack and ships them separately. With this option, the rack weighs approximately 423 kg (932 lbs.) and the rack can be tilted as much as necessary to fit under low doorways. This option reduces the load of the weight on floors and elevators that are used when a FlashSystem A9000R rack is moved from the truck to the final position in the installation site. An IBM SSR installs the components that were shipped separately during system installation.

5.3 Site requirements

This section summarizes additional site requirements and other optional features.

5.3.1 Power supply requirements

FlashSystem A9000 and FlashSystem A9000R require the following power supplies:

- Each of the three grid controllers and the flash enclosure must be connected to two separate, 10 A and 200 - 240 V AC power sources, for a total of eight power sources for a single FlashSystem A9000.
- FlashSystem A9000R features redundant main-power cables. A power supply must be provided from two independent sources of electricity.

  The following AC power source configurations are available:

  - (Model 415 only) Four 60/63 A, 200-240 V AC, North American, EMEA, and Japan single-phase receptacles, each connected to a different power source
  - (Model 425 only) Two or four (dependent on configuration's scaled-out power load) 60/63 A, 200-240 V AC, North American, EMEA, and Japan single-phase receptacles, each connected to a different power source
  - Two 60 A, 200 - 240 V AC, US, US and Japan delta three-phase receptacles, each connected to a different power source
  - Two 30/32 A, 200-240 V AC (Line-to-Neutral [LN]), EMEA WYE three-phase receptacles, each connected to a different power source

5.3.2 Power consumption and thermal dissipation

Before you start the physical installation, ensure that an electrician is available who can fulfill the power requirements for connecting IBM FlashSystem A9000 and FlashSystem A9000R.

For FlashSystem A9000, the typical power consumption is 1.69 KW, with a maximum of 2.91 KW. For FlashSystem A9000 model 415, the power consumption for different MicroLatency modules with a light load is shown in Table 5-3.

Table 5-3 Power consumption model 415 in kVA

FlashSystem A9000	kVA
1.2 TB*	1.79
2.9 TB*	1.97
5.7 TB*	2.17

* The numbers for power consumption of 1.2 TB and 2.9 TB MicroLatency modules are estimates.

For FlashSystem A9000 model 425, the power consumption for 12 x 18 TB MicroLatency modules with light load is 1.93 kVA.

For FlashSystem A9000R with two grid elements, the typical power consumption is 3.07 KW, with a maximum of 4.82 KW. A fully provisioned system model 415 with six grid elements has a typical power consumption of 8.81 KW, with a maximum of 13.91 KW.

Table 5-4 lists the power consumption for 5.7 TB MicroLatency modules with light load in kilovolt-ampere (kVA) of FlashSystem A9000R model 415 configurations.

Table 5-4 Power consumption model 415 in kVA

FlashSystem A9000R	**kVA**
Four grid controllers and two flash enclosures	3.82
Six grid controllers and three flash enclosures	5.63
Eight grid controllers and four flash enclosures	7.45
Ten grid controllers and five flash enclosures	9.26
Twelve grid controllers and six flash enclosures	11.08

Table 5-5 lists the power consumption for 18 TB MicroLatency modules with light load in kilovolt-ampere (kVA) of FlashSystem A9000R model 425 configurations.

Table 5-5 Power consumption model 425 in kVA

FlashSystem A9000R	**kVA**
Four grid controllers and two flash enclosures	3.42
Six grid controllers and three flash enclosures	5.03
Eight grid controllers and four flash enclosures	6.64

Table 5-6 indicates the cooling (thermal dissipation) for FlashSystem A9000 and for different configurations of FlashSystem A9000R. For FlashSystem A9000R to support capacity upgrades, the installation site must provide cooling capacity to support full-rack configurations.

Table 5-6 Thermal dissipation in kBTU/hour

FlashSystem A9000R	**kBTU/hour**
Model 415: Four grid controllers and two flash enclosures	16.40
Model 415: Six grid controllers and three flash enclosures	24.14
Model 415: Eight grid controllers and four flash enclosures	31.96
Model 415: Ten grid controllers and five flash enclosures	39.60
Model 415: Twelve grid controllers and six flash enclosures	47.30
Model 425: Four grid controllers and two flash enclosures	10.50
Model 425: Six grid controllers and three flash enclosures	15.80
Model 425: Eight grid controllers and four flash enclosures	21.00
FlashSystem A9000	**kBTU/hour**
Model 415: Three grid controllers and one flash enclosure	9.90
Model 425: Three grid controllers and one flash enclosure	6.70

5.3.3 Other equipment

The following sections summarize the main characteristics for other systems and servers that are typically used with FlashSystem A9000 and FlashSystem A9000R.

Hyper-Scale Manager Server

IBM Hyper-Scale Manager is a server application, which is used to run the graphical user interface (Hyper-Scale Manager UI or HSM UI) to manage FlashSystem A9000 and FlashSystem A9000R. It is installed on a separate host or virtual machine, which runs RedHat Enterprise Linux or CentOS as the operating system. See Table 5-7 for a list of system requirements.

Table 5-7 Hyper-Scale Manager server requirements

	Hyper-Scale Manager requirements
Environment	Physical host or virtual machine
Operating system	RedHat Enterprise Linux (RHEL) x64. Version 6.*x* or later or CentOS 6.8 or 7.2
CPU	At least two physical cores
Memory	At least 4 GB of memory or more
Disk space	A minimum of 76 GB in the installation path + a minimum of 4 GB where the installation file is located
Network	One 1 GbE (or faster) Ethernet connection.

Important: Do *not* install IBM Hyper-Scale Manager to the same IBM FlashSystem A9000 or FlashSystem A9000R that is managed by this software. In a system outage or planned downtime, Hyper-Scale Manager will be unusable and the system cannot be managed by this Hyper-Scale Manager.

Management by using the command-line interface (CLI) or another Hyper-Scale Manager is possible.

To be able to access the management web interface and for the Hyper-Scale Manager Server to be able to communicate with IBM FlashSystem A9000 and FlashSystem A9000R, it is necessary to open specific network ports.

Table 5-8 lists the ports that need to be opened in the linux firewall.

Table 5-8 Ports that must be opened

Ports that need to be opened	**Direction**	**Port number**
Outbound only: Simple Mail Transfer Protocol (SMTP)	Outbound	25
Communication with FlashSystem A9000/ R A9000R	Outbound	7778
HTTPS Access to Hyper-Scale Manager	Inbound	8443
HTTP Access to Hyper-Scale Manager	Inbound	8080
HTTPS Access to Cloud Integration feature	Inbound	8440

No other ports are required for HyperScale Manager to perform correctly. Any other ports still eventually open on the firewall can be closed.

During the installation, a user needs to be defined on the server on which the Hyper-Scale Manager application will run. The user might be required to have the related access rights to the linux system so that the user can install programs and create a user. For more information, talk to your operating system administrator.

LDAP server (optional)

IBM FlashSystem A9000 and FlashSystem A9000R support customer-supplied LDAP server-based user authentication.

When LDAP authentication is enabled, the storage system accesses a specified LDAP directory to authenticate users whose credentials are maintained in the LDAP directory (except for the admin, technician, maintenance, and development users, which remain locally administered and maintained).

The benefits of LDAP-based centralized user management can be substantial when you consider the size and complexity of the overall IT environment. Maintaining local user credentials repositories is straightforward and convenient when you work with only a few users and a few storage systems. However, as the number of users and interconnected systems grows, the complexity of user account management rapidly increases and managing this environment becomes a time-consuming task.

Although the benefits from using LDAP are significant, you must also evaluate the considerable planning effort and complexity of deploying LDAP infrastructure if it is not already in place.

IBM Security Key Lifecycle Manager Server

To take advantage of the external encryption solution of IBM FlashSystem A9000 and FlashSystem A9000R, at least one IBM Security Key Lifecycle Manager (SKLM) server is required. For the best data protection, it is better to have more than one key server installed, preferably in different locations. The SKLM server does not need to be dedicated to FlashSystem A9000 or FlashSystem A9000R, and it can be shared across multiple products in the data center.

The IBM SKLM provides an automated solution to centralize and automate encryption key management processes throughout the enterprise, helping minimize the risk of exposure and reduce operational costs.

The IBM SKLM offers a simple and a robust solution for external encryption key management through its lifecycle of key creation, secure storage, and protection of keys and key rotation, according to policy. It also provides key serving, including end-to-end authentication of encryption clients and security of the exchange of encryption keys.

After FlashSystem A9000 or FlashSystem A9000R starts, it must be able to communicate with at least one of the SKLM servers to obtain the encryption keys. Communication between FlashSystem A9000 and the SKLM server is through a KMIP over Secure Sockets Layer (SSL) protocol. The physical connection between FlashSystem A9000 or FlashSystem A9000R and the key server is through a Transmission Control Protocol (TCP)/Internet Protocol (IP) network.

Important: The IBM SKLM server can be installed as a virtual machine. In that case, make sure that it does not use the encrypted FlashSystem A9000 or FlashSystem A9000R as a storage device. This configuration can lead to an encryption deadlock situation, where an SKLM server cannot function because it depends on storage that cannot release data because it needs to communicate with that same server.

Also important: If the SKLM server is not reachable when IBM FlashSystem A9000 or FlashSystem A9000R (with external encryption activated) powers on (or is rebooted), the system will not be accessible to read or write data for the hosts. That is the reason why it is important to have at least two SKLM servers on your Internet Protocol (IP) network.

The Gemalto SafeNet KeySecure product can also be used as the external key server.

For more information see the IBM Redbooks publication, *Data-at-rest Encryption for the IBM Spectrum Accelerate Family*, REDP-5402.

5.4 Basic configuration

You must complete the configuration planning worksheets and checklists from the Planning topic of IBM FlashSystem A9000 or FlashSystem A9000R documentation in the IBM Knowledge Center so that the IBM SSR can physically install and configure the system.

You must provide the IBM SSR with the information that is required to attach the system to your network for operations and management, and enable remote connectivity for IBM Support and maintenance.

Enter all of the following required information in each worksheet to prevent further inquiry and delays during the installation:

- System Management IP Address

  Three IP addresses are required for the management of FlashSystem A9000 or FlashSystem A9000R and the appropriate netmask, and gateway. This information is necessary to manage and monitor the system by using either the GUI or the CLI. Each management interface needs a separate IP address to provide redundancy. If you require management to be resilient to a single network failure, be sure to connect these ports to at least two switches and make sure that the networking equipment that provides the management communication is protected by an uninterruptible power supply (UPS).

- Domain name server (DNS)

  If Domain Name System (DNS) is used in your environment, FlashSystem A9000 or FlashSystem A9000R must have the IP address of the primary DNS server and, if available, the secondary server.

- Simple Mail Transfer Protocol (SMTP) gateway

  An SMTP gateway is needed for event notification through email. FlashSystem A9000 or FlashSystem A9000R can initiate an email notification, which is sent out through the configured SMTP gateway (IP address or DNS name).

- Email sender address

  The email address to show as the sender in the email notification.

- Network Time Protocol (NTP) server

  FlashSystem A9000 or FlashSystem A9000R can be used with an NTP server to synchronize the system time with other systems. To use this time server, the IP address or server DNS name must be configured. It is a preferred practice to use an NTP server for time synchronization.

  Note that if for any reason, the system experiences NTP synchronization issues, the effect is limited to the timestamps of logs and events. There is no impact on IOs or on the storage system operations.

  When the NTP synchronization resumes, the system automatically adjusts its internal time to the NTP time again. However, to avoid issues with NTP corruption, the system will not do the automatic adjustment if the difference between the NTP time and the system time is greater than 2 hours. In that case, the user can adjust the system time manually to a time close to the NTP time. After the manual setting, the time difference between the system and the NTP should be less than 2 hours, triggering the system to perform the automatic adjustment. The XCLI command to adjust the system time is **`time_set`**.

- Time zone

  The time zone depends on the location where the system is installed. However, exceptions can occur for remote locations where the time zone equals the time of the host system location.

- Remote access

  FlashSystem A9000 and FlashSystem A9000R can use one of three methods for remote access:

  – Management IP ports
  – Virtual private network (VPN) ports
  – Modem (only for model 415)

  The remote support network connection must have outbound connectivity to the internet.

  For more information, see Chapter 7, “Monitoring and troubleshooting” on page 141.

- Contact information

  The customer contact for the site that can authorize remote support access and can enable FlashSystem A9000 and FlashSystem A9000R for remote support.

This basic configuration data is entered in the system by the IBM SSR after the physical installation.

Other configuration tasks, such as defining storage pools, volumes, and hosts, are the responsibility of the storage administrator; they are described in the IBM Redbooks publication, *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

5.5 Network connections

Network connection planning is also essential for installing FlashSystem A9000 or FlashSystem A9000R. To deploy and operate the system in your environment, several network connections are required:

- Gigabit Ethernet connections for management
- Gigabit Ethernet connections (VPN) and modem (model 415 only) for remote support
- Fibre Channel connections for host I/O over Fibre Channel
- Ten gigabit Ethernet connections for host I/O over iSCSI

On FlashSystem A9000R, the management, VPN, and modem remote support connections are through the patch panel in the IBM T42 rack, as explained in 3.3.3, “FlashSystem A9000R rack” on page 57.

5.5.1 Remote mirroring connectivity

Planning the physical network connections also includes considerations when FlashSystem A9000 or FlashSystem A9000R is installed in a Remote Copy environment. Contact your technical sales support team for assistance in planning remote mirroring connectivity to ensure the maximum resilience to hardware and connection failures.

Remote Copy links, which connect the direct primary system and secondary system, must also be planned for before the physical installation. The physical Remote Copy links can be Fibre Channel links, through a storage area network (SAN), or iSCSI port connections by using Ethernet.

5.5.2 Planning for growth

Consider growth and the future I/O demands of your business. Most applications and databases grow quickly and the need for greater storage capacity increases rapidly. Planning for growth before the implementation of the first FlashSystem A9000 in the environment can save time and effort in the future.

5.5.3 IPv6 addressing and planning

Internet Protocol Version 6 (IPv6) is intended to supplement and eventually replace the IPv4 protocol that is used today. Getting ready to make the transition to IPv6 is becoming critical for companies because the pool of available IPv4 addresses is depleted. IPv6 is described by the Internet Engineering Task Force (IETE), RFC2460.

IPv6 support is provided for Management and VPN ports only, and it is supported for network functions, such as DNS, SMTP, Lightweight Directory Access Protocol (LDAP), and Simple Network Management Protocol (SNMP).

Note: IPv6 address support is provided by both the GUI and CLI.

These specific functions do not support IPv6. (Only IPv4 is used.)

- Technician notebook port
- iSCSI port configuration
- Host iSCSI connections
- Mirroring iSCSI connections
- Support Center and remote support communication

IPv6 provides the following advantages over IPv4:

- Larger address space: 128 bit (2^128 ˜ 3.4×10^38 addresses)
- Fixed subnet size 64 bits (2^64 (~18.4 quintillion) addresses
- Elimination of the need for network address translation (NAT)
- Support for stateless auto-configuration and network renumbering
- Interoperability with IPv4 by fixed subnet size (64-bit addresses)
- Built-in protocol for multicasting
- Mandatory support for Internet Protocol Security (IPSec)

5.5.4 IP protocols

FlashSystem A9000 or FlashSystem A9000R is managed through dedicated management ports that run TCP/IP over Ethernet. Management is carried out through the following protocols (consider this design when you configure firewalls, other security protocols, and SMTP relay):

- Proprietary protocols are used to manage FlashSystem A9000 or FlashSystem A9000R from the Management GUI and the CLI. This management communication is performed over TCP port 7778, where the GUI/CLI, as the client, always initiates the connection, and FlashSystem A9000 or FlashSystem A9000R performs as the server.
- FlashSystem A9000 or FlashSystem A9000R sends and responds to SNMP management packets.
- FlashSystem A9000 or FlashSystem A9000R initiates SNMP packets when it sends traps to SNMP managers.
- FlashSystem A9000 or FlashSystem A9000R initiates SMTP traffic when it sends emails (for either event notification through email or for email-to-Short Message Service (SMS) gateways).
- FlashSystem A9000 or FlashSystem A9000R communicates with remote support SSH connections over standard TCP port 22. If remote support connection will be used, FlashSystem A9000 must be able to communicate on port 22 to the internet through either management or VPN ports.

5.5.5 SMTP server

For the correct operation of the Call Home function, the SMTP server must function in the following manner:

- Be reachable on port 25 for FlashSystem A9000 or FlashSystem A9000R customer-specified management IP addresses.
- Allow relaying from FlashSystem A9000 or FlashSystem A9000R customer-specified management IP addresses.
- Allow FlashSystem A9000 or FlashSystem A9000R to send emails. The default sender address is `xiv@il.ibm.com`, but this address can be changed.
- Allow recipient addresses of `xiv-callhome-western-hemisphere@vnet.ibm.com` and `xiv-callhome-eastern-hemisphere@vnet.ibm.com`. Additionally, you can choose the shorter alias recipient addresses of `xiv-west@vnet.ibm.com` and `xiv-east@vnet.ibm.com`.

5.5.6 IBM service ports (model 415 only)

FlashSystem A9000 or FlashSystem A9000R has two Ethernet ports that are dedicated for the use of IBM service personnel. The system serves as a Dynamic Host Configuration Protocol (DHCP) server and automatically configures the service personnel's notebook. These ports are only for IBM maintenance personnel, for example, an IBM service support representative (SSR) or IBM technical advisor (TA).

Restriction: Do not connect these ports to the user (customer) network.

5.6 Physical installation

After all previous planning steps are completed and the system is delivered to its final location, the physical installation can begin. An IBM SSR completes all necessary tasks and the first logical configuration steps, up to the point where you can connect to FlashSystem A9000 or FlashSystem A9000R through the Management GUI and the CLI. Configuring storage pools, logical unit numbers (LUNs), and attaching to the host are storage administrator responsibilities. For more information, see *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

The physical installation steps are shown.

Use these steps for FlashSystem A9000R:

1. Place and adjust the rack in its final position in the data center.
2. When the machine is delivered with the height or weight reduction feature, the IBM SSR installs the removed components into the rack.

Use these steps for both FlashSystem A9000 and FlashSystem A9000R:

1. Check the system hardware.
2. Connect the system power cords to the customer-provided power source and advise an electrician to switch on the power connections.
3. Perform the initial power-on of the machine and perform necessary checks according to the particular power-on procedure.
4. To complete the physical steps of the installation, the IBM SSR performs various final checks of the hardware before the SSR continues with the basic configuration.

5.6.1 Initial setup

After the completion of the physical installation steps, the IBM SSR establishes a connection to FlashSystem A9000 or FlashSystem A9000R through the technician port and completes the initial setup. You must provide the required completed information sheet that is referenced in 5.4, “Basic configuration” on page 117.

The initial setup steps are shown:

1. Set the Management IP addresses (customer network), gateway, and netmask.
2. Set the system name.
3. Set the email sender address and SMTP server address.
4. Set the primary DNS and the secondary DNS.
5. Set the SNMP management server address.
6. Set the time zone.
7. Set the NTP server address.
8. Configure the system to send events to IBM (Call Home).
9. Configure and test remote support.

5.6.2 Completing the physical installation

After the IBM SSR completes the physical installation and initial setup, the IBM SSR performs the final checks for FlashSystem A9000 or FlashSystem A9000R:

1. Power off and power on the machine by using the Management GUI or CLI.
2. Check the Events log carefully for problems.

3. Verify that all settings are correct and persistent.

The installation is complete, and the system is ready to be handed over to the customer to configure and use.

5.6.3 System shutdown/power-off and power-on

FlashSystem A9000 or FlashSystem A9000R must be shut down correctly before you power off the system. Otherwise, severe data loss can occur. You must be logged on as the storage administrator (the storageadmin role).

Important: Strictly follow these procedures to shut down and power off your FlashSystem A9000 or FlashSystem A9000R. Failure to do so can result in data loss.

System shutdown using the CLI

From the command prompt, issue the following command (where `ITSO_01` is the system name or IP address, <*storageadmin*> is the user name of a user with the storageadmin role, and ******* is the user's password):

```
xcli -c "ITSO_01" -u <storageadmin> -p ****** shutdown -y
```

You receive the response:

```
Command executed successfully
```

If you are using the CLI, use the shutdown procedure that is shown in Example 5-1.

Example 5-1 Executing a shutdown from the CLI session

```
User Name: itso
Password: ********
Machine IP/Hostname: 10.10.53.250
connecting.
ITSO_01>>shutdown
Password: ********

Warning:   ARE_YOU_SURE_YOU_WANT_TO_SHUT_DOWN y/n:
Command executed successfully
```

System shutdown using the GUI

You may shut down a FlashSystem A9000 or FlashSystem A9000R using the GUI and similar to the CLI procedure, you have to log in as a user with storageadmin role.

To shut down the system, go to **Systems & Domains views** →**Select a particular system** →**Actions** (a) →**Shutdown** (b) →**Shutdown system** (c) (See Figure 5-2).

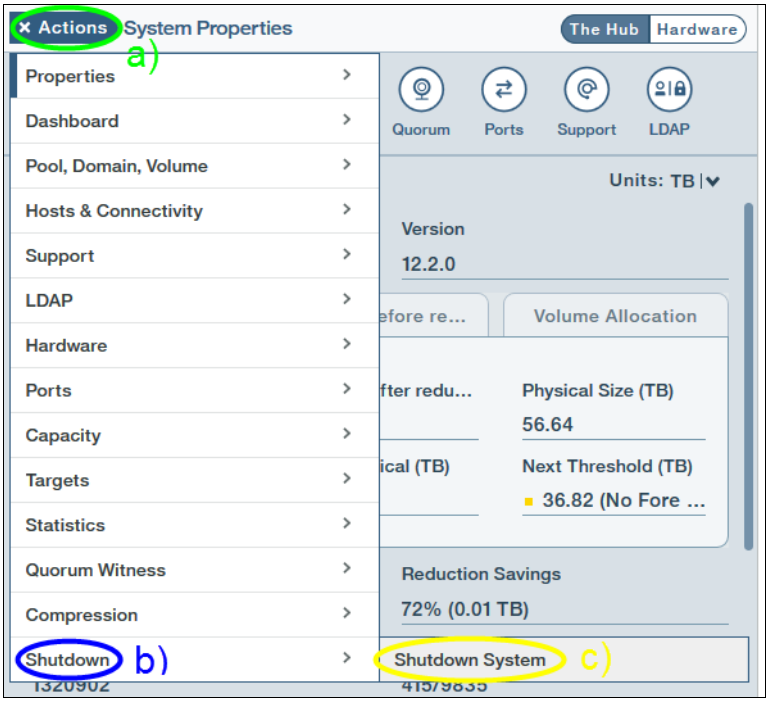


Figure 5-2 Shutdown System

Note: The minimum required GUI level for shutdown is 5.3.

The shutdown takes less than 5 minutes. When it is finished, all fans and front lights on all grid controllers and flash enclosures are off. Status LEDs might blink.

For FlashSystem A9000R, the InifiniBand (IB) switches, and the Power Distribution Units (PDUs) remain on, if the system power switch is set to ON.

Automatic shutdown

In case of a power loss on both main power sources, a loss in battery backup unit (BBU) redundancy or an overheating situation, IBM FlashSystemA9000 and FlashSystem A9000R will perform an automatic shutdown to protect the data. All active components (grid controller, flash enclosures and for FlashSystemA9000R, the IB switch) contain a redundant BBU. This BBU allows the system to stay online enough time to destage or vault all data that is not written to the system. During an automatic shutdown, the grid controller destages the following data:

- Data in write cache
- Metadata that changed since the last vault run

The redundant BBU included in the flash enclosure allows it to shutdown gracefully. During the shutdown process, it writes the following data:

- Data in the flash enclosure write cache
- Recent data distribution table

After all data is secured, the components power off. When the utility power recovers after the automatic shutdown, FlashSystem A9000 and FlashSystem A9000R automatically powers back on.

If the power recovers while the system is still shutting down, FlashSystem A9000 and FlashSystem A9000R can remain powered off and they must be powered on manually.

Power switch

FlashSystem A9000R has a power switch on the front panel of the rack that will control power output from the PDUs. If the power switch is put in the Off position while the system is operational, the system will perform a shutdown and power off as described previously.

Important: Use the power switch at the direction of IBM Support only.

The power switches for FlashSystem A9000R components are behind the front bezel and must never be used without direction from IBM Support.

Power on

After FlashSystem A9000 or FlashSystem A9000R was powered off, either due to an automatic shutdown or due to a planned shutdown, certain considerations affect whether the system is powered on gracefully.

After an automatic shutdown, the system powers on automatically if the power was off for more than 5 minutes. If the power outage was shorter (perhaps the input power returned while the system was still in the shutdown procedure), the system might remain off even though it has power.

Power on FlashSystem A9000R

Follow this procedure to power on FlashSystem A9000R manually:

1. Ensure that the power to both sources is ON and within the allowed voltage ranges.
2. Verify that the power cables are connected correctly. Consult with your site electrician if any power-related work is required.
3. Verify that the LEDs on both Power Distribution Units (PDUs) are lit. You can see them through the front bezel.
4. Check that no grid controller is powered on and that no flash enclosure is powered on. Check that no LEDs are illuminated on the front panel. The status LEDs might blink.

   Important: If any of the components (grid controller or flash enclosure) are powered on, do *not* proceed, but escalate to IBM Support immediately.

5. Check the position of the system power button in the front bezel. If it is in ON position, flip it to OFF. Wait at least 5 seconds before you proceed to the next step.
6. Turn on the system by switching the button to ON.

   Important: If any component (grid controller or flash enclosure) does not power on, escalate to IBM Support. Do not power on any component individually.

7. Check the system state through the GUI or CLI. The CLI command is **`state_list`**. Wait for the system state to show ON.

Note: If the system recovered from an automatic shutdown, it might return in the maintenance system state because the BBUs might need to charge first. The system state will not be ON if the battery power cannot sustain another power outage. Therefore, the system is waiting for the correct battery charge level first. If the system state does not return to ON after a few minutes, call IBM Support to dial in and check the system.

Power on FlashSystem A9000

Follow this procedure to power on the system manually:

1. Ensure that power to both sources is ON and within the allowed voltage ranges.
2. Verify that the power cables are connected correctly to every component (three grid controllers and one flash enclosure, and two power cords each). Consult with your site electrician if any power-related work is required.
3. Check that the grid controller and flash enclosure are off. Check that no LEDs are solid on the front. The status LEDs might blink.

   Important: If any of the components (grid controller or flash enclosure) are powered on, do *not* proceed, but escalate to IBM Support immediately.

4. If possible, turn off the wall power on both power lines that feed FlashSystem A9000. Wait for at least 5 seconds and turn on the power. FlashSystem A9000 will power on automatically. In this case, proceed to the last step. Otherwise, proceed with the next step.
5. Turn on the system by pressing the power button at the front of each component except the flash enclosures. To turn the flash enclosures on, make sure that at least 1 power cord is connected to each of them. Make sure that all components power on within 30 seconds.

   Important: If any component (grid controller or flash enclosure) does not power on or any component power button was not pressed within a 30-second time frame, do not attempt to power off or power on any component individually. Escalate to IBM Support.

6. Check the system state through the GUI or CLI. The CLI command is **`state_list`**. Wait for the system state to show ON.

   Note: If the system recovered from an emergency shutdown, it might return in the maintenance system state because the BBUs might need to charge first. The system will not be in the ON state if the battery power cannot sustain another power outage. Therefore, the system is waiting for the correct battery charge level first. If the system state does not return to ON after a few minutes, call IBM Support to dial in and check the system.

5.7 Host connections

FlashSystem A9000 or FlashSystem A9000R is equipped with Fibre Channel ports or iSCSI ports, depending on the configuration that is ordered. The external (customer-provided) Fibre Channel (FC) cables or 10 GbE optical cables are plugged into ports on the back of the grid controller modules. The ports are provided for connectivity to the switch network for host access or for use in remote mirroring or data migration scenarios. (The target or initiator role configuration can be changed dynamically.)

5.7.1 Fibre Channel connections

FlashSystem A9000 or FlashSystem A9000R supports 50-micron fiber cabling. If you have other requirements or special considerations, contact your IBM marketing representative.

5.7.2 iSCSI connections

FlashSystem A9000 or FlashSystem A9000R supports 10 GbE fiber cabling for iSCSI. If you have other requirements or special considerations, contact your IBM marketing representative.

5.7.3 Fibre Channel cabling and configuration

Fibre Channel cabling must be prepared based on the required fiber length and the selected configuration.

When you install FlashSystem A9000 or FlashSystem A9000R, complete the following Fibre Channel configuration procedures:

- You must configure Fibre Channel switches that are zoned correctly to allow access between the hosts and the storage system. The specific configuration to follow depends on the specific Fibre Channel switch. It is best to have a separate zone for each initiator.
- Hosts must be set up and configured with the correct multipathing software to balance the load over various paths. For multipathing software and setup, see the specific operating system section in the IBM Redbooks publication, *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368.

5.7.4 iSCSI cabling and configuration

Logical network configurations for iSCSI are equivalent to the logical configurations that are suggested for Fibre Channel networks.

The following options are available:

- Redundant configuration

  Each module connects through at least two ports to two Ethernet switches, and each host is connected to the two switches. This design provides a network architecture that is resilient to a failure of any individual network switch or module.
- Single switch configuration

  A single switch connects all modules and hosts.
- Single port host solution

  Each host connects to a single switch, and a switch is connected to two modules.

5.7.5 iSCSI IP configuration

The configuration of FlashSystem A9000 or FlashSystem A9000R iSCSI connection depends on your network. In the high availability configuration, the two customer-provided Ethernet switches that are used for redundancy can be configured as either two IP subnets or as part of the same subnet. FlashSystem A9000 or FlashSystem A9000R iSCSI configuration must match your network.

You must provide the following configuration information for each Ethernet port:

- IP address
- Netmask
- Optional: Maximum transmission unit (MTU), which is optional, depending on your network's MTU

  The MTU configuration is required if your network supports an MTU that differs from the standard one. The largest possible MTU must be specified. From FlashSystem A9000 or FlashSystem A9000R, the largest MTU size is 9,126 bytes. If the iSCSI hosts are on a different subnet than FlashSystem A9000 or FlashSystem A9000R, a default IP gateway for each port must be specified.
- Optional: Default gateway

  Because FlashSystem A9000 or FlashSystem A9000R always acts as a Transmission Control Protocol (TCP) server for iSCSI connections, packets are always routed through the port from which the iSCSI connection was initiated. The default gateways are required only if the hosts are not on the same Layer-2 subnet as FlashSystem A9000 or FlashSystem A9000R.

You can configure from the CLI by using the **`ipinterface_create`** or **`ipinterface_update`** command.

The IP network configuration must be ready to ensure connectivity between FlashSystem A9000 or FlashSystem A9000R and the host before the physical system installation:

- Ethernet virtual local area networks (VLANs), if required, must be configured correctly to enable access between hosts and the XIV Storage System.
- IP routers (if present) must be configured correctly to enable access between hosts and FlashSystem A9000 or FlashSystem A9000R.

Mixed iSCSI and Fibre Channel host access

FlashSystem A9000 or FlashSystem A9000R does not support mixed concurrent access from the same host to the same volume through FC and iSCSI simultaneously. A single host can access different volumes where certain volumes are accessed by using FC and other volumes are accessed by using iSCSI. You can also access the same volume from separate hosts by using various connection methods (if those hosts use a form of clustered file or operating system).

6

Performance

This chapter describes how IBM FlashSystem software and IBM FlashSystem A9000 or IBM FlashSystem A9000R hardware work together to deliver and monitor system performance. Performance characteristics are ingrained in the system design to deliver optimized and consistent performance. Little else is necessary on FlashSystem A9000 or FlashSystem A9000R that can contribute to performance gains beyond what the system automatically provides. However, several considerations and practices can help prioritize performance for business-critical applications or certain hosts or domains.

This chapter includes the following sections:

- Performance considerations
- Quality of service
- Performance monitoring

6.1 Performance considerations

The architecture of FlashSystem A9000 and FlashSystem A9000R is designed to deliver a high performance, hotspot-free storage system.

Real-world production environments involve multiple application servers that make multiple simultaneous I/O demands on storage systems. When customers decide to purchase a FlashSystem A9000 or FlashSystem A9000R, they have the reasonable expectation that they will migrate existing applications or install new applications to FlashSystem A9000 or FlashSystem A9000R and experience great flash performance.

6.1.1 Sizing

To get the required performance from a FlashSystem A9000 or FlashSystem A9000R, it is essential, as with any storage system, to size the system characteristics correctly. For FlashSystem A9000R, each grid controller adds an amount of cache and processing power to the grid, and each additional flash enclosure adds more capacity and performance capabilities. The overall system performance increases with the number of grid elements that are included in FlashSystem A9000R.

The best performance configuration might look different from the required capacity configuration. Storage requirements might indicate that a FlashSystem A9000R that is built of just two grid elements is sufficient to fulfill the capacity needs, but the performance requirements of the applications might indicate the need for a FlashSystem A9000R that consists of four grid elements. In this case, to satisfy the performance requirements, you need to select the larger configuration of four grid elements.

Performance requirements can be determined in several ways:

- Benchmark results
- Application vendor specifications (based on benchmark testing)
- Actual /O performance characteristics of the existing storage that is being replaced by FlashSystem A9000 or FlashSystem A9000R

6.1.2 Multipathing considerations

The optimum performance of FlashSystem A9000 and FlashSystem A9000R is realized by maximizing the use of the grid controllers. Ensuring this optimal usage and balancing the application workload among all of the grid controllers are the most important considerations when you deploy FlashSystem A9000 or FlashSystem A9000R in your environment.

Important: When you plan the host connections to FlashSystem A9000 and FlashSystem A9000R, it is important to ensure that all grid controllers are used.

One main multipathing goal, from a performance perspective, is for the host connectivity to create a balance of the I/O workload across all of the resources in FlashSystem A9000 or FlashSystem A9000R. The best way to achieve this balance is by distributing the host physical connections evenly across *all* of the grid controllers.

Providing host I/O access to every grid controller from every host bus adapter (HBA) has the following advantages:

- Uses the most cache
- Uses the maximum available processor power to handle I/O

- Fully uses the grid architecture
- Minimizes the impact of a host interface hardware failure

For different multipathing configurations, see *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368.

Important: To achieve a balance between port usage and performance, use at least one port per grid controller and an equal number of ports on each grid controller. The maximum number of paths must not exceed 12 because a higher number results in higher resource use on the host side and no additional performance gain.

6.2 Quality of service

The quality of service (QoS) feature allows FlashSystem A9000 and FlashSystem A9000R to deliver different service levels to hosts that are connected to the same FlashSystem.

The QoS feature is intended to favor the performance of critical business applications that run concurrently with less critical applications. Because FlashSystem A9000 and FlashSystem A9000R processing capacity and cache are shared among all applications and all hosts are attached to the same resources, equal allocation of these resources among both critical and less critical applications might negatively affect the performance of the business-critical applications.

The response to this issue is to limit the input/output operations per second (IOPS) rate and bandwidth of certain applications by specifying and then enforcing limits. As a result, the QoS feature in FlashSystem A9000 and FlashSystem A9000R enables better performance for the critical host applications that run on the same system, concurrently with the noncritical host applications.

See *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376 for information about how to set QoS by defining performance classes in terms of IOPS and bandwidth limitation. It also explains how to assign specific members to a particular performance class. Each member can be assigned to only a single performance class at a time. However, the number of members within a specified class is not limited.

6.2.1 Limitation by bandwidth

The interface service that is running on each grid controller enforces the configured limitations. The intended limitation value depends on the number of grid controllers that are used by the hosts within the same performance class. The maximum rate value that is specified is divided by the number of grid controllers that are installed in FlashSystem A9000 (always three) or FlashSystem A9000R (4 - 12 for model 415 and 4 - 8 for model 425) to determine the rate for the class.

For example, a noncritical host is connected to all three grid controllers on a FlashSystem A9000:

- If the application administrator intends to enforce a 300 MBps limit for that host, the administrator user must set the QoS bandwidth limit for that host to 300 and the Bandwidth Limit per grid controller is automatically set to 100.

- With three grid controllers, the enforcement is 100 MBps per grid controller, limiting the host to an aggregate bandwidth of 300 MBps (100 MBps x 3 modules = 300 MBps). If only two interface modules were used, the limit for the host is 200 MBps (100 MBps x 2 modules = 100 MBps).
- If the host has connections to only two of three grid controllers in a FlashSystem A9000, the actual host bandwidth limitation is only 200 MBps with this performance class setting (100 MBps x 2 modules = 200 MBps). Therefore, if the user intends to have a 300 MBps bandwidth limitation with two grid controllers that are connected in a full three-grid controller FlashSystem A9000, the bandwidth limit per interface is 150 MBps and the Bandwidth Limit must be set to `450`.

6.2.2 Limitation by input/output operations per second

If the intent is to set a limitation at 10,000 IOPS for a specified host in a FlashSystem A9000 configuration, the IOPS limit must be set to 10,000 and the enforcement is 3,333 (10,000/3) for each grid controller.

If the host is attached to only two interface modules in a three grid controller FlashSystem A9000, the host IOPS limitation is only 6666 with this performance class setting (3333 IOPS x 2 interface modules = 6666 IOPS).

If the intent is to have a 10 K IOPS limitation for a host that is connected to only two grid controllers in the specific scenario, “IOPS Limit Per Interface” must be set to `5000` or “IOPS Limit” for the performance class needs to be set to `15,000`.

Note: Users must consider these grid controller multiplication factors to meet their expected limitations correctly when a host is connected to only a few grid controllers.

Shared or independent limitation

During the creation of a performance class, the performance class can be given a parameter setting of either `Shared` or `Independent`. This parameter defines whether the specific limits are shared among all class members or apply individually to each class member.

When `Shared` is selected, the maximum limit (bandwidth, IOPS, or both) is shared among all members of a performance class. Each member can reach the maximum limited value, but not all at the same time. For example, with a performance class with a limit of 300 MBps and two assigned hosts, each host can reach a maximum of 300 MBps because 300 MBps is the maximum allowed limitation. However, when both hosts are performing I/O, they share the limit. In this situation, no division between the members is done. Every member gets as much as it can get, but not more than the defined limit.

When `Independent` is selected, the maximum limit (bandwidth, IOPS, or both) applies to each member of the performance class. For example, for a performance class with a limit of 300 MBps and two assigned members, each member can reach a maximum of 300 MBps, regardless of how much bandwidth another member in the performance class consumes at the same time.

Performance class members

A *performance class* can be defined for one of the following four levels:

- Domains
- Hosts
- Pools
- Volumes

When a performance class is created, either by using the graphical user interface (GUI) or through the command-line interface (CLI), it is assigned a name, the system where it is created, the shared or independent parameter, and either the bandwidth limit, IOPS limit, or both. Later on, the members are added to the class. The members can be one of the four types that are listed, but you cannot mix the types of members. Each performance class can contain one or multiple members of a certain type.

Domains

When a domain is added to a performance class, the total bandwidth or IOPS rate that is defined as the limit for this domain applies to all I/O operations of a domain member. Each host, volume, or pool within the domain cannot exceed the specific limit and all domain members share the maximum limitation. For example, if two hosts are in a domain, they both share the limitation that is given to the domain by the performance class.

If the performance class is defined with the shared parameter setting, the limitation is even shared among all domains that are part of the performance class. If it is defined as independent, each domain has its own limitation.

Hosts

When a host is added to a performance class, the total limit applies to all I/O operations of the member host, regardless of which volume the host has access to.

Pool

A performance class limitation on a pool level applies to all volumes in that particular pool. A host that is accessing a volume from that pool will be limited to the defined values. If the same host is accessing other volumes from a different pool, the limit does not apply because the pool might not be a member of a performance class.

Volume

The behavior for volumes is comparable to the behavior setting for pools. The only difference is that not all volumes in the pool will be limited, but only the particular volumes that are added to the performance class. All other volumes in the pool that are not members of a performance class can be accessed by hosts without limitation.

Layering quality of service

You can layer FlashSystem A9000 and FlashSystem A9000R performance classes.

For instance, a domain can be added to a performance class, and hosts that are part of the domain can be added to other performance classes. In this situation, the lower limit always applies. For example, if a domain has a limit of 1,000 MBps and a host, which is a member of this domain, is added to another performance class by the domain administrator with a limit of 200 MBps, that host will be limited to 200 MBps. Otherwise, the limit of the domain, which is 1,000 MBps, applies and the host can use up to that bandwidth.

To illustrate this possibility, consider the following scenario. A service provider has multiple customers that run their services on a single FlashSystem A9000R, and the service provider uses multi-tenancy. The service provider wants to provide different QoS levels to the tenants that run on this system and decides to create three service levels: Bronze, Silver, and Gold.

To achieve this goal, the service provider creates two performance classes with the following settings and one customer set with no settings:

- Bronze: Bandwidth limit: 1,000 MBps and Shared
- Silver: Bandwidth limit: 1,000 MBps and Individual
- Gold: No limit

For the Gold customers, the service provider does not create a performance class because the Gold customers are not limited.

Next, the service provider adds the correct customer domains to each performance class and leaves the Gold customer domains out of any performance class.

With this setting, the Bronze customers can reach up to 1,000 MBps, but they have to share the maximum bandwidth with all of the other customers at the Bronze level. The Silver customers can reach up to 1,000 MBps, and they do not have to share their limit with other customers in their performance class. The Gold customers are unlimited because they are not part of any performance class.

Each domain administrator now can create their own performance class, for example, to limit certain hosts, which are less critical than other hosts within their own domain. The limit can be either on bandwidth, on IOPS, or on both, regardless of what was defined in the domain level performance class. However, the lowest limit applies always. Figure 6-1 shows an illustration of this scenario.

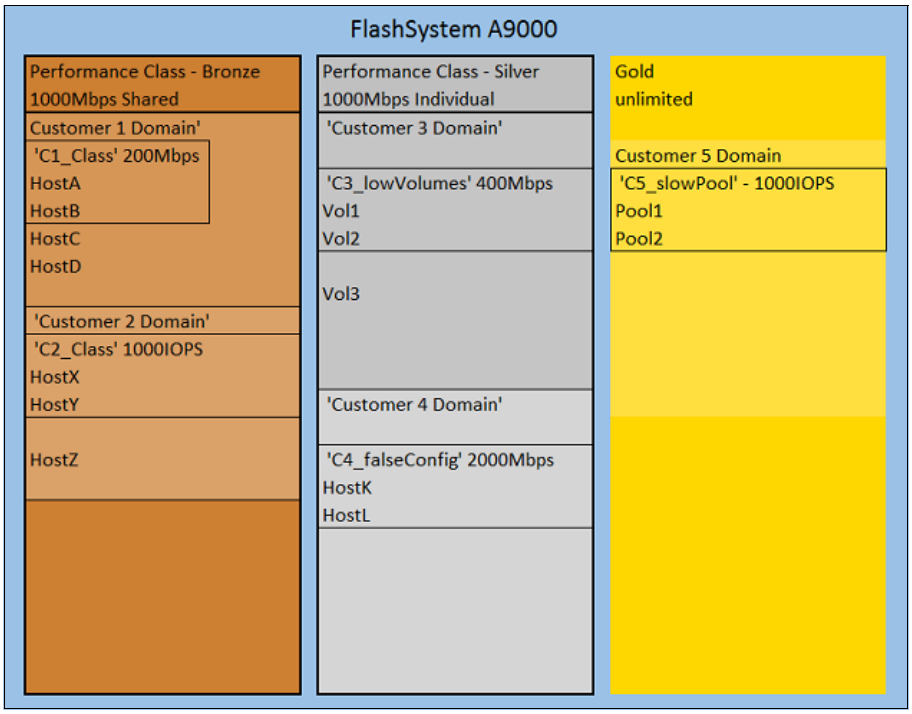


Figure 6-1 Multiple performance classes

In Figure 6-1, five domains are defined on the system, Customer 1 Domain to Customer 5 Domain. Two performance classes were defined to which domains can be added, one for Bronze customers and one for Silver customers:

- Customer 1 Domain and Customer 2 Domain are added to the Bronze performance class.
- Customer 3 Domain and Customer 4 Domain are added to the Silver performance class.
- Customer 5 Domain is not added to any class because the customer is a Gold level customer and will not be limited.

The domain administrator for Customer 1 Domain created a performance class (C1_Class) with a limit of 200 Mbps and added HostA and HostB to it. HostC and HostD were not added to the C1-Class. HostA and HostB are now limited to 200 Mbps. HostC and HostD are limited to a maximum of 1,000 Mbps due to their domain membership.

By looking at Customer 2 Domain, we see that the domain administrator created C2_Class. HostX and HostY are members of this class, so they are limited to 1,000 IOPS, but they are also limited to 1,000 Mbps due to their domain limit.

Furthermore, all I/Os to Customer 1 Domain and Customer 2 Domain must share 1,000 Mbps because the Bronze performance class is configured as shared.

The domain administrator for the Customer 3 Domain created a performance class (C4_lowVolumes) and added volumes to it, which is also possible. Vol1 and Vol2 are limited to 400 Mbps and Vol3 is limited to 1,000 Mbps, due to the domain limit in the Silver tier. All volumes of this domain share the limit of 1,000 Mbps even if Silver is set to shared. Shared means that Customer 3 Domain needs to share the bandwidth with Customer 4 Domain, which is also a member of the same performance class.

In Customer 4 Domain, we see a configuration that will not work correctly. This configuration can be created, but HostK and HostL will be limited to 1,000 Mbps instead of 2,000 Mbps, due to the lower limit of the Silver performance class.

Even if Customer 5 Domain is not a member of any performance class and has no bandwidth or IOPS limit, the domain administrator can create performance classes inside the domain and use its own QoS rules. The domain administrator created a class C5_slowPool with a limit of 1,000 IOPS and added Pool1 and Pool2 to it.

6.3 Performance monitoring

You monitor the performance of FlashSystem A9000 or FlashSystem A9000R by using different methods or tools.

The Hyper-Scale Storage Management GUI is the primary tool to monitor performance. Other tools, such as IBM Spectrum Control and IBM certified third-party tools, can be used for monitoring. The Hyper-Scale Storage Management GUI displays both current and historical performance statistics, and it allows historical statistics to be exported to files by using the CLI utility for further trending and analysis by the user.

6.3.1 Using the Storage Management GUI

Several views and panels are available for performance management.

Overall performance

The lower-right section of the Dashboard view displays the overall performance of all of the systems in the Hyper-Scale Manager inventory. The panel shows current statistics. You can switch between displaying IOPS and latency.

Starting with Storage Software V12.2.1 and Hyper-Scale Manager UI version 5.4, the system displays a breakdown of the total latency for a particular host, by separating network latency and storage system internal latency.

With HSM version 5.5 and system software version 12.3, you can now, *at a system level*, separate internal latency from external latency. This view can be obtained by placing the mouse cursor over the latency graph in the dashboard view.

System statistics

Users can navigate to the system statistics view by clicking the performance charts that are shown in the Dashboard or by clicking the **Statistics** icon on the left side panel and choosing **System & Ports Performance Statistics** from the Statistics Views menu, as shown in Figure 6-2.

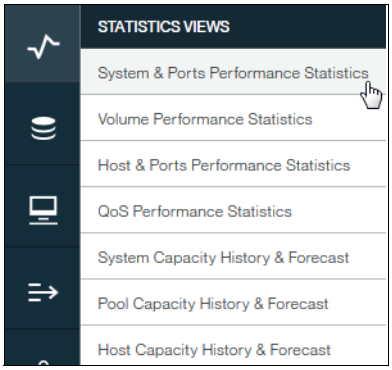


Figure 6-2 System and Ports Performance Statistics selection

The system statistics Workspace view is displayed. The top portion displays a list of systems that are managed under the Storage Management GUI, which are filtered according to your criteria. The bottom shows the average IOPS, latency, and bandwidth for systems that are selected in the list, as illustrated in Figure 6-3.

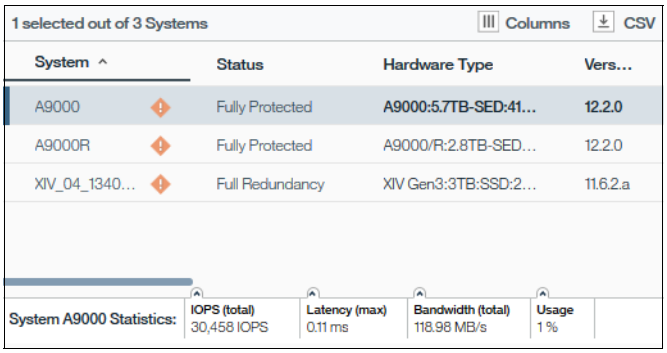


Figure 6-3 System statistics view

By clicking any of the vertical arrows at the bottom, you can display a chart that shows the historical statistics for the corresponding selection (IOPS, latency, or bandwidth). You can drill down to details about various aspects of the performance statistics. Figure 6-4 on page 136 shows an example for IOPS.

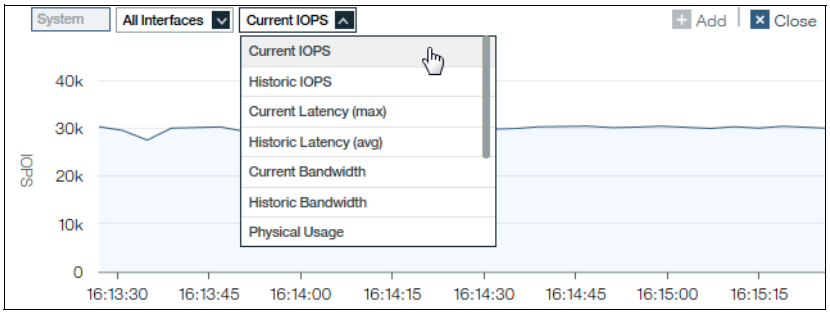


Figure 6-4 Filter options on the Statistics view

You can also refine the view or select to display historical statistics, as shown in the menu in Figure 6-4.

The system can display historical statistics for IOPS, bandwidth, and latency for various time ranges, up to one year, as illustrated in Figure 6-5.

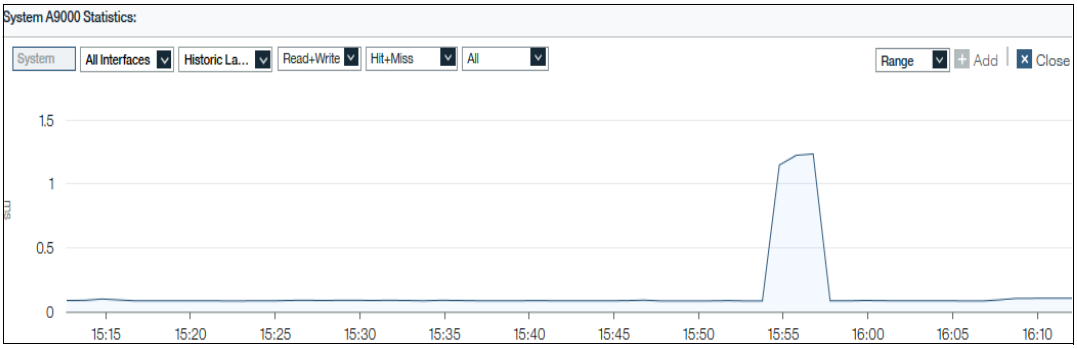


Figure 6-5 Historical latency

As previously indicated the total latency for a selected host is split between network latency and internal (storage system) latency, as shown in Figure 6-6.

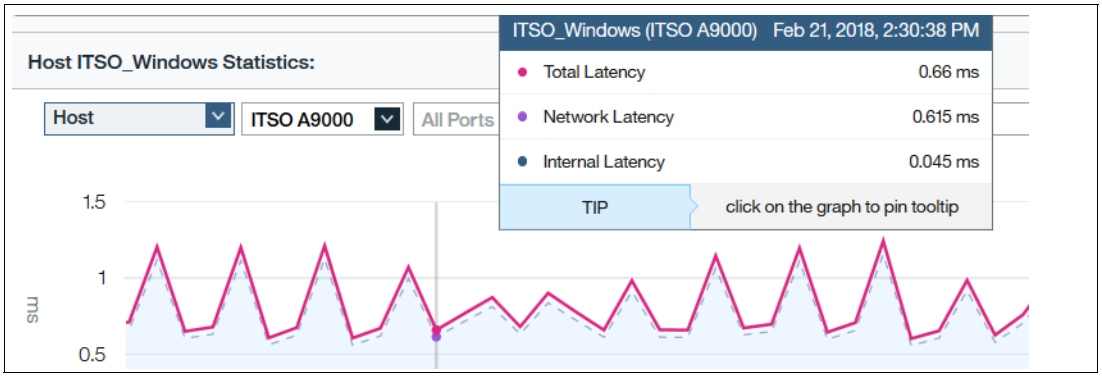


Figure 6-6 Total Latency break-out between Network and Internal Latency, per Host

With HSM version 5.5 and system software version 12.3, you can now, at a system level, separate internal latency from external latency. This view can be obtained by placing the mouse cursor over the latency graph in the dashboard view, as shown in Figure 6-7.

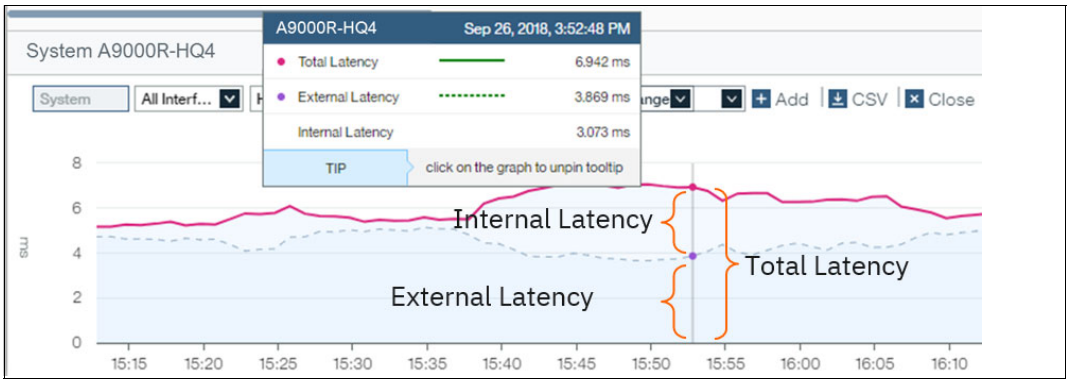


Figure 6-7 Total Latency break-out between Network and Internal Latency, per Storage System

You can choose various filters so that you can view the following measurements:

- Read + Write: Filters for Read, Write, and Read + Write to view historical IOPS, bandwidth, and latency by reads and writes.
- Hit + Miss: Filters Memory miss, Memory Hit, and Hit + Miss to view historical IOPS, bandwidth, and latency by hits and misses, and total hits plus misses.
- All: Filters under All provide a view of the block size breakdown, where you can choose 0 - 8 KB, > 8 - 64 KB, > 64 - 512 KB, > 512 KB, or All. The breakdown filters are useful in determining the I/O profile of the host and to further tune host-side queues and other parameters, such as size and coalesce.
- Range: You can use range filters to filter and view by time period. You can choose Last Year, Last Month, Last Week, Last Day, Last Hour, and From To to customize the time period.

6.3.2 Using the command-line interface

The second method to collect statistics is by using the command-line interface (CLI).

First, you must retrieve the system's time. To retrieve the system's time, issue the **time_list** command, as shown in Example 6-1.

Example 6-1 Retrieving the system time

```
A9000>>time_list
Time       Date         Time Zone       Daylight Saving Time
16:34:47   2017-10-24   Europe/Berlin   yes
```

After you obtain the system time, the **statistics_get** command can be formatted and issued. The **statistics_get** command requires various parameters to operate. It requires that you enter a starting or ending time point, a count for the number of intervals to collect, the size of the interval, and the units that are related to that size. The time stamp is modified by the previous **time_list** command. Example 6-2 provides a description of the command.

Example 6-2 Syntax of statistics_get command

```
statistics_get [ perf_class=perfClassName | host=HostName |
host_iscsi_name=initiatorName | host_fc_port=WWPN | target=RemoteTarget |
remote_fc_port=WWPN | remote_ipaddress=IPAddress | vol=VolName |
ipinterface=IPInterfaceName | local_fc_port=ComponentId ] < start=TimeStamp |
end=TimeStamp > [ module=ModuleNumber ] count=N interval=IntervalSize
resolution_unit=<minute|hour|day|week|month>
```

To further explain this command, assume that you want to collect 10 intervals, and each interval is for 1 minute. The point of interest occurred on 24 October 2017 roughly 5 minutes before 16:00:00.

Note: Use **statistics_get** to gather the performance data from any time period.

The time stamp is formatted as *YYYY-MM-DD.hh:mm:ss*, where the *YYYY* represents a four-digit year, *MM* is the two-digit month, and *DD* is the two-digit day. After the date portion of the time stamp is specified, you specify the time, where *hh* is the hour, *mm* is the minute, and *ss* represents the seconds. See Example 6-3.

Example 6-3 Using the statistics_get command

```
ITSO_2_A9000R>>statistics_get end=2017-10-24.16:00:00 count=10 interval=1
resolution_unit=minute
```

Figure 6-8 shows a sample output of the statistics. The output that is shown is a small portion of the data that was provided.

hput	Write Hit Medium - IOps	Write Hit Medium - Latency	Write Hit Medium - Throughp
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	262	1308	16854
	1758	1258	112683
	1695	1248	108612

Figure 6-8 Statistics output

7

Monitoring and troubleshooting

This chapter describes the various methods and functions that are available to monitor IBM FlashSystem A9000 and IBM FlashSystem A9000R systems. It shows how to gather information from the system in real time, in addition to the self-monitoring, self-healing, and automatic alert functions that are implemented with IBM Hyper-Scale Manager software.

This chapter includes the following sections:

- Monitoring
- Troubleshooting

7.1 Monitoring

This section describes the various methods and functions that are available to monitor FlashSystem A9000 and FlashSystem A9000R systems.

7.1.1 Monitoring events

FlashSystem A9000 and FlashSystem A9000R systems use a centralized event log. For any command that was executed that led to a change in the system, an event entry is generated and recorded in the event log. The object creation time and the user are also logged as object attributes.

The event log is implemented as a circular log. It can hold a set number of entries. When the log is full, the system wraps back to the beginning. To save the log entries beyond what the system normally holds, you can issue the **`event_list`** command-line interface (CLI) command and save the output to a file.

Event severity

The events are classified into levels of severity, depending on the event's effect on the system:

- *Informational* events are for information only without any impact or danger to system operation.
- *Warning* events are information for the user that something in the system was changed but the system was not affected.
- *Minor* events occur when a part failed but the system is still fully redundant and no operations were affected.
- *Major* events occur when a part failed and the redundancy is temporarily affected.
- *Critical* events occur when one or more parts fail and the redundancy and machine operation can be affected.

Event types

Table 7-1 lists event types that can be used as filters. They are specified with the **`object_type`** parameter in the CLI command.

Table 7-1 Event types

Event type	Description
cg	Consistency group
destgroup	Event destination group
dest	Event notification group
dm	Data migration
domain	Domain
host	Host
map	Volume mapping
mirror	Mirroring
pool	Pool

Event type	Description
rule	Rule
smsgw	Short Message Service (SMS) gateway
smtpgw	Simple Mail Transfer Protocol (SMTP) gateway
target	Fibre Channel or internet Small Computer System Interface (iSCSI) connection
volume	Volume mapping
cluster	Cluster
ipinterface	Internet Protocol (IP) interface
ldap	Lightweight Directory Access Protocol (LDAP) configuration
metadata	Metadata events
schedule	Schedules
user	User
user_group	User group
ldap	LDAP server
module	Modules status
ha	High Availability

Setting up notifications and rules by using the CLI

Configuration options are available with the CLI of FlashSystem A9000 and FlashSystem A9000R by giving the storage administrator flexibility to create a detailed events notification plan that is based on specific rules. The three-step process includes all of the required configurations to allow FlashSystem A9000 and FlashSystem A9000R to provide notification of events:

- Gateway
- Destination
- Rules

The gateway definition is used for Simple Mail Transfer Protocol (SMTP) and Short Message Service (SMS) messages. Various commands are used to create and manage the gateways for FlashSystem A9000 and FlashSystem A9000R. See Table 7-2 on page 145.

Example 7-1 shows the definition of an SMTP gateway. The gateway is named `test` and the messages from FlashSystem A9000 and FlashSystem A9000R are addressed `xiv@us.ibm.com`. When an SMTP gateway is added, the existing gateways are listed for confirmation. In addition to the gateway address and sender address, the port and reply-to address can also be specified.

Example 7-1 The smtpgw_define command

```
>>smtpgw_define smtpgw=test address=test.ibm.com from_address=a9000@us.ibm.com
Command executed successfully.

>>smtpgw_list
Name   Address        Priority
test   test.ibm.com   1
```

The SMS gateway is defined in a similar fashion. The difference is that the fields can use tokens to create variable text instead of static text. When you specify the address to send the SMS message, tokens can be used instead of hardcoded values. In addition, the message body also uses a token to send the error message instead of a hardcoded text.

Note: An SMTP gateway must be defined before you can define an SMS gateway as SMS messages are sent from FlashSystem A9000 and FlashSystem A9000R in an email.

The following tokens are available to be used for the SMS gateway definition. Example 7-2 provides an example of defining an SMS gateway.

- {*areacode*}: This escape sequence is replaced by the destination's mobile or cell phone number area code.
- {*number*}: This escape sequence is replaced by the destination's cell phone local number.
- {*message*}: This escape sequence is replaced by the text to show to the user.
- \{, \}, \\: These symbols are replaced by the {, } (for \{, \}) or \ (for \\).

Example 7-2 The smsgw_define command

```
>>smsgw_define smsgw=test email_address=1234567890@smstest.ibm.com
subject_line="A9000R System Event Notification" email_body="message here"
Command executed successfully.
>>smsgw_list
Name   Email Address                SMTP Gateways
test   1234567890@smstest.ibm.com   all
```

When the gateways are defined, the definition settings can be defined. Three types of destinations are available:

- SMTP or email
- SMS
- Simple Network Management Protocol (SNMP)

Example 7-3 provides an example of creating a destination for all three types of notifications. For the email notification, the destination receives a test message every Monday at 12:00. Each destination can be set to receive notifications on multiple days of the week at multiple times.

Example 7-3 Destination definitions

```
>> dest_define dest=emailtest type=EMAIL email_address=test@ibm.com smtpgws=ALL
heartbeat_test_hour=12:00 heartbeat_test_days=Mon
Command executed successfully.

>> dest_define dest=smstest type=SMS area_code=555 number=5555555 smsgws=ALL
Command executed successfully.

>> dest_define dest=snmptest type=SNMP snmp_manager=9.9.9.9
Command executed successfully.

>> dest_list
Name           Type   Email Address  Area Code Phone Number SNMP Manager
           User
ITSO_Catcher   SNMP
itsocatcher.us.ibm.com
smstest       SMS                  555      5555555
```

```
snmptest      SNMP                                    9.9.9.9
emailtest      EMAIL  test@ibm.com
```

Finally, the rules can be set for which messages can be sent. Example 7-4 provides two examples of setting up rules. The first rule is for SNMP and email messages, and all messages, even informational messages, are sent to the processing servers. The second example creates a rule for SMS messages. Only critical messages are sent to the SMS server, and they are sent every 15 minutes until the error condition is cleared.

Example 7-4 Rule definitions

```
>> rule_create rule=emailtest min_severity=informational dests=emailtest,snmptest
Command executed successfully.

>>rule_create rule=smstest min_severity=critical dests=smstest snooze_time=15
Command executed successfully.

>> rule_list
Name         Minimum Severity   Event Codes   Except Codes   Destinations         Active   Escalation Only
ITSO_Major   Major              all                          ITSO_Catcher         yes      no
emailtest    Informational      all                          emailtest,snmptest   yes      no
smstest      Critical           all                          smstest              yes      no
```

Example 7-5 shows how to delete rules, destinations, and gateways.

Note: It is not possible to delete a destination if a rule is using that destination. And, it is not possible to delete a gateway if a destination is pointing to that gateway.

Example 7-5 Deletion of notification setup

```
>> rule_delete -y rule=smstest
Command executed successfully.
>> dest_delete -y dest=smstest
Command executed successfully.
>> smsgw_delete -y smsgw=test
Command executed successfully.
```

Event-related CLI commands

Table 7-2 lists all of the event-related commands that are available in the CLI. This list includes setting up notifications and viewing the events in the system.

Table 7-2 CLI: Event-related commands

Command	Description
custom_event	Generates a custom event
dest_event	Defines a new destination for event notifications
dest_delete	Deletes an event notification destination
dest_list	Lists event notification destinations
dest_rename	Renames an event notification destination
dest_test	Sends a test message to an event notification destination
dest_update	Updates an event notification destination

Command	Description
destgroup_add_dest	Adds an event notification destination to a destination group
destgroup_create	Creates an event notification destination group
destgroup_delete	Deletes an event notification destination group
destgroup_list	Lists event notification destination groups
destgroup_remove_dest	Removes an event notification destination from a destination group
destgroup_rename	Renames an event notification destination group
destgroup_update	Updates an event notification destination group
event_clear	Clears alerting events
event_list	Lists system events
event_list_uncleared	Lists uncleared alerting events
event_redefine_threshold	Redefines the threshold of a parameterized event
event_threshold_list	Lists event thresholds
smsgw_define	Defines a Short Message Service (SMS) gateway
smsgw_delete	Deletes an SMS gateway
smsgw_list	Lists SMS gateways
smsgw_prioritize	Sets the priorities of the SMS gateways for sending SMS messages
smsgw_rename	Renames an SMS gateway
smsgw_update	Updates an SMS gateway
smtpgw_define	Defines an SMTP gateway
smtpgw_delete	Deletes a specified SMTP gateway
smtpgw_list	Lists SMTP gateways
smtpgw_prioritize	Sets the priority of the SMTP gateway to use to send emails
smtpgw_rename	Renames an SMTP gateway
smtpgw_update	Updates the configuration of an SMTP gateway
rule_activate	Activates an event notification rule
rule_create	Creates an event notification rule
rule_deactivate	Deactivates an event notification rule
rule_delete	Deletes an event notification rule
rule_list	Lists event notification rules
rule_rename	Renames an event notification rule
rule_update	Updates an event notification rule
domain_list	Lists all domains
domain_list_users	Lists users in the system per domain
domain_list_objects	Lists objects in the system per domain

7.1.2 Monitoring alerts and events

FlashSystem A9000 and FlashSystem A9000R allow users to monitor alerts in the GUI and monitor all events in the CLI.

Monitoring alerts by using the GUI

Clicking the **Alerts** icon in the upper-right corner of the GUI will display alerts that are currently reported from the systems in the inventory. Clicking an alert in the list displays corresponding information that is associated with the alert. To see a full list of system events, see “Monitoring alerts and events by using the CLI”.

Monitoring alerts and events by using the CLI

To monitor alerts and events by using the CLI, run the **event_list** command. The following syntax is for the **event_list** command, as shown in Example 7-6.

Example 7-6 Syntax for the event_list command

```
event_list
[ max_events=MaxEventsToList ] [ after=TimeStamp ] [ before=TimeStamp ]
[ min_severity=<INFORMATIONAL|WARNING|MINOR|MAJOR|CRITICAL> ]
[ alerting=<yes|no|all> ]
[ cleared=<yes|no|all> ]
[code=EventCode ]
[object_type=<cons_group|destgroup|dest|dm|host|map|mirror|pool|rule|smsgw|smtpgw|
target|volume|cluster|
ip_interface|ldap_conf|meta_data_object|sync_schedule|user|user_group|ldap_server|
modules_status|elicense|
ipsec_connection|ipsec_tunnel|cross_cons_group,... ]
[ internal=<yes|no|all> ]
[ beg=BeginIndex ]
[ end=EndIndex ]
[ count_all=<yes|no> ]
```

Events are always listed and sorted chronologically with the most recent events displayed last. Events are printed by default in their user-readable textual form. Alternatively, the CLI option for comma-separated values can be used to create an output file that serves as an input file for other applications. The output will always be all events that match all criteria, up to the size that is specified in **max_events**, which limits the size of the output (default is 300 events). See Example 7-7.

Example 7-7 Output example from the event_list command

```
>>event_list
Timestamp             Severity        Code
User            Description
2017-10-24 12:08:31  Informational  MODULE_BBU_IS_CHARGING
1:BBU:1:1 changed state from 'Not charging' to 'Charging'.
2017-10-24 12:21:27   Informational   MODULE_BBU_IS_FULL2016-01-29 10:12:00
Informational   USER_DEFINED                                  itso_stgadmin   A
user with name 'itso_secadmin' and category securityadmin was defined.
...
2017-10-26 09:15:01   Informational   HOST_DEFINE
admin             Host of type default was defined with name
'p7-770-04v7.mainz.de.ibm.com'.
```

```
2017-10-26 09:15:01   Informational   HOST_ADD_PORT
admin            Port of type FC and ID '10000090FA263482' was added to Host with
name 'p7-770-04v7.mainz.de.ibm.com'.
2017-10-26 09:15:01   Informational   HOST_ADD_PORT
admin            Port of type FC and ID '10000090FA263483' was added to Host with
name 'p7-770-04v7.mainz.de.ibm.com'.
...
```

7.1.3 Monitoring statistics

FlashSystem A9000 and FlashSystem A9000R statistics can be monitored by using the GUI or the CLI.

Monitoring statistics by using the GUI

The GUI provides the storage administrator with an option to monitor real-time and historical statistics, depending on the filter that is selected. Use the following GUI elements to access the system statistics:

- Dashboard view
- Workspace view
- Side menu

For additional information about monitoring statistics by using the GUI, see *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376.

Monitoring statistics by using the CLI

Use the following CLI commands to obtain system performance and usage history statistics:

- **statistics_get**: Retrieves performance statistics from FlashSystem A9000 and FlashSystem A9000R. The following syntax is for the **statistics_get** command:

  ```
  statistics_get [ perf_class=perfClassName | host=HostName |
  host_iscsi_name=initiatorName |
  host_fc_port=WWPN | target=RemoteTarget | remote_fc_port=WWPN |
  remote_ipaddress=IPAddress |
  vol=VolName | domain=DomainName | ipinterface=IPInterfaceName |
  local_fc_port=ComponentId ]
  < start=TimeStamp | end=TimeStamp > [ module=ModuleNumber ]
  count=N interval=IntervalSize resolution_unit=<minute|hour|day|week|month>
  ```

- **usage_get**: Shows the usage history of a volume or a storage pool. The following syntax is for the **usage_get** command:

  ```
  usage_get <vol=VolName | pool=PoolName> [ start=TimeStamp |
  start_in_seconds=StartTime ]
  [ end=TimeStamp ] [ max=MaxEntries ]
  ```

7.1.4 Monitoring hardware

FlashSystem A9000 and FlashSystem A9000R hardware can be monitored by using both the GUI and the CLI.

Monitoring hardware by using the GUI

The GUI displays hardware information in graphical and textual formats. Hardware information can be accessed by using one of the following GUI elements:

- Side menu: **Systems & Domains Views**
- Table view: **Hardware view**

Monitoring hardware by using the CLI

Use the following commands to monitor the hardware with the CLI:

- **`vault_device_list`**: Displays a list of special vault devices and statuses. *Vault devices* are solid-state drives (SSDs) that are used to store cache and metadata. In FlashSystem A9000 and FlashSystem A9000R Model Type 415, each grid controller contains two 250 GB enterprise-grade SSDs. In FlashSystem A9000 and FlashSystemA9000R, each grid controller contains two 400GB enterprise-grade SSDs. In addition, vault devices are used to save metadata and system configuration information regularly. For more information, see “Solid-state drive (SSD) vault devices” on page 48.
- **`flash_enclosure_list`**: Displays a list of all attached flash enclosures and their current status.
- **`flash_canister_list`**: Displays a list of canisters in attached flash enclosures and their current status.
- **`flash_control_connectivity_list`**: Displays the connectivity and its current status between the grid controllers and connected flash enclosures.
- **`flash_card_list`**: Displays a list of flash cards in the attached flash enclosures and their current status.
- **`flash_bbu_list`**: Displays a list of all flash enclosure battery backup units and their current status.
- **`flash_led_card_list`**: Displays a list of flash enclosure LED cards and their status.
- **`flash_psu_list`**: Displays a list of flash enclosure power supply units and their current status.
- **`flash_ib_adapter_list`**: Displays a list of flash enclosure InfiniBand adapters and their current status.
- **`flash_fan_list`**: Displays a list of flash enclosure fans and their current status.
- **`system_average_power_consumption`**: Displays the average system power consumption.
- **`system_average_temperature`**: Displays the average temperature of the system.
- **`component_list`**: Displays a list of all system components and their current status.
- **`service_list`**: Displays a list of services and their current status.
- **`module_list`**: Displays a list of grid controllers and their current status.
- **`component_service_required_list`**: Displays a list of system components that require service.
- **`system_power_management_get`**: Displays information about system power management, such as system run time, power consumption, and power supply units (PSUs) with no input power.

Tip: Use the **filter=notok** parameter with the CLI commands listed above to decrease the output and show only the components that are not in a good state and that need attention.

7.1.5 Simple Network Management Protocol monitoring

FlashSystem A9000 and FlashSystem A9000R systems support Simple Network Management Protocol (SNMP) for monitoring.

SNMP is an industry-standard set of functions for monitoring and managing TCP/IP-based networks and systems. SNMP includes a protocol, a database specification, and a set of data objects. A set of data objects forms a Management Information Base (MIB).

The SNMP protocol defines two terms, *agent* and *manager*, instead of the client and server terms that are used in many other TCP/IP protocols. An SNMP agent is implemented in FlashSystem A9000 and FlashSystem A9000R. The SNMP agent sends SNMP traps to an SNMP manager (such as IBM Systems Director) to indicate that an event occurred. By default, the trap is sent to User Datagram Protocol (UDP) port 162. The SNMP manager can also request certain information from FlashSystem A9000 and FlashSystem A9000R by using the SNMP **get** or **walk** command. These commands are sent to FlashSystem A9000 and FlashSystem A9000R on UDP port 161.

Management Information Base

The information that you can receive from FlashSystem A9000 and FlashSystem A9000R by sending an SNMP **get** request, or the contents of an SNMP trap that is sent from FlashSystem A9000 and FlashSystem A9000R, are defined in a MIB. The structure of a MIB is defined as an internet standard in Request for Comments (RFC) 1155. The MIB forms a tree structure that consists of object identifiers (OIDs).

Most hardware and software vendors provide you with extended MIB objects to support their own requirements. The SNMP standards allow this extension by using the private subtree, which is called an *enterprise-specific MIB*. Because each vendor has a unique MIB subtree under the private subtree, no conflict occurs among the vendors' original MIB extensions.

You can use the **mib_get** command from the CLI.

Monitoring System Events is performed using SNMP traps only.

7.1.6 Call Home

The *Call Home* function allows FlashSystem A9000 and FlashSystem A9000R to send event notifications to the IBM Remote Support Center. This function allows both proactive and failure notifications to be sent directly to IBM for analysis.

The IBM Remote Support Center takes corrective actions, up to dispatching an IBM service support representative (SSR) with a replacement part or engaging higher level of support to ensure complete problem determination and a solution.

Important: The configuration of Call Home and remote support facilities is highly advised to assist with failure detection, diagnosis, and resolution.

Call Home is always configured to use SMTP. It is configured by qualified IBM SSRs only, typically when FlashSystem A9000 or FlashSystem A9000R is first installed. If an event is

received by the IBM Remote Support Center that requires service or investigation, it will trigger an IBM problem management record (PMR).

As the Call Home feature uses customer's network and SMTP service, IBM cannot guarantee the delivery of events. Therefore, events must be monitored by the customer, as described in "Setting up notifications and rules by using the CLI" on page 143.

When the Call Home feature is configured, events are received by the IBM Remote Support Center. Periodic heartbeat events are also received. The heartbeats are monitored by IBM and the customer is notified if the heartbeats are no longer received.

The SMTP address for Call Home is configured separately from the general FlashSystem A9000 and FlashSystem A9000R SMTP setting. If the customer mail server gateway changes, a service call must be logged so that the internal Call Home SMTP setting is changed as well to reflect the new information.

Tip: Email-relaying on the SMTP gateway server needs to be enabled to allow Call Home events to be sent to IBM.

7.1.7 Encrypting Call Home and heartbeat notifications

As just explained, FlashSystem A9000 and A9000R use email to send Call Home and heartbeat notifications to IBM. These emails are sent using standard SMTP protocol through a customer provided email gateway server.

There might be customer requirements to send call home information to IBM only via a secured channel.

It is possible to configure the customer email gateway to send the emails from the customer mail server to the IBM mail server via a TLS encrypted path and the IBM receiving mail server is capable of accepting mail transfer via SMTP protocol over TLS.

Configuration example for a Postfix SMTP server

Example 7-8 shows a sample configuration for a Postfix SMTP server. The configuration may vary for each customer environment and gateway.

Note: Mails from the storage system to the mail gateway are still being sent in plain text. Only the path on the public internet between the customer mail gateway and IBM receiving mail server is TLS encrypted.

Example 7-8 Configuring Postfix SMTP server for TLS

```
smtpd_use_tls = yes
smtp_tls_security_level = encrypt
smtpd_tls_loglevel = 2
smtpd_tls_key_file = /etc/postfix/smtpd.key
smtpd_tls_cert_file = /etc/postfix/smtpd.cert
relayhost = vnet.ibm.com
inet_protocols = all
```

The smtpd.key and smtpd.cert key and certificate files need to be generated. You can use the OpenSSL tool for that purpose:

```
openssl req -new -outform PEM -out smtpd.cert -newkey rsa:2048 -nodes -keyout
smtpd.key -keyform PEM -days 365 -x509
```

7.1.8 Subscribe to support notifications

Stay informed of critical IBM software support updates by signing up for *My notifications*. By signing up for My notifications, you will receive the following benefits:

- A proactive approach to problem prevention.
- Receive support content that is tailored to your needs and delivered directly to you.
- Receive immediate notifications of Security Bulletins and Flashes.
- Receive daily or weekly notifications of technical support information, such as downloads, tips, technical notes, and publications.

Signing up is fast and easy with a simple subscription interface. You can register on the following site:

https://ibm.biz/Bd4WSs

After you log in, you can select the IBM products which you are interested in, choose between daily and weekly notifications, and select a delivery method. The delivery methods are email and RSS news feed.

7.2 Troubleshooting

This section explains what steps to take in case you encounter an issue with FlashSystem A9000 and FlashSystem A9000R, in order to receive the best service in a timely manner. Both FlashSystem A9000 and FlashSystem A9000R are serviceable products. All diagnostics, repairs and upgrades are performed by IBM SSRs and/or IBM Support Representatives. It is important for them to have key information provided by you that will help them with problem determination and resolution.

When an issue occurs, it is important to know the type of support you are entitled to. Be familiar with your IBM warranty, maintenance agreements, and the Customer Support Plan (CSP) that is provided during the installation planning phase and Technical Delivery Assessment of your implementation. The CSP document will provide the information that you need to start the support process.

Important: To avoid encountering known and fixed issues, keep your software up to date to take advantage of all of the latest software fixes and enhancements.

Develop a disaster recovery plan and back up your configuration settings regularly.

7.2.1 IBM Support process

This section provides information about how to interact with IBM Support if an issue arises.

Calling IBM Support

Consider this information when you call IBM support:

- For problems that are known to be hardware-related, place calls against the affected FlashSystem A9000 or FlashSystem A9000R machine type and serial number. Using the correct machine type, model, and serial number, avoids service delays.
- For software problems, navigate through the automated voice response for software and provide your IBM customer number, product description, plus machine type and serial number.
- If you are unsure whether the issue is hardware or software related, call Storage Support (option 3 is US only). Provide your IBM customer number, product description, plus machine type and serial number.

Before you contact IBM Support

To resolve your support service request in the fastest way possible, take the following steps before you contact the IBM Support Center. You must gather information about the problem, and have it available when you describe the situation to the IBM Support specialist.

The following steps are an example of the type of information that is required:

- Define the problem

  Describing the issue and the symptoms that you are experiencing before you contact support, can expedite the problem-solving process. Be as specific as possible when you explain an issue or a question to our IBM Support specialists.

- Gather relevant diagnostic information

  In many instances, our technical support specialists need to analyze specific diagnostic data, such as relevant system logs, storage dumps and traces, to address your problem. Product-specific diagnostic documentation can be extremely helpful to gather information that is typically required to resolve specific issues.

  If you are unsure about what type of documentation is required, you can always contact IBM Support for assistance in gathering the needed diagnostic information.

- Determine the severity level

  You have to assign a severity level to the problem when you initially report it. After a problem record is created, the severity level can be discussed with the support analyst, based on the business impact caused by the specific problem. If you designated a problem as a severity 1, IBM will work on it 7 days a week, 24 hours a day, providing you are also available to work during those hours. You can change the severity level of a problem if circumstances change from when it was first entered to match current business impact conditions.

When you communicate with a technical support specialist, you must have the following information available as well:

- IBM customer number
- Company name
- Machine type/model/serial number (needed for IBM to see your entitlements for support)
- Primary contact name and secondary contact name
- Preferred means of contact (electronic or voice)
- Alternate ways to reach you, such as multiple phone numbers, pager, or email address
- Your availability (that is, when you can work with IBM Service and Support)
- Related storage area network (SAN) fabric information
- Related operating system information
- Other problems (PMRs/incidents) with IBM that are currently active and related to your service request

You need to research this situation before you contact IBM Support so that you can provide the detailed information and documentation to help identify and solve the problem.

Tip: Machine type, model, and serial number can be found on the back and front of the rack doors (A9000R) and on the front of the pod (A9000) bezel.

Support call flow

When a support call is initiated, either by you or through the automated Call Home process, a problem management record (PMR) is opened to track the call. The call is routed to the correct group to address the specific issue, as depicted in Figure 7-1.

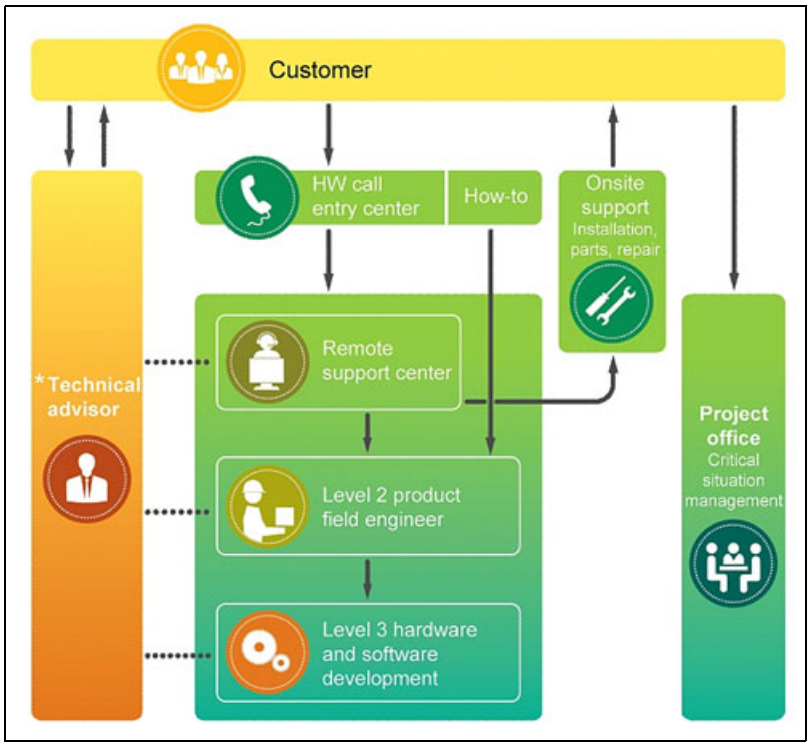


Figure 7-1 IBM Support process flow

IBM Storage Technical Advisor

The IBM Technical Advisor for Storage enhances the end-to-end support for complex IT solutions. Each FlashSystem A9000 or FlashSystem A9000R purchased with a three-year contract includes Storage Technical Advisor (TA) support for the initial hardware warranty period. TA support can also be purchased separately.

This section describes the IBM TA program in general with specifics about how customers can use their TA.

The TA service is built around three value propositions (benefits):

- Proactive approach to ensure high availability for vital IT services
- A Client Advocate that manages the problem resolution through the entire support process
- A trusted consultant for both storage hardware and software

The IBM Technical Advisor for Storage benefits customers by providing a consultant for questions and requests concerning FlashSystem A9000 and FlashSystem A9000R. Most

customers are introduced to their Technical Advisor during a Technical Delivery Assessment (Solution Assurance Meeting) before the initial installation.

After the initial meeting, the TA is considered focal point for support-related activities by performing the following services:

- Maintaining a customer support plan that is specific to each customer. This support plan contains an inventory of equipment, including customer numbers and serial numbers.
- Coordinating service activities by working with your support team in the background. Monitoring the progress of open service requests, escalation, and expert consultation about problem avoidance.
- Communicating issues with IBM customers, IBM Business Partners, and IBM Sales teams.
- Performing periodic reviews and reports about hardware inventories and service requests, including the use of Call Home information to provide customer reports about the status of their systems.
- Managing IBM Support activities to help companies anticipate and respond to new problems and challenges faster.
- Offering proactive planning, advice, and guidance to improve the availability and reliability of your systems.

The IBM Storage Technical Advisor is an effective way to improve your total cost of ownership and to free up customer resources. Customers can optionally extend the Technical Advisor service beyond the initial hardware warranty by using the IBM Technical Support Services (TSS) offerings. Contact your IBM Sales Team or IBM Business Partner for details.

7.2.2 Collecting logs

System support logs help IBM storage support diagnose problems that can arise on FlashSystem A9000 and FlashSystem A9000R. Several established processes can greatly help in problem determination. They allow many events to be cataloged and stored for reference. The log collection and upload can be accomplished by using the GUI.

The following prerequisites are necessary to successfully collect and send FlashSystem A9000 and FlashSystem A9000R system logs:

- A user that is assigned to the Storage Administrator or Operation Administrator category on FlashSystem A9000 and FlashSystem A9000R
- Access to FlashSystem A9000 and FlashSystem A9000R by using the CLI or the GUI

Collecting support logs by using the GUI

Complete the following steps to collect the support logs by using the GUI:

1. Log in to the GUI.
2. In the side menu, click **Systems & Domains Views** to display a list of filters.
3. Click **Systems** to display a list of the systems that are in the inventory.
4. Select the system which you want to collect a log from.

5. Click **Actions** →**Support** →**Collect System Logs** to display the System support section. See Figure 7-2.

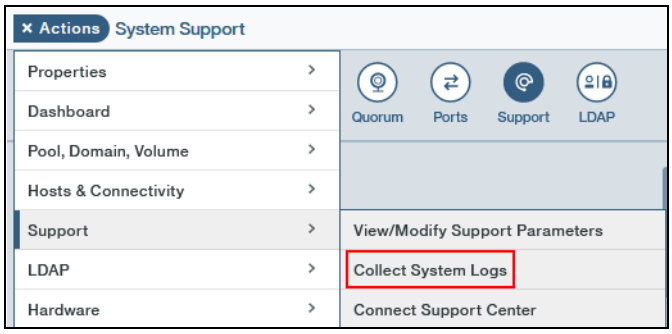


Figure 7-2 Collect System Logs option in the Actions Support menu

6. From the Support Logs section, click **Collect Logs** to download the support logs. See Figure 7-3.

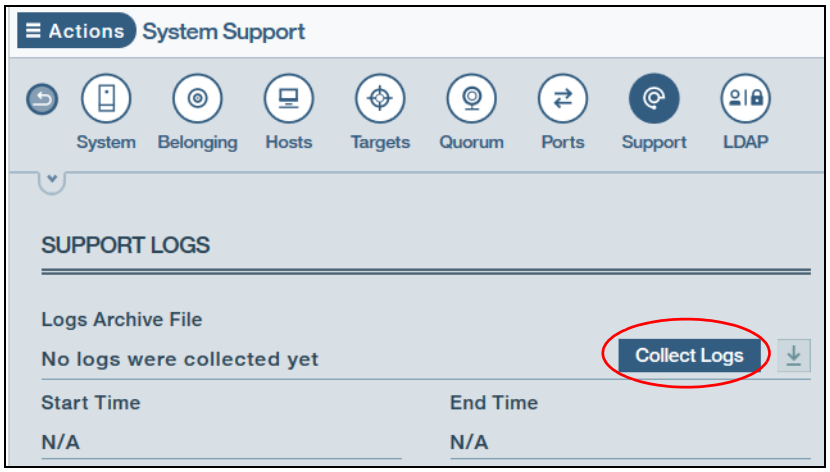


Figure 7-3 Collect Logs option

7. When the download of the log is complete, click the log archive file **Download** icon to transfer the log file to your computer. See Figure 7-4.

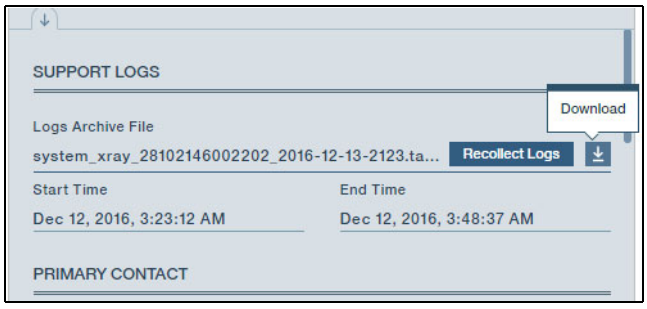


Figure 7-4 Log Download icon

Uploading logs to IBM ECuRep

IBM Enhanced Customer Data Repository (ECuRep) is a secure and fully supported data repository with problem determination tools and functions. It updates problem management records (PMRs) and maintains full data lifecycle management.

This server-based solution is used to exchange data between IBM customers and IBM Technical Support. Do not place files on this server or download files from this server without authorization *in advance* from an IBM representative. The IBM representative can provide further instructions, as needed.

To use ECuRep, you need a documented PMR number that is either provided by the IBM Support team with a Call Home or it is issued by using the IBM Service Request tool on the IBM Support portal. IBM provides the Service Request (SR) problem submission tool to electronically submit and manage service requests on the web. This tool replaces the Electronic Service Request (ESR) tool:

`http://www.ecurep.ibm.com/`

To provide logs to IBM ECuRep, complete the following steps:

1. The ECuRep opening page is shown in Figure 7-5. This page provides information about the repository, instructions for preparing files for upload, and multiple alternatives for sending data. Click **Send data**.

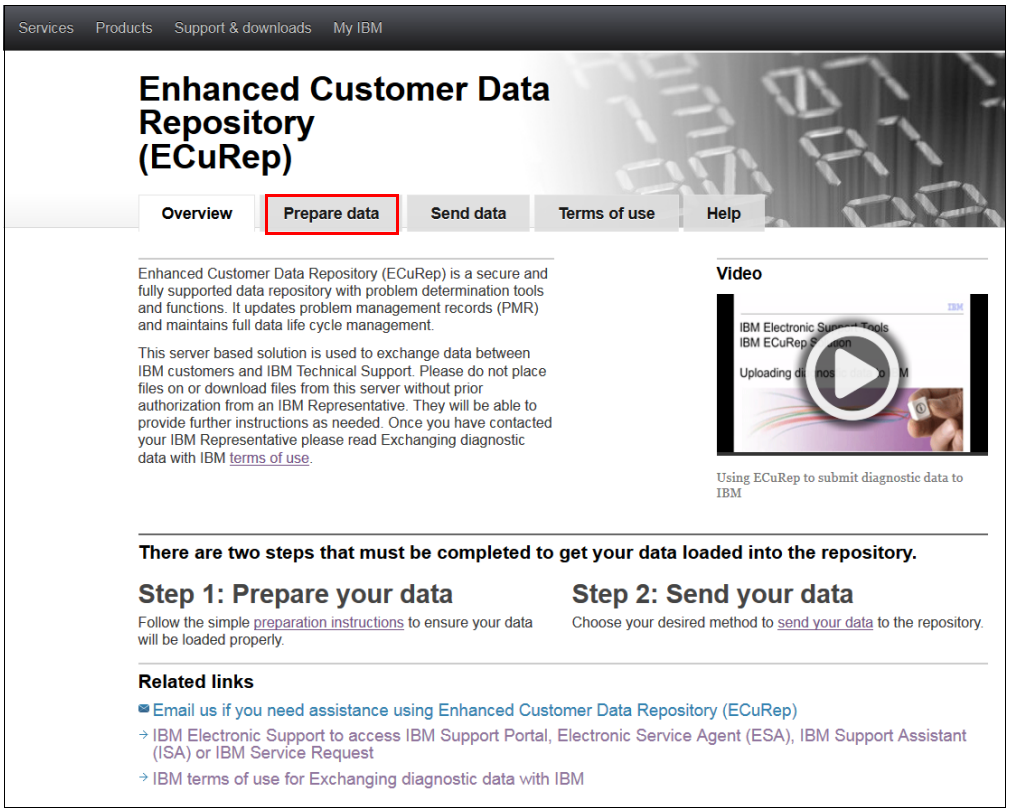


Figure 7-5 Main page of ECuRep

Tip: The system that you are uploading logs from is connected to an IBM Problem Management Record. Support tickets are automatically updated when the files are uploaded for IBM Support review.

2. IBM provides multiple options for uploading data (Figure 7-6). The Java utility is the most efficient method to upload a file. When you use FTP (number 1 in the figure) or the Java utility (2), also select **Prepare data** (3) to see the details about file-naming conventions. The HTTPS upload option (4) eliminates the file-naming requirement.

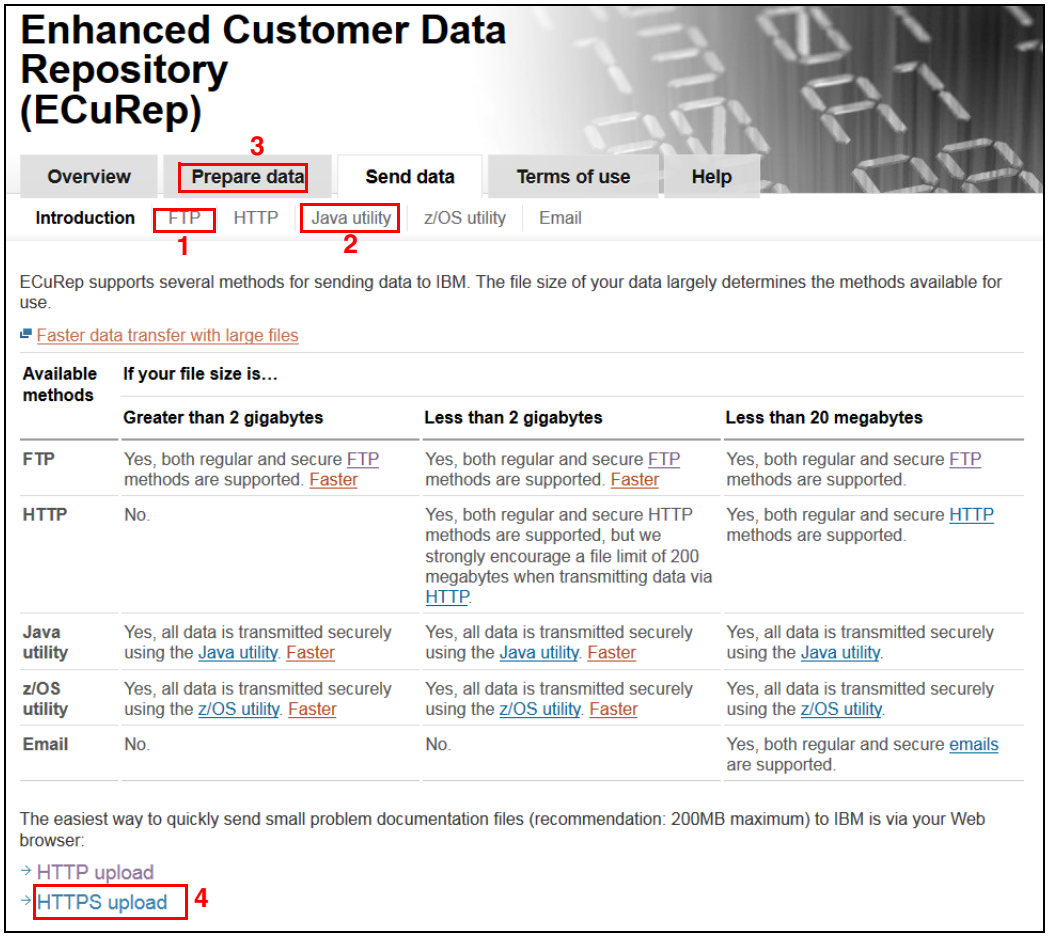


Figure 7-6 Options for sending data

3. The file naming requirement is shown in Figure 7-7. Put your PMR number on the relevant field in this form to accurately log the files that you upload to the correct problem record. For the “Upload is for” field, select **Hardware** from the drop down menu. Optionally, provide your email address for a confirmation. Click **Continue**.

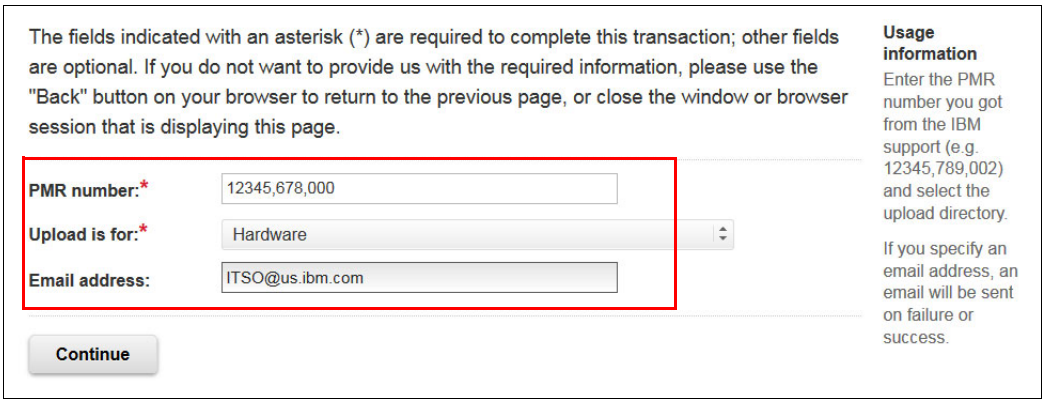


Figure 7-7 Use the PMR number with the HTTP option

4. The file selection panel opens, and you can select the files to upload (Figure 7-8).

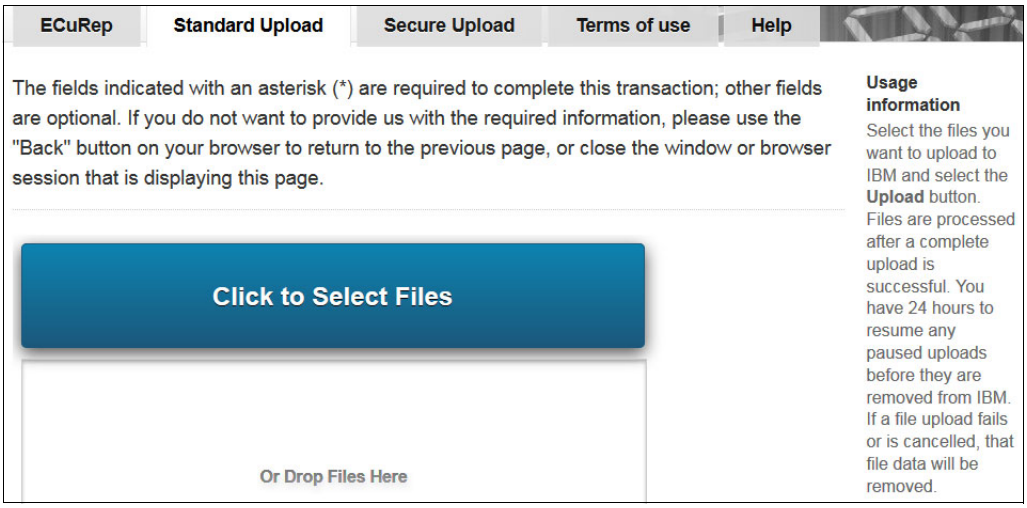


Figure 7-8 File selection panel

Tip: This, for most customers, is the most effective method for system logs upload. IBM suggests that you understand the best method for your organization *in advance* and *document* the process to save time during a crisis.

Uploading Logs to Blue Diamond Lab

The Blue Diamond Lab is a portal where you can upload your diagnostic data, which provides extra layers of security. If you are a Blue Diamond customer, IBM Remote Support has no authorisation to collect any sort of logs from your system. The log files can only be uploaded from customer's end to the secure Blue Diamond FTP server for IBM Remote Support review. To upload data to the secure FTP server, you need an active PMR number.

Follow the guide steps listed below in order to upload support logs to the BlueDiamond FTP:

1. Log into the Blue Diamond Secure FTP Portal at `https://msciportal.im-ies.ibm.com/Portal.aspx`, as shown in Figure 7-9.

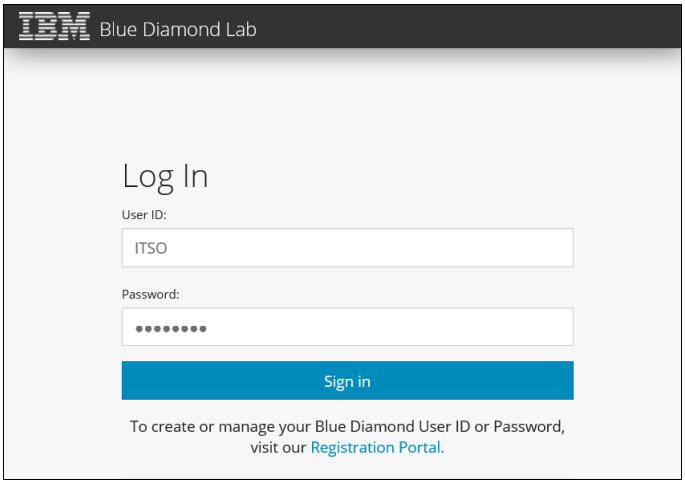


Figure 7-9 Blue Diamond Lab home page

2. Log into our secure FTP server, using your Blue Diamond credentials, as shown in Figure 7-10.

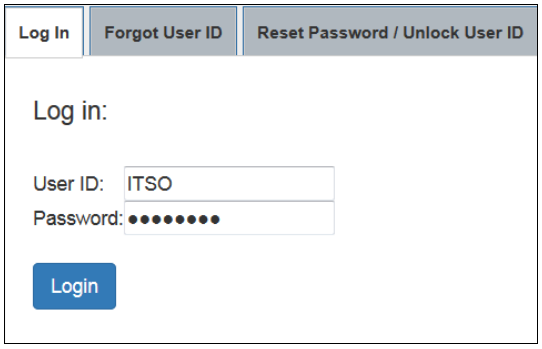


Figure 7-10 Blue Diamond FTP Log in page

3. When you are logged in, click the **Upload** button, as shown in Figure 7-11.

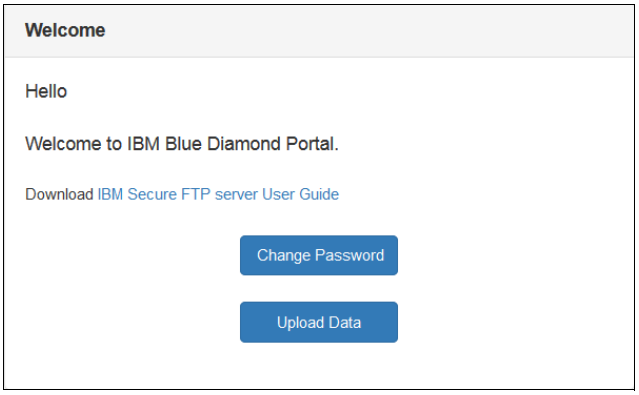


Figure 7-11 Blue Diamond Data Upload Page

4. Navigate to the folder that contains your company name (IBM/YourCompanyName) See (Figure 7-12).

Note: You are only permitted to see your own company folder.

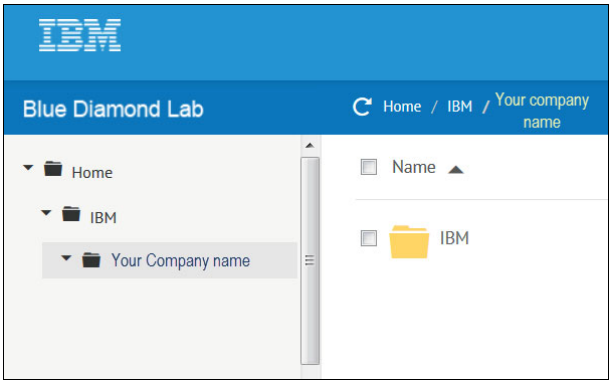


Figure 7-12 Navigate your company folder

5. Create a sub-folder with the name of your PMR number (Example, 12345,678,000) (Figure 7-13).

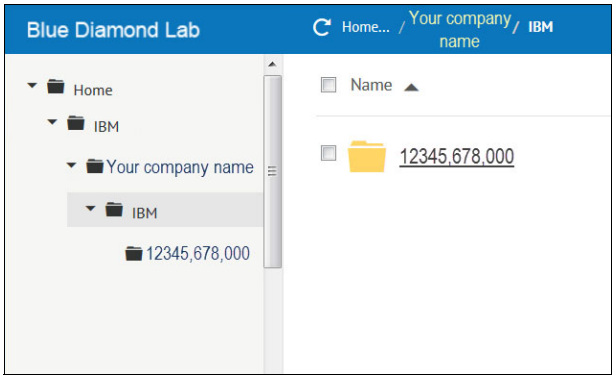


Figure 7-13 Sub-folder with PMR number

6. Upload the diagnostic data to the sub-folder that you have created.

   When the file is successfully uploaded to the Blue Diamond Lab, IBM Remote Support will receive a notification through your active PMR and can then review the logs.

7.2.3 IBM Remote Support

FlashSystem A9000 and FlashSystem A9000R are maintained and repaired by trained IBM service personnel, either remotely with the help of the remote support connection (RSC) or onsite by an IBM service support representative (SSR). By using the RSC, the customer initiates a secure connection from FlashSystem A9000 or FlashSystem A9000R to IBM when problems arise. An IBM remote support specialist can then connect to the system to analyze the problem, repair it remotely if possible, or assist an IBM SSR who is onsite.

Important: IBM encourages all customers to use the secure, high-speed remote support solution that is enabled by the RSC. Problem analysis and repair actions without a remote connection can be complicated and time-consuming.

If problems occur, the remote support specialist can analyze the problems and also assist an IBM SSR that is dispatched onsite to repair the system or replace field-replaceable units (FRUs).

To enable remote support, you must allow an internet connection through your firewall that allows IBM to use a Secure Shell (SSH) connection to your FlashSystem A9000 and FlashSystem A9000R.

Remote support connection (RSC)

The RSC uses a high-speed internet connection, but it gives the customer the ability to initiate an outbound SSH call to a secure IBM server.

Firewall rules might need to be configured at the customer's firewall to allow FlashSystem A9000 and FlashSystem A9000R virtual private network (VPN)/Management ports to connect to the RSC.

Tip: The type of access that is required for a remote support connection is `outbound port 22/ssh` from FlashSystem A9000 and FlashSystem A9000R network ports.

The RSC consists of FlashSystem A9000 and FlashSystem A9000R internal functions with a set of globally deployed supporting servers. Together, they provide secure IBM Support access to FlashSystem A9000 and FlashSystem A9000R when necessary and when authorized by the customer's personnel.

Objectives and components

The RSC meets three major objectives:

- Security from a malicious attack

  The RSC was designed with security as a major concern, but the system architecture is simple and easy to deploy. It relies on standard, proven technologies and minimizes the logic (code) that must be located either on the external RSC server or on customer machines. The architecture of the RSC anticipates and handles various potential attack vectors, including encryption protocol and authentication logic. The design of the RSC identifies, analyzes, and defends against these attack vectors and other possible attack routes.

- Simplicity of implementation and design

  Simplicity enhances security, facilitates implementation at the customer's site, decreases staff training costs, and ultimately lowers the total cost of ownership (TCO) by allowing more issues to be resolved quickly and with fewer unnecessary visits by support personnel to the customer's premises.

- Accessibility by support personnel

  The auditing features of the RSC both enhance security and enable easy reassignment of a support case among support personnel. The RSC can expedite the escalation of support cases to higher-level support, and it allows IBM SSRs to pool knowledge to resolve complex customer issues.

Underlying architecture

The remote support mechanism has four major components:

- Remote Support Client (machine internal)

  The Remote Support Client is a software component inside FlashSystem A9000 and FlashSystem A9000R that handles remote support connectivity. It relies only on a single outgoing Transmission Control Protocol (TCP) connection, and it cannot receive inbound connections of any kind. The Remote Support Client is controlled by using CLI and the Remote Support Connection is used for starting a connection, terminating a connection (because of timeout or customer request), and trying the connection again in case the connection terminates unexpectedly.

- Optional Remote Support Proxy

  The Remote Support Client can access the Remote Support Center Front Server directly, or through an optional proxy server. The optional Remote Support Proxy can be used when one or more FlashSystem A9000 and FlashSystem A9000R systems do not have direct access to the internet (for example, because of firewall restrictions).

  You can use the Remote Support Proxy to facilitate the connection to the RSC.

- Remote Support Center Front Server (internet)

  Front Servers are on an IBM DMZ of the internet and receive connections from the Remote Support Client and the IBM Remote Support Back Server. Front Servers are security-hardened machines that provide a minimal set of services, such as maintaining connectivity to connected Clients and to the Back Server. They are strictly inbound, and never initiate anything on their own accord.

No sensitive information is ever stored on the Front Server, and all data that passes through the Front Server from the Client to the Back Server is encrypted so that the Front Server cannot access this data.

- Remote Support Center Back Server (IBM intranet)

  The Back Server manages most of the logic of the system. It is located within the IBM intranet. The Back Server is access-controlled. Only IBM employees that are authorized to perform remote support of FlashSystem A9000 and FlashSystem A9000R are allowed to use it, and only through specific support interfaces, not with a CLI or a GUI shell.

  The Back Server is in charge of authenticating a support person. It provides the support person with a user interface (UI) through which to choose a system to support based on the support person's permissions.

  It also provides the list of systems that are currently connected to the Front Servers, and it manages the remote support session as it progresses (logging it, allowing additional support persons to join the session, and so on). The Back Server maintains connection to all Front Servers. Support people connect to the Back Server by using any SSH client or an HTTPS connection with any browser.

Follow these steps to initiate the remote connection process:

1. The customer initiates an internet-based SSH connection to the RSC through the CLI of FlashSystem A9000 and FlashSystem A9000R.
2. The RSC identifies FlashSystem A9000 or FlashSystem A9000R and marks it as "connected".
3. The Support person connects to the RSC by using SSH over the IBM intranet.
4. The RSC authenticates the Support person against the IBM intranet.
5. The RSC then shows the connected customer systems that are available to the Support person.
6. The IBM Support person then chooses the system to support and to which system to connect:
   - Only permitted FlashSystem A9000 and FlashSystem A9000R systems are shown.
   - IBM Support personnel log their intended activity.
7. A fully recorded support session begins.
8. When the session is complete, the support person terminates the session and the RSC disconnects FlashSystem A9000 and FlashSystem A9000R from the remote support system.

Customer use scenario

After a customer decides that remote support is needed, possibly because of a telephone request from a remote support professional, communication between the RSC and the customer site must be established. This process is always initiated by the customer and always remains under the customer's control. It proceeds in the following manner:

- The customer instructs FlashSystem A9000 or FlashSystem A9000R to connect to the RSC by using a specific CLI command. Until this command is issued, any FlashSystem A9000 and FlashSystem A9000R will not connect to the RSC for any reason.
- The customer can set a regular timeout and an "idle timeout" for the connection by using an optional parameter in the connection command. A *timeout* causes the connection to be dropped after a specified length of time whether a support session is in progress or not. An *idle timeout* causes the connection to be dropped after a certain period of inactivity. By specifying the keyword **never** for these two parameters, the connection remains established indefinitely.

- The customer can also choose to set a unique password for a session when the session is initiated. This password must be provided to the IBM Support representative for them to complete the authentication process and access FlashSystem A9000 and FlashSystem A9000R.
- In addition, the customer can see the RSC connectivity status at any time (disconnected, idle, or in-session) and can choose to forcibly disconnect the connection regardless of whether a support session is in progress or not.

Although forcibly terminating a support session by using the timeout or force-disconnect mechanisms is not good, these mechanisms exist and the customer has full control over whether to proceed with a support session.

The initiation or loss of connection to the RSC causes the system to generate events that can be seen in the machine's event log. These events can be forwarded to any destination of the customer's choice (just as any other event that uses the event-rules mechanism for FlashSystem A9000 and FlashSystem A9000R).

Connection loss events, whether because of a transport error, timeout, or due to specific customer action, specify whether the connection was idle or in-use at the time of disconnection. A warning event is issued 15 minutes before the **`timeout`** parameter disconnects a busy support session. If the connection is lost because of a network error before the timeout for the session expires, the system automatically tries to reconnect to any of the configured RSC servers.

While a support session is in progress, FlashSystem A9000 or FlashSystem A9000R generates events and shows the machine's status on the GUI window as usual. Therefore, for example, the customer can see the process of phasing out a module or the restart of customer-visible machine services as they happen.

Starting an RSC connection by using the CLI

The following CLI commands are used to manage Remote Support Center connections:

- `support_center_list`
- `support_center_status`
- `support_center_connect`
- `support_center_disconnect`

To start an RSC connection by using the CLI, issue the **`support_center_list`** CLI command, as shown in Example 7-9.

Note: For information on recent changes regarding the support centers, see the following article: http://www-01.ibm.com/support/docview.wss?uid=ssg1S1010513.

Example 7-9 List available support centers

```
>>support_center_list
Name           Address          Port   Priority
front_bld     204.146.30.139     22        11
front_pok     129.33.206.139     22        11
```

Contact your IBM Support representative if the **`support_center_list`** commands do not return a value and you want to use this feature.

Next, you can see the status of the connection by running the **support_center_status** command (Example 7-10).

Example 7-10 Check support center connection status

```
>>support_center_status
State          Connected sessions  Timeout (min)  Module  Connected since
no connection
```

In this example, no connections are currently established. To start a session, run **support_center_connect** and then run **support_center_status**, as shown in Example 7-11.

Example 7-11 Connect to support center

```
>> support_center_connect
command 0:
administrator:
    command:
        code = "SUCCESS"
        status = "0"
        status_str = "Command completed successfully"
aserver = "DELIVERY_SUCCESSFUL"

>> support_center_status
State   Connected sessions   Timeout (min)   Module       Connected since
------- -------------------- --------------- ------------ --------------------
idle    0                    no timeout      1:Module:4   2010-10-08 10:45:35
```

If no parameters are specified with the **support_center_connect** command, the timeout values default to `never` and no password is used. You can specify the parameters for **timeout**, **idle_timeout**, and **password**, as shown in Example 7-12. The **timeout** values are in the format *hh*:*mm* when you use the CLI commands.

Example 7-12 Specify the connection parameters

```
>> support_center_connect timeout=01:00 idle_timeout=00:30 password=4support
command 0:
administrator:
    command:
        code = "SUCCESS"
        status = "0"
        status_str = "Command completed successfully"
aserver = "DELIVERY_SUCCESSFUL"

>> support_center_status
State   Connected sessions   Timeout (min)   Module       Connected since
------- -------------------- --------------- ------------ --------------------
idle    0                    27.6            1:Module:4   2010-10-08 10:49:40
```

The status shows an idle state until an RSC representative establishes a connection, at which time it shows a state of `busy`, as shown in Example 7-13.

Example 7-13 Connection busy

```
>> support_center_status
State   Connected sessions   Timeout (min)   Module       Connected since
------- ------------------- --------------- ------------ --------------------
busy    1                   40.2            1:Module:4   2010-10-08 10:49:40
```

To terminate the Remote Support Center connection, run the **`support_center_disconnect`** command, as shown in Example 7-14.

Example 7-14 Disconnect from the support center

```
>> support_center_disconnect
command 0:
administrator:
    command:
        code = "SUCCESS"
        status = "0"
        status_str = "Command completed successfully"
aserver = "DELIVERY_SUCCESSFUL"

>> support_center_status
State           Connected sessions   Timeout (min)   Module   Connected since
--------------- ------------------- --------------- -------- -----------------
no connection   ?                   ?               ?        ?
```

Remote support for severe system conditions

A remote support connection may be established automatically, when activated, if severe system conditions, or critical issues, are seen on the system and host access is blocked.

Note: Enabling the remote support for severe system conditions is strongly recommended. It can drastically improve the recovery time during critical situations. This feature allows IBM service representatives to access the system remotely and start the repair actions immediately.

Remote support for severe system conditions permits a remote support connection without the need for the customer to initiate a RSC session. When the system is configured for remote support it also eliminates the need to dispatch an IBM service representative to the customer site to enable RSC.

The remote support for severe system conditions can be configured by the IBM service representative at the time of the storage system installation or the customer can allow it using the GUI or CLI at any time afterwards. This information is communicated through the Technical and Delivery Assessment (TDA) checklist and worksheets.

Configuration can be done through the following

- An IBM technician, or support representative can define RSC automatic connection by connecting to the system either on site, or remotely via a RSC session.
- XCLI utility: RSC automatic connection to a support center can be done using the **`support_center_config`** command, setting `automatically_connect` to `yes`. See Example 7-15 on page 167.

Example 7-15 Applying remote support for severe conditions

```
A9000>>support_center_config_list
Enable Auto Conn   First Module   Second Module   Third Module
no                 -1             -1              -1
A9000>>support_center_config automatically_connect=yes
Command executed successfully.
A9000>>support_center_config_list
Enable Auto Conn   First Module   Second Module   Third Module
yes                1              3               1
```

- GUI: Starting with version 5.3, the customer can enable RSC for severe system conditions using the GUI.

  To enable the remote connection for severe conditions, go to **System and Domains Views -> Systems-> Select the system that you want-> Actions (a) -> Support (b) -> Allow my IBM Support Center to proactively connect in case of severe conditions (c)->Apply.** See Figure 7-14.

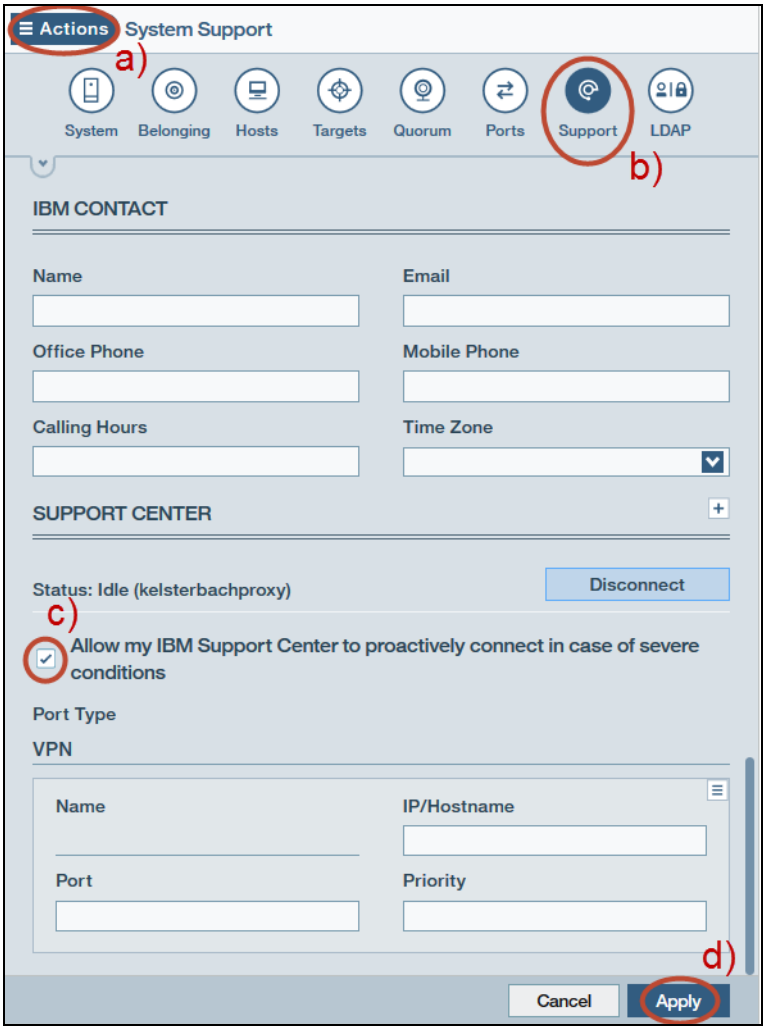


Figure 7-14 Enabling RSC for severe conditions in the GUI

Note: If RSC gets automatically connected after a system rebooted to 'ON' state (and the RSC was not set to always be connected) – the connection will automatically disconnect after 12 hours.

In all other automatic RSC connection cases (including after reboot to 'Maintenance' mode) – RSC will remain connected. In this instance the connection will have to be manually disconnected.

Remote Support Proxy

The optional Remote Support Proxy agent can be used when one or more FlashSystem A9000 and FlashSystem A9000R systems do not have direct access to the internet (for example, due to firewall restrictions).

After the Remote Support Proxy agent is configured, the connection to the RSC is performed normally from FlashSystem A9000 and FlashSystem A9000R, as described in "Remote support connection (RSC)" on page 161.

The agent is a small program that runs on the following Linux versions:

- Red Hat Enterprise Linux, Version 6.0 or later, for x86 and x86-64 systems
- Red Hat Enterprise Linux, Version 5.1 or later, for x86 and x86-64 systems
- Red Hat Enterprise Linux, Version 4.6 or later, for x86 and x86-64 systems
- SUSE Linux Enterprise Server 11 or later, for x86 and x86-64 systems

Important: Review the latest release notes for updates.

The host that is running the agent must have TCP/443 outbound access to RSC addresses (information that is supplied by IBM Support) and it must listen for inbound connections from FlashSystem A9000 or FlashSystem A9000R.

For more information about the Remote Support Proxy, see the documentation for FlashSystem A9000 and FlashSystem A9000R in IBM Knowledge Center.

Installation

The installation files and documentation are at the storage portal website for FlashSystem A9000 and FlashSystem A9000R:

After you download the correct package to the Linux host, you can run the file as root and it starts the installation wizard.

For more information about the Remote Support Proxy, see IBM Knowledge Center.

Related publications

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

IBM Redbooks publications

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *IBM Hyper-Scale Manager for IBM Spectrum Accelerate Family: IBM XIV, IBM FlashSystem A9000 and A9000R, and IBM Spectrum Accelerate*, SG24-8376
- *IBM FlashSystem A9000 and A9000R Business Continuity Solutions*, REDP-5401
- *IBM HyperSwap and Multi-site HA/DR for IBM FlashSystem A9000 and A9000R*, REDP-5434
- *Data-at-rest Encryption for the IBM Spectrum Accelerate Family*, REDP-5402
- *IBM FlashSystem A9000, IBM FlashSystem A9000R, and IBM XIV Storage System: Host Attachment and Interoperability*, SG24-8368

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications and online resources

These publications are also relevant as further information sources:

- IBM FlashSystem A9000 on the IBM Knowledge Center:

  http://www.ibm.com/support/knowledgecenter/STJKN5

  The following publications are at this website:

  - *IBM FlashSystem A9000 Command-Line Interface (CLI) Reference Guide*, SC27-8559
  - *IBM FlashSystem A9000 Product Overview,* GC27-8583
  - *Hyper-Scale Manager 5.3 REST API Specification*, SC27-6440
  - *Hyper-Scale Manager 5.3 User Guide*, SC27-8560
  - *IBM FlashSystem A9000 Models 9836-415, 9838-415, 9836-425, 9838-425, and 9838-U25 Deployment Guide,* GC27-8564
- IBM FlashSystem A9000R on the IBM Knowledge Center:

  http://www.ibm.com/support/knowledgecenter/STJKMM

  The following publications are at this website:

  - *IBM FlashSystem A9000R Command-Line Interface (CLI) Reference Guide*, SC27-8711
  - *IBM FlashSystem A9000R Product Overview,* GC27-8558

  - *Hyper-Scale Manager 5.3 REST API Specification*, SC27-6440
  - *Hyper-Scale Manager 5.3 User Guide*, SC27-8560
  - *IBM FlashSystem A9000R Models 9835-415, 9837-415, 9835-425, 9837-425, and 9837-U25 Deployment Guide,* GC27-8565
- IBM Hyper-Scale Manager Knowledge Center:

  https://www.ibm.com/support/knowledgecenter/SSUMNQ_5.3.0
- IBM FlashSystem A9000 product page:

  http://www.ibm.com/systems/storage/flash/a9000
- IBM FlashSystem A9000R product page:

  http://www.ibm.com/systems/storage/flash/a9000r
- IBM Fix Central:

  http://www.ibm.com/support/fixcentral/
- For the latest hardware and software requirements, visit the IBM System Storage Interoperation Center (SSIC) website:

  http://www.ibm.com/systems/support/storage/ssic/interoperability.wss
- IBM Offering Information page (announcement letters and sales manuals):

  http://www.ibm.com/common/ssi/index.wss?request_locale=en

  On this page, enter A9000, select the information type, and click **Search**. On the next page, narrow your search results by geography and language.

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

IBM FlashSystem A9000 and IBM FlashSystem A9000R Architecture and Implementation

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

SG24-8345-03

ISBN 073845740x

Printed in U.S.A.