IBM

Implementing the IBM Storwize V7000 Gen2

- Learn about the latest addition to the IBM SAN Volume Controller/Storwize family
- Understand the new functions and features
- Benefit from an uncomplicated implementation

Jon Tate
Morten Dannemand
Nancy Kinney
Massimo Rosati
Lev Sturmer

Redbooks

ibm.com/redbooks

International Technical Support Organization

Implementing the IBM Storwize V7000 Gen2

January 2015

SG24-8244-00

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

First Edition (January 2015)

This edition applies to the IBM Storwize V7000 Gen2 running software version 7.3.

Contents

Notices vii
Trademarks viii

IBM Redbooks promotions ix

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

Chapter 1. Introduction to IBM storage virtualization 1
1.1 Storage virtualization terminology 2
1.2 Requirements driving storage virtualization 5
1.2.1 Benefits of using IBM Storwize V7000 Gen2 5
1.3 IBM Storwize V7000 Gen2 architecture 6
1.3.1 IBM Storwize V7000 Gen2 components 8
1.4 More information 20

Chapter 2. IBM Storwize V7000 Gen2 Hardware 21
2.1 Overview 22
2.1.1 Scalable write cache 23
2.1.2 Increase in performance 23
2.1.3 Support for 20 Expansion Enclosures 24
2.2 New hardware on the Storwize V7000 Gen2 24
2.2.1 Controller hardware overview 24
2.2.2 Base configuration 26
2.2.3 Controller SAS Topology 29
2.2.4 Compression Accelerator card 29
2.2.5 Host interface cards 29
2.2.6 Expansion Enclosure 30
2.2.7 SES firmware 30
2.3 Expansion Enclosures 31
2.3.1 Control enclosure midplane 32
2.3.2 Battery backup 32
2.3.3 Battery Pack 33
2.3.4 Control Enclosure power supplies 33
2.3.5 Control enclosure fan modules 34
2.3.6 Compression Accelerator card 35
2.3.7 FCoE and iSCSI 35
2.4 LEDs 36
2.4.1 Battery LED 36
2.4.2 Control Enclosure LED 38
2.4.3 Expansion Enclosure LED 42
2.5 Technician Port 45
2.5.1 Technician port is marked with a T (Ethernet port 4) 45

Chapter 3. Planning and configuration 47
3.1 General planning rules 48

3.1.1 Base software 49
3.2 Physical planning 54
3.2.1 Single rack or dual rack configuration 55
3.2.2 Cable connections 57
3.3 Logical planning 58
3.3.1 Management IP addressing plan 58
3.3.2 SAN zoning and SAN connections 60
3.3.3 iSCSI IP addressing plan 62
3.3.4 Native IP replication 63
3.3.5 Back-end storage subsystem configuration 63
3.3.6 Real-time Compression. 66
3.3.7 Easy Tier version 3 66
3.3.8 Storwize V7000 Gen2 clustered system configuration 67
3.3.9 Volume configuration 68
3.3.10 Host mapping (LUN masking) 70
3.3.11 Advanced Copy Services 72
3.3.12 SAN boot support 77
3.3.13 Data migration from a non-virtualized storage subsystem 78
3.3.14 Storwize V7000 Gen2 configuration backup procedure 79
3.4 Performance considerations 79
3.4.1 SAN. 80
3.4.2 Disk subsystems 80
3.4.3 Cache 80
3.4.4 Port Configuration 82
3.4.5 Performance monitoring 83

Chapter 4. IBM Storwize V7000 Gen2 Easy Tier 85
4.1 IBM Storwize family software Easy Tier history 86
4.1.1 New features in Easy Tier 3 86
4.1.2 Storage pool balancing 88
4.2 Performance and monitoring considerations. 88
4.2.1 Considerations for optimal performance 88
4.2.2 Monitoring tools. 92

Chapter 5. The IBM Storwize V7000 Gen2 Initial Configuration 95
5.1 Managing the Storwize V7000 Gen2 96
5.1.1 Network requirements for the IBM Storwize V7000 Gen2 96
5.1.2 Prerequisites 97
5.2 Initial Configuration of the IBM Storwize V7000 Gen2 98
5.2.1 How to make the first connection to the Storwize V7000 Gen2 98
5.2.2 System Setup wizard 101

Chapter 6. IBM Real-time Compression and the IBM Storwize V7000 Gen2 117
6.1 Real-time Compression background, overview, and value proposition. 118
6.1.1 The solution: IBM Real-time Compression 118
6.1.2 Common use cases 119
6.2 IBM Real-time Compression technology 120
6.2.1 Random Access Compression Engine 121
6.2.2 RACE in Storwize V7000 Gen2 software stack 126
6.2.3 Data write flow 127
6.2.4 Data read flow. 127
6.2.5 Compression of existing data 127
6.3 Storwize V7000 Gen2 software and hardware updates that enhance Real-time Compression 127

6.3.1 Software enhancements 128
6.3.2 Hardware enhancements 128

Chapter 7. Performance overview of the IBM Storwize V7000 Gen2 131
7.1 IBM Storwize V7000 Gen2 performance overview 132
7.1.1 Drive Performance 132
7.1.2 RAID performance 133
7.1.3 Drive, RAID, and storage pool best practices 134
7.2 SAN Performance considerations 134
7.3 Performance monitoring 135
7.3.1 Collecting performance statistics 135
7.3.2 Real-time performance monitoring 136
7.3.3 Performance data collection and Tivoli Storage Productivity Center for Disk . . . 142

Chapter 8. IBM Storwize V7000 Gen2 command-line interface 143
8.1 How to set up the CLI 144
8.2 Configuring the IBM Storwize V7000 Gen2 using the CLI 155
8.2.1 Viewing the Node Canister details 155
8.2.2 Configuring event notification settings 157
8.2.3 Configuring internal storage 157
8.2.4 Configuring external storage 159
8.2.5 Configuring volumes 162
8.2.6 Configuring remote copy 163
8.2.7 Configuring hosts 165
8.3 New Commands introduced with the IBM Storwize V7000 Gen2 166

Chapter 9. IBM Storwize V7000 Gen2 operations using the GUI 169
9.1 Introduction to IBM Storwize V7000 GUI 170
9.1.1 Overview 171
9.1.2 Monitoring 172
9.1.3 Pools 173
9.1.4 Volumes 174
9.1.5 Hosts 175
9.1.6 Copy services 176
9.1.7 Access 177
9.1.8 Settings 178
9.2 IBM Storage Tier Advisor Tool 178

Related publications 183
IBM Redbooks 183
Other resources 184
Referenced websites 185
Help from IBM 185

Notices

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

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

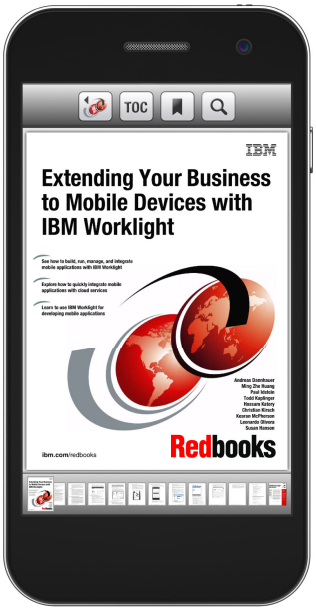
Trademarks

Find and read thousands of IBM Redbooks publications

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

Get the latest version of the Redbooks Mobile App

iOS **Download Now** Android



Promote your business in an IBM Redbooks publication

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

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

ibm.com/Redbooks
About Redbooks → Business Partner Programs

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

Data is the new currency of business, the most critical asset of the modern organization. In fact, enterprises that can gain business insights from their data are twice as likely to outperform their competitors. Nevertheless, 72% of them have not started, or are only planning, big data activities. In addition, organizations often spend too much money and time managing where their data is stored. The average firm purchases 24% more storage every year, but uses less than half of the capacity that it already has.

The IBM® Storwize® family, including the IBM SAN Volume Controller Data Platform, is a storage virtualization system that enables a single point of control for storage resources. This functionality helps support improved business application availability and greater resource use. The following list describes the business objectives of this system:

- To manage storage resources in your information technology (IT) infrastructure
- To make sure that those resources are used to the advantage of your business
- To do it quickly, efficiently, and in real time, while avoiding increases in administrative costs

Virtualizing storage with Storwize helps make new and existing storage more effective. Storwize includes many functions traditionally deployed separately in disk systems. By including these functions in a virtualization system, Storwize standardizes them across virtualized storage for greater flexibility and potentially lower costs.

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

Integrating these functions into Storwize also means that they are designed to operate smoothly together, reducing management effort.

This IBM Redbooks® publication provides information about the latest features and functions of the Storwize V7000 Gen2 and software version 7.3 implementation, architectural improvements, and Easy Tier.

Authors

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

Jon Tate is a Project Manager for IBM System Storage® storage area network (SAN) Solutions at the ITSO, San Jose Center. Before joining the ITSO in 1999, he worked in the IBM Technical Support Center, providing Level 2 support for IBM storage products. Jon has 28 years of experience in storage software and management, services, and support, and is both an IBM Certified IT Specialist and an IBM SAN Certified Specialist. He is also the UK Chairman of the Storage Networking Industry Association.

Morten Dannemand is a Certified Client Technical Specialist with IBM Technical Support Services in IBM Denmark. He has been with IBM for 17 years, and has worked primarily with SAN since 2005. Morten works across a wide range of platforms, for example, EMC, SUN, IBM System x, and IBM System z®. He is also an Account Advocate for Storage, advising customers in managing their IBM storage portfolio.

Nancy Kinney is a Cloud Architect responsible for the migration of large quantities of data from various locations into the Cloud. She works with other architects globally, and assists other IBM business units and the client in the development and delivery of data migration solutions. She has significant knowledge of storage technologies across multiple platforms, processes, and architectures. Previously, she held the position of Infrastructure Architect, and used engineering infrastructure solutions to make use of technology guidelines to meet technical requirements and business needs. She has also worked as an IBM US Remote Technical Support Engineer for IBM Global Technology Services® (GTS) in the Austin IBM AIX® system and IBM System Storage center, acquiring end-to-end experience in troubleshooting Fibre Channel (FC) technologies. She holds an IBM Midrange Specialist Certification, and NetApp/N-Series Certified Data Management Administrator (NCDA) Certification.

Massimo Rosati is a Certified Senior Storage IT Specialist for IBM Italy. He has 29 years of experience in the delivery of Professional Services and Software Support. His areas of expertise include storage hardware, SAN, storage virtualization, disaster recovery (DR), and business continuity solutions. He has achieved Brocade and Cisco SAN Design Certifications, and supports critical and complex client engagements in the SAN and storage areas. Massimo has written extensively about SAN and virtualization products in several IBM Redbooks publications.

Lev Sturmer is a Level 3 Support Engineer in the Real-time Compression division of IBM Storage and Technology Group for the past 3 years, and works for IBM Israel. Before joining IBM, Lev worked as a Security and Networking Engineer where he gained 10 years experience across various environments.

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

Alan Dawson
Steven White
Chris Canto
Barry Whyte
Evelyn Perez
Lee Sanders
Katja Gebuhr
Paul Merrison
Gareth Nicholls
Ian Boden
John Fairhurst

IBM Hursley, UK

Special thanks to the Brocade Communications Systems staff in San Jose, California for their unparalleled support of this residency in terms of equipment and support in many areas:

Madu Amajor
Silviano Gaona
Sangam Racherla
Brian Steffler
Marcus Thordal
Jim Baldyga

Brocade Communications Systems

Now you can become a published author, too

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

Obtain more information 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:

  ibm.com/redbooks

- Send your comments in an email:

  redbooks@us.ibm.com

- Mail your comments:

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

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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

1

Introduction to IBM storage virtualization

In this chapter, we introduce the concept of storage virtualization. Then, we present an overview explaining how you can apply virtualization to help address challenging storage requirements. Finally, we describe in detail all changes and enhancements of the IBM Storwize V7000 Gen2 version 7.3, a key milestone in the product development roadmap.

1.1 Storage virtualization terminology

Although *storage virtualization* is a term that is used extensively throughout the storage industry, it can be applied to a wide range of technologies and underlying capabilities. In reality, most storage devices can technically claim to be virtualized in one form or another. Therefore, we must start by defining the concept of storage virtualization as used in this publication.

IBM interprets storage virtualization in the following manner:

- Storage virtualization is a technology that makes one set of resources look and feel like another set of resources, preferably with more desirable characteristics.
- It is a logical representation of resources that are not constrained by physical limitations:
  - It hides part of the complexity.
  - It adds or integrates new functionality with existing services.
  - It can be nested or applied to multiple layers of a system.

When considering storage virtualization, it is important to understand that virtualization can be implemented at various layers within the input/output (I/O) stack. We must clearly distinguish between virtualization at the disk layer and virtualization at the file system layer.

The focus of this book is virtualization at the disk layer, which is referred to as *block-level virtualization*, or the *block aggregation layer*. A description of file system virtualization is beyond the scope of this book.

However, if you are interested in file system virtualization, see the following information about IBM General Parallel File System (GPFS™) or IBM Scale Out Network Attached Storage (SONAS), which is based on GPFS.

To obtain more information and an overview of GPFS and IBM Elastic Storage™, see the following website:

`http://www.ibm.com/systems/technicalcomputing/platformcomputing/products/gpfs/`

More information about IBM SONAS is available on the following website:

`http://www.ibm.com/systems/storage/network/sonas/`

The Storage Networking Industry Association's (SNIA) block aggregation model (Figure 1-1 on page 3) provides a useful overview of the storage domain and its layers. The figure shows the three layers of a storage domain:

- The file
- The block aggregation
- The block subsystem layers

The model splits the block aggregation layer into three sublayers. Block aggregation can be realized within hosts (servers), in the storage network (storage routers and storage controllers), or in storage devices (intelligent disk arrays).

One of the IBM implementations of a block aggregation solution is IBM Storwize V7000 Gen2. Storwize V7000 Gen2 is implemented as a clustered appliance in the storage network layer.

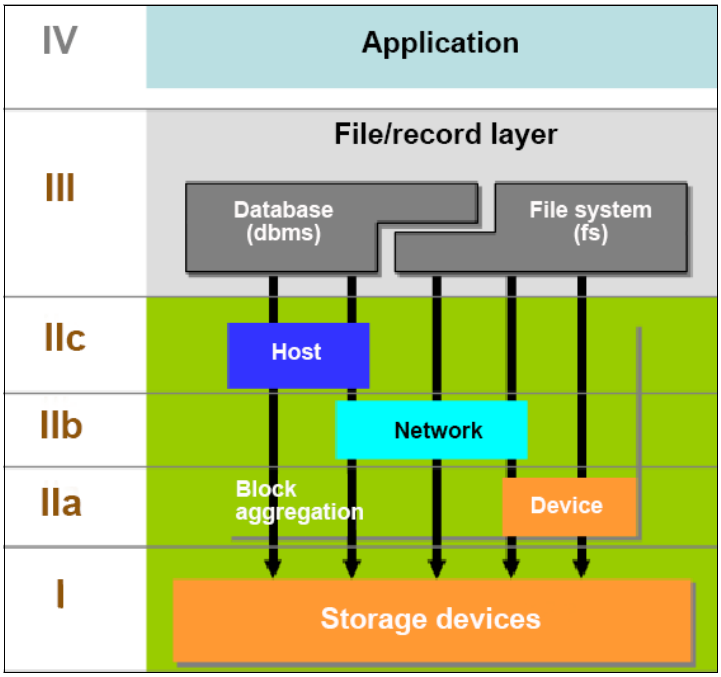


Figure 1-1 SNIA block aggregation model[1]

The key concept of virtualization is to decouple the storage from the storage functions that are required in the storage area network (SAN) environment. *Decoupling* means abstracting the physical location of data from the logical representation of that data. The virtualization engine presents logical entities to the user, and internally manages the process of mapping these entities to the actual location of the physical storage.

The actual mapping that is performed depends on the specific implementation, such as the granularity of the mapping, which can range from a small fraction of a physical disk, up to the full capacity of a physical disk.

A single block of information in this environment is identified by its *logical unit number* (LUN), which is the physical disk, and an offset within that LUN, which is known as a *logical block address* (LBA).

The term *physical disk* is used in this context to describe a piece of storage that might be carved out of a Redundant Array of Independent Disks (RAID) in the underlying disk subsystem.

Specific to the Storwize V7000 Gen2 implementation, the address space that is mapped between the logical entity is referred to as *volume*, and the physical disk is referred to as *managed disks* (MDisks).

[1] This figure is produced by the Storage Networking Industry Association.

Figure 1-2 shows an overview of block-level virtualization.

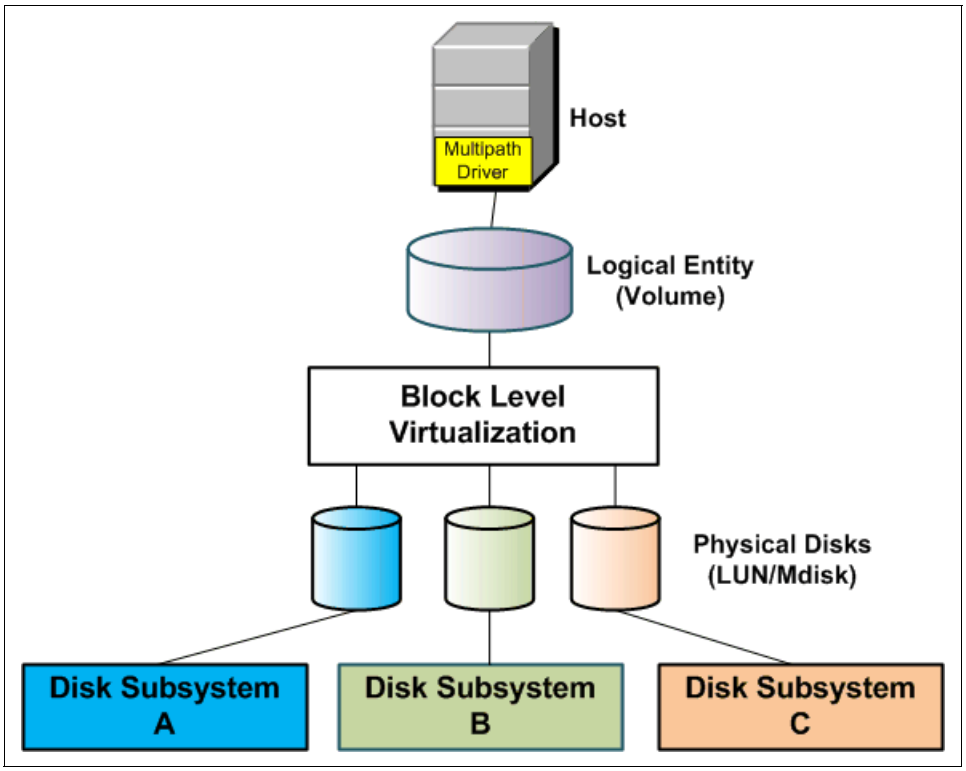


Figure 1-2 Block-level virtualization overview

The server and application are only aware of the logical entities, and they access these entities using a consistent interface that is provided by the virtualization layer.

The functionality of a volume that is presented to a server, such as expanding or reducing the size of a volume, mirroring a volume, creating an IBM FlashCopy®, and thin provisioning, is implemented in the virtualization layer. It does not rely in any way on the functionality that is provided by the underlying disk subsystem. Data that is stored in a virtualized environment is stored in a location-independent way, which enables a user to move or migrate data between physical locations (referred to as *storage pools*).

We refer to block-level storage virtualization as the *cornerstones of virtualization*. These cornerstones of virtualization are the core benefits that a product, such as Storwize V7000 Gen2, can provide over the traditional directly attached or SAN-attached storage.

Storwize V7000 Gen2 provides the following benefits:

- Provide online volume migration while applications are running, which is possibly the greatest single benefit for storage virtualization. This capability enables data to be migrated on and between the underlying storage subsystems without any affect to the servers and applications. In fact, this migration is performed without the servers and applications “knowing” that it even occurred.
- Simplify storage management by providing a single image for multiple controllers, and a consistent user interface (UI) for provisioning heterogeneous storage.
- Provide enterprise-level copy services functions. Performing the copy services functions in Storwize V7000 Gen2 removes dependencies on the storage subsystems, therefore enabling the source and target copies to be on other storage subsystem types.

- Increase storage use by pooling storage across the SAN, and by enabling space reduction techniques such an IBM Real-time Compression or thin provisioning.
- Improve system performance as a result of volume striping across multiple arrays or controllers, and the additional cache that it provides.

Storwize V7000 Gen2 delivers these functions in a homogeneous way on a scalable and highly available platform, over any attached storage, and to any attached server.

1.2 Requirements driving storage virtualization

Today, an emphasis exists on a dynamic infrastructure being able to adapt to company requirements for diversified data operations. Therefore, there is a need for a storage environment that is as flexible as the application and server mobility. Business demands change quickly.

These key client concerns drive storage virtualization:

- Growth in data center costs
- Inability of information technology (IT) organizations to respond quickly to business demands
- Poor asset use
- Poor availability or service levels
- Lack of skilled staff for storage administration

You can see the importance of addressing the complexity of managing storage networks by applying the total cost of ownership (TCO) metric to storage networks. Industry analyses show that storage acquisition costs are only about 20% of the TCO. Most of the remaining costs relate to managing the storage system.

However, how much of the management of multiple systems, with separate interfaces, can be handled as a single entity? In a non-virtualized storage environment, every system is an “island” that needs to be managed separately.

1.2.1 Benefits of using IBM Storwize V7000 Gen2

Because storage virtualization is no longer merely a concept or an unproven technology, all major storage vendors offer storage virtualization products. Using storage virtualization as the foundation for a flexible and reliable storage solution helps enterprises to better align business and IT by optimizing the storage infrastructure and storage management to meet business demands.

Storwize V7000 Gen2 is a mature, second-generation storage system. It is based on the software of IBM SAN Volume Controller, an eighth-generation virtualization solution that uses open standards and is consistent with the SNIA storage model.

Storwize V7000 Gen2 is a storage system-based in-band block virtualization process, in which intelligent functionality, including advanced storage functions, is available for internal storage and any virtualized storage device.

Storwize V7000 Gen2 can improve the use of your storage resources, simplify your storage management, and improve the availability of your applications. It can also reduce the number of separate environments that need to managed down to a single environment. It provides a single interface for storage management. After the initial configuration of the storage subsystems that you are going to virtualize, all of the day-to-day storage management operations are performed from Storwize V7000 Gen2.

In addition, because Storwize V7000 Gen2 provides advanced functions, such as mirroring and FlashCopy, there is no need to purchase them again for each new virtualized disk subsystem.

Today, it is typical that open systems run at less than 50% of the usable capacity that is provided by the RAID subsystems. Using the installed raw capacity in the disk subsystems will, depending on the RAID level that is used, show usage numbers of less than 35%.

A block-level virtualization solution, such as Storwize V7000 Gen2, can enable capacity usage to increase to approximately 75 - 80%. With Storwize V7000 Gen2, free space does not need to be maintained and managed within each storage subsystem, which further increases capacity use.

1.3 IBM Storwize V7000 Gen2 architecture

Storwize V7000 Gen2 is a SAN block aggregation virtualization appliance that is designed for attachment to various host computer systems.

There are two major approaches in use today to consider for the implementation of block-level aggregation and virtualization:

- Symmetric: In-band appliance

  The device is a SAN appliance that sits in the data path, and all I/O flows through the device. This implementation is referred to as *symmetric virtualization* or *in-band*.

  The device is both target and initiator. It is the target of I/O requests from the host perspective, and the initiator of I/O requests from the storage perspective. The redirection is performed by issuing new I/O requests to the storage. Storwize V7000 Gen2 uses symmetric virtualization.

- Asymmetric: Out-of-band or controller-based

  The device is usually a storage controller that provides an internal switch for external storage attachment. In this approach, the storage controller intercepts and redirects I/O requests to the external storage as it does for internal storage. The actual I/O requests are themselves redirected. This implementation is referred to as *asymmetric virtualization* or *out-of-band*.

Figure 1-3 shows variations of the two virtualization approaches.

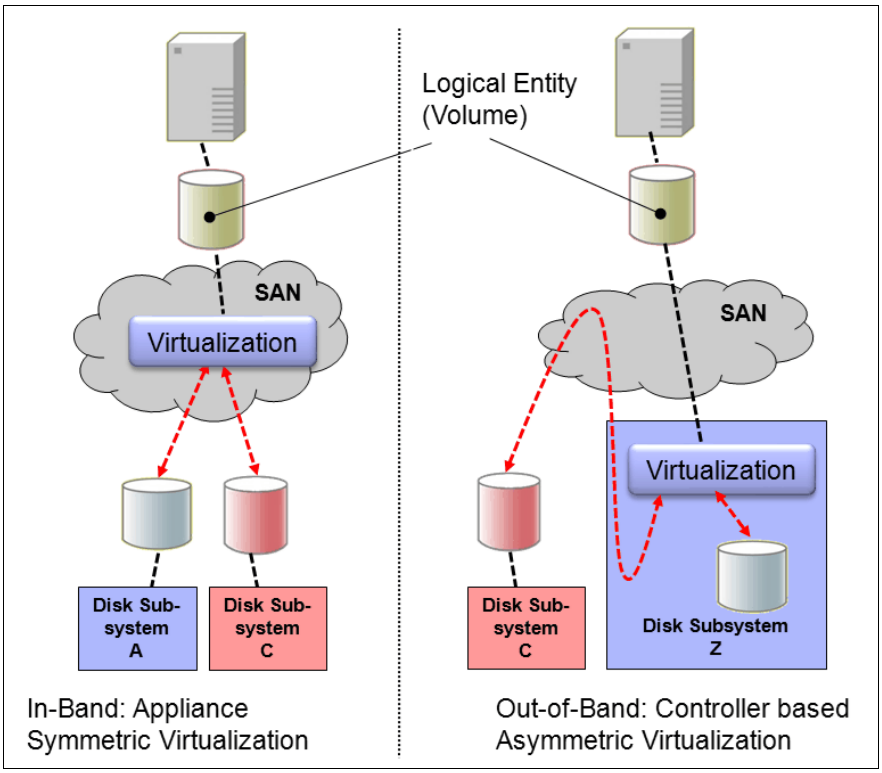


Figure 1-3 Overview of block-level virtualization architectures

The Storwize V7000 Gen2 solution provides a modular storage system that includes the capability to virtualize both external SAN-attached storage and its own internal storage. The Storwize V7000 Gen2 solution is built upon the IBM SAN Volume Controller technology base, and uses technology from the IBM System Storage DS8000® family.

A Storwize V7000 Gen2 system provides several configuration options that are aimed at simplifying the implementation process. It also provides automated wizards, called *Directed Maintenance Procedures* (DMP), to help resolve any events that might occur. A Storwize V7000 Gen2 system is a midrange, clustered, scalable, and external virtualization device.

Included with a Storwize V7000 Gen2 system is a graphical user interface (GUI) that enables storage to be deployed quickly and efficiently. The GUI runs on the Storwize V7000 Gen2 system, so there is no need for a separate console. The management GUI contains a series of preestablished configuration options that are called *presets*, and that use common settings to quickly configure objects on the system. Presets are available for creating volumes and FlashCopy mappings, and for setting up a RAID configuration.

The Storwize V7000 Gen2 solution provides a choice of up to 1056 serial-attached Small Computer System Interface (SCSI), called SAS, drives for the internal storage in a clustered system. It uses SAS cables and connectors to attach to the optional expansion enclosures. In a clustered system, the Storwize V7000 Gen2 can provide about 4 pebibytes (PiB) of internal raw capacity.

When virtualizing external storage arrays, a Storwize V7000 Gen2 system can provide up to 32 PiB of usable capacity. A Storwize V7000 Gen2 system supports a range of external disk systems, similar to what the IBM SAN Volume Controller supports today.

The Storwize V7000 Gen2 solution consists of 1 - 4 control enclosures and, optionally, up to 80 expansion enclosures. It also supports the intermixing of the different expansion enclosures. Within each enclosure are two canisters. Control enclosures contain two node canisters, and expansion enclosures contain two expansion canisters.

1.3.1 IBM Storwize V7000 Gen2 components

Storwize V7000 Gen2 has the following benefits:

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

The Storwize V7000 Gen2 subsystem consists of a set of drive enclosures. Control enclosures contain disk drives and two nodes (an I/O Group), which are attached to the SAN fabric. Expansion enclosures contain drives, and are attached to control enclosures.

The simplest use of Storwize V7000 Gen2 is as a traditional RAID subsystem. The internal drives are configured into RAID arrays and virtual disks created from those arrays. Storwize V7000 Gen2 can also be used to virtualize other storage controllers.

Storwize V7000 Gen2 supports regular and solid-state drives (SSDs) and uses IBM System Storage Easy Tier to automatically place volume hot spots on better-performing storage. In this section, we briefly explain the basic architecture components of Storwize V7000 Gen2.

Nodes

Each Storwize V7000 Gen2 hardware controller is called a *node* or *nodecanister*. The node provides the virtualization for a set of volumes, cache, and copy services functions. Nodes are deployed in pairs, and multiple pairs make up a *clustered system* or *system*. A system can consist of 1 - 4 Storwize V7000 Gen2 node pairs.

One of the nodes within the system is known as the *configuration node*. The configuration node manages the configuration activity for the system. If this node fails, the system chooses a new node to become the configuration node.

Because the nodes are installed in pairs, each node provides a failover function to its partner node in the event of a node failure.

I/O Groups

In Storwize V7000 Gen2, there are 1 - 4 pairs of node canisters known as *I/O Groups*. Storwize V7000 Gen2 supports eight node canisters in the clustered system, which provides four I/O Groups.

When a host server performs I/O to one of its volumes, all of the I/Os for a specific volume are directed to the I/O Group. Also, under normal conditions, the I/Os for that specific volume are always processed by the same node within the I/O Group.

Both nodes of the I/O Group act as preferred nodes for their own specific subset of the total number of volumes that the I/O Group presents to the host servers (a maximum of 2048 volumes per I/O Group). However, each node also acts as a failover node for its partner node within the I/O Group, so a node takes over the I/O workload from its partner node, if required, with no effect to the server's application.

In a Storwize V7000 Gen2 environment that is using an active/active architecture, the I/O handling for a volume can be managed by both nodes of the I/O Group. Therefore, it is mandatory for servers that are connected through Fibre Channel (FC) connectors to use multipath device drivers to be able to handle this capability.

The Storwize V7000 Gen2 I/O Groups are connected to the SAN so that all application servers accessing volumes from the I/O Group have access to them. Up to 2048 host server objects can be defined in four I/O Groups.

If required, host servers can be mapped to more than one I/O Group in the Storwize V7000 Gen2 system. Therefore, they can access volumes from separate I/O Groups. You can move volumes between I/O Groups to redistribute the load between the I/O Groups.

However, moving volumes between I/O Groups cannot always be done concurrently with host I/O, and requires in some cases a brief interruption to remap the host. On the following website, you can check the compatibility of Storwize V7000 Gen2 non-disruptive volume move (NDVM) function with your hosts:

http://www-01.ibm.com/support/docview.wss?uid=ssg1S1004622

Important: The active/active architecture provides availability to process I/Os for both controller nodes, and enables the application to continue running smoothly, even if the server has only one access route or path to the storage controller. This type of architecture eliminates the path and LUN thrashing typical of an active/passive architecture.

System

The system or clustered system consists of 1 - 4 I/O Groups. Certain configuration limitations are then set for the individual system. For example, the maximum number of volumes supported per system is 8192 (having a maximum of 2048 volumes per I/O Group), or the maximum managed disk supported is 32 petabytes (PB) per system.

All configuration, monitoring, and service tasks are performed at the system level. Configuration settings are replicated to all nodes in the system. To facilitate these tasks, a management Internet Protocol (IP) address is set for the system.

A process is provided to back up the system configuration data onto disk so that it can be restored in the event of a disaster. Note that this method does *not* back up application data. Only Storwize V7000 Gen2 system configuration information is backed up. For the purposes of remote data mirroring, two or more systems must form a *partnership* before creating relationships between mirrored volumes.

System configuration backup: After backing up the system configuration, save the backup data on your hard disk (or at the least outside of the SAN). If you are unable to access Storwize V7000 Gen2, you do not have access to the backup data if it is on the SAN.

For details about the maximum configurations that are applicable to the system, I/O Group, and nodes, see the following link:

http://www-01.ibm.com/support/docview.wss?uid=ssg1S1004628

RAID

The Storwize V7000 Gen2 setup contains several internal drive objects, but these drives cannot be directly added to storage pools. The drives need to be included in a RAID to provide protection against the failure of individual drives.

These drives are referred to as members of the array. Each array has a RAID level. RAID levels provide various degrees of redundancy and performance, and have various restrictions regarding the number of members in the array.

Storwize V7000 Gen2 supports hot spare drives. When an array member drive fails, the system automatically replaces the failed member with a hot spare drive and rebuilds the array to restore its redundancy. Candidate and spare drives can be manually exchanged with array members.

Each array has a set of goals that describe the location and performance of each array. A sequence of drive failures and hot spare takeovers can leave an array unbalanced (with members that do not match these goals). The system automatically rebalances such arrays when the appropriate drives are available.

The following RAID levels are available:

- RAID 0 (striping, no redundancy)

  RAID 0 arrays stripe data across the drives. The system supports RAID 0 arrays with just one member, which is similar to a traditional just a bunch of disks (JBOD) unconfigured attach. RAID 0 arrays have no redundancy, so they do not support hot spare takeover or immediate exchange. A RAID 0 array can be formed by 1 - 8 drives.

- RAID 1 (mirroring between two drives)

  RAID 1 arrays stripe data over mirrored pairs of drives. A RAID 1 array mirrored pair is rebuilt independently. A RAID 1 array can be formed by two drives only.

- RAID 5 (striping, can survive one drive fault)

  RAID 5 arrays stripe data over the member drives with one parity strip on every stripe. RAID 5 arrays have single redundancy. The parity algorithm means that an array can tolerate no more than one member drive failure. A RAID 5 array can be formed by 3 - 16 drives.

- RAID 6 (striping, can survive two drive faults)

  RAID 6 arrays stripe data over the member drives with two parity stripes (known as the *P-parity* and the *Q-parity*) on every stripe. The two parity strips are calculated using different algorithms, which give the array double redundancy. A RAID 6 array can be formed by 5 to 16 drives.

- RAID 10 (RAID 0 on top of RAID 1)

  RAID 10 arrays have single redundancy. Although they can tolerate one failure from every mirrored pair, they cannot tolerate two-disk failures. One member out of every pair can be rebuilding or missing at the same time. A RAID 10 array can be formed by 2 - 16 drives.

MDisks

A managed disk (MDisk) is the unit of storage that Storwize V7000 Gen2 virtualizes. This unit could be a logical volume on an external storage array presented to Storwize V7000 Gen2, or a RAID array consisting of internal drives. Storwize V7000 Gen2 can then allocate these MDisks into various storage pools. An MDisk is not visible to a host system on the SAN, because it is internal or zoned only to the Storwize V7000 Gen2 system.

The MDisks are placed into storage pools where they are divided into several extents, which can range in size from 16 megabytes (MB) - 8182 MB, as defined by the storage administrator. See the following link for an overview of the total storage capacity that is manageable per system regarding the selection of extents:

http://www-01.ibm.com/support/docview.wss?uid=ssg1S1004628#_Extents

A volume is host-accessible storage that has been provisioned out of one *storage pool* or, if it is a mirrored volume, out of two storage pools. The maximum size of an MDisk is 1 PB. A Storwize V7000 Gen2 system supports up to 4096 MDisks (including internal RAID arrays).

At any point in time, an MDisk is in one of the following four modes:

- Array

  Array mode MDisks are constructed from drives using the RAID function. Array MDisks are always associated with storage pools.

- Unmanaged MDisk

  An MDisk is reported as *unmanaged* when it is not a member of any storage pool. An unmanaged MDisk is not associated with any volumes, and has no metadata stored on it. Storwize V7000 Gen2 does not write to an MDisk that is in unmanaged mode, except when it attempts to change the mode of the MDisk to one of the other modes. Storwize V7000 Gen2 can see the resource, but the resource is not assigned to a storage pool.

- Managed MDisk

  Managed mode MDisks are always members of a storage pool, and they contribute extents to the storage pool. Volumes (if not operated in image mode) are created from these extents. MDisks operating in managed mode might have metadata extents allocated from them, and can be used as *quorum disks*. This mode is the most common and normal mode for an MDisk.

- Image mode MDisk

  Image mode provides a direct block-for-block translation from the MDisk to the volume by using virtualization. This mode is provided to satisfy three major usage scenarios:

  – Image mode enables the virtualization of MDisks already containing data that was written directly, and not through Storwize V7000 Gen2. Rather, it was created by a direct-connected host. This mode enables a client to insert Storwize V7000 Gen2 into the data path of an existing storage volume or LUN with minimal downtime. The image mode is typically used for data migration from old storage systems to new.

  – Image mode enables a volume that is managed by Storwize V7000 Gen2 to be used with the native copy services function provided by the underlying RAID controller. To avoid the loss of data integrity when Storwize V7000 Gen2 is used in this way, it is important that you disable the Storwize V7000 Gen2 cache for the volume.

  – Storwize V7000 Gen2 provides the ability to migrate to image mode, which enables Storwize V7000 Gen2 to export volumes and access them directly from a host without the Storwize V7000 Gen2 in the path.

  Each MDisk presented from an external disk controller has an online path count that is the number of nodes having access to that MDisk. The *maximum count* is the maximum number of paths detected at any point in time by the system. The *current count* is what the system sees currently. A current value less than the maximum can indicate that SAN fabric paths have been lost.

  SSDs (flash drives) that are in Storwize V7000 Gen2 are presented to the cluster as MDisks. To determine whether the selected MDisk is a flash drive, click the link on the MDisk name to display the Viewing MDisk Details pane. The Viewing MDisk Details pane displays values for the Node ID, Node Name, and Node Location attributes.

Quorum disk

A *quorum disk* is an MDisk that contains a reserved area for use exclusively by the system. The system uses quorum disks to break a tie when exactly half the nodes in the system remain after a SAN failure. This situation is referred to as *split brain*. Quorum functionality is not supported on flash drives in Storwize V7000 Gen2. There are three candidate quorum disks. However, only one quorum disk is active at any time.

Disk tier

It is likely that the MDisks (LUNs) presented to the Storwize V7000 Gen2 system have various performance attributes due to the type of disk or RAID on which they reside. The MDisks can be on 15,000 disk revolutions per minute (RPMs) Fibre Channel or SAS disks, Nearline SAS or Serial Advanced Technology Attachment (SATA) disks, or even flash drives.

Therefore, a storage tier attribute is assigned to each MDisk, with the default being **`enterprise`**. A tier 0 (zero)-level disk attribute (**`ssd`**) is available for flash drives, and a tier 2-level disk attribute (**`nearline`**) is available for nl-sas.

Storage pool

A *storage pool* is a collection of up to 128 MDisks that provides the pool of storage from which volumes are provisioned. A single system can manage up to 128 storage pools. The size of these pools can be changed (expanded or shrunk) at run time by adding or removing MDisks, without taking the storage pool or the volumes offline.

At any point in time, an MDisk can only be a member in one storage pool, except for image mode volumes.

Each MDisk in the storage pool is divided into several extents. The size of the extent is selected by the administrator when the storage pool is created, and cannot be changed later. The size of the extent ranges from 16 MB - 8192 MB.

It is a leading practice to use the same extent size for all storage pools in a system. This approach is a prerequisite for supporting volume migration between two storage pools. If the storage pool extent sizes are not the same, you must use volume mirroring.

Figure 1-4 shows an overview of a Storwize clustered system with an I/O Group.

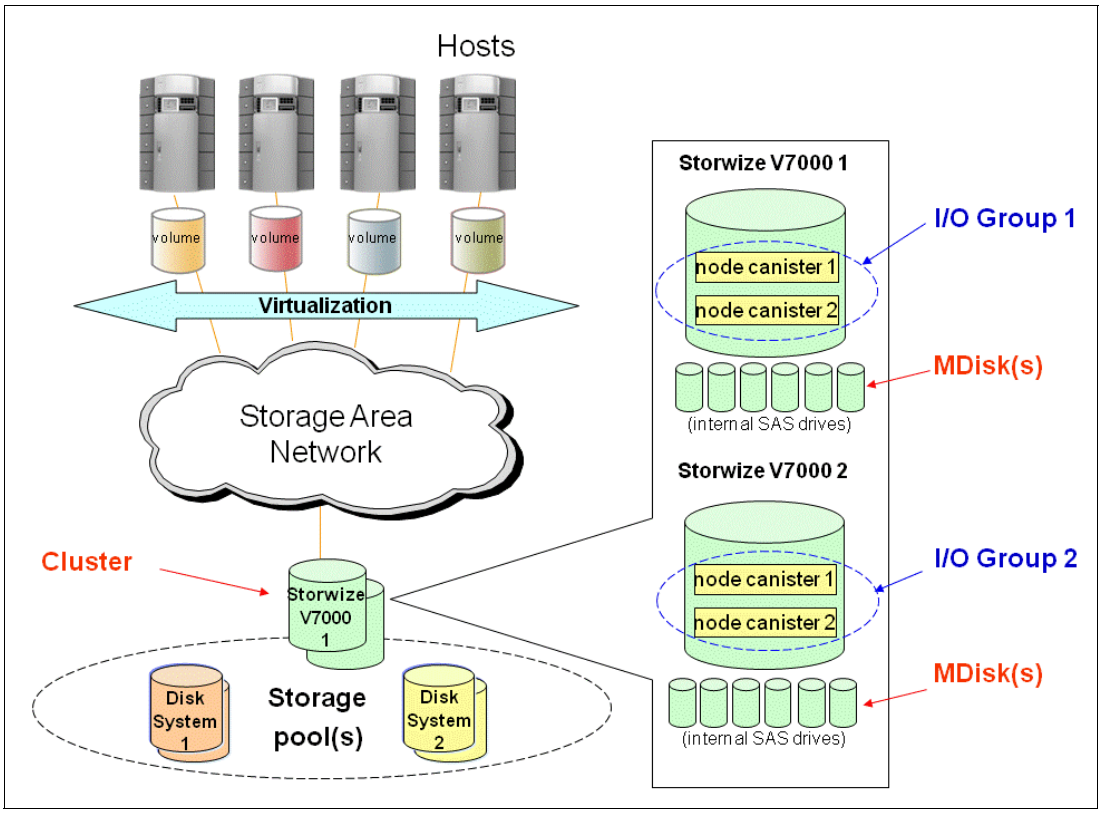


Figure 1-4 IBM Storwize V7000 Gen2 clustered system

Storwize V7000 Gen2 limits the number of extents in a system to 2^{22} = ~4 million. Because the number of addressable extents is limited, the total capacity of a Storwize V7000 Gen2 system depends on the extent size that is chosen by the Storwize V7000 Gen2 administrator. The capacity numbers that are specified in Table 1-1 for a Storwize V7000 Gen2 system assume that all of the defined storage pools have been created with the same extent size.

Table 1-1 Extent size-to-addressability matrix

Extent size maximum	System capacity	Extent size maximum	System capacity
16 MB	64 terabytes (TB)	512 MB	2 PB
32 MB	128 TB	1024 MB	4 PB
64 MB	256 TB	2048 MB	8 PB
128 MB	512 TB	4096 MB	16 PB
256 MB	1 PB	8192 MB	32 PB

For most systems, a capacity of 1 - 2 PB is sufficient. In a Storwize V7000 Gen2 environment, a leading practice is to use the default size of 1 gigabyte (GB) as the standard extent size.

Volumes

Volumes are logical disks that are presented to the host or application servers by Storwize V7000 Gen2. The hosts cannot see the MDisks. They can only see the logical volumes created from combining extents from a storage pool.

There are three types of volumes: Striped, sequential, and image. These types are determined by how the extents are allocated from the storage pool, as explained in the following list:

- A volume created in striped mode has extents allocated from each MDisk in the storage pool in a round-robin fashion.
- With a sequential mode volume, extents are allocated sequentially from an MDisk.
- Image mode is a one-to-one mapped extent mode volume.

Using striped mode is the best method to use for most cases. However, sequential extent allocation mode can slightly increase the sequential performance for certain workloads.

Figure 1-5 shows the striped volume mode and sequential volume mode, and it illustrates how the extent allocation from the storage pool differs.

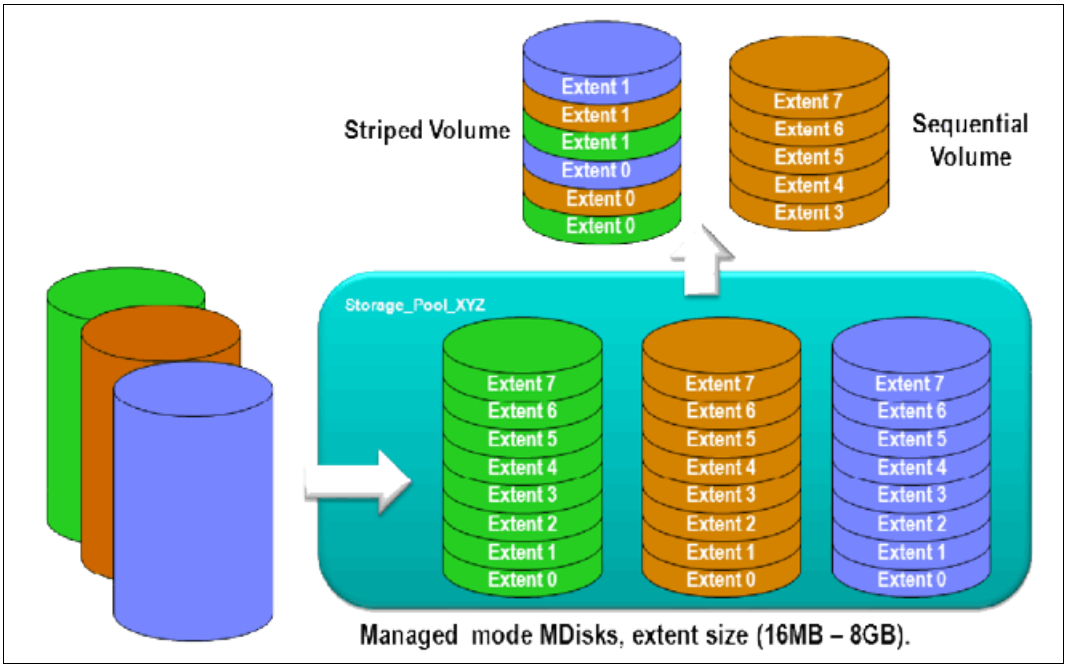


Figure 1-5 Storage pool extents overview

You can allocate the extents for a volume in many ways. The process is under full user control at volume creation time, and can be changed at any time by migrating single extents of a volume to another MDisk within the storage pool.

Hosts

Volumes can be mapped to a *host* to allow access for a specific server to a set of volumes. A host in Storwize V7000 Gen2 is a collection of host bus adapter (HBA) worldwide port names (WWPNs) or Internet SCSI (iSCSI) qualified names (IQNs) defined on the specific server. Note that iSCSI names are internally identified by “fake” WWPNs, or WWPNs that are generated by the Storwize V7000 Gen2.

Volumes can be mapped to multiple hosts, for example, a volume that is accessed by multiple hosts of a server system. iSCSI is an alternative means of attaching hosts. However, all communication with back-end storage subsystems, and with other Storwize V7000 Gen2 systems, is still through FC.

Node failover can be handled without having a multipath driver installed on the iSCSI server. An iSCSI-attached server can simply reconnect after a node failover to the original target IP address, which is now presented by the partner node. To protect the server against link failures in the network or HBA failures, using a multipath driver is mandatory.

Volumes are LUN-masked to the host's HBA WWPNs by a process called *host mapping*. Mapping a volume to the host makes it accessible to the WWPNs or IQNs that are configured on the host object. For a SCSI over Ethernet connection, the IQN identifies the iSCSI target (destination) adapter. Host objects can have both IQNs and WWPNs.

Easy Tier

Easy Tier is a performance function that automatically migrates or moves extents off a volume to, or from, one MDisk storage tier to another MDisk storage tier. In Storwize V7000 Gen2, the Easy Tier automatically moves extents between highly used and less-used MDisks within the same storage tier. This function is called *Storage Pool Balancing*, and it is enabled by default without any need for licensing.

Easy Tier monitors the host I/O activity and latency on the extents of all volumes with the Easy Tier function turned on in a multitier storage pool, over a 24-hour period.

New in Storwize family software V7.3: Easy Tier V3 integrates the automatic functionality to balance the workloads between highly used and less-used MDisks within the same tier. This function is turned on by default at storage pool creation, however it can be modified by the administrator (by turning off Easy Tier at pool level).

Next, it creates an extent migration plan based on this activity, and then dynamically moves high-activity (or *hot*) extents to a higher disk tier in the storage pool. It also moves extents whose activity has dropped off (or *cooled*) from the high-tier MDisks back to a lower-tiered MDisk.

Easy Tier: The Easy Tier function can be turned on or off at the storage pool level and the volume level. It supports any combination of three tiers within the system. Flash drives are always marked as Tier 0.

To experience the potential benefits of using Easy Tier in your environment before installing expensive flash drives, you can turn on the Easy Tier function for a *single-level* storage pool. Next, turn on the Easy Tier function for the volumes within that pool. Easy Tier then starts monitoring activity on the volume extents in the pool.

Easy Tier creates a report every 24 hours, providing information about how Easy Tier behaves if the tier were a multitiered storage pool. So, even though Easy Tier extent migration is not possible within a single-tiered pool, the Easy Tier statistical measurement function is available.

The Easy Tier function can make it more appropriate to use smaller storage pool extent sizes. The usage statistics file can be off-loaded from Storwize V7000 Gen2. Then, you can use the IBM Storage Tier Advisor Tool to create a summary report.

Thin provisioning

Volumes can be configured to be either *thin-provisioned* or *fully allocated*. A thin-provisioned volume behaves regarding application reads and writes as though they were fully allocated. When creating a thin-provisioned volume, the user specifies two capacities:

- The real physical capacity allocated to the volume from the storage pool

- The virtual capacity available to the host

In a fully allocated volume, these two values are the same.

Therefore, the real capacity determines the quantity of MDisk extents that is initially allocated to the volume. The virtual capacity is the capacity of the volume reported to all other Storwize V7000 Gen2 components (for example, FlashCopy, Cache, and remote copy), and to the host servers. The real capacity is used to store both the user data and the metadata for the thin-provisioned volume. The real capacity can be specified as an absolute value, or a percentage of the virtual capacity.

Thin-provisioned volumes can be used as volumes assigned to the host, by FlashCopy to implement thin-provisioned FlashCopy targets, and also with the mirrored volumes feature. When a thin-provisioned volume is initially created, a small amount of the real capacity is used for initial metadata.

Write I/Os to grains of the thin volume that *were not* previously written to cause grains of the real capacity to be used to store metadata and the actual user data. Write I/Os to grains that *were* previously written to update the grain where data was previously written. The grain size is defined when the volume is created, and can be 32 kilobytes (KB), 64 KB, 128 KB, or 256 KB. The default grain size is 256 KB, which is the strongly suggested option. If you select 32 KB for the grain size, the volume size cannot exceed 260,000 GB.

The grain size cannot be changed after the thin-provisioned volume has been created. Generally, smaller grain sizes save space but require more metadata access, which can adversely affect performance. If you are not going to use the thin-provisioned volume as a FlashCopy source or target volume, use 256 KB to maximize performance. If you are going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function.

Figure 1-6 illustrates the thin-provisioning concept.

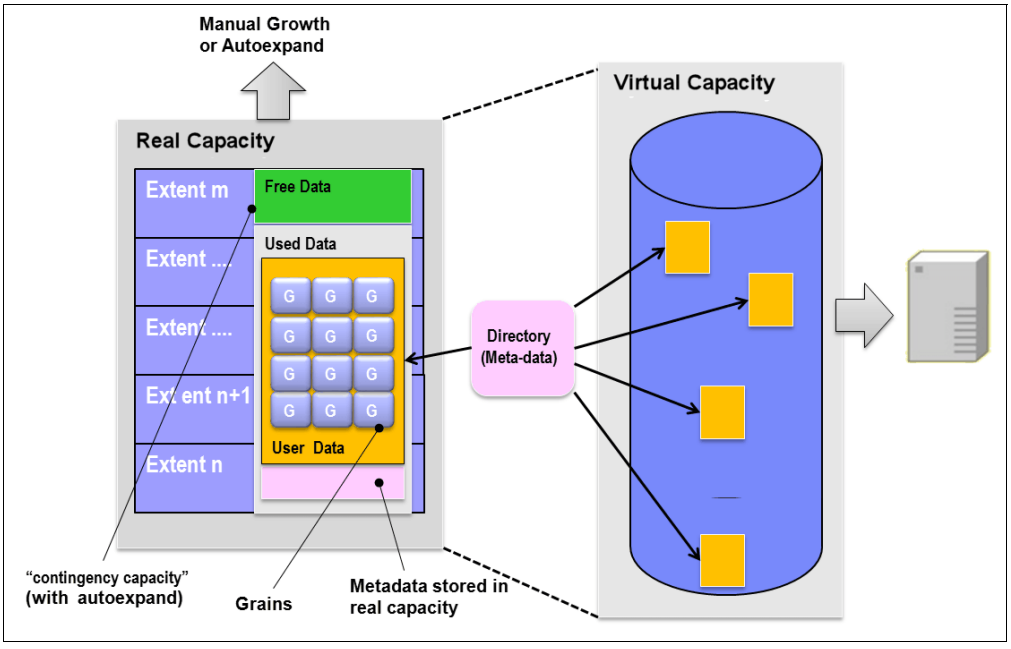


Figure 1-6 Conceptual diagram of a thin-provisioned volume

Thin-provisioned volumes store both user data and metadata. Each grain of data requires metadata to be stored. Therefore, the I/O rates that are obtained from thin-provisioned volumes are less than the I/O rates that are obtained from fully allocated volumes.

The metadata storage fixed use is never greater than 0.1% of the user data. The fixed resource use is independent of the virtual capacity of the volume. If you are using thin-provisioned volumes in a FlashCopy map, for the best performance, use the same grain size as the map grain size. If you are using the thin-provisioned volume directly with a host system, use a small grain size.

The real capacity of a thin volume can be changed if the volume is not in image mode. Increasing the real capacity enables a larger amount of data and metadata to be stored on the volume. Thin-provisioned volumes use the real capacity that is provided in ascending order as new data is written to the volume. If the user initially assigns too much real capacity to the volume, the real capacity can be reduced to free storage for other uses.

A thin-provisioned volume can be configured to *autoexpand*. This feature causes Storwize V7000 Gen2 to automatically add a fixed amount of additional real capacity to the thin volume as required. Autoexpand therefore attempts to maintain a fixed amount of unused real capacity for the volume. This amount is known as the *contingency capacity*.

The contingency capacity is initially set to the real capacity that is assigned when the volume is created. If the user modifies the real capacity, the contingency capacity is reset to be the difference between the used capacity and the real capacity. A volume that is created without the autoexpand feature, and therefore has a zero contingency capacity, goes offline as soon as the real capacity is fully used and needs to expand.

Autoexpand does not cause the real capacity to grow much beyond the virtual capacity. The real capacity can be manually expanded to more than the maximum that is required by the current virtual capacity, and the contingency capacity is recalculated.

To support the auto expansion of thin-provisioned volumes, the storage pools from which they are allocated have a configurable capacity warning. When the used capacity of the pool exceeds the warning capacity, a warning event is logged. For example, if a warning of 80% has been specified, the event is logged when 20% of the free capacity remains.

A thin-provisioned volume can be converted nondisruptively to a fully allocated volume, or vice versa, by using the volume mirroring function. For example, you can add a thin-provisioned copy to a fully allocated primary volume, and then remove the fully allocated copy from the volume after they are synchronized.

The fully allocated to thin-provisioned migration procedure uses a zero-detection algorithm, so that grains containing all zeros do not cause any real capacity to be used.

Real-time Compression

Compressed volumes are a special type of volume where data is compressed as it is written to disk, saving additional space. To use the compression function, you must obtain the IBM Real-time Compression license. With the IBM Storwize V7000 Gen2 model (2076-524), you already have one compression acceleration adapter included in the base product, and you can get one more optional.

It is also suggested to upgrade your memory to 64 GB for best use of Real-time Compression. Enabling compression on Storwize V7000 Gen2 nodes does not affect non-compressed host-to-disk I/O performance. Like thin-provisioned volumes, compressed volumes have virtual, real, and used capacities. Use the following guidelines before working with compressed volumes:

- Real capacity is the extent space that is allocated from the storage pool. The real capacity is also set when the volume is created and, like thin-provisioned volumes, can be expanded or shrunk down to the used capacity.
- Virtual capacity is available to hosts. The virtual capacity is set when the volume is created and can be expanded or shrunk afterward.
- Used capacity is the amount of real capacity used to store client data and metadata after compression.
- Capacity before compression is the amount of client data that has been written to the volume and then compressed. The capacity before compression does not include regions where zero data is written to unallocated space.
- An I/O Group can contain a maximum of 200 compressed volumes and compressed volume mirrors.
- You can also monitor information about compression usage to determine the savings to your storage capacity when volumes are compressed. To monitor system-wide compression savings and capacity, select **Monitoring** → **System** and either select the system name or Compression View. You can compare the amount of capacity used before compression is applied to the capacity that is used for all compressed volumes.

  In addition, you can view the total percentage of capacity savings when compression is used on the system. Furthermore, you can also monitor compression savings across individual pools and volumes. For volumes, you can use these compression values to determine which volumes have achieved the highest compression savings.

Cache

The primary benefit of storage cache is to improve I/O response time. Reads and writes to a magnetic disk drive suffer from both seek and latency time at the drive level, which can result in from 1 millisecond (ms) - 10 ms of response time (for an enterprise-class disk).

The Storwize V7000 Gen2 nodes combined with Storwize family software V7.3 provide 32 GB memory per node, or 64 GB (128 GB) per I/O Group, or 256 GB (512 GB) per system. The Storwize V7000 Gen2 provides a semi-flexible cache model, and the node's memory can be used as read or write cache, either one as an I/O workload cache. The size of the write cache is maximally 12 GB of the node's memory. The remaining part of the memory is split between read cache allocation and compression allocation.

When data is written by the host, the preferred node saves the data in its cache. Before the cache returns completion to the host, the write must be mirrored to the partner node, or copied into the cache of its partner node, for availability reasons. After having a copy of the written data, the cache returns completion to the host. A volume that has not received a write update during the last two minutes will automatically have all modified data destaged to disk.

Note: Optional cache upgrade of 32 GB on Storwize V7000 Gen2 is reserved for RtC and it is not used when RtC is disabled.

Starting with Storwize V7000 Gen2 the concept of the cache architecture has been changed. Storwize V7000 Gen2 now distinguishes between upper and lower cache that enables the system to be more scalable:

- Required for support beyond 8192 volumes
- Required for support beyond 8 node clusters
- Required for 64-bit addressing beyond 28 GB
- Required for larger memory in nodes
- Required for more processor cores

- Required for improved performance and stability

The architectural overview is shown in Figure 1-7.

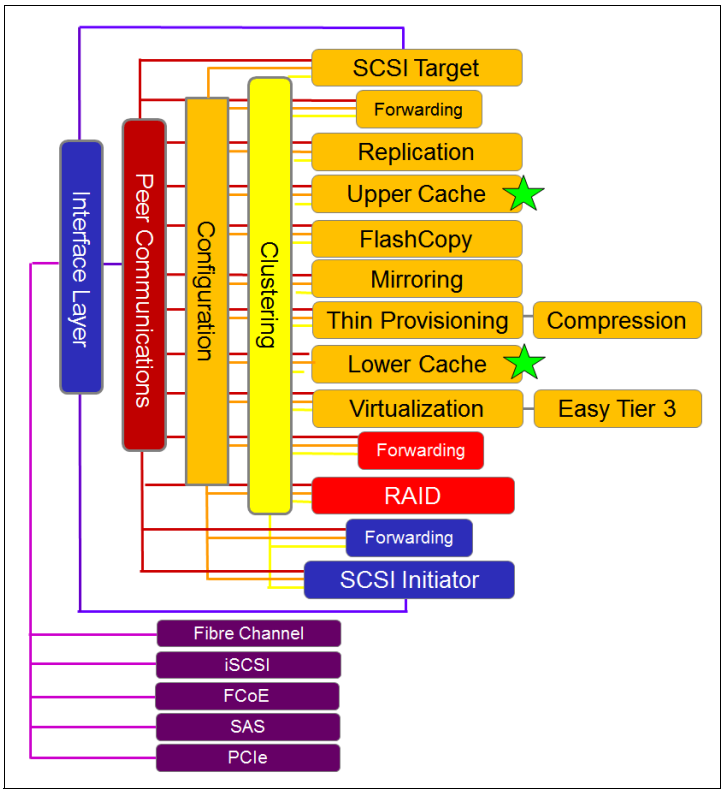


Figure 1-7 New cache architecture

If one node of an I/O Group is missing, due to a restart or a hardware failure, the remaining node empties all of its write cache and proceeds in operation mode, which is referred to as *write-through mode*. A node operating in write-through mode writes data directly to the disk subsystem before sending an `I/O complete` status message back to the host. Running in this mode can degrade the performance of the specific I/O Group.

Write cache is partitioned by storage pool. This feature restricts the maximum amount of write cache that a single storage pool can allocate in a system. Table 1-2 shows the upper limit of write-cache data that a single storage pool in a system can occupy.

Table 1-2 Upper limit of write cache per storage pool

One storage pool	Two storage pools	Three storage pools	Four storage pools	More than four storage pools
100%	66%	40%	33%	25%

Storwize V7000 Gen2 will treat part of its physical memory as non-volatile. *Non-volatile* means that its contents are preserved across power losses and resets. Bitmaps for FlashCopy and Remote Mirroring relationships, the virtualization table, and the write cache are items in the non-volatile memory.

In the event of a disruption or external power loss, the physical memory is copied to a file in the file system on the node's internal disk drive, so that the contents can be recovered when

external power is restored. The functionality of uninterruptible power supply units is provided by internal batteries, which are delivered with each node's hardware.

They ensure that there is sufficient internal power to keep a node operational to perform this dump when the external power is removed. After dumping the content of the non-volatile part of the memory to disk, Storwize V7000 Gen2 shuts down.

1.4 More information

This IBM Redbooks publication covers key aspects of the implementation of IBM Storwize V7000 Gen2, focused on version 7.3. For more detailed description about the concept of storage virtualization and how IBM uses V7000 and V7000 Gen2 in data centers, study the following materials:

- *Implementing the IBM Storwize V7000 V7.2*, SG24-7938
- *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- *IBM System Storage SAN Volume Controller and Storwize V7000 Replication Family Services*, SG24-7574
- *Implementing IBM Real-time Compression in SAN Volume Controller and IBM Storwize V7000*, TIPS1083
- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072

2

IBM Storwize V7000 Gen2 Hardware

In this chapter we describe the hardware changes introduced with IBM Storwize V7000 Gen2 which include the following topics:

- Controller and Expansion Enclosure hardware
- Scalable cache
- Performance
- PCIe technology
- Compression Accelerator card
- Support up to 20 Expansion Enclosures
- New integrated battery
- Light-emitting diodes (LEDs)
- New Technician Port

2.1 Overview

In conjunction with the release of the version 7.3 code, IBM introduced a hardware refresh for the IBM Storwize V7000 7.3 platform. In this chapter, we focus on the details of the specific hardware 2076-524 changes:

- Integrated battery pack
- Enhanced scalability and flexibility
- Expansion unit 2076 for 24 flash drives
- Improved Real-time Compression engine with hardware assistance

To meet these objectives, the base hardware configuration of the Storwize V7000 Gen2 was improved substantially to support more advanced processors, more memory and faster interconnects. This is also the first time that the IBM storage area network (SAN) Volume Controller platform and the Storwize platform share the same processors.

To learn more about the changes to the new Storwize V7000 Gen2, see Table 2-1.

Table 2-1 Storwize V7000 Gen1 Gen2 Comparison.

Feature	Storwize V7000 Gen1	Storwize V7000 Gen2
Canister node	2 * 4-core processor	Ivy Bridge 2 * 8-core processor and integrated hardware assisted compression acceleration
Expansion	Maximum of nine enclosures, 6 gigabit (Gb) serial-attached Small Computer System Interface (SCSI), called SAS	Maximum of 20 enclosures, 12 Gb SAS
SAS cable lengths available (meters)	1, 3, and 6	0.6, 1.5, 3, and 6
Drives	► 2076-112 Storwize V7000 control enclosure for 3.5-inch drives ► 2076-124 Storwize V7000 control enclosure for 2.5-inch drives ► 2076-312 Storwize V7000 control enclosure for 3.5-inch drives (with two 10 gigabits per second (Gbps) Internet SCSI (iSCSI)/Fibre Channel over Ethernet (FCoE) Ethernet ports) ► 2076-324 Storwize V7000 control enclosure for 2.5-inch drives (with two 10 Gbps iSCSI/FCoE Ethernet ports) ► 2076-212 Storwize V7000 Expansion Enclosure for 3.5-inch drives ► 2076-224 Storwize V7000 Expansion Enclosure for 2.5-inch drives	► 2076-524 Storwize V7000 control enclosure ► 2076-12F Storwize V7000 Expansion Enclosure for 3.5-inch drives ► 2076-24F Storwize V7000 Expansion Enclosure for 2.5-inch drives

Feature	Storwize V7000 Gen1	Storwize V7000 Gen2
Scaling	Up to 960 drives	Up to 1056
Capacity	1.92 petabytes (PB)	4 PB
Standard host interface	8 Gb FC, 1 Gb iSCSI	1 Gb iSCSI
Optional host interface	10 Gb iSCSI/FCoE	2x (8 Gb Fibre Channel (FC) or 10 gigabyte (GB) iSCSI/FCoE)
Random access memory (RAM) per Node Canister	8 GB	32 GB or 64 GB
Solid-state drive (SSD) support	Yes	Yes
Compression	No	Yes

The following items provide details that should help understand the changes made across both of the platforms to meet the goals:

- Processors

  IBM SAN Volume Controller DH8 and Storwize V7000 Gen2. Both platforms use the Ivy Bridge processors from Intel, which has eight cores.

- Memory

  32 GB or cache and compression with an option to have another 32 GB is added for Real-time Compression workloads.

- Peripheral Component Interconnect (PCI) Express (PCIe) technology

  Both the platforms have multiple PCIe Gen3 slots, as compared to dual PCIe Gen 2 slots in previous versions. This shift to PCIe Gen3 enables each PCIe lane to get a maximum speed of 1000 megabytes per second (MBps).

- Optional adapters

  In previous models, the only option that customers had was to add a dual port 10 Gbps converged network adapter. The Storwize V7000 Gen1 base model has dual port 1 Gbps adapters, plus quad port 8 Gbps FC adapters. In both of the new platforms, the base models come with three 1 Gbps Ethernet ports onboard. However, customers have an option to select multiple add-on adapters for driving host input/output (I/O) and off-loading compression workloads.

2.1.1 Scalable write cache

In Storwize V7000 Gen1, write cache was limited to 8 GB. However, Storwize V7000 Gen2 includes a 32 GB cache, with an optional upgrade of up to 64 GB (in which the additional 32 GB is available for Real-time Compression workloads).

2.1.2 Increase in performance

IBM Storwize V7000 Gen1 came with a standard quad port 1 Gbps Ethernet onboard adapter and quad port 8 Gbps FC adapter. It enabled clients to add optional dual port 10 Gbps converged network adapter (CNA) for running FCoE and 10Gb iSCSI traffic. So, with the optional card in place, clients would get a 4 Gbps pipe from 1 Gbps Ethernet ports, 32 Gbps of pipe from the FC adapter, and an additional 20 Gbps of pipe from the CNA.

Storwize V7000 Gen2 includes, standard, three 1 Gbps Ethernet onboard ports (the fourth 1 Gbps Ethernet port is dedicated as a Technician Port). However, due to the higher form factor (2U), it enables inserting three cards per enclosure. The system board has three PCIe Gen3 slots, out of which two slots can be used to install either quad port 8 Gbps FC adapters or one quad port 10 Gbps CNA. This effectively increases the available pipe to 64 Gbps when using the same 8 Gbps FC adapter. For the 10 Gbps CNA, the pipe increases to 40 Gbps.

The increase in the performance of input/output operations per second (IOPS) is due to the two-fold increase in the disks that can be attached behind a Gen 2 Control Enclosure.

2.1.3 Support for 20 Expansion Enclosures

Storwize V7000 Gen2 Control Enclosure can attach more than two times of disk possible with Storwize V7000 Gen1.

Storwize V7000 Gen1 supported a maximum of nine Expansion Enclosures, each with 24 Small Form Factor (SFF) drives. This enables for a maximum of 240 drives per controller.

Storwize V7000 Gen2 supports having a maximum of 20 Expansion Enclosures allowing a maximum of 504 drives per controller. The system can cluster with up to four Control Enclosures, which enables a large storage space provided by 84 Enclosures (80 Expansion Enclosures plus four Control Enclosures).

The design of Storwize V7000 Gen2 is geared toward making the platform more scalable, flexible, and gives higher performance without using more space in customer's data centers. Each Storwize V7000 Gen2 has the ability to handle maximum capacity up to 4 PB with the capability to virtualize external storage, and to enable storage administrators to provide more bandwidth for applications by adding more I/O adapters, memory, and quick assist card.

2.2 New hardware on the Storwize V7000 Gen2

In this section we go over the Storwize V7000 Gen2 new hardware:

- Controller hardware
- Base configuration
- Controller SAS topology
- Compression Accelerator card
- HIC
- Expansion Enclosure

2.2.1 Controller hardware overview

The IBM Storwize V7000 Gen2 storage engine 2076-524 and Expansion Enclosure Model 12F or 24F deliver increased performance, expanded connectivity, compression acceleration, and additional internal flash memory capacity.

Storwize V7000 Gen2 SFF Control Enclosure Model 524 includes the following components:

- Two node Enclosures, each with an eight-core processor and integrated hardware-assisted compression acceleration
- 64 GB cache (32 GB per Enclosure) with optional 128 GB cache (64 GB per Enclosure)
- 8 Gb FC, 10 Gb Ethernet, and 1 Gb Ethernet ports for FC, iSCSI, and FCoE connectivity

- Optional Compression Accelerator card for additional hardware-assisted compression acceleration
- 12 Gb SAS ports for Expansion Enclosure attachment
- Twenty-four slots for 2.5-inch SAS drives
- 2U, 19-inch rack mount Enclosure with alternating current (ac) power supplies

The front and back views of the two-node cluster based on the 2076-524 are shown in Figure 2-1.

Figure 2-1 Front and rear of controller

The Storwize V7000 Gen2 brings with it several significant changes and enhancements over the previous generation hardware. The IBM Storwize V7000 Gen2 2076-524 includes preinstalled V7.3 software.

Be aware that it is not supported to downgrade the software to version 7.2 or earlier, and the Storwize V7000 Gen2 will reject any attempt to install a version earlier than 7.3. See the following links for integration with existing clustered systems, compatibility, and interoperability with installed nodes and other system components:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003850
http://www.ibm.com/support/docview.wss?rs=591&uid=ssg1S1003705

2.2.2 Base configuration

A detailed rear view of the Storwize V7000 Gen2 Control Enclosure is shown in Figure 2-2.

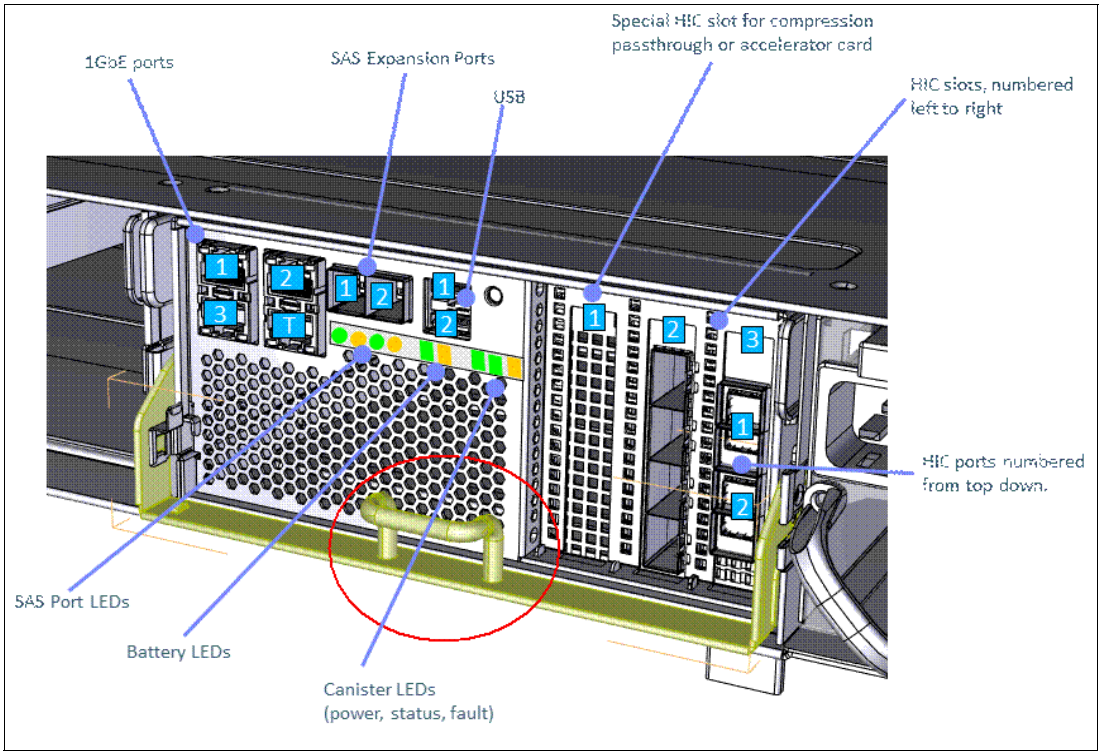


Figure 2-2 Rear of Control Enclosure

All models are delivered in a 2U, 19-inch rack mount Enclosure and include a three-year warranty with customer-replaceable unit (CRU) and onsite service. Optional warranty service upgrades are available for enhanced levels of warranty service.

With the migration to the 2U platform and integration of previous external components into a single chassis, no more rack space is required. The Storwize V7000 Gen2 integrates a redundant battery backup system, therefore eliminating the need for external rack-mount uninterruptible power supply (UPS), optional power switch, and related cabling.

The IBM Storwize V7000 Gen2 seamlessly integrates into the existing infrastructure, and enables nondisruptive hardware upgrades from previous generations. We start with the base configuration of the following components:

- Two node Enclosures, each with the following components:
  - Three 1 Gbps Ethernet ports (iSCSI and management)
  - 1 Gbps technician Ethernet port for immediate availability and emergency access
  - 8-core Intel processor
  - 32 GB RAM
  - 2 x 12 Gbps SAS drive expansion ports
  - Hardware compression
  - Three expansion slots (2 x Host Interface, 1 x for Compression)
  - Two Universal Serial Bus (USB) ports for debug and emergency access
  - Battery
- SAS and Nearline-SAS hard disk drives (HDDs), and Flash drives

- Up to 20 Expansion Enclosures per Control Enclosure, which can be 2U12 or 2U24:
  - Ten Expansion Enclosures on each of two chains
  - Maximum 21 Enclosures, maximum 21*24=504 drives, per Control Enclosure
- Up to four Control Enclosures per clustered system
- Maximum of 1056 drives per clustered system (cannot fully populate every I/O Group)

The base card contains the following major components of Storwize V7000 Gen2:

- Intel Ivy Bridge 64-bit processor (see Table 2-2)

Table 2-2 IVY Bridge processor

Model 2076-524	
Intel designator	E5-2628Lv2
Number of cores	8
Core frequency	1.9 gigahertz (GHz)
Lower-level cache	20 MB
Double data rate (DDR) memory speed	1600 megahertz (MHz)
QuickPath Interconnect (QPI) speed	7.2 gigatransfers per second (GT/s)
Power consumption	70 W

- One dual inline memory module (DIMM) slot per channel for registered error correction code (ECC) DDR3L DIMMs running at 1600 megatransfers per second (MT/s):
  - Support for 32 GB total using commodity memory DIMMs of 8 GB
  - Support for 64 GB total using 16 GB DIMMs
  - Memory support follows the processor-based model designations:
    - DIMM size supported 8 GB DDR3, 16 GB DDR3
    - Base memory fitted 16 GB, 32 GB
    - Optional additional memory +16 GB, +32 GB
- Intel Platform Controller Hub (PCH)
- Coleto Creek 8926 stock keeping unit (SKU) with hardware assists for compression
- Serial peripheral interface (SPI) flash memory to power-on self-test (POST)
- Onboard SSD for boot, Linux file system, and Fire-Hose Dump (FHD):
  - The SSD capacity of 64 GB with a sustained bandwidth of 50 MBps for sequential reads and sequential write to support the FHD
  - Multi-level cell (MLC) flash technology

Definition: The node Enclosure, which contains a battery, is able to power the Enclosure while it stores cache and system data to an internal drive in the event of a power failure. This process is referred to as a *Fire-Hose Dump*.

- Integrated, battery-backed, Real Time Clock (RTC), and complementary metal-oxide semiconductor (CMOS) memory for Basic Input/Output System (BIOS) configuration settings, with 10 years battery life
- Two USB 2.0 ports on the back-panel to attach a serial console or memory stick for initial configuration, dumps, and so on

- Four 10/100/1000 Ethernet ports for management:
  - Registered Jack-45 (RJ45) connectors on the back panel
  - Single root I/O virtualization (SR-IOV) support
- A direct inter-integrated circuit (I2C) link between the Intel processor and the SCSI Enclosure Services (SES) processor to enable communications for hardware management purposes
- PCIe switch
- Host Bus Adapter (HBA) slots:
  - Two slots for HBAs (implemented as Enclosure Slots 2 &3)
  - 8x PCIe Gen 3 link from the processor to each slot
- SAS 3.0 controller:
  - PMC-Sierra SPCve 16 x 12G (PM8073) chip
  - 8x PCIe Gen 3 link from the processor
  - T10 data integrity field (DIF) support for block sizes of 512, 520, 4096, or 4160 bytes (excluding the 8-byte DIF)
  - 2-off 4x links to Mini-SAS half-duplex (HD) connectors (SFF-8644) on the back panel for Expansion Enclosures
  - 8x link to the SAS expander, using 4 Physical Layer interfaces (PHYs) from each of the two reduced instruction set computer (RISC) cores to ensure even performance distribution
- SAS 3.0 expander:
  - PMC-Sierra SXP 12G rev B PM8054, 36 ports
  - Speed-matching buffers for optimum performance with 6G devices
  - 8x link to the SAS controller
  - 1x links to the midplane for the internal drive slots
  - Two Wire Interface (TWI) busses 1 and 2 to the midplane to access the vital product data (VPD) erasable programmable read-only memory (EPROMs) and the VPD and status of the power supplies
  - TWI bus 3 used to access Enclosure complex programmable logic device (CPLD) and temperature sensors
  - Enclosure management processor and firmware with external memories
- A Hardware Expansion socket for a custom expansion card
  - Implemented as an x16 PCIe add-in card socket:
    - 8x PCIe Gen 3 link from the processor
    - A pass-through card, which passes the PCIe bus to the on-board Coleto Creek chip
    - A Compression Accelerator card, to provide additional compression and decompression performance
- DC power input from the midplane:
  - An early power-off warning (EPOW) provided for Storwize V7000 when ac power fails

2.2.3 Controller SAS Topology

The 16 SAS lanes from the SPCve chip in the Node Enclosure are divided into four groups of four lanes each. Two groups of four lanes are routed to the internal SAS expanders for local

drive attach. This provides maximum bandwidth for supporting local SSDs. The other two groups of four lanes are each routed to one of the external Mini-SAS HD connectors for connection to Expansion Enclosures.

Dividing the attached Expansion into two separate chains of Expansion Enclosures helps to reduce the resources required to set up a connection and improve network stability.

Internally, the SPCve uses two RISC cores. Each one is responsible for a group of eight SAS lanes. To provide a balanced network, four lanes from a RISC group must be routed to an external port, and the other four lanes to the internal expander.

2.2.4 Compression Accelerator card

The optional Compression Accelerator card is fitted in Node Enclosure PCIe Slot 1. The Compression Accelerator card conforms to the form-factor for a half-length, low-profile PCIe card. It uses a 16-lane PCIe edge finger connector.

The card has two chips:

- An Intel Coleto Creek (8926 SKU without encryption support). This is identical to the chip used as the base board PCH, but here only the Compression function is used.
- A 24-lane PLX PCIe switch. This is configured as three 8-lane busses:
  - A Gen3 bus from the Intel processor
  - A Gen2 bus to the on-card Coleto Creek chip
  - A Gen2 bus passed back through the rear 8 lanes of the PCIe socket

2.2.5 Host interface cards

The base node can be configured with I/O ports according to application needs by fitting one or more host interface cards (HICs). Each HIC attaches using an 8x PCIe Gen 3 bus, which also enables access to its VPD. HIC cards are fitted in Node Enclosure PCIe Slots 2 and 3.

The HIC slots support two card form factors:

- HIC slot 1 is for compression, and contains one of the following cards:
  - A compression pass-through card to enable the base hardware compression
  - A Compression Accelerator card, which adds to the base hardware compression
- The other HIC slots are flexible and independent of RAM, processor, and compression:
  - The first HIC slot must be populated at ordering time.
  - The second HIC slot can be optionally populated at any time.

Enclosure and 12V power supplies to the HIC card are turned off during FHD to preserve battery life. However, during the 5-second ride-through period, the supply remains within normal PCIe tolerances.

Fire-Hose Dump: When the system detects loss of input power, it enters a Ride-Through mode, which lasts for 10 seconds. If input power returns during this time, the system does not power off. Power is immediately cut to the drives, which have to be powered on again. If input power is still unavailable after the 10-second ride-through period, the software places the hardware into a low power-consumption mode, stops the I/O process, and performs the FHD.

After the FHD is complete, the node powers off. The node restarts automatically when input power returns. If this is after the ride-through period but before the end of the FHD, the node still shuts down but immediately restarts.

The following HICs are supported:

- 0 - 2 4-port 8 Gbps FC adapters
- 0 - 1 4-port 10 Gbps Ethernet adapter for iSCSI, FCoE, and Internet Protocol (IP) replication

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

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

2.2.6 Expansion Enclosure

The Expansion Enclosure contains the following major components:

- SAS 3.0 expander:
  - PMC-Sierra SXP 12G chip
  - Speed-matching buffers for optimum performance with 6G devices
  - 1x links to the midplane for the internal drive slots
  - 2-off 4x Mini-SAS HD connectors on the back-panel:
    - One port is connected to the upstream enclosure.
    - The remaining port is available for attaching a downstream enclosure.
  - TWI busses 1 and 2 to the mid-plane to access the VPD EPROMs and the VPD and status of the power supplies
  - TWI bus 3, used to access canister CPLD and temperature sensors
  - Enclosure management processor and SES firmware with external memories:
    - 16 MB SPI Flash with cyclic redundancy check (CRC) protection
    - 8 MB static random access memory (SRAM)
    - At least 64 KB non-volatile SRAM (NVSRAM)
- RS232 port on the back panel (3.5 millimeter (mm) stereo jack). This is used for configuration during manufacturing.

2.2.7 SES firmware

The Storwize code manages the enclosures over the SAS fabric using SCSI-3 Enclosure Services (SES) commands. Each SAS expander has an integrated management processor and associated SES firmware.

IBM Storwize V7000 Gen2 uses the IBM SES firmware. IBM requires the Object Data Manager (ODM) to provide details of the hardware interfaces and fan control algorithms and assist with resolving integration problems.

2.3 Expansion Enclosures

There are two types of available Expansion Enclosures, Storwize V7000 large form factor (LFF) Expansion Enclosure Model 12F and small form factor (SFF) 24F.

Storwize V7000 Gen2 LFF 12F includes the following components:

- Two expansion canisters
- 12 Gb SAS ports for control enclosure and Expansion Enclosure attachment
- Twelve slots for 3.5-inch SAS drives
- 2U, 19-inch rack mount enclosure with ac power supplies

Storwize V7000 SFF Expansion Enclosure Model 24F includes the following components:

- Two expansion canisters
- 12 Gb SAS ports for control enclosure and expansion enclosure attachment
- Twenty-four slots for 2.5-inch SAS drives
- 2U, 19-inch rack mount enclosure with AC power supplies

The Expansion Enclosure is a 2U enclosure, containing the following components:

- 24 2.5 in. drives (HDDs or SSDs).
- 2 Storage Bridge Bay (SBB)-compliant enclosure services manager (ESM) canisters.
- 2 fan assemblies. These mount between the drive midplane and the Node Canisters. Each fan module is removable when the Node Canister is removed.
- 2 Power supplies.
- RS232 port on the back panel (3.5 mm stereo jack). This is used for configuration during manufacturing.

The front of an Expansion Enclosure is shown in Figure 2-3.

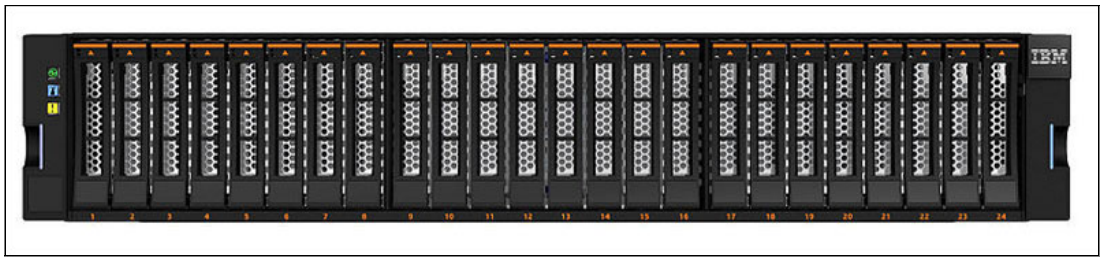


Figure 2-3 Front of Expansion Enclosure

Figure 2-4 shows a rear view of an enclosure.

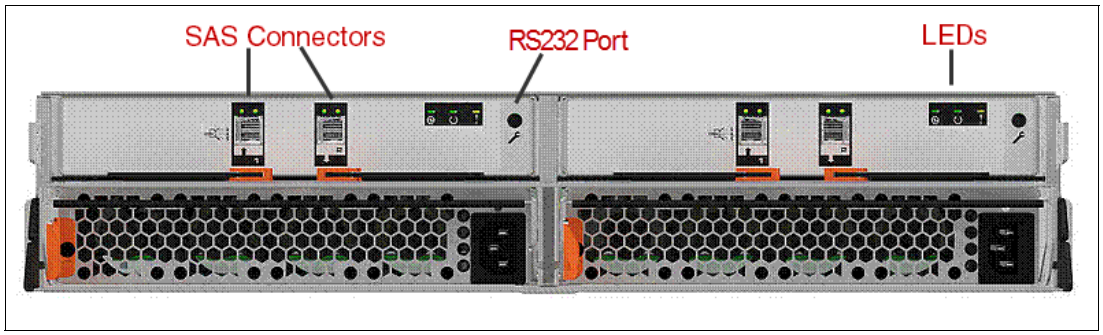


Figure 2-4 Expansion Enclosure rear view

Restriction: The Storwize V7000 Gen2 expansion enclosures can only be used with a Storwize V7000 Gen2 control enclosure. The Storwize V7000 Gen1 expansion enclosures cannot be used with a Storwize V7000 Gen2 control enclosure.

2.3.1 Control enclosure midplane

The midplane is a printed-circuit board, which provides power and signal connections. It contains the following blind-mating connectors:

- One SFF 8680 connector for each drive slot:
  - Each slot has two 1x SAS 3.0 links, one to each canister.
  - Each drive has independent power control, as supported in the SAS-3 standard (using pin P3). Midplane power field-effect transistors (FETs) per drive are not required. Power control is used to recover a hung drive by power cycling.
- Two canister connectors
- Connectors for each power supply

2.3.2 Battery backup

To retain space requirements for the IBM Storwize V7000 Gen2 and its associated components within two rack units, and to simplify the cabling and rack layout, the Storwize V7000 Gen2 node integrates the uninterruptible power supplies inside the node chassis. The control enclosure provides battery backup to support a non-volatile write cache and protect persistent metadata.

In the event of a power failure, each node performs an independent FHD from memory to the on-board SSD under software control. This method is compatible with the existing Storwize code, and it retains the data indefinitely. The memory is not required to be persistent across a node reset.

The drives do not require battery backup, so the control enclosure removes power from the drive slots immediately before the EPOW expires (approximately within 5 ms after an ac failure). This function is implemented in hardware. Battery packs are not required in Expansion Enclosures.

The battery pack powers the processor and memory for a few minutes while the Storwize code copies the memory contents to the onboard SSD. The FHD code runs on a single processor core to minimize power requirements. The I/O chips and HIC slot are powered down to save energy. The fans run with the node components remaining within thermal limits.

Each node switches from ac power to battery and back again without interruption. The battery supports a 5-second ride-thru delay with all node electronics active before the dump starts, in case power comes back quickly. When started, the dump always runs to completion. To allow the system to be brought online immediately after a longer power outage, the total energy stored in each battery pack supports two consecutive cycles without an intervening recharge. Each cycle incorporates a ride-thru delay plus a 16 GB FHD.

The onboard SSD has a high write bandwidth to allow the dump to complete quickly. For example, if the SSD has a sustained write bandwidth of 100 MBps, each battery pack would need to power its node canister for just under three minutes.

The operational status of batteries and their VPD are available from the IBM Storwize V7000 Gen2 command-line interface (CLI) using the **`saninfo lsservicestatus`** command.

2.3.3 Battery Pack

This is a CRU containing a rechargeable battery pack and management electronics. The Storwize code running on the Intel CPU reads the battery status and VPD directly using an I2C interface.

The high-discharge-rate battery solution uses Lithium Nickel Manganese Colbalt Oxide (NMC) cells (Sony US18650VTC4), in a three cells in a series made up of two cells connected in parallel (3S2P) configuration. The nominal battery capacity is derated to allow for operating temperature and degradation at end of life. The batteries are tested for conformance with IBM and agency standards for safety.

The battery packs recharge from flat to 98% capacity in one hour or less. Provisions are made to monitor their state with an electronic “gas gauge,” and should be tested by periodically discharging one battery pack at a time.

Each battery pack is maintainable concurrently with system operation. The battery packs should be replaced as required, with an average life expectancy of at least five years assuming a one FHD cycle per month.

2.3.4 Control Enclosure power supplies

Each Control Enclosure contains redundant hot-pluggable, n+1 redundant power supplies. The design point for Storwize V7000 has power supplies that do not incorporate system cooling fans. The following list describes each enclosure hardware specification:

- Single-phase ac input, with 90V - 264V and 47 Hz - 63 Hz, or 180V - 240V direct current (dc) input via standard International Electrotechnical Commission (IEC) grounded connector (auto-ranging):
  - On/off switch (preferred)
  - Power factor correction
  - 20 ms minimum error-free ride-through time
  - 5 ms minimum EPOW (after the error-free ride-through period)
- +5V redundant STANDBY_PWR @ 2.5A.
- +12V regulated dc output to each canister (supports current sharing).
- 200 watts minimum for each canister in the control enclosure.
- +5V and +12V regulated dc outputs to the drive slots (supports current sharing):
  - 14 watts total per SFF drive slot.
  - The drives spin-up in groups to minimize the surge at power-on.
  - The power system tolerates a short-circuit on any single drive slot without damage. All other dc voltages remain within tolerance. The operation of other drives and the canisters are not affected.
- Energy Star Climate Saver Gold certification:
  - 90% efficiency at 50% of rated load
  - Input power monitoring
- TWI link to midplane for VPD and management, including fault reporting, fan control, and monitoring output voltage, output current, and temperature.
- All enclosures power back on automatically after an AC interruption.
- Input power monitoring.

- Output voltage and current monitoring.
- Output voltage margining.
- Three rear LED indicators for input ok, output ok, and fault. (More information in Table 2-4 on page 37).

2.3.5 Control enclosure fan modules

The Storwize V7000 Gen2 control enclosure has a new component called a *Fan Module*. This part contains the fans that cool the system.

The Storwize V7000 Gen2 has fan cooling modules, which are not housed within the power supplies. Rather, they sit between the Node Canisters and the midplane. The reason for the Fan Module being separated is so that when the canister is removed, the fan continues to cool drives. There are two of these per enclosure, which means that you have to pull out the canister to service the fan assembly after first removing the relevant Node Canister.

There are two cam levers to eject the Fan Module, which are accessible when the Node Canister is removed. The Fan Module must be reinserted into the Storwize V7000 Gen2 within 3 minutes of removal to maintain adequate system cooling.

The Fan Module is shown in Figure 2-5.

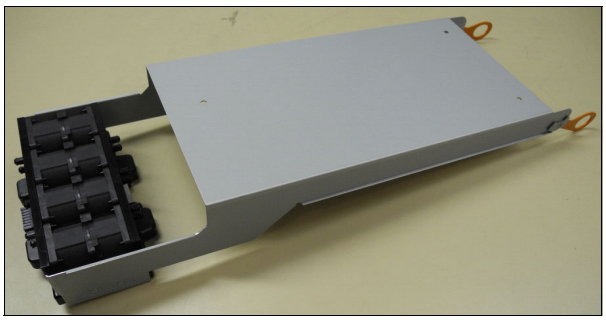


Figure 2-5 Fan Module

Each Storwize V7000 Gen2 control enclosure contains two Fan Modules for cooling. Each Fan Module contains eight individual fans in four banks of two, as shown in Figure 2-6.

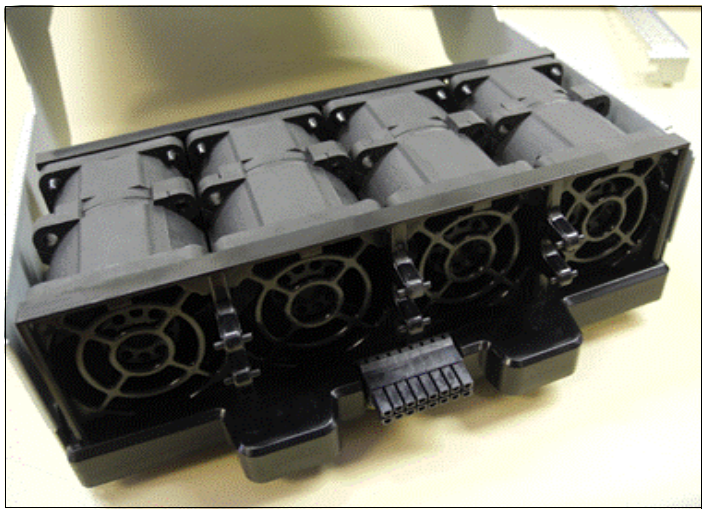


Figure 2-6 Fan Module

The Fan Module as a whole is a replaceable component, but the individual fans are not. There is a new CLI command, **`lsenclosurefanmodule`**, which is explained in detail in Chapter 8, “IBM Storwize V7000 Gen2 command-line interface” on page 143.

2.3.6 Compression Accelerator card

Compressed volumes are a special type of volume where data is compressed as it is written to disk, saving additional space. To use the compression function, you must obtain the IBM Real-time Compression license and the hardware level for both nodes within the I/O group of the Storwize V7000 Gen2, for that I/O group to support compression. Refer to Chapter 3, “Planning and configuration” on page 47 for more information.

Storwize V7000 Gen2 nodes must have two processors, 64 GB of memory, and at least one *Compression Accelerator* card installed to use compression. Enabling compression on Storwize V7000 Gen2 nodes does not affect non-compressed host-to-disk I/O performance. The compression card is shown in Figure 2-7.

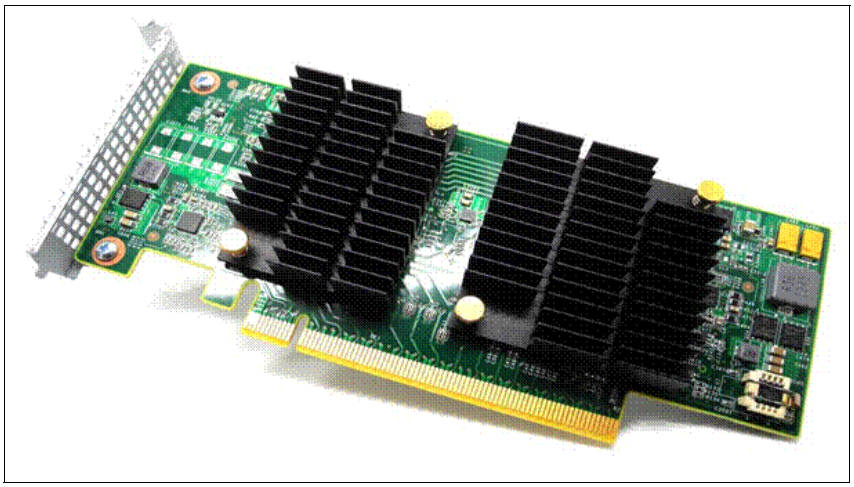


Figure 2-7 Compression accelerator card

The Compression Accelerator card conforms to the form-factor for a half length, low-profile PCIe card. It uses a 16-lane PCIe edge finger connector, although the usage of this is non-standard. The card has two chips:

- An Intel Coleto Creek (8926 SKU without encryption support) used for compression functions only.
- A 24-lane PLX PCIe switch. This is configured as three 8-lane busses as follows:
  - A Gen3 bus from the Intel CPU
  - A Gen2 bus to the on-card Coleto Creek chip
  - A Gen2 bus passed back through the rear 8 lanes of the PCIe socket

2.3.7 FCoE and iSCSI

The new 4x port 10 gigabit Ethernet (GbE) card is only supported in the IBM System Storage SAN Volume Controller, 2145-DH8, and in the Storwize V7000 Gen2, 2076-524.

In Storwize V7000 Gen2, we support 1 x 10 GbE adapter in each of the platforms. Only IBM-supported 10 Gb SFPs are used. Each adapter port has amber and green colored LEDs to indicate port status.

FCoE frame routing should be done by an FCoE switch. Storwize V7000 Gen2 makes it possible to use both FCoE and iSCSI protocols simultaneously on the same 10 Gb port. For best practices, however, it is suggested that separate ports be used for each protocol.

iSCSI is an alternative means of attaching hosts to the Storwize V7000 Gen2. All communications with back-end storage subsystems, and with other Storwize V7000 Gen2, only occur through an FC connection.

The iSCSI function is a *software function* that is provided by the Storwize software, and not hardware. Refer to the IBM Knowledge Center for Storwize V7000 7.3 for iSCSI for more information:

http://www.ibm.com/support/knowledgecenter/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/fab1_hic_installing.html?lang=en

2.4 LEDs

LED indicators on the system let you know the status of the system. The indicators have changed from previous generations of the Control Enclosure models. We are going to discuss Storwize V7000 Gen2, which refers to the newer generation of enclosures in Table 2-3 and their respective LEDs and meanings.

Table 2-3 LED Control Enclosure

Machine type and model	Description
2076-524 Storwize V7000	Control Enclosure
2076-12F Storwize V7000	Expansion Enclosure for 3.5-inch drives
2076-24F Storwize V7000	Expansion Enclosure for 2.5-inch drives

2.4.1 Battery LED

Storwize V7000 node canisters cache volume data and hold state information in volatile memory. If power to a node canister fails, the node canister uses battery power to write cache and state data to its boot drive. The battery is maintained in a fully charged state by the battery subsystem. At maximum power, the battery can save critical data and state information in two back-to-back power failures.

If power to a node canister is lost, saving critical data starts after a 5-second wait. (If the outage is shorter than five seconds, the battery continues to support the node and critical data is not saved.) The node canister stops handling I/O requests from host applications. The saving of critical data runs to completion, even if power is restored during this time. The loss of power might be because the input power to the enclosure is lost, or because the node canister is removed from the enclosure.

When power is restored to the node canister, the system restarts without operator intervention. How quickly it restarts depends on whether there is a history of previous power failures. The system restarts only when the battery has sufficient charge for the node canister to save the cache and state data again. A node canister with multiple power failures might not have sufficient charge to save critical data. In such a case, the system starts in service state and waits to start I/O operations until the battery has sufficient charge.

Two LED indicators denote the state of the battery:

- Status LED: Green
- Fault LED: Amber

Figure 2-8 shows the Control Enclosure LEDs that indicate battery status.

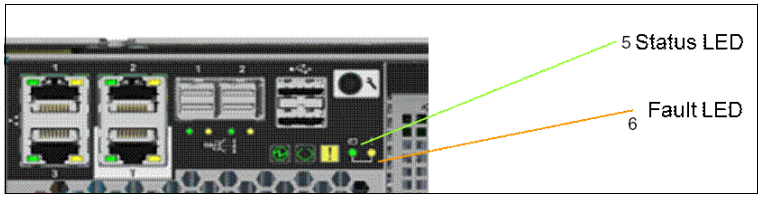


Figure 2-8 Controller battery LEDs

A detailed view of the system state is provided in the Monitoring sections of the management GUI, and by the service assistant. If neither the management GUI nor the service assistant is accessible, use this table to determine the system status using the LED indicators on the Control Enclosures.

The system status LEDs visible at the rear of each control enclosure can show one of several states, as described in Table 2-4.

Table 2-4 Storwize V7000 2076-524 battery status LEDs

Name	Callout	Color/State	Meaning
Battery Status	5	Green/OFF	Indicates that the battery is not available for use (for example, battery is missing or there is a fault in the battery).
		Green/Fast flash	The battery has insufficient charge to perform an FHD.
		Green/Slow flash	The battery has sufficient charge to perform a single FHD.
		Green/ON	The battery has sufficient charge to perform at least two FHDs.
Battery Fault	6	Amber/OFF	No fault. An exception to this would be where a battery has insufficient charge to complete a single FHD.
		Amber/ON	There is a fault in the battery.

2.4.2 Control Enclosure LED

Each Storwize V7000 2076-524 node canister has indicator LEDs that provide status information about the canister. Figure 2-9 shows the front of the Control Enclosure where three lights are displayed. These are discussed in Table 2-5.

Figure 2-9 Front Control Enclosure LEDs

Table 2-5 Front Control Enclosure LEDs description

Name	Description	Color	Symbol
Power	Indicates whether the Control Enclosure has power: ► If the LED is on, the enclosure has power. ► If the LED if off, the enclosure does not have power.	Green	
Status	Indicates whether the Control Enclosure is active: ► If the LED is on, the enclosure is active. ► If the LED is off, the enclosure is not active. ► If the LED is flashing, there is a VPD error.	Green	
Fault	Indicates whether a fault is present, and identifies the Control Enclosure: ► If the LED is on, a fault exists. ► If the LED is off, no fault exists. ► If the LED is flashing, the Control Enclosure is being identified. This status might or might not be a fault.	Amber	

Using the callout numbers in Figure 2-10, refer to the following tables for a listing of the Storwize V7000 2076-524 node canister LEDs, and a description of the meaning of the LED activity.

Figure 2-10 shows the rear Control Enclosure LEDs and their meanings.

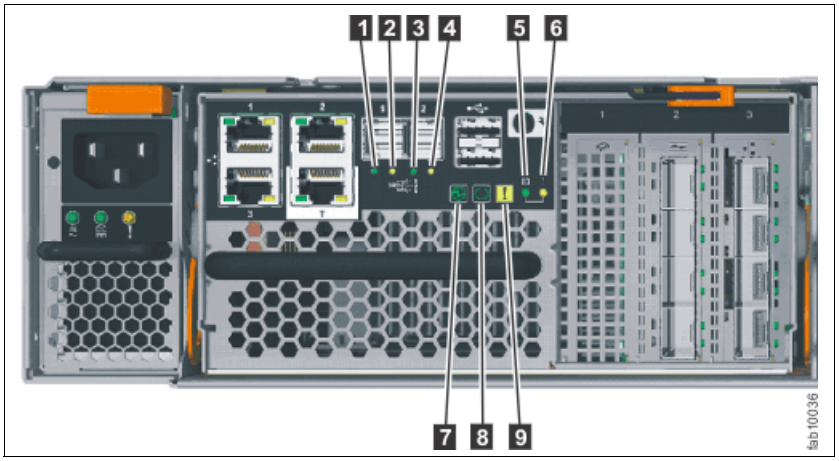


Figure 2-10 Rear Control Enclosure LED

Table 2-6 provides a cross-reference back to Figure 2-10 on page 39.

Table 2-6 LEDs for SAS ports 1 and 2 Indicators for Control Enclosure

Name	Call out	Symbol	Color/State	Meaning
SAS Port 1 Link	1	None	Green/OFF	No link connection on any PHYs. The connection is down.
			Green/ON	There is a connection on at least one PHY. At least one of the PHYs to that connector is up.
SAS Port 1 Fault	2	None	Amber/OFF	No fault. All four PHYs have a link connection.
			Amber/ON	This can indicate a number of different error conditions: ► One or more, but not all, of the four PHYs are connected. ► Not all four PHYs are at the same speed. ► One or more of the connected PHYs are attached to an address different from the others
SAS Port 2 Link	3	None	Green/OFF	No fault. All four PHYs have a link connection.
			Green/ON	There is a connection on at least one PHY. At least one of the lanes to that connector is up.
SAS Port 2 Fault	4	None	Amber/OFF	No fault. All four PHYs have a link connection.
			Amber/ON	This can indicate a number of different error conditions: ► One or more, but not all, of the four PHYs are connected. ► Not all four PHYs are at the same speed. One or more of the connected PHYs are attached to an address different from the others.

For Callouts 5 and 6, refer to Table 2-4 on page 37.

Table 2-7 provides a cross-reference to Figure 2-10 on page 39.

Table 2-7 Storwize V7000 2076-524 node canister system status LEDs

Name	Callout	Color/State	Meaning
Power	7	Green/OFF	No power is available or power is coming from the battery.
		Green/Slow flash	Power is available but the main CPU is not running. This is called *standby mode*.
		Green/Fast flash	In Self Test mode.
		Green/ON	Power is available, and the system code is running.
Status	8	Green/OFF	The system code has not started. The system is off, in standby, or self test.
		Green/Blink	The canister is in candidate or service state. It is not performing I/O. It is safe to remove the node.
		Green/Fast flash	The canister is active, able to perform I/O, or starting.
		Green/ON	The canister is active, able to perform I/O, or starting. The node is part of a cluster.
Canister Fault	9	Amber/OFF	The canister is able to function as an active member of the system. If there is a problem on the node canister, it is not severe enough to stop the node canister form performing I/O.
		Amber/Blink	The canister is being identified. There might or might not be a fault condition.
		Amber/ON	The node is in service state or an error exists that might be stopping the system code from starting. The node canister cannot become active in the system until the problem is resolved. You must determine the cause of the error before replacing the node canister. The error may be due to insufficient battery charge; in this event, resolving the error simply requires waiting for the battery to charge.

To understand in more detail the status of the I/O port at the rear of a control enclosure, refer to the topic about Storwize V7000 2076-524 node canister ports and indicators that is linked at the IBM Knowledge Center on the following website:

http://www.ibm.com/support/knowledgecenter/api/content/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/tbrd_sysstsleds.html

2.4.3 Expansion Enclosure LED

Each Storwize V7000 Gen2 Expansion Enclosure has three LEDs that provide status and identification for the Expansion Enclosure and two SAS ports.

Figure 2-11 shows the Expansion Enclosure LEDs.

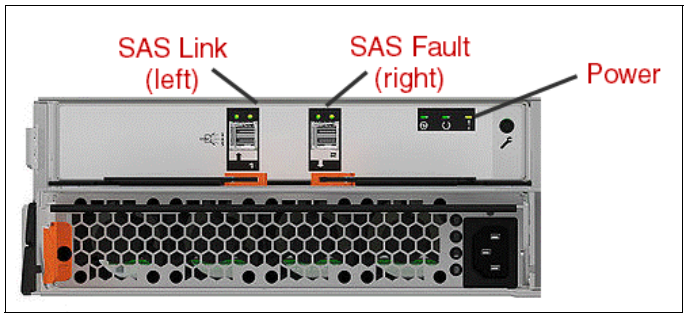


Figure 2-11 SAS ports and Power LEDs at rear of Expansion Enclosure

Three LEDs are in a horizontal row on the right side (when viewed from the rear) of the expansion canister, and there are two SAS Link ports on either side of the center.

Figure 2-12 shows the Expansion Enclosure LEDs.

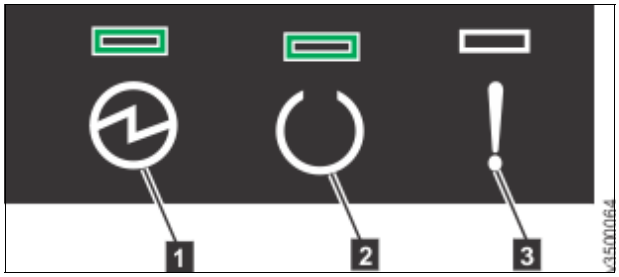


Figure 2-12 Expansion Enclosure LED

Table 2-8 cross-references Figure 2-12 on page 42.

Table 2-8 Expansion Enclosure LED descriptions

Callout/Name	Description	Color	Symbol
1. Power	Indicates whether the Expansion Enclosure has power: ► If the LED is on, the enclosure has power. ► If the LED is off, the enclosure does not have power.	Green	
2. Status	Indicates whether the Expansion Enclosure is active: ► If the LED is on, the enclosure is active. ► If the LED is off, the enclosure is not active. ► If the LED is flashing, there is a vital product data (VPD) error.	Green	
3. Fault	Indicates whether a fault is present, and identifies the Expansion Enclosure: ► If the LED is on, a fault exists. ► If the LED is off, no fault exists. ► If the LED is flashing, the Expansion Enclosure is being identified. This status might or might not be a fault.	Amber	

Two SAS ports are in the rear of the Storwize V7000 Gen2 Expansion Enclosure.

SAS ports are identified at the bottom of the port, with 1 on the left and 2 on the right, as shown in Figure 2-11 on page 42. Use of port 1 is required. Use of port 2 is optional. Each port connects four data channels.

Table 2-9 describes LED states for each of the two LEDs per SAS port. The link LED is on the left of each set of ports.

Table 2-9 SAS port LEDs on the Expansion Enclosure

Name	Color	State	Meaning
SAS Port 1 Link	Green	OFF	No link connection on any PHY. The connection is down.
		ON	There is a connection on at least one PHY. At least one of the PHYs to that connector is up.
SAS Port 1 Fault	Amber	OFF	No fault. All four PHYs have a link connection.
		ON	This can indicate a number of different error conditions: ► One or more, but not all, of the four PHYs are connected. ► Not all four PHYs are at the same speed. ► One or more of the connected PHYs are attached to an address different from the others.
SAS Port 2 Link	Green	OFF	No link connection on any PHYs (lanes). The connection is down.
		ON	There is a connection on at least one PHY. At least one of the lanes to that connector is up.
SAS Port 2 Fault	Amber	OFF	No fault. All four PHYs have a link connection.
		ON	This can indicate a number of different error conditions: ► One or more, but not all, of the four PHYs are connected. ► Not all four PHYs are at the same speed. ► One or more of the connected PHYs are attached to an address different from the others.

More detailed information is available from the Storwize IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/api/content/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/fab1_system_leds.html

2.5 Technician Port

A technician port is available on the rear of the IBM Storwize V7000 Gen2 system for service support. The purpose and key benefit of the Storwize V7000 Gen2 technician port is to simplify and ease the initial basic configuration of the Storwize V7000 system by the local administrator or by service personnel. The technician port provides direct access to the service assistant GUI and CLI.

The technician port can be used by directly connecting a computer that has web browsing software and is configured for Dynamic Host Configuration Protocol (DHCP) via a standard 1 Gbps Ethernet cable. On uninitialized systems, the technician port provides access to the system initialization wizard instead of the service assistant.

After a system has been initialized, the technician port provides access to the following components:

- The service assistant
- The password reset facility (if enabled)

2.5.1 Technician port is marked with a T (Ethernet port 4)

The technician port is used for the initial installation of the system. As soon as the system is installed and the user connects to the Technician Port, the user is directed to the new Init tool. Verify that you are using a supported operating system. The initialization tool is valid for the following operating systems:

- Microsoft Windows 8.1 (64-bit), or Microsoft Windows 7 (64-bit)
- Apple MacOS X 10.71
- Red Hat Enterprise Linux (RHEL) Server 5 and 6
- Ubuntu desktop 11.04 and 13.10

The Init tool is not displayed if there is a problem that prevents the system from clustering. For example, this occurs if Node canister is in Service state because of an error, or if there is a stored System ID (the system was set up before, and the user forgot to remove the ID using the **`chenclosurevpd -resetclusterid`** command). If there is a problem, then the Service Assistant GUI is shown, where the customer can log on and check the node canisters status.

To initialize the system, complete the following steps:

1. Ensure that the system is powered on.
2. Configure an Ethernet port on the personal computer to enable DHCP configuration of its IP address and Domain Name System (DNS) settings.
3. Locate the Ethernet port that is labeled T on the rear of a node canister. This is the Technician port.
4. Connect an Ethernet cable between the port of the personal computer that is configured in step 2 and the technician port.
5. A few moments after the connection is made, the node uses DHCP to configure IP and DNS settings of the personal computer.
6. After the Ethernet port of the personal computer is connected, open a supported browser and browse to the `http://install` address. The browser automatically opens the initialization tool.

Note: If the user's machine has DHCP, it connects directly to the Storwize V7000 Gen2. If they do not have DHCP, they need to set the IP of the Ethernet adapter to `192.168.0.2`.

7. Follow the instructions that are presented by the initialization tool to configure the system with a management IP address. After you complete the initialization process, the system can be reached by opening a supported web browser and entering the following adress:

   http://<*management_IP_address*>

   Figure 2-13 shows the setup flow for the technician port.

Figure 2-13 Technician port setup

8. After the system is set up and the user connects to the Technician port, they are directed to the Service GUI.
9. Only the Technician port has a Password Reset option from the Service Assistant available. The **`sainfo lsservicestatus`** command displays the current status of the node, and there is a new information field to indicate if password reset is enabled.

For more information on CLI commands, see Chapter 8, "IBM Storwize V7000 Gen2 command-line interface" on page 143.

3

Planning and configuration

In this chapter, we describe the steps that are required when you plan the installation of the IBM Storwize V7000 Gen2, using the new 2076-524 hardware in your environment. We look at the implications for your storage network, and also describe performance considerations.

3.1 General planning rules

Important: At the time of writing, the statements made in this book are correct, but they might change over time. Always verify any statements that have been made with the Storwize V7000 Gen2 supported hardware list, device driver, firmware, and suggested software levels on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004622

To achieve the most benefit from Storwize V7000 Gen2, pre-installation planning must include several important steps. These steps ensure that the Storwize V7000 Gen2 provides the best possible performance, reliability, and ease of management for your application needs. Proper configuration also helps minimize downtime by avoiding changes to the Storwize V7000 Gen2 and the storage area network (SAN) environment to meet future growth needs.

Tip: For comprehensive information about the topics that are described here, see *IBM Storwize V7000 Gen2 Product Manuals* on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S7003318

Follow these steps when planning for the Storwize V7000 Gen2:

1. Collect and document the number of hosts (application servers) to attach to the Storwize V7000 Gen2, the traffic profile activity (read or write, sequential or random), and the performance requirements in terms of input/output (I/O) operations per second (IOPS).
2. Collect and document the storage requirements and capacities:
   - The total existing back-end storage to be provisioned on Storwize V7000 Gen2 (if any)
   - The total new back-end storage to be provisioned on the Storwize V7000 Gen2 (if any)
   - The required storage capacity for local mirror copy (volume mirroring)
   - The required storage capacity for point-in-time copy (IBM FlashCopy)
   - The required storage capacity for remote copy (Metro Mirror and Global Mirror)
   - The required storage capacity for compressed volumes
   - Per host:
     - Storage capacity
     - Host logical unit number (LUN) quantity and sizes
   - The required virtual storage capacity that is used as a fully managed volume, and used as a thin-provisioned volume
3. Define the local and remote SAN fabrics and clustered systems, if a remote copy or a secondary site is needed.
4. Define the number of clustered system systems, and the number of pairs of nodes for each system. Each pair of nodes (an I/O Group) is the container for the volume. The number of necessary I/O Groups depends on the overall performance requirements.
5. Design the SAN according to the requirement for high availability (HA) and best performance. Consider the total number of ports. Also consider the bandwidth that is needed between the host and the Storwize V7000 Gen2, the Storwize V7000 Gen2 and the disk subsystem, between the Storwize V7000 Gen2 nodes, and for the inter-switch link (ISL) between the local and remote fabric.
6. Design the Internet Small Computer System Interface (iSCSI) network according to the requirements for HA and best performance. Consider the total number of ports and bandwidth that is needed between the host and the Storwize V7000 Gen2.

7. Determine the Storwize V7000 Gen2 service Internet Protocol (IP) address.
8. Determine the IP addresses for the Storwize V7000 Gen2 system and for the host that connects through iSCSI.
9. Determine the IP addresses for IP replication.
10. Define a naming convention for the Storwize V7000 Gen2 nodes, host, and storage subsystem.
11. Define the managed disks (MDisks) in the disk subsystem.
12. Define the storage pools. The storage pools depend on the disk subsystem in place, and on the data migration requirements.
13. Plan the logical configuration of the volume within the I/O Groups and the storage pools to optimize the I/O load between the hosts and the Storwize V7000 Gen2.
14. Plan for the physical location of the equipment in the rack.

Storwize V7000 Gen2 planning can be categorized into two types:

- Physical planning
- Logical planning

We describe these planning types in more detail in the following sections.

3.1.1 Base software

IBM Storwize family software V7.3 introduces new software licenses for Storwize V7000 Gen2. This new license and pricing structure provides intuitive licensing based on the functions customers want to enable and use the most.

Although Storwize V7000 Gen2 used to provide additional licenses for purchase to enable additional functionality, the new structure and pricing model ties the software licenses closely to the enclosure, and provides the capability to enable additional functionality by purchasing feature codes under that license. The Storwize V7000 Gen2 base software includes the following functions:

- Software Redundant Array of Independent DisksRAID (0/1/5/6/10 with global spares and rebalancing)
- Thin provisioning
- Volume mirroring
- Read/write cache
- Automatic pool balancing
- Volume and host limits per Storwize V7000: 2048 volumes and 512 host objects per Control Enclosure, and so on
- Embedded management and service graphical user interfaces (GUIs), Storage Networking Industry Association (SNIA) Storage Management Initiative Specification (SMI-S)-compliant Common Information Model (CIM) Object Manager (CIMOM)
- Management command-line interface (CLI) over Secure Shell (SSH)
- Four-way system clustering
- Environmental statistics reporting for Energy Star compliance

Storwize V7000 Gen2 Licensing Structure

With the broad range of capabilities of the Storwize V7000 Gen2, including IBM Easy Tier and IBM Real-time Compression, we have simplified the licensing to include these new features.

We noted two key areas to improve on:

- Take away the multiplicity of licenses
- Take the counting out of licensing, and resolve those concerns through pricing

We have a solution to address those complexities and challenges, and we are happy to have arrived at an intuitive and straightforward way to order Storwize V7000 Gen2 licenses, and to maintain them going forward. In this section, we review the licensing structure, as shown in Figure 3-1.

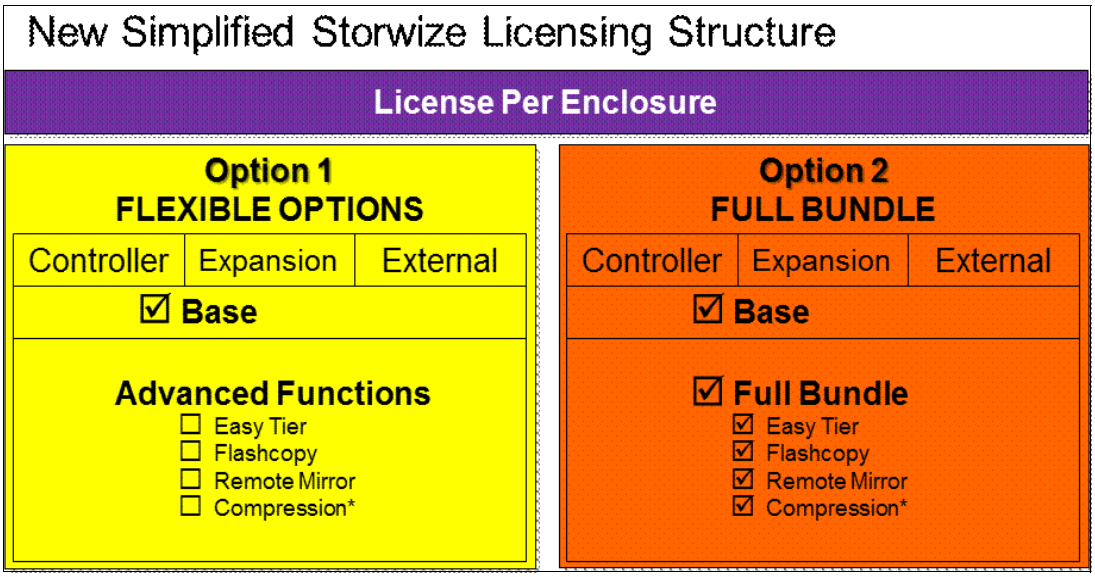


Figure 3-1 New Licensing Structure

How the new licensing works

Storwize family software V7.3 announced the new licensing features on May 6, 2014. The new licensing codes, also known as product identifiers (PIDs), are shown in Figure 3-2.

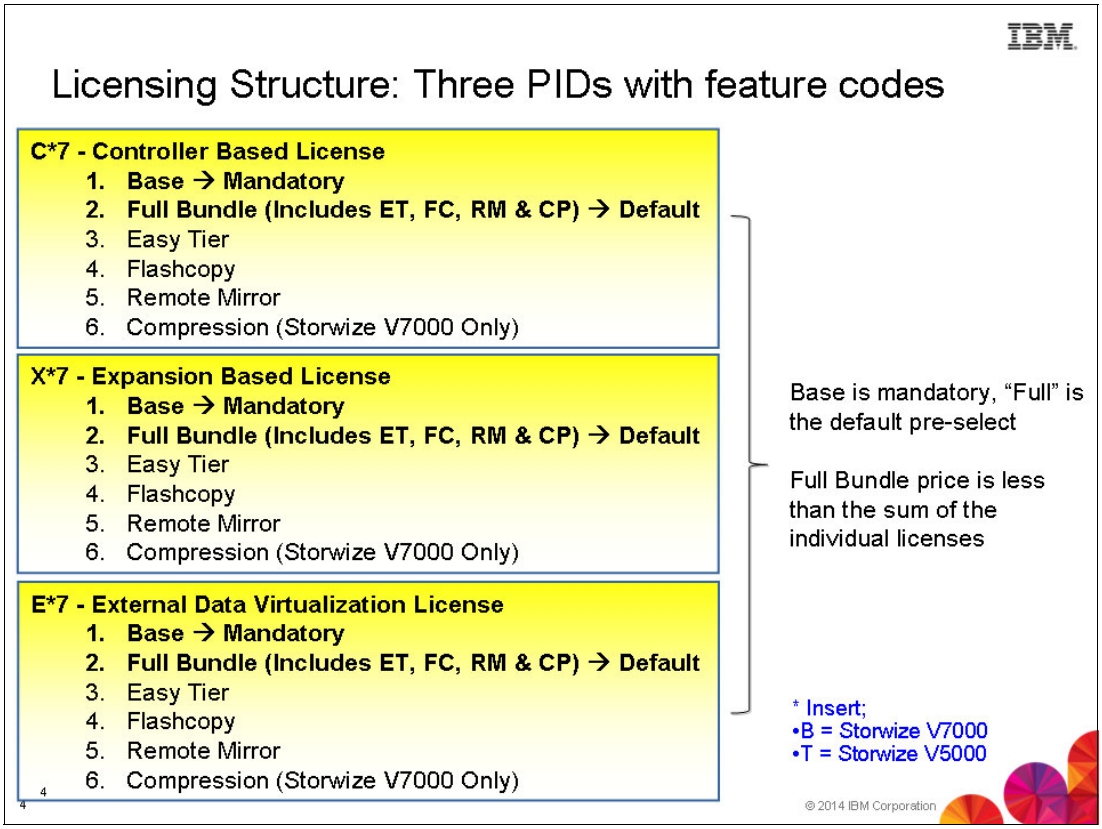


Figure 3-2 Licensing Codes

IBM Storwize family software V7.3 introduces new software licenses for Storwize V7000 Gen2. This new license and pricing structure provides intuitive licensing based on the functions customers want to enable and use the most.

Storwize V7000 Gen2 used to provide additional licenses for purchase to enable additional functionality. The new structure and pricing model ties the software licenses closely to the enclosure, and provides the capability to enable additional functionality by purchasing feature codes under that license.

Each Storwize Family Software for Storwize V7000 Gen2 5639-CB7, 5639-XB7, and 5639-EB7 license has the following feature codes:

- Base software
- Full Feature Set
- Easy Tier
- FlashCopy
- Remote Mirroring
- Compression

The following sections describe each new software license in detail, explains when and where it applies, and provides examples.

IBM Storwize Family Software for Storwize V7000 Gen2 Controller V7 (5639-CB7)

Each IBM Storwize V7000 Gen2 (2076-524) Disk Control Enclosure uses IBM Storwize Family Software for Storwize V7000 Gen2 Controller Software, 5639-CB7. A 5639-CB7 IBM Storwize Family Software for Storwize V7000 Gen2 Controller software license is required for each Control Enclosure. In addition to this Control Enclosure license, advanced functions are available for purchase as feature codes of the software.

For example, if you are running a Storwize V7000 Gen2 system and want to improve performance efficiencies with Easy Tier, you might purchase that feature code to use Easy Tier across all of the Control Enclosures, Expansion Enclosures, and externally virtualized enclosures configured with that system. You can then use Easy Tier for that Storwize V7000 Gen2 system. Alternatively, you can enable Easy Tier and all other optional advanced functions available by purchasing the single feature code labeled *Full Feature Set*.

IBM Storwize Family Software for Storwize V7000 Gen2 Expansion V7 (5639-XB7)

Each IBM Storwize V7000 Gen2 (2076-12F/24F) Expansion Enclosure uses IBM Storwize Family Software for Storwize V7000 Gen2 Expansion Software, 5639-XB7. A 5639-XB7 IBM Storwize Family Software for Storwize V7000 Gen2 Expansion software license is required for each Expansion Enclosure. In addition to this Expansion Enclosure license, advanced functions are available for purchase as feature codes of the software.

For example, if you are running a Storwize V7000 Gen2 system and want to improve performance efficiencies with Easy Tier, you might purchase that feature code to use Easy Tier across all of the Control Enclosures, Expansion Enclosures, and externally virtualized enclosures configured with that system. You can then use Easy Tier for that Storwize V7000 Gen2 system. Alternatively, you can enable Easy Tier and all other optional advanced functions available by purchasing the single feature code labeled *Full Feature Set*.

IBM Storwize Family Software for V7000 External Data Virtualization V7 (5639-EB7)

Each IBM Storwize V7000 Gen2 (2076-524) Disk Control Enclosure can attach and manage external storage devices. IBM Storwize Family Software for Storwize V7000 Gen2 External Data Virtualization Software, 5639-EB7, must be licensed to authorize use of this function.

This license must be purchased for each storage enclosure that is attached to and externally managed by the Storwize V7000 Gen2. IBM Storwize V7000 Gen2 (2076-524) Disk Control Enclosures and IBM Storwize V7000 Gen2 (2076-12F/24F) Disk Expansion Enclosures are not included in this External Virtualization license.

A storage enclosure externally managed by the Storwize V7000 Gen2 is defined as an independently powered, channel-attached device that stores data on magnetic disks or solid-state drives (SSDs), such as disk controllers and their respective expansion units, each constituting separate enclosures. Therefore, an enclosure can be either the main controller housing disk or SSDs, or the expansion chassis that houses additional disk or SSDs for expanding the total capacity of the storage system.

Consult an IBM sales representative with any questions regarding storage controllers. For example, adding an IBM System Storage DS5020 consisting of two enclosures to an IBM Storwize V7000 Gen2 consisting of one Control Enclosure and three Expansion Enclosures requires the purchase of the external virtualization license with a feature code quantity of two enclosures.

In addition to this external enclosure license, advanced functions are available for purchase as feature codes of the software. For example, if you are running a Storwize V7000 Gen2 system and want to improve performance efficiencies with Easy Tier, you might purchase that feature code for Easy Tier across all of the Control Enclosures, Expansion Enclosures, and externally virtualized enclosures configured with that system.

You can then use Easy Tier for that Storwize V7000 Gen2 system. Alternatively, you can enable Easy Tier and all other optional advanced functions available by purchasing the single feature code labeled *Full Feature Set*.

Full Feature Set with mixed configurations

We provide a few examples of licensing in mixed configurations.

Mirroring a Storwize V7000 Gen2 system to an existing Storwize V7000 Gen1

In this first scenario, at the primary site, you are managing a new Storwize V7000 Gen2 2076-524 Disk Control Enclosure that has four 2076-24F Expansion Enclosures attached. In addition, this Storwize V7000 Gen2 system is managing a System Storage DS5020 consisting of four enclosures. At the secondary site, you have a Storwize V7000 Gen1 system consisting of one 2076-124 Disk Control Enclosure that has five 2076-224 Expansion Enclosures attached. You want to use the Remote Mirroring features in this configuration.

This configuration requires the following licenses:

- A quantity of one IBM Storwize Family Software for Storwize V7000 Gen2 Controller Software license (5639-CB7), with the Remote Mirroring feature code selected
- A quantity of four IBM Storwize Family Software for Storwize V7000 Gen2 Expansion Software licenses (5639-XB7), each with the Remote Mirroring feature code selected
- A quantity of one IBM Storwize Family Software for Storwize V7000 Gen2 External Data Virtualization Software license (5639-EB7), with a feature code quantity of four of the Base software, and a feature code quantity of four of the Remote Mirroring feature code (or, of course, this could be a feature code quantity of four of the *Full Bundle* feature code)
- A quantity of six IBM Storwize Family Software for Storwize V7000 Gen2 Software licenses (5639-VM7)
- A quantity of six IBM Storwize Family Software for Storwize V7000 Gen2 Remote Mirroring Software licenses (5639-RM7)

Mixing a new Storwize V7000 Gen 2 system to an existing Storwize V7000 Gen1 in the same cluster

In this second scenario, at the primary site, you are managing a clustered Storwize V7000 Gen2 system consisting of one 2076-124 Disk Control Enclosure that has two 2076-224 Expansion Enclosures attached, and a 2076-524 Disk Control Enclosure that has four 2076-24F Expansion Enclosures attached.

In addition, this Storwize V7000 Gen2 system is managing a System Storage DS5020 consisting of four enclosures. You want to use the Remote Mirroring features in this configuration. At the secondary site, you have the exact same configuration set up.

This configuration requires the following licenses at each site:

- A quantity of three IBM Storwize Family Software for Storwize V7000 Gen2 Software licenses (5639-VM7)
- A quantity of three IBM Storwize Family Software for Storwize V7000 Gen2 Remote Mirroring Software licenses (5639-RM7)
- A quantity of one IBM Storwize Family Software for Storwize V7000 Gen2 Controller Software license (5639-CB7), each with the Remote Mirroring feature code selected
- A quantity of four IBM Storwize Family Software for Storwize V7000 Gen2 Expansion Software licenses (5639-XB7), each with the Remote Mirroring feature code selected
- One of the following configurations:
  - A quantity of one IBM Storwize Family Software for Storwize V7000 Gen2 External Data Virtualization Software license (5639-EB7), with a feature code quantity of four of the Base software, and a feature code quantity of four of the Remote Mirroring feature code
  - A quantity of four IBM Storwize Family Software for Storwize V7000 Gen2 External Virtualization licenses (5639-EV7), plus an additional four IBM Storwize Family Software for Storwize V7000 Gen2 Remote Mirroring Software licenses (5639-RM7), to properly license Remote Mirroring for the externally managed System Storage DS5020

For more detailed information see:

http://www.ibm.com/storage/support/storwize/v7000

3.2 Physical planning

This section of the document is intended to provide guidance about the cabinet elevation layouts to use for physically installing your IBM Storwize V7000 Gen2 in racks.

One of the first factors to consider is whether you are building *a brand new cluster* of Storwize V7000 Gen2 with only 2076-524s in it, or if you are adding the 2076-524 to *an existing cluster* having older model Storwize V7000 Gen1 or IBM SAN Volume Controller nodes in it. A second factor is, if it is a brand new Storwize V7000 Gen2 cluster, you need to determine if you are racking your Storwize V7000 Gen2s in a *single cabinet* layout or a *dual cabinet* layout.

Additionally, when using the optional 2076-24F flash arrays as part of your Storwize V7000 Gen2 cluster implementation, the distance that you can separate the 2076-524 nodes in the I/O Group away from their shared 2076-24F flash array is limited by the *maximum length of the 6-meter serial-attached SCSI (SAS) cable* used to attach the array to the Storwize V7000 Gen2 units.

Important: You must consider the maximum power rating of the rack. *Do not exceed it*. For more information about the power requirements, see the following website:

http://www.ibm.com/support/knowledgecenter/api/content/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/tbrd_physicalconfig.html

3.2.1 Single rack or dual rack configuration

The Storwize V7000 Gen2 system must be installed in pairs to provide HA, and each pair makes up an I/O Group. A Storwize V7000 Gen2 cluster can contain up to four I/O Groups, or a total of eight nodes. Each node requires a power cable connection to connect to the 750w power supplies. Limitations and restrictions for the Storwize V7000 Gen2 can be found on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004628

Figure 3-3 shows the rear view of a Storwize V7000 Gen2 Control Enclosure with the power supplies.

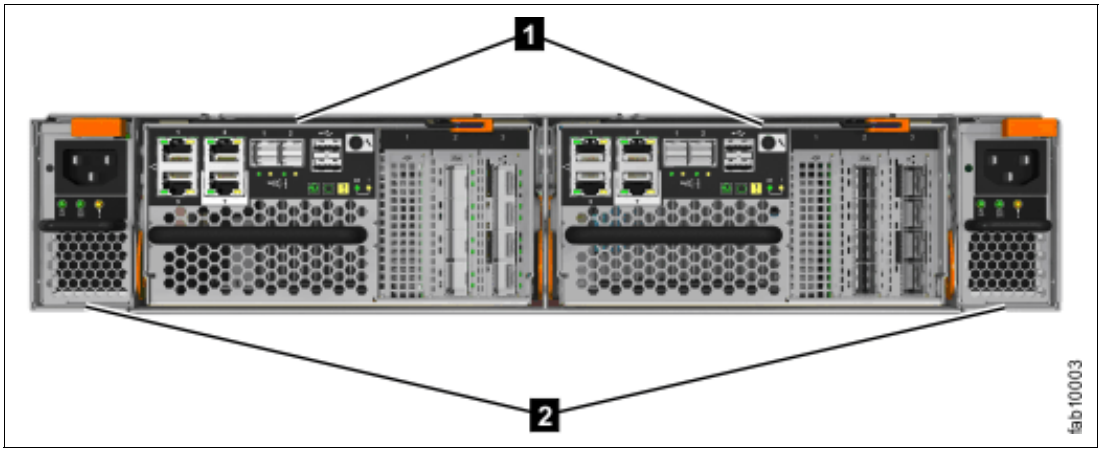


Figure 3-3 2076-524 Power Connectors

The Control Enclosure rear layout includes the following items:

- Node canisters (1)
- Power supply units (2):
  - Power supply 1 (left)
  - Power supply 2 (right)

A rear view of a Storwize V7000 Gen2 Expansion Enclosure is shown in Figure 3-4.

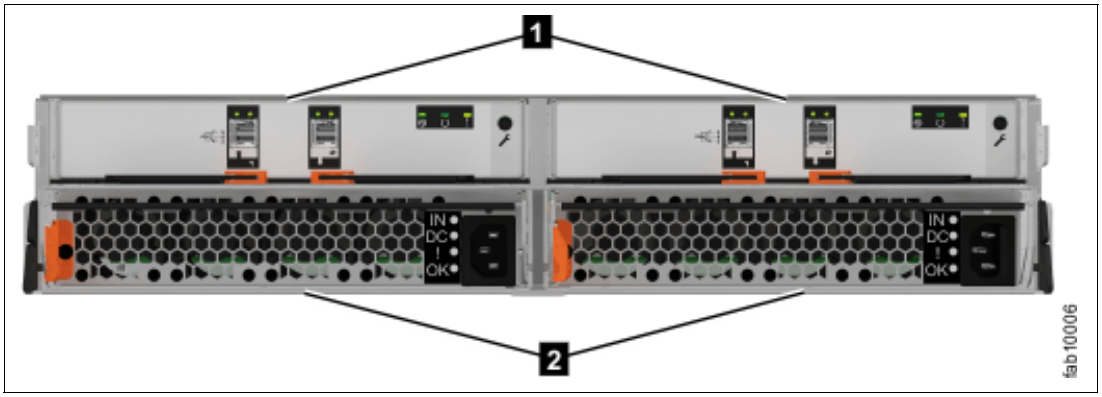


Figure 3-4 Rear view of Enclosure Expansion

The Enclosure Expansion rear layout includes the following items:

- Expansion Canisters (1)
- Power supply units (2)

If multiple racks are required, locate the racks next to one another.

Plan for one of these installations:

- Control Enclosure only
- Control Enclosure plus one or more Expansion Enclosures
- Expansion Enclosure only
- Another Control Enclosure plus one or more Expansion Enclosures

Each Control Enclosure contains two node canisters, forming an I/O Group. The guidelines apply on an I/O Group by I/O Group basis:

- Control Enclosure only

  The Control Enclosure requires two standard rack units of space in a rack. If you plan to add Expansion Enclosures in the future, follow the guidelines for a Control Enclosure plus one or more Expansion Enclosures.

- Control Enclosure plus one or more Expansion Enclosures

  If you have one or more Expansion Enclosures, position the Control Enclosure in the center of the rack to make cabling easier. Balance the number of Expansion Enclosures above and below the Control Enclosure.

- Storwize V7000 Gen2 has the following requirements:
  - Each enclosure requires two standard rack units of space in a rack.
  - Attach no more than 10 Expansion Enclosures to port 1 of the Control Enclosure.
  - Attach no more than 10 Expansion Enclosures to port 2 of the Control Enclosure.
- There is support for 10 expansions per chain, for a total of 21 enclosures to be included:
  - Position the enclosures together. Avoid adding other equipment between enclosures.
  - Position the enclosures in the rack so that you can easily view them and access them for servicing. This action also enables the rack to remain stable, and enables two or more people to install and remove the enclosures.

The following considerations are also part of your decision:

- Upstream redundancy of the power to your cabinet, such as power circuit panels and on-floor Power Distribution Units (PDUs)
- Within cabinet power redundancy, such as dual power strips or in-cabinet PDUs
- Upstream HA structures, such as uninterruptible power supply (UPS), generators, and so on

Many data centers today are at an Uptime Tier 3 or higher level, so power redundancy concerns that would require a dual cabinet Storwize V7000 Gen2 implementation are no longer an issue.

However, Fire Protection Systems Type, such as overhead wet pipe sprinkler systems, should be considered. In association to these items, you should also consider physical separation and location of other key storage environment components.

If you are implementing your entire storage environment with multiple redundant devices physically separated across multiple cabinets and strings, you need to provide sufficient physical distance to ensure that your redundant components are in different fire protection zones and different power-sourced zones. Otherwise, your end-to-end storage environment can be compromised in case of a zonal facilities failure.

If the data center you are moving into has the proper power redundancy attributes, and your storage environment design strategy does not have fully redundant components placed at sufficient distances apart, a single cabinet implementation saves you the costs associated with the second cabinet.

If the data center does not have a robust enough power redundancy infrastructure, or your storage environment design strategy does not have fully redundant components placed at sufficient distances apart, the investment in a dual cabinet implementation is justified in furthering the level of HA and redundancy for your overall storage environment.

Another consideration would be that if you anticipate that you will be adding another Storwize V7000 Gen2 cluster to your storage environment in the future, by implementing a dual cabinet approach from the start, and reserving remaining space in each cabinet for nodes from the second cluster, you accomplish both objectives.

3.2.2 Cable connections

Create a cable connection table or similar documentation to track all of the connections that are required for the setup:

- Nodes
- Ethernet
- Fibre Channel over Ethernet (FCoE) and iSCSI connections
- Fibre Channel (FC) ports

Figure 3-5 shows the rear view of a 2076-524 Node with the two PCIe adapter slots identified.

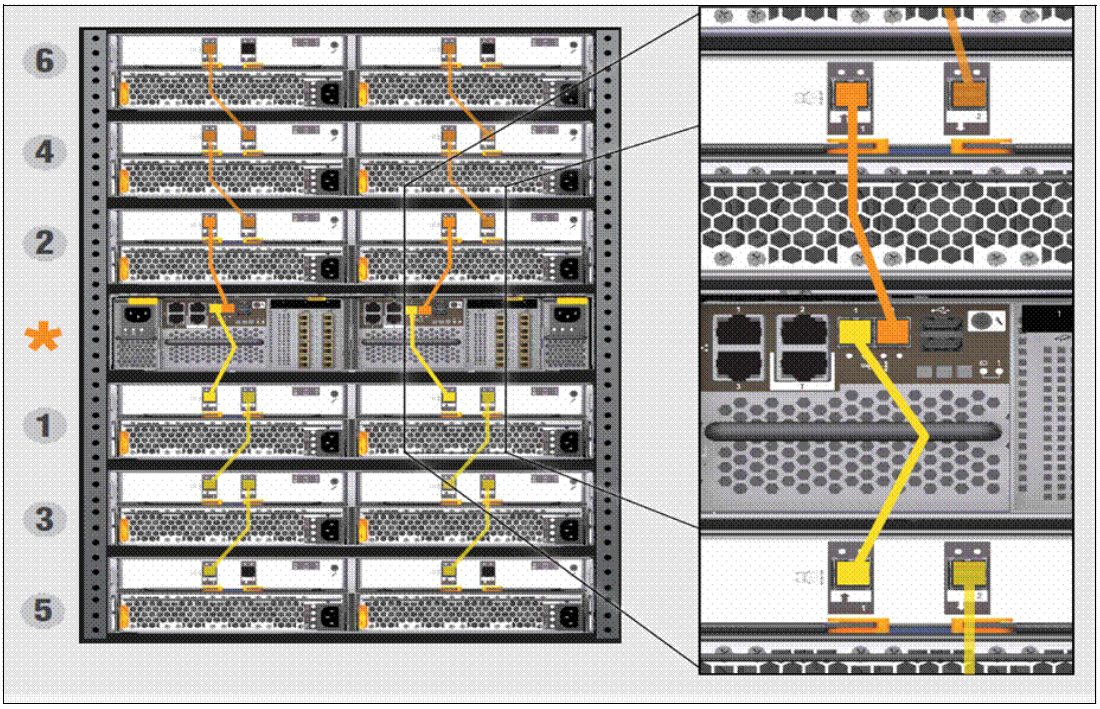


Figure 3-5 2076-524 PCIe Expansion Slots

Each 2076-524 node can support up to two PCIe expansion I/O cards, as identified in Table 3-1, to provide a range of connectivity and capacity expansion options.

Table 3-1 Layout of Expansion Card options for 2076-524 Node.

Slot	Supported Cards	
1	Compression pass-through	Compression Acceleration card
2	None	8 gigabit (Gb) FC, 10 Gb Ethernet
3	None	8 Gb FC, 10 Gb Ethernet

A sample cable connection table can be downloaded by following these steps:

1. Go to: `http://www.ibm.com/storage/support/`
2. Under Product Support Content, click **Plan** and install documentation.
3. In the search box, type and search for `IBM Storwize V7000 Gen2 (2076)`. Then, click the search result.
4. Click the wanted link to download it to your computer.

3.3 Logical planning

For logical planning, this section provides information about the following topics:

- Management IP addressing plan
- SAN zoning and SAN connections
- iSCSI IP addressing plan
- IP replication
- Back-end storage subsystem configuration
- Storwize V7000 Gen2 system configuration
- Storage pool configuration
- Volume configuration
- Host mapping (LUN masking)
- Advanced Copy Services functions
- SAN boot support
- Data migration from non-virtualized storage subsystems
- Storwize V7000 Gen2 configuration backup procedure

3.3.1 Management IP addressing plan

Starting with Storwize V7000 6.1, the system management is performed through an embedded GUI running on the nodes. A separate console, such as the traditional Storwize V7000 Hardware Management Console (HMC) or IBM System Storage Productivity Center (SSPC), is no longer required to access the management interface. To access the management GUI, you direct a web browser to the system management IP address.

The Storwize V7000 Gen2 2076-524 node introduces a new feature called a *Technician Port*. Ethernet port 4 is allocated as the Technician service port, and is marked with a T. All initial configuration for each node is performed through the Technician Port.

The port broadcasts a Dynamic Host Configuration Protocol (DHCP) service so that a notebook or computer is automatically assigned an IP address on connection to the port. More details about this are described in 2.5.1, “Technician port is marked with a T (Ethernet port 4)” on page 45.

Figure 3-6 shows the location of the Technician Port.

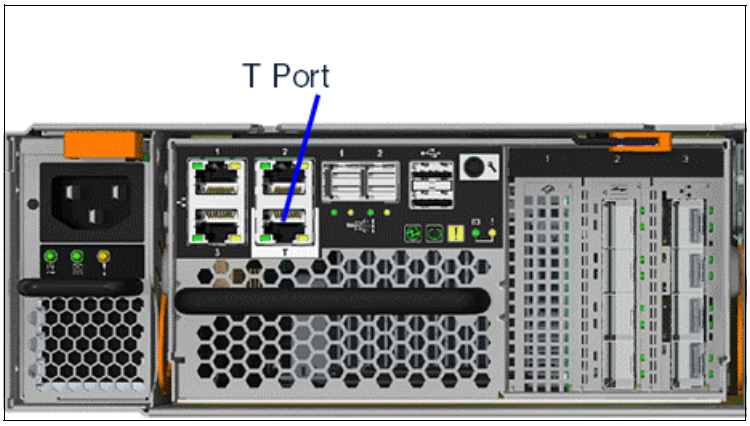


Figure 3-6 Technician Port

After the cluster configuration has been completed, the Technician Port automatically routes the connected user directly to the service GUI.

Information: The default IP address for the Technician Port on a 2076-524 Node is 192.168.0.1. If the Technician Port is connected to a switch, it is disabled and an error is logged.

Each Storwize V7000 Gen2 node requires one Ethernet cable to connect it to an Ethernet switch or hub. The cable must be connected to port 1. A 10/100/1000 megabit (Mb) Ethernet connection is required for each cable. Both Internet Protocol Version 4 (IPv4) and Internet Protocol Version 6 (IPv6) are supported.

To ensure system failover operations, Ethernet port 1 on all nodes must be connected to the same set of subnets. Each Storwize V7000 Gen2 cluster has a Cluster Management IP address and a Service IP address for each node in the cluster. See Example 3-1 for details.

Example 3-1 Management IP address sample

```
management IP add. 10.11.12.120
node 1 service IP add. 10.11.12.121
node 2 service IP add. 10.11.12.122
```

Each node in a Storwize V7000 Gen2 clustered system needs to have at least one Ethernet connection.

Support for iSCSI provides one additional IPv4 and one additional IPv6 address for each Ethernet port on every node. These IP addresses are independent of the clustered system configuration IP addresses.

When accessing the Storwize V7000 Gen2 through the GUI or SSH, choose one of the available IP addresses to which to connect. No automatic failover capability is available. If one network is down, use an IP address on the alternative network. Clients might be able to use the intelligence in domain name servers (DNS) to provide partial failover.

3.3.2 SAN zoning and SAN connections

SAN storage systems using the Storwize V7000 Gen2 can be configured with two, or up to eight, Storwize V7000 Gen2 node canisters, arranged in an Storwize V7000 Gen2 clustered system. This Storwize V7000 Gen2 cluster is attached to the SAN fabric, along with disk subsystems and host systems. The Storwize V7000 Gen2 node canisters within a Storwize V7000 Gen2 system must be able to see each other and all of the storage that is assigned to the Storwize V7000 Gen2 system.

The zoning capabilities of the SAN switch are used to create three distinct zones. Storwize V7000 Gen2 7.3 supports 2 gigabits per second (Gbps), 4 Gbps, or 8 Gbps FC fabric, (16 Gbps connects, but it only uses the 8 Gbps speed by default) depending on the hardware platform and on the switch where the Storwize V7000 Gen2 is connected. In an environment where you have a fabric with multiple-speed switches, the preferred practice is to connect the Storwize V7000 Gen2 and the disk subsystem to the switch operating at the highest speed.

All Storwize V7000 Gen2 nodes in the Storwize V7000 Gen2 clustered system are connected to the same SANs, and they present volumes to the hosts. These volumes are created from storage pools that are composed of MDisks presented by the disk subsystems.

The fabric must have three distinct zones:

- Storwize V7000 Gen2 cluster system zones

  Create one cluster zone per fabric, and include any port per node that is designated for intra-cluster traffic. No more than four ports per node should be allocated to intra-cluster traffic.

- Host zones

  Create a host zone for each server host bus adapter (HBA) port accessing Storwize V7000 Gen2.

- Storage zone

  Create one Storwize V7000 Gen2 storage zone for each storage subsystem that is virtualized by the Storwize V7000 Gen2. Some storage control systems need two separate zones (one per controller) so that they do not “see” each other.

Zoning considerations for Metro Mirror and Global Mirror

Ensure that you are familiar with the constraints on zoning a switch to support Metro Mirror and Global Mirror partnerships.

SAN configurations that use inter-cluster Metro Mirror and Global Mirror relationships require the following additional switch zoning considerations:

- For each node in a clustered system, it is preferred to zone two FC ports from the source system to two FC ports on the target system.
- If dual-redundant ISLs are available, split the two ports from each node evenly between the two ISLs.
- Local clustered system zoning continues to follow the standard requirement for all ports on all nodes in a clustered system to be zoned to one another.

Attention: Failure to follow these configuration rules exposes the clustered system to the following condition, and can result in the loss of host access to volumes.

If an inter-cluster link becomes severely and abruptly overloaded, the local FC fabric can become congested to the extent that no FC ports on the local Storwize V7000 Gen2 nodes are able to perform local intra-cluster heartbeat communication. This situation can, in turn, result in the nodes experiencing lease expiry events.

In a lease expiry event, a node reboots to attempt to re-establish communication with the other nodes in the clustered system. If the leases for all nodes expire simultaneously, a loss of host access to volumes can occur for the duration of the reboot events.

- Configure your SAN so that FC traffic can be passed between the two clustered systems. To configure the SAN this way, you can connect the clustered systems to the same SAN, merge the SANs, or use routing technologies.
- Configure zoning to enable all of the nodes in the local fabric to communicate with all of the nodes in the remote fabric.

You must also observe the following guidelines:

- LUNs (MDisks) must have exclusive access to a single Storwize V7000 Gen2 clustered system, and cannot be shared between other Storwize V7000 clustered systems or hosts.
- A storage controller can present LUNs to both the Storwize V7000 Gen2 (as MDisks) and to other hosts in the SAN. However, in this case it is better to avoid the Storwize V7000 and hosts that share the same storage ports.
- Mixed port speeds are not permitted for intra-cluster communication. All node ports within a clustered system must be running at the same speed.
- The switch configuration in a Storwize V7000 Gen2 fabric must comply with the switch manufacturer's configuration rules, which can impose restrictions on the switch configuration. For example, a switch manufacturer might limit the number of supported switches in a SAN. Operation outside of the switch manufacturer's rules is not supported.
- Host zones are to contain only one initiator (HBA) each, and one or two Storwize V7000 Gen2 node ports, depending on the HA and performance that you want from your configuration.

Important: Be aware of the following considerations:

- The use of ISLs for intra-cluster node communication could possibly negatively affect the availability of the system due to the high dependency on the quality of these links to maintain heartbeat and other system management services. Therefore, we strongly advise that you only use them as part of an interim configuration to facilitate SAN migrations, and not as part of the architected solution.
- The use of ISLs for Storwize V7000 Gen2 node to storage controller access can lead to port congestion, which could negatively affect the performance and resiliency of the SAN. Therefore, we strongly advise that you only use them as part of an interim configuration to facilitate SAN migrations, and not as part of the designed solution.

  With Storwize V7000 Gen2 V7.3, you can use ISLs between nodes, but they must be in a dedicated SAN, virtual SAN (CISCO technology), or logical SAN (Brocade technology).
- The use of mixed port speeds for inter-cluster communication can lead to port congestion, which could negatively affect the performance and resiliency of the SAN, and is therefore not supported.

You can use the **`lsfabric`** command to generate a report that displays the connectivity between nodes and other controllers and hosts. This report is helpful for diagnosing SAN problems.

For more information about zoning and configuration, see the following website:

`http://www.ibm.com/support/knowledgecenter/api/content/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/svc_configrulessummary_02171530.html`

3.3.3 iSCSI IP addressing plan

Since version 6.3, Storwize V7000 has supported host access through iSCSI (as an alternative to FC). The following considerations apply:

- Storwize V7000 Gen2 uses the built-in Ethernet ports for iSCSI traffic. If the optional 10 Gbps Ethernet feature is installed, you can connect host systems through the two 10 Gbps Ethernet ports per node.
- All node types that can run Storwize V7000 6.1 or later can use the iSCSI feature.
- Storwize V7000 Gen2 supports the Challenge Handshake Authentication Protocol (CHAP) authentication methods for iSCSI.
- iSCSI IP addresses can fail over to the partner node in the I/O Group if a node fails. This design reduces the need for multipathing support in the iSCSI host.
- iSCSI IP addresses can be configured for one or more nodes.
- iSCSI simple name server (iSNS) addresses can be configured in Storwize V7000 Gen2.
- The iSCSI qualified name (IQN) for a Storwize V7000 Gen2 node is `iqn.1986-03.com.ibm:2076.<cluster_name>.<node_name>`. Because the IQN contains the clustered system name and the node name, it is important not to change these names after iSCSI is deployed.
- Each node can be given an iSCSI alias, as an alternative to the IQN.

- The IQN of the host to a Storwize V7000 Gen2 host object is added in the same way that you add FC worldwide port names (WWPNs).
- Host objects can have both WWPNs and IQNs.
- Standard iSCSI host connection procedures can be used to discover and configure Storwize V7000 Gen2 as an iSCSI target.

3.3.4 Native IP replication

One of the most important new functions introduced in version 7.2 of the Storwize family was native IP replication, which enables the use of lower-cost Ethernet connections for remote mirroring. The capability is available as an option (Metro Mirror or Global Mirror) on all Storwize family systems. The new function is transparent to servers and applications in the same way that traditional FC-based mirroring is. All remote mirroring modes (Metro Mirror, Global Mirror, and Global Mirror with Changed Volumes) are supported.

Configuration of the system is straightforward: Storwize family systems can normally find each other in the network, and can be selected from the GUI. IP replication includes Bridgeworks SANSlide network optimization technology, and is available at no additional charge. Remote mirror is a chargeable option, but the price does not change with IP replication. Existing remote mirror users can access the new function at no additional charge.

Information: Full details of how to set up and configure IP replication are available in the *IBM SAN Volume Controller and Storwize Family Native IP Replication* publication:

http://www.redbooks.ibm.com/abstracts/redp5103.html

3.3.5 Back-end storage subsystem configuration

Back-end storage subsystem configuration planning must be applied to all storage controllers that are attached to the Storwize V7000 Gen2.

See the following website for a list of currently supported storage subsystems:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004622

Apply the following general guidelines for back-end storage subsystem configuration planning:

- In the SAN, storage controllers that are used by the Storwize V7000 Gen2 clustered system must be connected through SAN switches. Direct connection between the Storwize V7000 Gen2 and the storage controller is not supported.
- Multiple connections are enabled from the redundant controllers in the disk subsystem to improve data bandwidth performance. It is not mandatory to have a connection from each redundant controller in the disk subsystem to each counterpart SAN, but it is a preferred practice. Therefore, both canisters port 1 & 3 in a Storwize V3700 subsystem can be connected to SAN A, and port 2 & 4 to SAN B.
- All Storwize V7000 Gen2 nodes in an Storwize V7000 clustered system must be able to see the same set of ports from each storage subsystem controller. Violating this guideline causes the paths to become degraded. This degradation can occur as a result of applying inappropriate zoning and LUN masking. This guideline has important implications for a disk subsystem, such as DS3000, Storwize V3700, Storwize V5000, or Storwize V7000, which imposes exclusivity rules as to which HBA WWPNs a storage partition can be mapped.

MDisks within storage pools:

- Storwize V7000 6.1 and later provide for better load distribution across paths within storage pools.

  In previous code levels, the path to MDisk assignment was made in a round-robin fashion across all MDisks configured to the clustered system. With that method, no attention is paid to how MDisks within storage pools are distributed across paths. Therefore, it is possible and even likely to have certain paths that are more heavily loaded than others.

  This condition is more likely to occur with a smaller number of MDisks contained in the storage pool. Starting with Storwize V7000 6.1, the code contains logic that considers MDisks within storage pools. Therefore, the code more effectively distributes their active paths that are based on the storage controller ports that are available.

- The **`Detect Mdisk`** command must be run after the creation or modification (add or remove MDisk) of storage pools for paths to be redistributed.

If you do not have a storage subsystem that supports the Storwize V7000 Gen2 round-robin algorithm, make the number of MDisks per storage pool a multiple of the number of storage ports that are available. This approach ensures sufficient bandwidth to the storage controller and an even balance across storage controller ports.

In general, configure disk subsystems as though no Storwize V7000 exists. However, we suggest the following specific guidelines:

- Disk drives:
  - Exercise caution with large disk drives so that you do not have too few spindles to handle the load.
  - RAID 5 is suggested for most workloads.
- Array sizes:
  - An array size of 8+P or 4+P is suggested for the IBM DS4000® and IBM DS5000™ families, if possible.
  - Use the DS4000 segment size of 128 kilobytes (KB) or larger to help the sequential performance.
  - Upgrade to EXP810 drawers, if possible.
  - Create LUN sizes that are equal to the RAID array and rank size. If the array size is greater than 2 terabytes (TB) and the disk subsystem does not support MDisks larger than 2 TB, create the minimum number of LUNs of equal size.
  - An array size of 7+P is suggested for the V3700, V5000, and V7000 Storwize families.
  - When adding more disks to a subsystem, consider adding the new MDisks to existing storage pools versus creating additional small storage pools.
- Maximum of 1024 worldwide node names (WWNNs) per cluster:
  - EMC DMX/SYMM, all HDS, and SUN/HP HDS clones use one WWNN per port. Each WWNN appears as a separate controller to the Storwize V7000 Gen2.
  - IBM, EMC CLARiiON, and HP use one WWNN per subsystem. Each WWNN appears as a single controller with multiple ports and WWPNs, for a maximum of 16 ports and WWPNs per WWNN.

- DS8000 using four of, or eight of, the 4-port HA cards:
  - Use ports 1 and 3 or ports 2 and 4 on each card (it does not matter for 8 Gb cards).
  - This setup provides 8 or 16 ports for Storwize V7000 Gen2 use.
  - Use eight ports minimum, up to 40 ranks.
  - Use 16 ports for 40 or more ranks. Sixteen is the maximum number of ports.
- DS4000/DS5000 and EMC CLARiiON/CX:
  - Both systems have the preferred controller architecture, and Storwize V7000 Gen2 supports this configuration.
  - Use a minimum of four ports, and preferably eight or more ports, up to a maximum of 16 ports, so that more ports equate to more concurrent I/O that is driven by the Storwize V7000 Gen2.
  - Support is available for mapping controller A ports to Fabric A and controller B ports to Fabric B, or cross-connecting ports to both fabrics from both controllers. The cross-connecting approach is preferred to avoid auto volume transfer (AVT)/Trespass occurring if a fabric or all paths to a fabric fail.
- DS3400 subsystems:
  - Use a minimum of four ports.
- Storwize family:
  - Use a minimum of four ports, and preferably eight ports.
- IBM XIV® requirements and restrictions:
  - The use of XIV extended functions, including snaps, thin provisioning, synchronous replication (native copy services), and LUN expansion of LUNs presented to the Storwize V7000 Gen2 is not supported.
  - A maximum of 511 LUNs from one XIV system can be mapped to a Storwize V7000 Gen2 clustered system.
- Full 15 module XIV suggestions (161 usable TB):
  - Use two interface host ports from each of the six interface modules.
  - Use ports 1 and 3 from each interface module, and zone these 12 ports with all Storwize V7000 Gen2 node ports.
  - Create 48 LUNs of equal size, each of which is a multiple of 17 gigabytes (GB). This creates approximately 1632 GB if you are using the entire full frame XIV with the Storwize V7000 Gen2.
  - Map LUNs to the Storwize V7000 Gen2 as 48 MDisks, and add all of them to the single XIV storage pool so that the Storwize V7000 Gen2 drives the I/O to four MDisks and LUNs for each of the 12 XIV FC ports. This design provides a good queue depth on the Storwize V7000 Gen2 to drive XIV adequately.
- Six module XIV suggestions (55 TB usable):
  - Use two interface host ports from each of the two active interface modules.
  - Use ports 1 and 3 from interface modules 4 and 5. (Interface module 6 is inactive). Also zone these four ports with all Storwize V7000 Gen2 node ports.
  - Create 16 LUNs of equal size, each of which is a multiple of 17 GB. This creates approximately 1632 GB if you are using the entire XIV with the Storwize V7000 Gen2.
  - Map the LUNs to the Storwize V7000 Gen2 as 16 MDisks, and add all of them to the single XIV storage pool, so that the Storwize V7000 Gen2 drives I/O to four MDisks and LUNs per each of the four XIV FC ports. This design provides a good queue depth on the Storwize V7000 Gen2 to drive the XIV adequately.

- Nine module XIV suggestions (87 usable TB):
  - Use two interface host ports from each of the four active interface modules.
  - Use ports 1 and 3 from interface modules 4, 5, 7, and 8 (interface modules 6 and 9 are inactive). Also, zone these eight ports with all of the Storwize V7000 Gen2 node ports.
  - Create 26 LUNs of equal size, each of which is a multiple of 17 GB. This creates approximately 1632 GB if you are using the entire XIV with the Storwize V7000 Gen2.
  - Map the LUNs to the Storwize V7000 Gen2 as 26 MDisks, and map all of them to the single XIV storage pool, so that the Storwize V7000 Gen2 drives I/O to three MDisks and LUNs on each of the six ports and four MDisks and LUNs on the other two XIV FC ports. This design provides a useful queue depth on Storwize V7000 Gen2 to drive XIV adequately.
- Configure XIV host connectivity for the Storwize V7000 Gen2 clustered system:
  - Create one host definition on XIV, and include all Storwize V7000 Gen2 node WWPNs.
  - You can create clustered system host definitions (one per I/O Group), but the preceding method is easier.
  - Map all LUNs to all Storwize V7000 Gen2 node WWPNs.

3.3.6 Real-time Compression

The 2076-524 introduces additional hardware dedicated to the improvement of the Real-time Compression functionality in Storwize V7000 Gen2. It is a mandatory requirement that each node has the second processor and cache upgrade, and a minimum of one Compression Acceleration Card, for the I/O Group to support compressed volumes. For more information, see Chapter 6, “IBM Real-time Compression and the IBM Storwize V7000 Gen2” on page 117.

Note: Active Data Workload is typically 5 - 8% of the total managed capacity. In a single I/O Group, 8 TB of active data equates to approximately 160 TB managed. In an eight-node cluster, this equates to 32 TB of active data (8 TB per I/O Group).

3.3.7 Easy Tier version 3

With the release of Storwize V7000 Gen2 V7.3, Easy Tier has been enhanced to support several new features:

- Easy Tier with three tiers in a pool:
  - Nearline (NL-SAS)
  - Enterprise (SAS)
  - Flash (SSD or Flash)
- Easy Tier puts hot extents on faster storage, and cold extents on slower storage.
- Easy Tier with any two tiers in a pool:
  - Nearline + Enterprise
  - (anything) + Flash
- Drive and storage system sensitivity:
  - Easy Tier understands exactly what type of drive and RAID level, and what class of storage system, is being used.
  - It knows how much performance to expect from an MDisk and avoids overloading.

- Major enhancements to the STAT tool to support the previously mentioned functionality and add more metrics:
  - STAT tool outputs three sets of data.
  - Detailed logging can be uploaded to Disk Magic to validate skew curve.

Figure 3-7 shows the basic layout of how Easy Tier works. With Storwize V7000 Gen2 the user must manually define flash disk MDisk or Nearline MDisks. All MDisks are classed as Enterprise by default.

By default the following features are enabled:

- Easy Tier – On any pool that contains more than one tier
- Storage Pool Balancing – On all pools
- Storwize mdisks are automatically assigned a tier based on the drive type

Tier 0	Tier 1	Tier 2
Three Tier Pools:		
SSD	Enterprise	Nearline
Two Tier Pools:		
SSD	Enterprise	-
SSD	Nearline	-
-	Enterprise	Nearline
Single Tier Pools:		
SSD	-	-
-	Enterprise	-
-	-	Nearline

Figure 3-7 Identifies the three tiers of disk accessible by Easy Tier

For more information about Easy Tier, see Chapter 4, “IBM Storwize V7000 Gen2 Easy Tier” on page 85.

3.3.8 Storwize V7000 Gen2 clustered system configuration

To ensure HA in Storwize V7000 Gen2 installations, consider the following guidelines when you design a SAN with the Storwize V7000 Gen2:

- All nodes in a clustered system must be in the same LAN segment, because the nodes in the clustered system must be able to assume the same clustered system or service IP address. Make sure that the network configuration enables any of the nodes to use these IP addresses. If you plan to use the second Ethernet port on each node, it is possible to have two LAN segments. However, port 1 of every node must be in one LAN segment, and port 2 of every node must be in the other LAN segment.
- To maintain application uptime in the unlikely event of an individual Storwize V7000 Gen2 node failing, Storwize V7000 Gen2 nodes are always deployed in pairs (I/O Groups). If a node fails or is removed from the configuration, the remaining node operates in a degraded mode, but it is still a valid configuration. The remaining node operates in write-through mode, meaning that the data is written directly to the disk subsystem (the cache is disabled for the write).
- The FC SAN connections between the Storwize V7000 Gen2 node and the switches are optical fiber. These connections can run at either 2 Gbps, 4 Gbps, or 8 Gbps, depending on your Storwize V7000 Gen2 and switch hardware.

- Two Storwize V7000 Gen2 clustered systems *cannot* have access to the same LUNs in a disk subsystem. Configuring zoning so that two Storwize V7000 Gen2 clustered systems have access to the same LUNs (MDisks) can, and will likely, *result in data corruption*.
- The Storwize V7000 Gen2 uses three MDisks as quorum disks for the clustered system. A preferred practice for redundancy is to have each quorum disk in a separate storage subsystem, where possible. The current locations of the quorum disks can be displayed using the **`lsquorum`** command, and can be relocated using the **`chquorum`** command.

  Figure 3-8 displays the quorum disk layout.

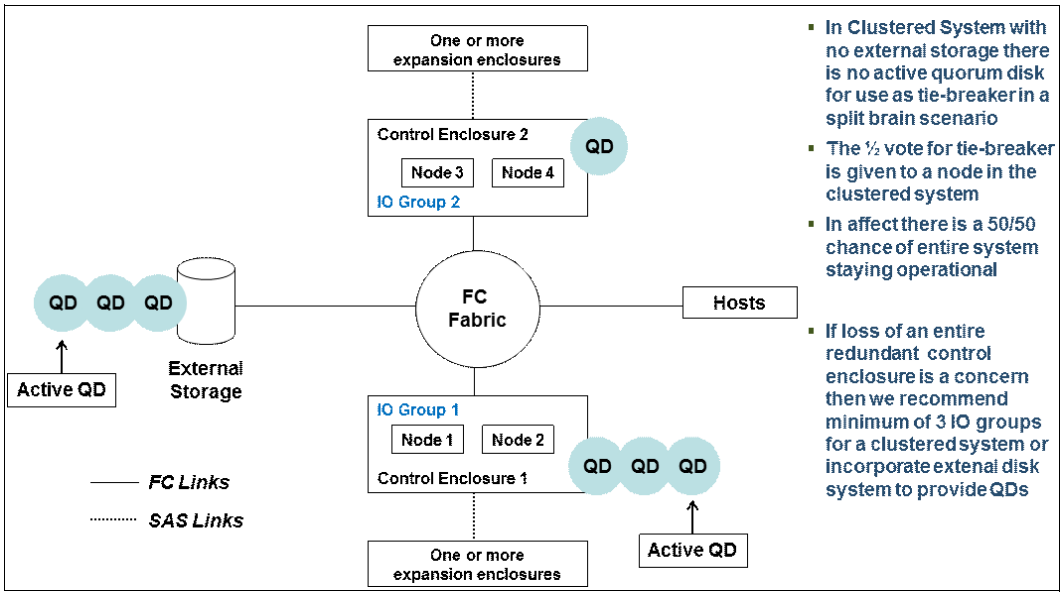


Figure 3-8 Quorum disk layout

3.3.9 Volume configuration

An individual volume is a member of one storage pool and one I/O Group. When creating a volume, you first identify the wanted performance, availability, and cost requirements for that volume, and then select the storage pool accordingly:

- The storage pool defines which disk subsystem MDisks make up the volume.
- The I/O Group (two nodes make an I/O Group) defines which Storwize V7000 Gen2 nodes provide I/O access to the volume.

Important: There is no fixed relationship between I/O Groups and storage pools.

Perform volume allocation based on the following considerations:

- Optimize performance between the hosts and the Storwize V7000 Gen2 by attempting to distribute volumes evenly across available I/O Groups and nodes within the clustered system.
- Reach the level of performance, reliability, and capacity that you require by using the storage pool that corresponds to your needs (you can access any storage pool from any node). That is, choose the storage pool that fulfills the demands for your volumes concerning performance, reliability, and capacity.

- I/O Group considerations:
  - When you create a volume, it is associated with one node of an I/O Group. By default, every time that you create a new volume, it is associated with the next node using a round-robin algorithm. You can specify a *preferred access node*, which is the node through which you send I/O to the volume rather than using the round-robin algorithm. A volume is defined for an I/O Group.
  - Even if you have eight paths for each volume, all I/O traffic flows only toward one node (the preferred node). Therefore, only four paths are used by the IBM Subsystem Device Driver (SDD). The other four paths are used only in the case of a failure of the preferred node, or when concurrent code upgrade is running.
- Creating image mode volumes:
  - Use image mode volumes when an MDisk already has data on it from a non-virtualized disk subsystem. When an image mode volume is created, it directly corresponds to the MDisk from which it is created. Therefore, volume logical block address (LBA) *x* = MDisk LBA *x*. The capacity of image mode volumes defaults to the capacity of the supplied MDisk.
  - When you create an image mode disk, the MDisk does have a mode of *unmanaged*, and should be put into a separate storage pool when you import the volume. A capacity of 0 is not allowed. Image mode volumes can be created in sizes with a minimum granularity of 512 bytes, and they must be at least one block (512 bytes) in size.

    Attention: Image mode disks should be imported into a storage pool of like disks, otherwise you risk the possibility of data corruption or loss.

  - When creating a managed mode volume with sequential or striped policy, you must use several MDisks containing extents that are available and of a size that is equal to or greater than the size of the volume that you want to create. There might be sufficient extents available on the MDisk, but a contiguous block large enough to satisfy the request might not be available.
- Thin-Provisioned volume considerations:
  - When creating the Thin-Provisioned volume, you need to understand the use patterns by the applications or group users accessing this volume. You must consider items, such as the actual size of the data, the rate of creation of new data, modifying or deleting existing data, and so on.
  - Two operating modes for Thin-Provisioned volumes are available:
    - *Autoexpand volumes* allocate storage from a storage pool on demand, with minimal required user intervention. However, a malfunctioning application can cause a volume to expand until it has used all of the storage in a storage pool.
    - *Non-autoexpand volumes* have a fixed amount of assigned storage. In this case, the user must monitor the volume and assign additional capacity when required. A malfunctioning application can only cause the volume that it uses to fill up.
  - Depending on the initial size for the real capacity, the grain size and a warning level can be set. If a volume goes offline, either through a lack of available physical storage for autoexpand, or because a volume that is marked as non-expand had not been expanded in time, a danger exists of data being left in the cache until storage is made available. This situation is not a data integrity or data loss issue, but you must not rely on the Storwize V7000 Gen2 cache as a backup storage mechanism.

Important:

- Keep a warning level on the used capacity so that it provides adequate time to respond and provision more physical capacity.
- Warnings must not be ignored by an administrator.
- Use the autoexpand feature of the Thin-Provisioned volumes.

  - When you create a thin-provisioned volume, you can choose the grain size for allocating space in 32 KB, 64 KB, 128 KB, or 256 KB chunks. The grain size that you select affects the maximum virtual capacity for the thin-provisioned volume. The default grain size is 256 KB, and is the strongly suggested option. If you select 32 KB for the grain size, the volume size cannot exceed 260,000 GB. The grain size cannot be changed after the thin-provisioned volume is created.

    Generally, smaller grain sizes save space but require more metadata access, which could adversely affect performance. If you are not going to use the thin-provisioned volume as a FlashCopy source or target volume, use 256 KB to maximize performance. If you are going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function.
  - Thin-provisioned volumes require more I/Os because of directory accesses. For truly random workloads with 70% read and 30% write, a thin-provisioned volume requires approximately one directory I/O for every user I/O.
  - The directory is two-way write-back-cached (just like the Storwize V7000 Gen2 fast-write cache), so certain applications perform better.
  - Thin-provisioned volumes require more processor processing, so the performance per I/O Group can also be reduced.
  - A thin-provisioned volume feature called *zero detect* provides clients with the ability to reclaim unused allocated disk space (zeros) when converting a fully allocated volume to a Thin-Provisioned volume using volume mirroring.
- Volume mirroring guidelines:
  - Create or identify 2 separate storage pools to allocate space for your mirrored volume.
  - Allocate the storage pools containing the mirrors from separate storage controllers.
  - If possible, use a storage pool with MDisks that share the same characteristics. Otherwise, the volume performance can be affected by the poorer-performing MDisk.

3.3.10 Host mapping (LUN masking)

For the host and application servers, the following guidelines apply:

- Each Storwize V7000 Gen2 node presents a volume to the SAN through four ports. Because two nodes are used in normal operations to provide redundant paths to the same storage, a host with two HBAs can see multiple paths to each LUN that is presented by the Storwize V7000 Gen2. Use zoning to limit the pathing from a minimum of two paths to the maximum that is available of eight paths, depending on the kind of HA and performance that you want to have in your configuration.

  It is best to use zoning to limit the pathing to four paths. The hosts must run a multipathing device driver to limit the pathing back to a single device. The multipathing driver supported and delivered by Storwize V7000 Gen2 is SDD. Native multipath I/O (MPIO) drivers on selected hosts are supported.

For operating system-specific information about MPIO support, see the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003741

You can find the actual version of the IBM Subsystem Device Driver Device Specific Module (SDDDSM) for IBM products on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004622

- The number of paths to a volume from a host to the nodes in the I/O Group that owns the volume must not exceed eight, even if eight is not the maximum number of paths supported by the multipath driver (SDD supports up to 32). To restrict the number of paths to a host volume, the fabrics must be zoned so that each host FC port is zoned to no more than two ports from each Storwize V7000 Gen2 node in the I/O Group that owns the volume:
  - If the suggested number of paths to a volume is exceeded, a path failure might not be recovered in the required amount of time.
  - Too many paths can cause excessive I/O waits, resulting in application failures.
  - Under certain circumstances, it can reduce performance.
  - Eight paths are supported, but four are optimum for SDD, SDDDSM, and IBM Subsystem Device Driver Path Control Module (SDDPCM).

Note: It is a supported configuration to have eight paths to each volume, but this design provides no performance benefit, and it does not improve reliability or availability by any significant degree.

Hosts with four (or more) HBAs take a little more planning, because eight paths are not an optimum number, so you must instead configure your IBM SAN Volume Controller Host Definitions (and zoning) as though the single host is two or more separate hosts.

During VDisk assignment, alternate which VDisk is assigned to one of the pseudo-hosts, in a round robin fashion. (A pseudo-host is nothing more than another regular host definition in the IBM SAN Volume Controller host configuration. Each pseudo-host contains two unique host WWPNs, one WWPN mapped to each fabric.)

A pseudo-host is not a defined function or feature of the IBM SAN Volume Controller. If you need to define a pseudo-host, you are simply adding another host ID to the IBM SAN Volume Controller host configuration. Rather than creating one host ID with four WWPNs, you would define two hosts with two WWPNs. This is now the reference for the term *pseudo-host*.

Be careful not to share the volume to more than two adapters per host, to not oversubscribe the number of datapaths per volumes per host.

- If a host has multiple HBA ports, each port must be zoned to a separate set of Storwize V7000 Gen2 ports to maximize HA and performance.

Note: We use the term *HBA port* to describe the *SCSI initiator*. We use the term *Storwize V7000 port* to describe the *SCSI target*.

The maximum number of host paths per volume must not exceed eight.

- To configure more than 256 hosts, you must configure the host to I/O Group mappings on the Storwize V7000 Gen2. Each I/O Group can contain a maximum of 256 hosts, so it is possible to create 1024 host objects on an eight-node Storwize V7000 Gen2 clustered system. Volumes can only be mapped to a host that is associated with the I/O Group to which the volume belongs.
- Port masking

  You can use a *port mask* to control the node target ports that a host can access, which satisfies two requirements:

  - As part of a security policy, to limit the set of WWPNs that are able to obtain access to any volumes through a given Storwize V7000 Gen2 port
  - As part of a scheme to limit the number of logins with mapped volumes visible to a host multipathing driver, such as SDD, and therefore limit the number of host objects configured without resorting to switch zoning
- The port mask is an optional parameter of the **mkhost** and **chhost** commands. The port mask is four binary bits. Valid mask values range from 0000 (no ports enabled) to 1111 (all ports enabled). For example, a mask of 0011 enables port 1 and port 2. The default value is 1111 (all ports enabled).
- The Storwize V7000 Gen2 supports connection to the Cisco MDS family and Brocade family. See the following website for current support information:

  http://www.ibm.com/systems/storage/software/virtualization/StorwizeV7000/interop.html

3.3.11 Advanced Copy Services

The Storwize V7000 Gen2 offers these Advanced Copy Services:

- FlashCopy
- Metro Mirror
- Global Mirror

Storwize V7000 Gen2 Advanced Copy Services must apply the following guidelines.

FlashCopy guidelines

Consider these FlashCopy guidelines:

- Identify each application that must have a FlashCopy function implemented for its volume.
- FlashCopy is a relationship between volumes. Those volumes can belong to separate storage pools and separate storage subsystems.
- You can use FlashCopy for backup purposes by interacting with the IBM Tivoli® Storage Manager Agent, or for cloning a particular environment.
- Define which FlashCopy best fits your requirements: No copy, Full copy, Thin-Provisioned, or Incremental.
- Define which FlashCopy rate best fits your requirement in terms of the performance and the amount of time to complete the FlashCopy. Table 3-2 on page 73 shows the relationship of the background copy rate value to the attempted number of grains to be split per second.
- Define the grain size that you want to use. A *grain* is the unit of data that is represented by a single bit in the FlashCopy bitmap table. Larger grain sizes can cause a longer FlashCopy elapsed time, and a higher space usage in the FlashCopy target volume.

  Smaller grain sizes can have the opposite effect. Remember that the data structure and the source data location can modify those effects.

In an actual environment, check the results of your FlashCopy procedure in terms of the data that is copied at every run and in terms of elapsed time, comparing them to the new Storwize V7000 Gen2 FlashCopy results. Eventually, adapt the grain per second and the copy rate parameter to fit your environment's requirements.

Table 3-2 Grain splits per second

User percentage	Data copied per second	256 KB grain per second	64 KB grain per second
1% - 10%	128 KB	0.5	2
11% - 20%	256 KB	1	4
21% - 30%	512 KB	2	8
31% - 40%	1 MB	4	16
41% - 50%	2 MB	8	32
51% - 60%	4 MB	16	64
61% - 70%	8 MB	32	128
71% - 80%	16 Mb	64	256
81% - 90%	32 MB	128	512
91% - 100%	64 MB	256	1024

Metro Mirror and Global Mirror guidelines

Storwize V7000 Gen2 supports inter-cluster Metro Mirror and Global Mirror.

Inter-cluster operation needs at least two clustered systems that are separated by several moderately high-bandwidth links.

Figure 3-9 shows a schematic of Metro Mirror connections.

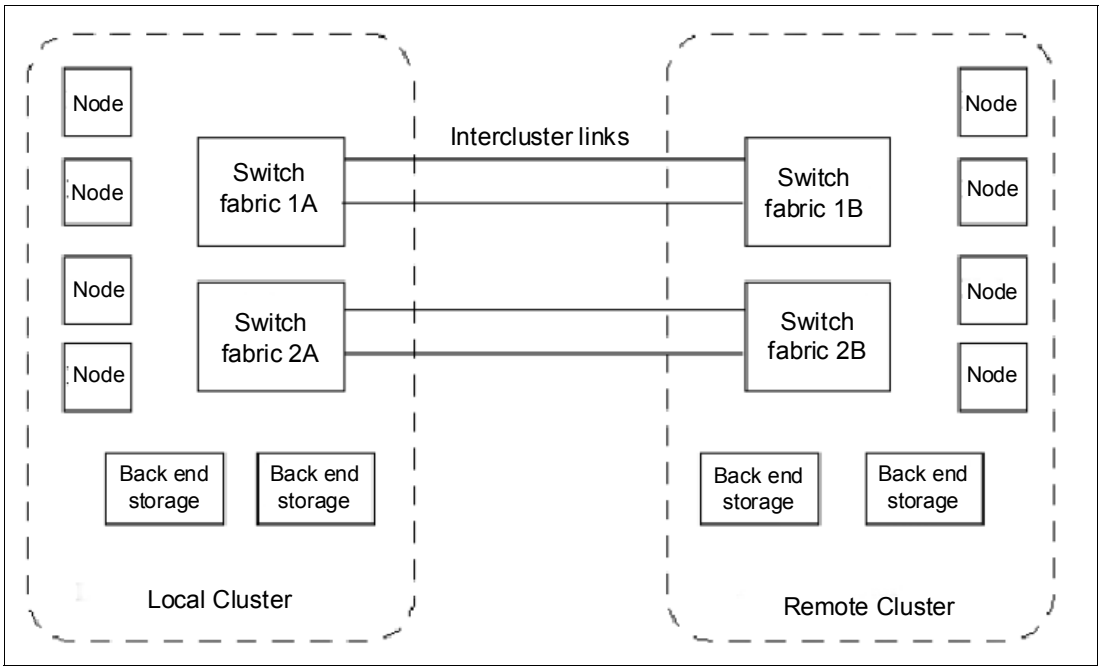


Figure 3-9 Metro Mirror connections

Figure 3-9 on page 73 contains two redundant fabrics. Part of each fabric exists at the local clustered system and at the remote clustered system. No direct connection exists between the two fabrics.

Technologies for extending the distance between two Storwize V7000 Gen2 clustered systems can be broadly divided into two categories: FC extenders and SAN multiprotocol routers.

Due to the more complex interactions involved, IBM explicitly tests products of this class for interoperability with the Storwize V7000 Gen2. You can obtain the current list of supported SAN routers in the supported hardware list on the Storwize V7000 Gen2 support website:

https://www.ibm.com/support/entry/myportal/product/system_storage/disk_systems/mid-range_disk_systems/ibm_storwize_v7000_(2076)

IBM has tested several FC extenders and SAN router technologies with the Storwize V7000 Gen2. You must plan, install, and test FC extenders and SAN router technologies with the Storwize V7000 Gen2 so that the following requirements are met:

- The round-trip latency between sites must not exceed 80 milliseconds (ms), 40 ms one way. For Global Mirror, this limit enables a distance between the primary and secondary sites of up to 8000 kilometers (km), 4970.96 miles, using a planning assumption of 100 km (62.13 miles) per 1 ms of round-trip link latency.
- The latency of long-distance links depends on the technology that is used to implement them. A point-to-point dark fiber-based link typically provides a round-trip latency of 1 ms per 100 km (62.13 miles) or better. Other technologies provide longer round-trip latencies, which affects the maximum supported distance.
- The configuration must be tested with the expected peak workloads.
- When Metro Mirror or Global Mirror is used, a certain amount of bandwidth is required for Storwize V7000 Gen2 inter-cluster heartbeat traffic. The amount of traffic depends on how many nodes are in each of the two clustered systems.
- The bandwidth between sites must, at the least, be sized to meet the peak workload requirements, in addition to maintaining the maximum latency that has been specified previously. You must evaluate the peak workload requirement by considering the average write workload over a period of one minute or less, plus the required synchronization copy bandwidth.

  Determine the true bandwidth that is required for the link by considering the peak write bandwidth to volumes participating in Metro Mirror or Global Mirror relationships, and adding it to the peak synchronization copy bandwidth.
- If the link between the sites is configured with redundancy so that it can tolerate single failures, you must size the link so that the bandwidth and latency statements continue to be true even during single failure conditions.
- The configuration is tested to simulate the failure of the primary site (to test the recovery capabilities and procedures), including eventual failback to the primary site from the secondary.
- The configuration must be tested to confirm that any failover mechanisms in the inter-cluster links interoperate satisfactorily with the Storwize V7000 Gen2.
- The FC extender must be treated as a normal link.
- The bandwidth and latency measurements must be made by, or on behalf of, the client. They are *not* part of the standard installation of the Storwize V7000 Gen2 by IBM. Make these measurements during installation, and record the measurements. Testing must be repeated after any significant changes to the equipment that provides the inter-cluster link.

Global Mirror guidelines

For Global Mirror, the following guidelines apply:

- When using Storwize V7000 Gen2 Global Mirror, all components in the SAN must be capable of sustaining the workload that is generated by application hosts and the Global Mirror background copy workload. Otherwise, Global Mirror can automatically stop your relationships to protect your application hosts from increased response times. Therefore, it is important to configure each component correctly.
- Use a SAN performance monitoring tool, such as IBM Tivoli Storage Productivity Center, which enables you to continuously monitor the SAN components for error conditions and performance problems. This tool helps you detect potential issues before they affect your disaster recovery solution.
- The long-distance link between the two clustered systems must be provisioned to provide for the peak application write workload to the Global Mirror source volumes, plus the client-defined level of background copy.
- The peak application write workload ideally must be determined by analyzing the Storwize V7000 Gen2 performance statistics.
- Statistics must be gathered over a typical application I/O workload cycle, which might be days, weeks, or months, depending on the environment on which the Storwize V7000 Gen2 is used. These statistics must be used to find the peak write workload that the link must be able to support.
- Characteristics of the link can change with use. For example, latency can increase as the link is used to carry an increased bandwidth. The user must be aware of the link's behavior in such situations, and ensure that the link remains within the specified limits. If the characteristics are not known, testing must be performed to gain confidence of the link's suitability.
- Users of Global Mirror must consider how to optimize the performance of the long-distance link, which depends on the technology that is used to implement the link. For example, when transmitting FC traffic over an IP link, it can be desirable to enable jumbo frames to improve efficiency.
- Using Global Mirror and Metro Mirror between the same two clustered systems is supported.
- Using Global Mirror and Metro Mirror between the Storwize V7000 Gen2 clustered system and IBM Storwize systems with a minimum code level of 6.3 is supported.
- It is supported for cache-disabled volumes to participate in a Global Mirror relationship; however, it not a preferred practice to do so.
- The **gmlinktolerance** parameter of the remote copy partnership must be set to an appropriate value. The default value is 300 seconds (five minutes), which is appropriate for most clients.
- During SAN maintenance, the user must choose to reduce the application I/O workload for the duration of the maintenance (so that the degraded SAN components are capable of the new workload):
  - Disable the **gmlinktolerance** feature.
  - Increase the **gmlinktolerance** value (meaning that application hosts might see extended response times from Global Mirror volumes).
  - Stop the Global Mirror relationships.

  If the **gmlinktolerance** value is increased for maintenance lasting *x* minutes, it must only be reset to the normal value *x* minutes after the end of the maintenance activity.

  If **gmlinktolerance** is disabled for the duration of the maintenance, it must be re-enabled after the maintenance is complete.

- Global Mirror volumes must have their preferred nodes evenly distributed between the nodes of the clustered systems. Each volume within an I/O Group has a preferred node property that can be used to balance the I/O load between nodes in that group.

  Figure 3-10 shows the correct relationship between volumes in a Metro Mirror or Global Mirror solution.

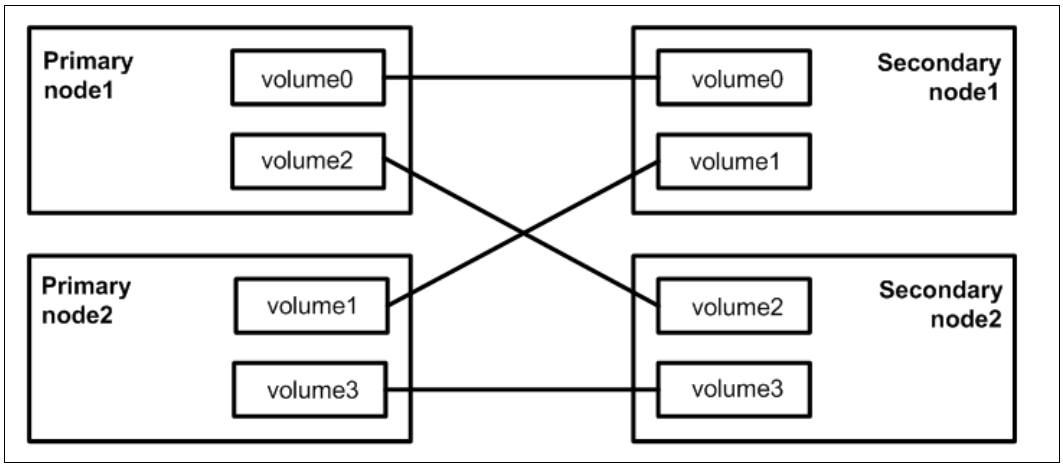


Figure 3-10 Correct volume relationship

- The capabilities of the storage controllers at the secondary clustered system must be provisioned to provide for the peak application workload to the Global Mirror volumes, plus the client-defined level of background copy, plus any other I/O being performed at the secondary site.

  The performance of applications at the primary clustered system can be limited by the performance of the back-end storage controllers at the secondary clustered system to maximize the amount of I/O that applications can perform to Global Mirror volumes.
- It is necessary to perform a complete review before using Serial Advanced Technology Attachment (SATA) for Metro Mirror or Global Mirror secondary volumes. Using a slower disk subsystem for the secondary volumes for high-performance primary volumes can mean that the Storwize V7000 Gen2 cache might not be able to buffer all of the writes, and flushing cache writes to SATA might slow I/O at the production site.
- Storage controllers must be configured to support the Global Mirror workload that is required of them:
  - Dedicate storage controllers to only Global Mirror volumes.
  - Configure the controller to ensure sufficient quality of service (QoS) for the disks being used by Global Mirror.
  - Ensure that physical disks are not shared between Global Mirror volumes and other I/O (for example, by not splitting an individual RAID array).
- MDisks in a Global Mirror storage pool must be similar in their characteristics, for example, RAID level, physical disk count, and disk speed. This requirement is true of all storage pools, but it is particularly important to maintain performance when using Global Mirror.
- When a consistent relationship is stopped, for example, by a persistent I/O error on the intercluster link, the relationship enters the `consistent_stopped` state. I/O at the primary site continues, but the updates are not mirrored to the secondary site. Restarting the relationship begins the process of synchronizing new data to the secondary disk.

While this synchronization is in progress, the relationship is in the `inconsistent_copying` state. Therefore, the Global Mirror secondary volume is not in a usable state until the copy has completed and the relationship has returned to a Consistent state.

For this reason, it is highly advisable to create a FlashCopy of the secondary volume before restarting the relationship. When started, the FlashCopy provides a consistent copy of the data, even while the Global Mirror relationship is copying. If the Global Mirror relationship does not reach the Synchronized state (if, for example, the intercluster link experiences further persistent I/O errors), the FlashCopy target can be used at the secondary site for disaster recovery purposes.

- If you plan to use a Fibre Channel over IP (FCIP) intercluster link, it is extremely important to design and size the pipe correctly.

  Example 3-2 shows a best-guess bandwidth sizing formula, assuming that the write/change rate is consistent.

  Example 3-2 WAN link calculation example

```
Amount of write data within 24 hours times 4 to allow for peaks
Translate into MB/s to determine WAN link needed
Example:
250 GB a day
250 GB * 4 = 1 TB
24 hours * 3600 secs/hr. = 86400 secs
1,000,000,000,000/ 86400 = approximately 12 MB/s,
Which means OC3 or higher is needed (155 Mbps or higher)
```

- If compression is available on routers or wide area network (WAN) communication devices, smaller pipelines might be adequate. Note that workload is probably not evenly spread across 24 hours. If there are extended periods of high data change rates, consider suspending Global Mirror during that time frame.
- If the network bandwidth is too small to handle the traffic, the application write I/O response times might be elongated. For the Storwize V7000 Gen2, Global Mirror must support short-term *Peak Write* bandwidth requirements.
- You must also consider the initial sync and resync workload. The Global Mirror partnership's background copy rate must be set to a value that is appropriate to the link and secondary back-end storage. The more bandwidth that you give to the sync and resync operation, the less workload can be delivered by the Storwize V7000 Gen2 for the regular data traffic.
- Do not propose Global Mirror if the data change rate exceeds the communication bandwidth, or if the round-trip latency exceeds 80 - 120 ms. A greater than 80 ms round-trip latency requires Solution for Compliance in a Regulated Environment and request for price quotation (SCORE/RPQ) submission.

3.3.12 SAN boot support

The Storwize V7000 Gen2 supports SAN boot or startup for IBM AIX, Microsoft Windows Server, and other operating systems. SAN boot support can change from time to time, so check the following websites regularly:

http://www.ibm.com/systems/storage/software/virtualization/StorwizeV7000/interop.html
http://www.ibm.com/systems/support/storage/ssic/interoperability.wss

3.3.13 Data migration from a non-virtualized storage subsystem

Data migration is an extremely important part of a Storwize V7000 Gen2 implementation. Therefore, you must accurately prepare a data migration plan. You might need to migrate your data for one of these reasons:

- To redistribute workload within a clustered system across the disk subsystem
- To move workload onto newly installed storage
- To move workload off old or failing storage, ahead of decommissioning it
- To move workload to rebalance a changed workload
- To migrate data from an older disk subsystem to Storwize V7000 Gen2-managed storage
- To migrate data from one disk subsystem to another disk subsystem

Because multiple data migration methods are available, choose the method that best fits your environment, your operating system platform, your kind of data, and your application's service-level agreement (SLA).

We can define data migration as belonging to three groups:

- Based on operating system Logical Volume Manager (LVM) or commands
- Based on special data migration software
- Based on the Storwize V7000 Gen2 data migration feature

With data migration, apply the following guidelines:

- Choose which data migration method best fits your operating system platform, your kind of data, and your SLA.
- Check the interoperability matrix for the storage subsystem to which your data is being migrated:

  http://www.ibm.com/systems/storage/software/virtualization/StorwizeV7000/interop.html
- Choose where you want to place your data after migration, in terms of the storage pools that relate to a specific storage subsystem tier.
- Determine whether enough free space or extents are available in the target storage pool.
- Decide if your data is critical and must be protected by a volume mirroring option, or if it must be replicated in a remote site for disaster recovery.
- Prepare offline all of the zone and LUN masking and host mappings that you might need, to minimize downtime during the migration.
- Prepare a detailed operation plan so that you do not overlook anything at data migration time.
- Run a data backup before you start any data migration. Data backup must be part of the regular data management process.
- You might want to use the Storwize V7000 Gen2 as a data mover to migrate data from a non-virtualized storage subsystem to another non-virtualized storage subsystem. In this case, you might have to add additional checks that relate to the specific storage subsystem to which you want to migrate.

  Be careful using slower disk subsystems for the secondary volumes for high-performance primary volumes, because the Storwize V7000 Gen2 cache might not be able to buffer all of the writes, and flushing cache writes to SATA might slow I/O at the production site.

3.3.14 Storwize V7000 Gen2 configuration backup procedure

Save the configuration externally when changes, such as adding new nodes, disk subsystems, and so on, have been performed on the clustered system. Saving the configuration is a crucial part of Storwize V7000 Gen2 management, and various methods can be applied to back up your Storwize V7000 Gen2 configuration.

The preferred practice is to implement an automatic configuration backup by applying the configuration backup command. We describe this command for the CLI in Chapter 8, “IBM Storwize V7000 Gen2 command-line interface” on page 143, and we describe the GUI operation in Chapter 9, “IBM Storwize V7000 Gen2 operations using the GUI” on page 169.

3.4 Performance considerations

Storage virtualization with the Storwize V7000 Gen2 improves flexibility and provides simpler management of a storage infrastructure, and it can also provide a substantial performance advantage for various workloads. The Storwize V7000 Gen2 caching capability, and its ability to stripe volumes across multiple disk arrays, are the reasons why performance improvement is significant when implemented with midrange disk subsystems. This technology is often only provided with high-end enterprise disk subsystems.

Tip: Technically, almost all storage controllers provide both striping (RAID 5 or RAID 10) and a form of caching. The real benefit is the degree to which you can stripe the data across all MDisks in a storage pool, and therefore have the maximum number of active spindles at one time. The caching is secondary. The Storwize V7000 Gen2 provides additional caching to the caching that midrange controllers provide (usually several GB), but enterprise systems have much larger caches.

To ensure the wanted performance and capacity of your storage infrastructure, undertake a performance and capacity analysis to reveal the business requirements of your storage environment. When this analysis is done, you can use the guidelines in this chapter to design a solution that meets the business requirements.

When considering performance for a system, always identify the bottleneck and, therefore, the limiting factor of a given system. You must also consider the component for whose workload you identify a limiting factor. The component might not be the same component that is identified as the limiting factor for other workloads.

When designing a storage infrastructure with Storwize V7000 Gen2, or implementing Storwize V7000 Gen2 in an existing storage infrastructure, you must consider the performance and capacity of the following components:

- The SAN
- The Storwize V7000 Gen2
- The disk subsystems
- The known or expected workload

The Storwize V7000 Gen2 is designed to handle large quantities of multiple paths from the back-end storage.

In most cases, the Storwize V7000 Gen2 can improve performance, especially on mid-sized to low-end disk subsystems, older disk subsystems with slow controllers, or uncached disk systems, for these reasons:

- The Storwize V7000 Gen2 can stripe across disk arrays, and it can stripe across the entire set of supported physical disk resources.
- Each Storwize V7000 Gen2 2076-524 node has 32 GB of base cache and 64 GB when the second processor and cache upgrade are added for Real-time Compression, providing a total of 64 GB/128 GB per I/O Group. (An 8-node cluster contains 512 GB base cache/10,240 GB with Real-time Compression.)

The Storwize V7000 Gen2 is capable of providing automated performance optimization of hot spots by using flash drives and Easy Tier.

3.4.1 SAN

The currently available Storwize V7000 Gen2 models have connection to 2 Gbps, 4 Gbps, 8 Gbps, and 16 Gbps switches. From a performance point of view, connecting the Storwize V7000 Gen2 to 8 Gbps or 16 Gbps switches is better to maximize the benefits of the performance and I/O speed.

Correct zoning on the SAN switch brings security and performance together. Implement a dual-HBA approach at the host to access the Storwize V7000 Gen2.

3.4.2 Disk subsystems

Each MDisk presented to Storwize V7000 Gen2 should consist of a single RAID group, or MDisk, of a single type of drive, from the underlying storage controller.

Advanced features, such as Disk Tiering, should be disabled on the underlying storage controller, because they will skew the results of the performance of the MDisk expected by Storwize V7000 Gen2.

Storwize family controllers should not use MDisk pooling, but should present a single MDisk, as a single pool (as a single volume), because the Storage Pool Balancing feature affects the way the MDisk behaves to Storwize V7000 Gen2.

3.4.3 Cache

The Storwize V7000 Gen2 clustered system is scalable up to four control enclosures and therefore four I/O groups, and the performance is nearly linear when adding more control enclosures into a Storwize V7000 Gen2 clustered system.

The large cache and advanced cache management algorithms in Storwize V7000 Gen2 enable it to improve on the performance of many types of underlying disk technologies. The Storwize V7000 Gen2 capability to manage, in the background, the destaging operations that are incurred by writes (in addition to still supporting full data integrity), assists with Storwize V7000 Gen2's capability in achieving good database performance.

There are several changes to how Storwize V7000 Gen2 uses its cache in the 7.3 code level. The cache is separated into two layers, an upper cache, and a lower cache.

Figure 3-11 shows the separation of the upper and lower cache.

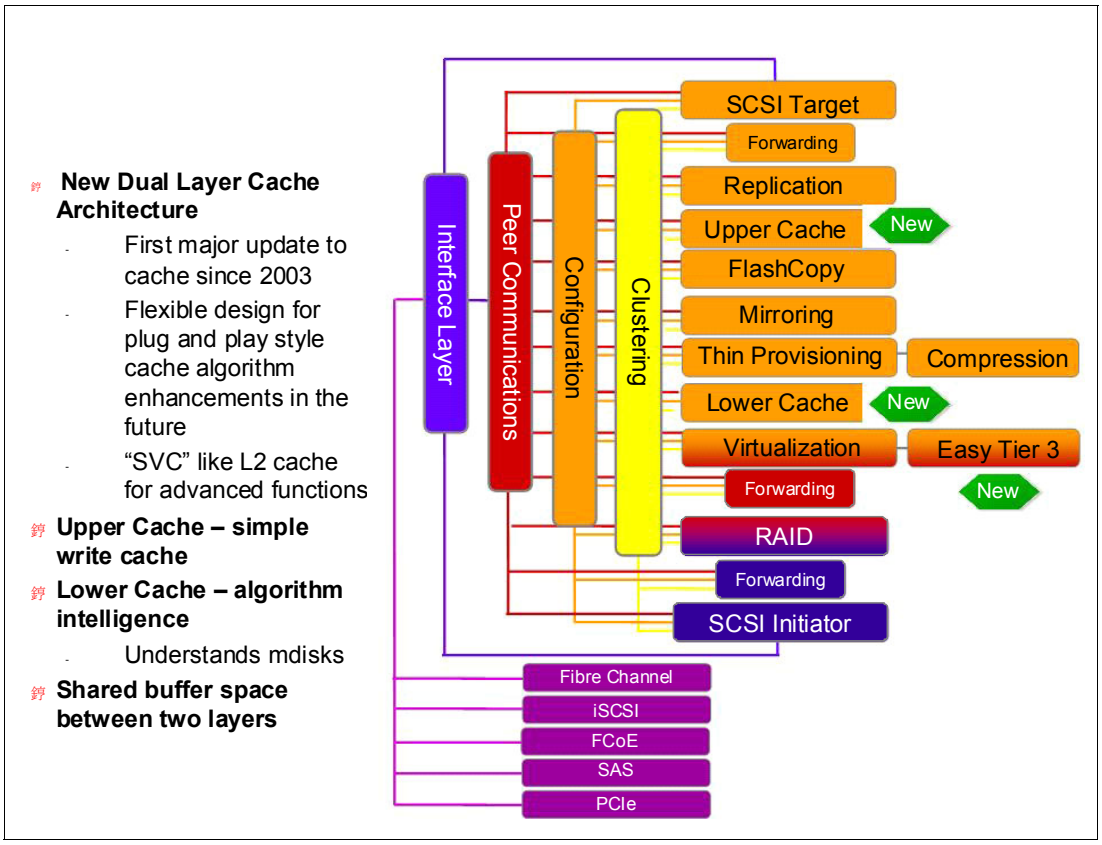


Figure 3-11 Separation of upper and lower cache

The upper cache delivers the following functionality enabling Storwize V7000 Gen2 to streamline data write performance:

- Provides fast write response times to the host by being as high up in the I/O stack as possible
- Provides partitioning

The lower cache delivers the following additional functionality:

- Ensuring write cache between two nodes is in sync
- Cache partitioning to ensure that a slow back end cannot consume the entire cache
- A destage algorithm that adapts to the amount of data and the back-end performance
- Providing read caching and prefetching

Combined together, the two levels of cache also deliver the following functionality:

- Pin data when LUN goes offline.
- Provide enhanced statistics for Tivoli Storage Productivity Center for Replication while maintaining compatibility with an earlier version.
- Provide trace for debugging.
- Report medium errors.
- Correctly resync cache and provide the atomic write functionality.
- Ensure that other partitions continue operation where one partition becomes 100% full of pinned data.

- Support fast-write (two-way and one-way), flush-through, and write-through.
- Integrate with T3 recovery procedures.
- Support two-way operation.
- Support none, read-only, and read/write as user-exposed caching policies.
- Support flush-when-idle.
- Support expanding cache as more memory becomes available to the platform.
- Enable credit throttling to avoid I/O skew and fairness/balanced I/O between the two nodes of the I/O Group.
- Enable switching of the preferred node without needing to move Volumes between I/O Groups.

Depending on the size, age, and technology level of the disk storage system, the total cache available in the Storwize V7000 Gen2 can be larger, smaller, or about the same as that associated with the disk storage. Because hits to the cache can occur in either the Storwize V7000 Gen2 or the disk controller level of the overall system, the system as a whole can take advantage of the larger amount of cache wherever it is located.

Therefore, if the storage controller level of the cache has the greater capacity, expect hits to this cache to occur, in addition to hits in the Storwize V7000 Gen2 cache.

Also, regardless of their relative capacities, both levels of cache tend to play an important role in enabling sequentially organized data to flow smoothly through the system. The Storwize V7000 Gen2 cannot increase the throughput potential of the underlying disks in all cases, because this increase depends on both the underlying storage technology and the degree to which the workload exhibits *hot spots* or sensitivity to cache size or cache algorithms.

IBM SAN Volume Controller 4.2.1 Cache Partitioning, REDP-4426, explains the IBM SAN Volume Controller and Storwize V7000 Gen2 cache partitioning capability:

`http://www.redbooks.ibm.com/abstracts/redp4426.html?Open`

3.4.4 Port Configuration

With the introduction of the 2076-524 nodes, and the ability to have up to eight FC ports per node, there are several different options that are valid for attaching storage to the Storwize V7000 Gen2 cluster. See section 3.3.2, “SAN zoning and SAN connections” on page 60 for configurations based on a single or dual four-port FC HBA adapter port allocation suggestions:

- If you require a high-throughput environment greater than 10 Gbps, then zoning all ports on the disk back-end storage to all ports on the Storwize V7000 Gen2 nodes in the cluster is a valid option.
- If you are looking to achieve the lowest latency storage environment, the allocation of four ports per node to inter-cluster traffic and inter-I/O Group traffic is the best suggestion. Each of the four ports should be zoned so that it only sees one other port in the same I/O Group. The same ports should be used for the remote traffic zoning. The remaining four ports per node can be separated into host and storage attachment roles.

Note: A port should be used for the same purpose (attached to the same switch and zones) for every node.

The only exception to this is when using mixed hardware types, in which case the lowest ports should be used for the same purposes, and the remaining ports can be allocated as required. (The lowest ports are the lowest numbered adapter slots, or the rightmost bits in the mask.)

Although virtualization with the Storwize V7000 Gen2 provides a great deal of flexibility, it does not diminish the necessity to have a SAN and disk subsystems that can deliver the wanted performance. Essentially, Storwize V7000 Gen2 performance improvements are gained by having as many MDisks as possible, therefore creating a greater level of concurrent I/O to the back-end without overloading a single disk or array.

Assuming that no bottlenecks exist in the SAN or on the disk subsystem, remember that you must follow specific guidelines when you perform these tasks:

- Creating a storage pool
- Creating volumes
- Connecting to or configuring hosts that must receive disk space from a Storwize V7000 Gen2 clustered system

You can obtain more detailed information about performance and preferred practices for the Storwize V7000 Gen2 in *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521:

http://www.redbooks.ibm.com/abstracts/sg247521.html?Open

3.4.5 Performance monitoring

Performance monitoring must be an integral part of the overall information technology (IT) environment.

This topic is covered in more detail in *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521:

http://www.redbooks.ibm.com/abstracts/sg247521.html?Open

For the Storwize V7000 Gen2, as for the other IBM storage subsystems, the official IBM product to collect performance statistics and supply a performance report is the IBM Tivoli Storage Productivity Center.

You can obtain more information about using the IBM Tivoli Storage Productivity Center to monitor your storage subsystem in *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364:

http://www.redbooks.ibm.com/abstracts/sg247364.html?Open

More reference links for IBM Storwize V7000 Gen2 Support Portal to download code, download manuals, and review current information for planning and installation, can be viewed on the following website:

http://www.ibm.com/storage/support/storwize/v7000

IBM Storwize V7000 Gen2 Supported Hardware List, Device Driver, Firmware, and Suggested Software Levels V7.x can be viewed on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003741

IBM Storwize V7000 Gen2 is listed in the IBM System Storage Interoperation Center (SSIC):

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

IBM Storwize V7000 Gen2 Configuration Limits and Restrictions can be viewed on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003741

The IBM Storwize V7000 Gen2 Knowledge Center is on the following website:

http://www.ibm.com/support/knowledgecenter/ST3FR7/landing/V7000_welcome.html

View IBM Storwize V7000 Gen2 Power and Cooling Requirements on the following website:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003711

4

IBM Storwize V7000 Gen2 Easy Tier

This chapter describes the history of IBM Storwize V7000 Easy Tier, describes the changes in versions, and the enhancements in IBM Storwize family software V7.3. In addition, we assess the capabilities of the new functionality, compare usage cases, and identify configuration and deployment considerations that should be taken into account during the planning stage of your Storwize V7000 Gen2 deployment.

In the following chapter, our intent is to provide only a basic technical overview, and focus on the benefits with the new version of Easy Tier. More details for planning and configuration are available in the following IBM Redbooks publications:

- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072
- *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- *IBM DS8000 Easy Tier*, REDP-4667 (this concept is similar to Storwize V7000 Gen2 Easy Tier)

4.1 IBM Storwize family software Easy Tier history

IBM Easy Tier is a performance function that automatically and non-disruptively migrates frequently accessed data from magnetic media to solid-state drives (SSD or flash drives). In that way, the most frequently accessed data is stored on the fastest storage tier, and the overall performance is improved.

IBM Storwize family software has benefited from the software development work for the IBM System Storage DS8000 product, in which there have been six versions of Easy Tier. Of those versions, versions 1 and 3 have been implemented in the 7.3 IBM Storwize family software.

The first generation of Easy Tier introduced automated storage performance management by efficiently boosting enterprise-class performance with flash drives (SSD), and automating storage tiering from enterprise-class drives to flash drives. These changes optimized flash deployments with minimal costs. Easy Tier also introduced dynamic volume relocation and dynamic extent pool merge.

The second generation of Easy Tier was only implemented in DS8000.

The third generation of Easy Tier introduces further enhancements that provide automated storage performance and storage economics management across all three drive tiers (flash, enterprise, and Nearline storage tiers) as outlined in Figure 4-1. It enables you to consolidate and efficiently manage more workloads on a single IBM Storwize V7000 Gen2 system. It also introduces support for *storage pool balancing* in homogeneous pools. It is based on performance, not capacity.

4.1.1 New features in Easy Tier 3

The following enhancements are included in Easy Tier 3:

- Support for three tiers of disk, or a mixture of any two tiers
- Storage pool balancing
- Enhancements to the IBM Storage Tier Advisor Tool (STAT) tool, including additional graphing from the STAT utility.

Figure 4-1 shows the supported easy tier pools now available in Easy Tier 3.

hree Tier Pools :		
SSD	Enterprise	Nearline
wo Tier Pools :		
SSD	Enterprise	-
SSD	Nearline	-
-	Enterprise	Nearline
ingle tier pools :		
SSD	-	-
-	Enterprise	-
-	-	Nearline

Figure 4-1 Easy Tier 3 multi-level tiering

By default, Easy Tier is enabled on any pool that contains more than one class of disk drive. Easy Tier manages extent migration using the following processes:

- Promote / swap

  Moves the relevant hot extents to a higher-performing tier

- Warm demote:
  - Prevents performance overload of a tier by demoting a warm extent to the lower tier
  - Triggered when bandwidth or input/output operations per second (IOPS) exceeds a predefined threshold

- Cold demote

  Coldest data moved to lower hard disk drive (HDD) tier

- Expanded cold demote

  Demotes appropriate sequential workloads to the lowest tier to better use Nearline bandwidth

- Storage pool balancing:
  - Redistribute extents within a tier to balance use across managed disks (MDisks) for maximum performance
  - Either move or swap

Note: Extent migrations occur only between adjacent tiers.

Figure 4-2 shows the Easy Tier process for extent migration.

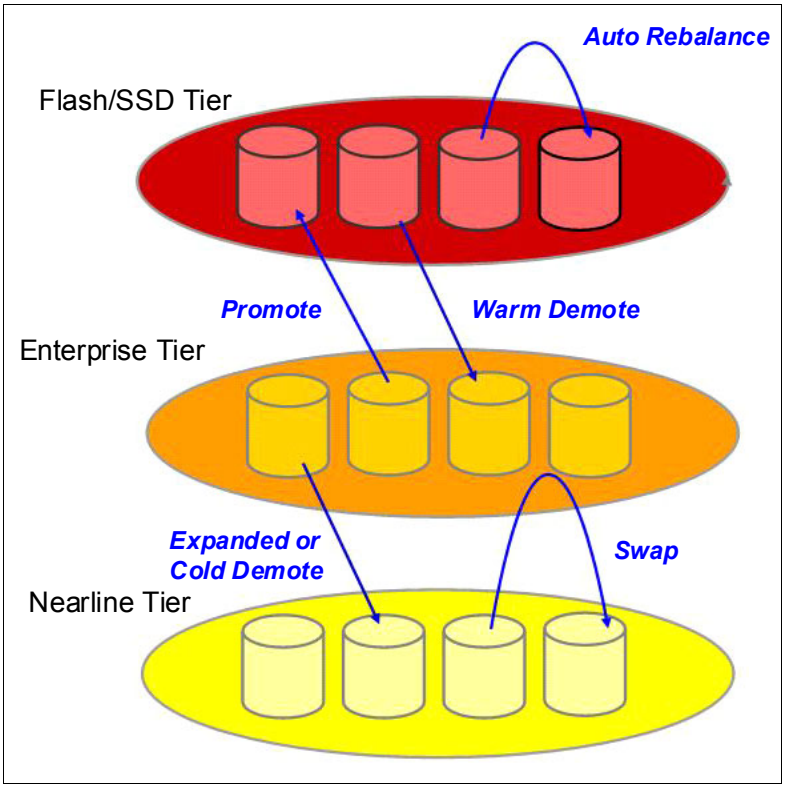


Figure 4-2 Easy Tier extent migration

4.1.2 Storage pool balancing

Storage pool balancing is a new feature in 7.3 Storwize family software that, although associated with Easy Tier, operates independently of Easy Tier. For Storwize V7000 Gen2, storage pool balancing does not require an Easy Tier license. This feature assesses the extents that are written in a pool, and balances them automatically across all MDisks in the pool. This process works with Easy Tier when multiple classes of disks exist in a single pool.

The process automatically balances existing data when new MDisks are added into an existing pool, even if the pool only contains a single type of drive.

Note: Storage pool balancing is used to balance extents across a storage Pool with the same performance tier. For example, when adding new drives of the same class to an existing storage pool, storage pool balancing redistributes the extents based on performance factors, not capacity.

If a pool contains a single type of MDisk, Easy Tier goes into balancing mode (status is *balanced*). When the pool contains multiple types of MDisks, Easy Tier is automatically turned on (status is *active*).

The Storwize V7000 Gen2 does not automatically identify external flash drive MDisks. All external MDisks are put into the enterprise tier by default. You must manually identify external flash drive MDisks and change their tiers. Local (internal) MDisks are automatically classified as flash, enterprise, or Nearline, and are placed in the appropriate tier without user intervention.

4.2 Performance and monitoring considerations

In this section, we briefly explain the effect of Easy Tier on performance, and introduce its monitoring tools.

4.2.1 Considerations for optimal performance

With the availability of Storwize family software version 7.3, Storwize V7000 Gen2 can now classify and identify performance profiles based on the category of drive within a storage pool. A Storwize V7000 Gen2 recognizes three types of disk: Flash drives, enterprise drives, and Nearline drives.

However, when a new external MDisk is added to Storwize V7000 Gen2, Storwize V7000 Gen2 does not automatically classify the MDisk by the type of drive that the MDisk consists of. You need to manually select the MDisk, choose the type of drive, and allocate it to the MDisk.

In Figure 4-3 on page 90, `md_v7kgen1-2-001` is a serial-attached SCSI (SAS) MDisk (enterprise tier) allocated from an IBM Storwize V7000 Gen1 storage controller virtualized to the Storwize V7000 Gen2. When you right-click the required MDisk, you can choose the option **Select Tier**.

The graphical user interface (GUI) of Storwize V7000 Gen2 does not have the same behavior for internal MDisks. The option appears in gray because you can't change it from the GUI, but you can do it from the command-line interface (CLI) using the **chmdisk** command, as shown in Example 4-1.

Example 4-1 The chmdisk command to change the tier of an internal MDisk

```
IBM_Storwize:ITSO_V7000Gen2:superuser>lsmdisk md_v7kgen2-2-001
id 0
name md_v7kgen2-2-001
status online
mode array
mdisk_grp_id 0
mdisk_grp_name INT_V7KGEN2
capacity 558.4GB
quorum_index
block_size
controller_name
ctrl_type
ctrl_WWNN
controller_id
path_count
max_path_count
ctrl_LUN_#
UID
preferred_WWPN
active_WWPN
fast_write_state empty
raid_status online
raid_level raid1
redundancy 1
strip_size 256
spare_goal 1
spare_protection_min 1
balanced exact
tier enterprise
slow_write_priority latency
fabric_type
site_id
site_name
easy_tier_load
IBM_Storwize:ITSO_V7000Gen2:superuser>chmdisk -tier nearline md_v7kgen2-2-001
IBM_Storwize:ITSO_V7000Gen2:superuser>
IBM_Storwize:ITSO_V7000Gen2:superuser>lsmdisk md_v7kgen2-2-001
id 0
name md_v7kgen2-2-001
status online
mode array
mdisk_grp_id 0
mdisk_grp_name INT_V7KGEN2
capacity 558.4GB
quorum_index
block_size
controller_name
ctrl_type
ctrl_WWNN
```

```
controller_id
path_count
max_path_count
ctrl_LUN_#
UID
preferred_WWPN
active_WWPN
fast_write_state empty
raid_status online
raid_level raid1
redundancy 1
strip_size 256
spare_goal 1
spare_protection_min 1
balanced exact
tier nearline
slow_write_priority latency
fabric_type
site_id
site_name
easy_tier_load
IBM_Storwize:ITSO_V7000Gen2:superuser>
```

Figure 4-3 shows how to select a tier.

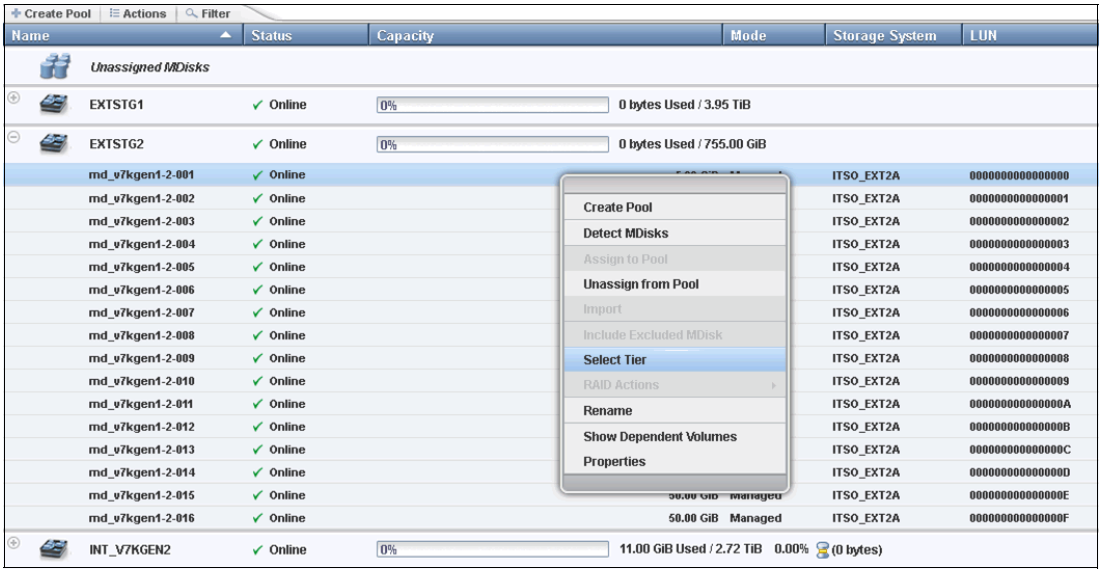


Figure 4-3 Shows the option to select a tier for a specific external MDisk

Figure 4-4 shows the options that are available for each MDisk to define the class of drive within the MDisk.

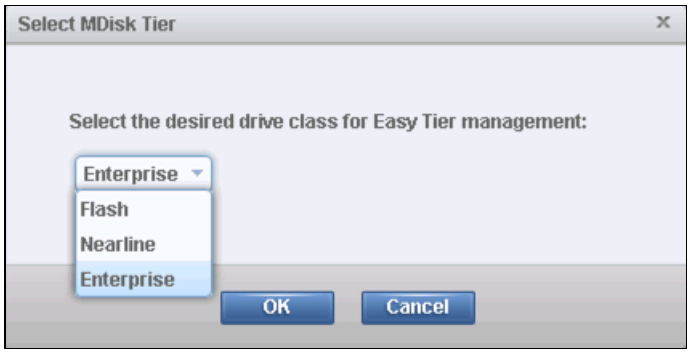


Figure 4-4 Shows the class of drives available for Storwize V7000 Gen2 MDisks

Figure 4-5 shows an example of the properties window for a three-tier storage pool.

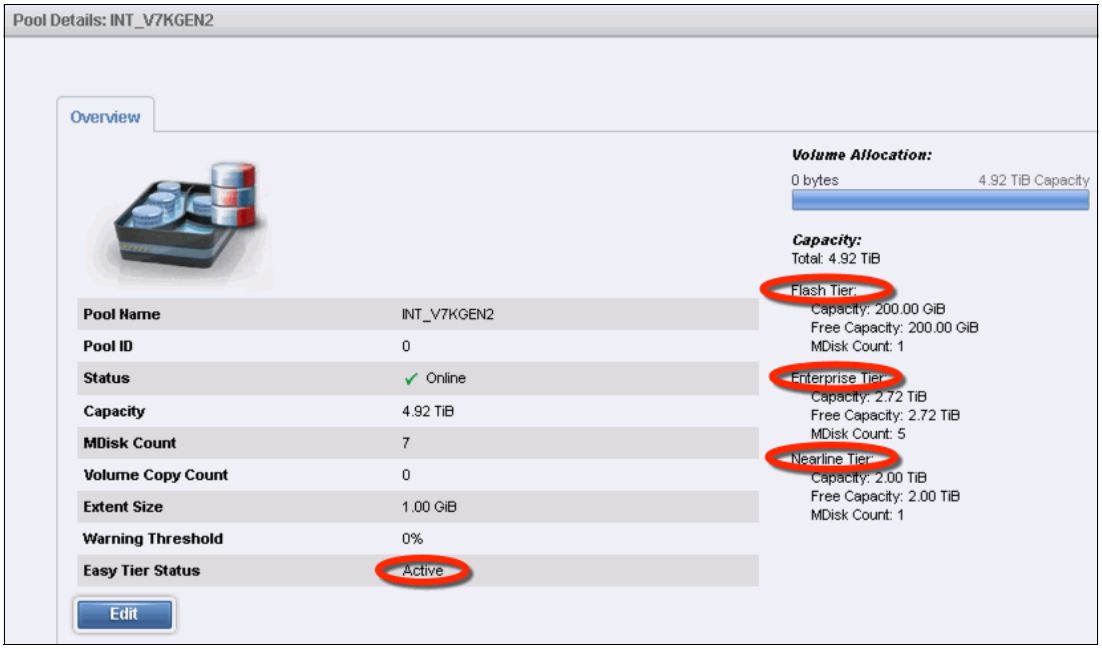


Figure 4-5 Properties window of a three-tier storage pool

Important: When virtualizing any Storwize family storage controller that supports storage pool balancing, you must disable storage pool balancing on the virtualized Storwize family storage controller by using the **`chmdiskgrp`** command. Failure to do so means that storage pool balancing on MDisks within Storwize V7000 Gen2 competes with storage pool balancing on the Storwize virtualized controller, causing performance degradation at both levels.

Be sure that you are using the same extent size on your storage pools.

If you have any flash drives into your virtualized Storwize family storage controller, use them for Easy Tier at the higher level (Storwize V7000 Gen2).

4.2.2 Monitoring tools

The IBM STAT utility is a Microsoft Windows console application that analyzes heat data files produced by Easy Tier, and produces a graphical display of the amount of “hot” data per volume. It also displays predictions of how additional flash drive (SSD) capacity, enterprise drive, and Nearline drive could benefit performance, for the system and by storage pool.

Heat data files are produced approximately once a day (every 24 hours) when Easy Tier is active on one or more storage pools, and summarizes the activity per volume since the prior heat data file was produced. On Storwize family products, the heat data file is in the `/dumps` directory on the configuration node, and is named `dpa_heat.`*`node_name`*`.`*`time_stamp`*`.data`.

Any existing heat data file is erased when it has existed for longer than seven days. The user must off-load the file, and start STAT from a Windows command-line interface (CLI) with the file specified as a parameter. The user can also specify the output directory. The STAT creates a set of Hypertext Markup Language (HTML) files, and the user can open the resulting `index.html` file in a browser to view the results.

Updates to the STAT for Storwize V7000 Gen2 have added additional capability for reporting. As a result, when the STAT is run on a heat map file, an additional three comma-separated values (CSV) files are created and placed in the `Data_files` directory.

The IBM STAT utility can be downloaded from the IBM Support website:

http://www.ibm.com/support/docview.wss?uid=ssg1S4000935

Figure 4-6 shows the CSV files highlighted in the `Data_files` directory after running the STAT over an IBM storage area network (SAN) Volume Controller heatmap.

```
C:\Program Files (x86)\IBM\STAT\Data_files>dir
 Volume in drive C has no label.
 Volume Serial Number is 14D1-6D4C

 Directory of C:\Program Files (x86)\IBM\STAT\Data_files

09/01/2014  09:55 AM    <DIR>          .
09/01/2014  09:55 AM    <DIR>          ..
09/01/2014  09:55 AM               161 7836640-2_data_movement.csv
09/01/2014  09:55 AM             3,596 7836640-2_skew_curve.csv
09/01/2014  09:55 AM            14,912 7836640-2_workload_ctg.csv
05/11/2014  07:49 PM               271 banner_background.gif
05/11/2014  07:42 PM             2,819 banner_right.gif
05/11/2014  07:42 PM             8,248 banner_title.gif
05/11/2014  07:42 PM               942 head.html
09/01/2014  09:55 AM               886 innerBottom.html
05/11/2014  07:42 PM               669 innerTop.html
05/11/2014  07:42 PM             2,515 main.css
09/01/2014  09:55 AM            51,380 pool_rec_p0000.html
09/01/2014  09:55 AM            48,247 pool_rec_p0001.html
09/01/2014  09:55 AM            57,172 pool_rec_p0002.html
09/01/2014  09:55 AM            17,439 pool_rec_p0003.html
05/11/2014  07:49 PM             8,273 product.jpg
09/01/2014  09:55 AM            15,119 System Summary.html
09/01/2014  09:55 AM             4,499 Systemwide Recommendation.html
              17 File(s)        237,148 bytes
               2 Dir(s)  20,779,122,688 bytes free

C:\Program Files (x86)\IBM\STAT\Data_files>
```

Figure 4-6 CSV files created by the STAT for Easy Tier

In addition to the STAT, Storwize family software V7.3 now has an additional utility, which is a Microsoft Structured Query Language (SQL) file for creating additional graphical reports of the workload that Easy Tier is performing. The IBM STAT Charting Utility takes the output of the three CSV files and turns them into graphs for simple reporting.

The three new graphs display the folowing information:

- Workload categorization

  New workload visuals help you compare activity across tiers within and across pools, to help determine optimal drive mix for current workloads. The output is illustrated in Figure 4-7.

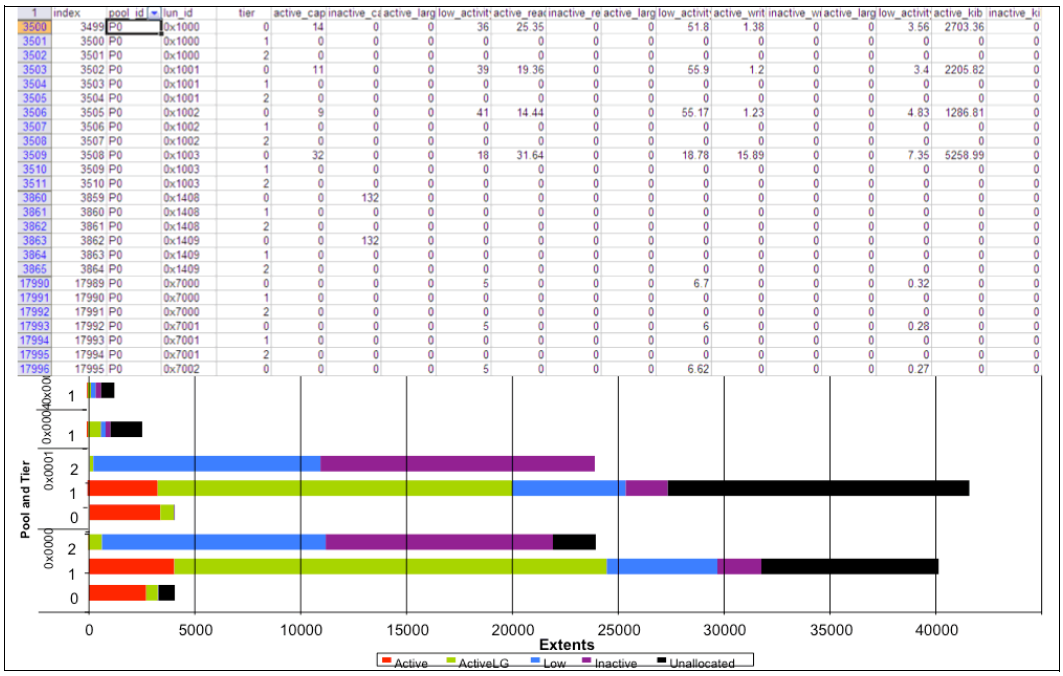
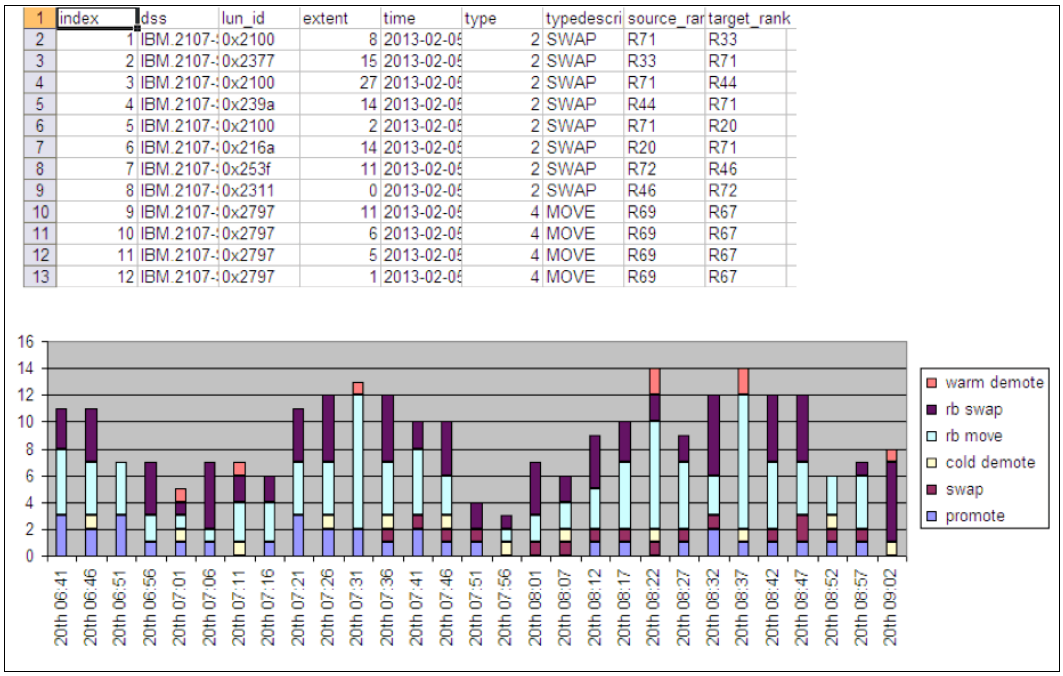


Figure 4-7 STAT Charting Utility Workload Categorization report

- Daily movement reporting

  This graph is a new Easy Tier summary report illustrating data migration activity with five-minute intervals, and can help visualize migration types and patterns for current workloads.

The output is illustrated in Figure 4-8.

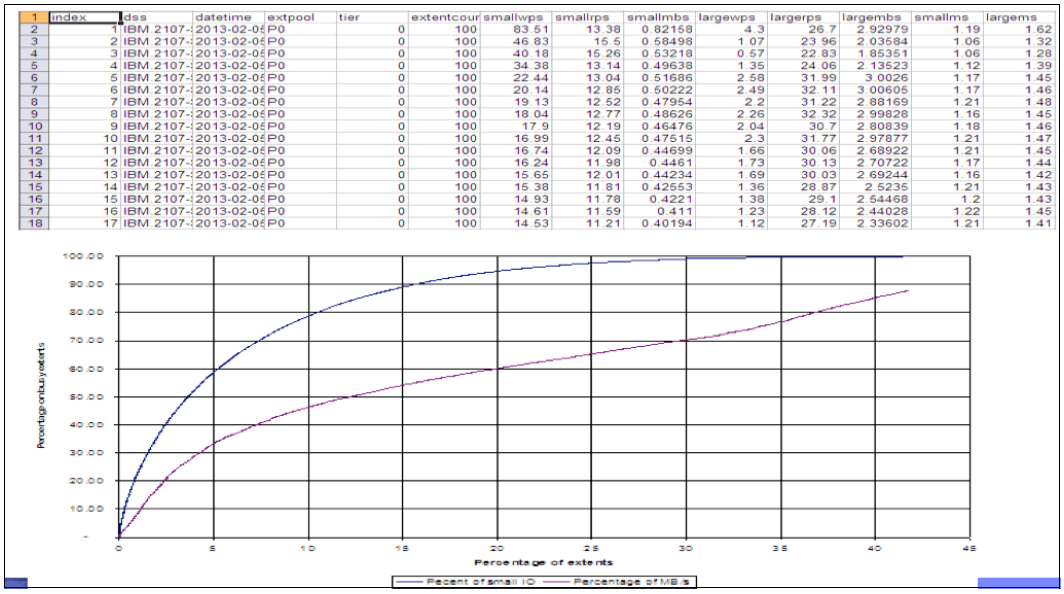


	index	dss	lun_id	extent	time	type	typedescri	source_rar	target_rank
2	1	IBM.2107-(	0x2100	8	2013-02-0	2	SWAP	R71	R33
3	2	IBM.2107-(	0x2377	15	2013-02-0	2	SWAP	R33	R71
4	3	IBM.2107-(	0x2100	27	2013-02-0	2	SWAP	R71	R44
5	4	IBM.2107-(	0x239a	14	2013-02-0	2	SWAP	R44	R71
6	5	IBM.2107-(	0x2100	2	2013-02-0	2	SWAP	R71	R20
7	6	IBM.2107-(	0x216a	14	2013-02-0	2	SWAP	R20	R71
8	7	IBM.2107-(	0x253f	11	2013-02-0	2	SWAP	R72	R46
9	8	IBM.2107-(	0x2311	0	2013-02-0	2	SWAP	R46	R72
10	9	IBM.2107-(	0x2797	11	2013-02-0	4	MOVE	R69	R67
11	10	IBM.2107-(	0x2797	6	2013-02-0	4	MOVE	R69	R67
12	11	IBM.2107-(	0x2797	5	2013-02-0	4	MOVE	R69	R67
13	12	IBM.2107-(	0x2797	1	2013-02-0	4	MOVE	R69	R67

Figure 4-8 STAT Charting Utility Daily Summary report

- Workload skew

  This graph shows the skew of all workloads across the system, to help clients visualize and accurately tier configurations when adding capacity or a new system. The output is illustrated in Figure 4-9.



	index	dss	datetime	extpool	tier	extentcour	smallwps	smallrps	smallmbs	largewps	largerps	largembs	smallms	largems
2	1	IBM.2107-(	2013-02-0	P0	0	100	83.51	13.38	0.82158	4.3	26.7	2.92979	1.19	1.62
3	2	IBM.2107-(	2013-02-0	P0	0	100	46.83	15.5	0.58498	1.07	23.96	2.03584	1.06	1.32
4	3	IBM.2107-(	2013-02-0	P0	0	100	40.18	15.26	0.53218	0.57	22.83	1.85351	1.06	1.28
5	4	IBM.2107-(	2013-02-0	P0	0	100	34.38	13.14	0.49638	1.35	24.06	2.13523	1.12	1.39
6	5	IBM.2107-(	2013-02-0	P0	0	100	22.44	13.04	0.51686	2.58	31.99	3.0026	1.17	1.45
7	6	IBM.2107-(	2013-02-0	P0	0	100	20.14	12.85	0.50222	2.49	32.11	3.00605	1.17	1.46
8	7	IBM.2107-(	2013-02-0	P0	0	100	19.13	12.52	0.47954	2.2	31.22	2.88169	1.21	1.48
9	8	IBM.2107-(	2013-02-0	P0	0	100	18.04	12.77	0.48626	2.26	32.32	2.99828	1.16	1.45
10	9	IBM.2107-(	2013-02-0	P0	0	100	17.9	12.19	0.46476	2.04	30.7	2.80839	1.18	1.46
11	10	IBM.2107-(	2013-02-0	P0	0	100	16.99	12.45	0.47515	2.3	31.77	2.97877	1.21	1.47
12	11	IBM.2107-(	2013-02-0	P0	0	100	16.74	12.09	0.44699	1.66	30.06	2.68922	1.21	1.45
13	12	IBM.2107-(	2013-02-0	P0	0	100	16.24	11.98	0.4461	1.73	30.13	2.70722	1.17	1.44
14	13	IBM.2107-(	2013-02-0	P0	0	100	15.65	12.01	0.44234	1.69	30.03	2.69244	1.16	1.42
15	14	IBM.2107-(	2013-02-0	P0	0	100	15.38	11.81	0.42553	1.36	28.87	2.5235	1.21	1.43
16	15	IBM.2107-(	2013-02-0	P0	0	100	14.93	11.78	0.4221	1.38	29.1	2.64468	1.2	1.43
17	16	IBM.2107-(	2013-02-0	P0	0	100	14.61	11.59	0.411	1.23	28.12	2.44028	1.22	1.45
18	17	IBM.2107-(	2013-02-0	P0	0	100	14.53	11.21	0.40194	1.12	27.19	2.33602	1.21	1.41

Figure 4-9 STAT Charting Utility Workload Skew report

The STAT Charting Utility can be downloaded from the IBM support website:

http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5251

5

The IBM Storwize V7000 Gen2 Initial Configuration

In this chapter, we describe the following topics:

- Managing the IBM Storwize V7000 Gen2:
  - Network requirements
  - Prerequisites
- The Storwize V7000 Gen2 initial configuration:
  - How-to make the first connection to the Storwize V7000 Gen2
  - System Setup wizard

5.1 Managing the Storwize V7000 Gen2

You can manage the Storwize V7000 Gen2 in many ways. The following methods are the most common:

- Using the Storwize V7000 Gen2 Management graphical user interface (GUI)
- Using a PuTTY-based command-line interface (CLI)
- Using IBM Tivoli Storage Productivity Center for Replication

Figure 5-1 shows the various ways to manage the Storwize V7000 Gen2.

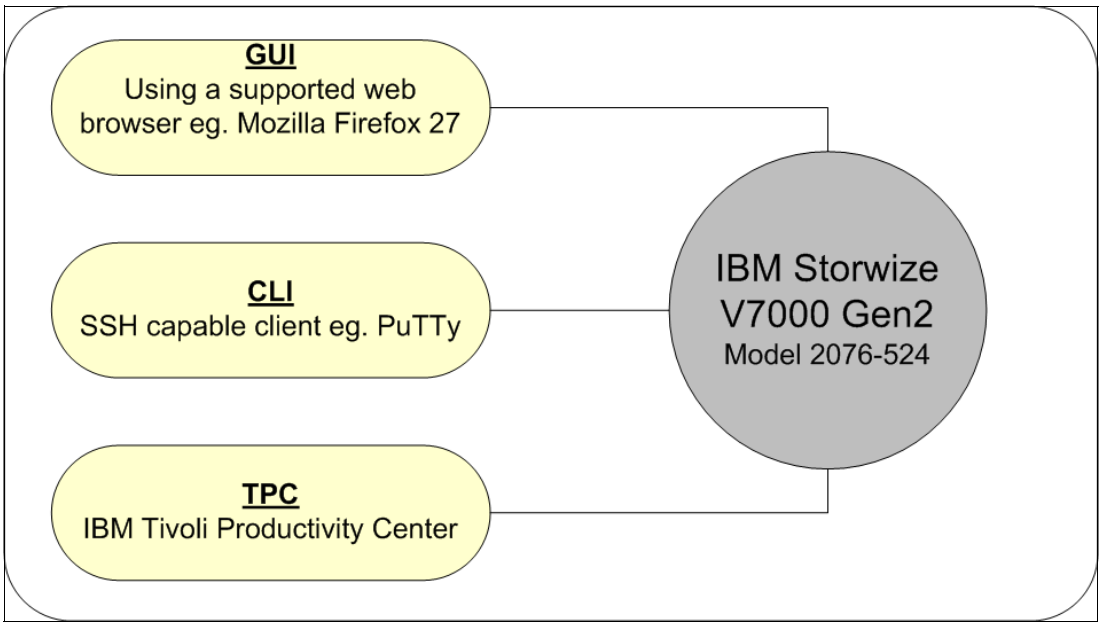


Figure 5-1 The Storwize V7000 Gen2 management

Information: For information about supported web browsers, see the following IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/v7000_ichome_730.html

Note that you have full management control of the Storwize V7000 Gen2, regardless of which method you choose. IBM Tivoli Storage Productivity Center is a robust software product with various functions that needs to be purchased separately. You can learn more about IBM Tivoli Storage Productivity Center on the following website:

http://www.ibm.com/software/products/en/tivostorprodcent

5.1.1 Network requirements for the IBM Storwize V7000 Gen2

To plan your installation, you need to take into account the Transmission Control Protocol/Internet Protocol (TCP/IP) address requirements of the Storwize V7000 Gen2, and the requirements for the Storwize v7000 Gen2 to access other services. You must also plan the address allocation and the Ethernet router, gateway, and firewall configuration to provide the required access and network security.

Figure 5-2 shows the TCP/IP ports and services that are used by the Storwize V7000 Gen2.

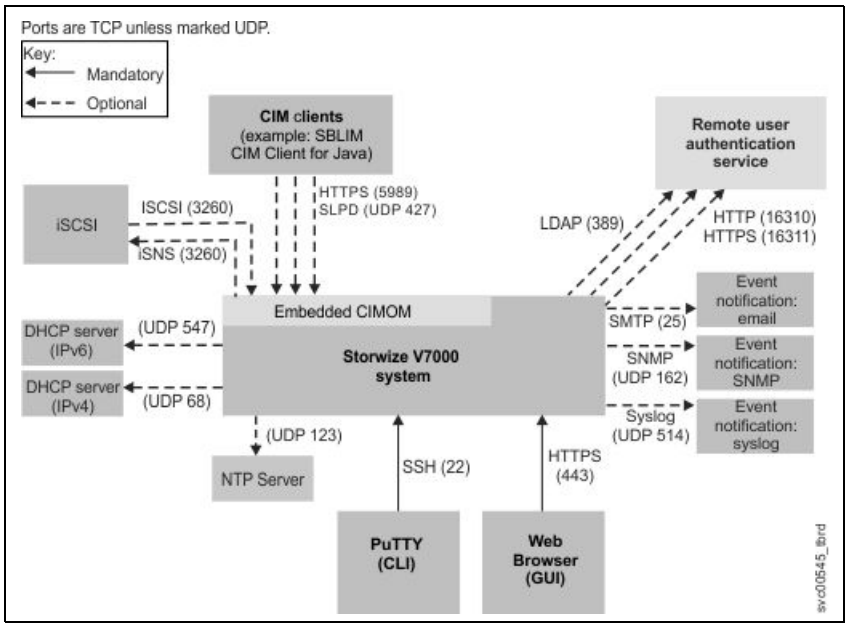


Figure 5-2 TCP/IP ports

For more information about TCP/IP prerequisites, see Chapter 3, “Planning and configuration” on page 47.

5.1.2 Prerequisites

Ensure that the Storwize V7000 Gen2 has been physically installed, and that Ethernet and Fibre Channel (FC) connectivity has been correctly configured. For information about physical connectivity, see Chapter 3, “Planning and configuration” on page 47.

Before configuring the Storwize V7000 Gen2, ensure that the following information is available:

- Licenses

  The licenses indicate whether the client is permitted to use IBM Storwize V7000 Easy Tier, IBM FlashCopy, IBM Storwize V7000 External Virtualization, remote copy, and IBM Real-time Compression. For details about Licensing, see Chapter 3, “Planning and configuration” on page 47.

- IPv4 addressing:
  - Cluster IPv4 address (one IP address for management)
  - Service IPv4 addresses (two addresses for the service interfaces)
  - IPv4 subnet mask
  - Gateway IPv4 address
- IPv6 addressing:
  - Cluster IPv6 address (one address for management)
  - Service IPv6 addresses (two addresses for the service interface, one for each node)
  - IPv6 prefix
  - Gateway IPv6 address

5.2 Initial Configuration of the IBM Storwize V7000 Gen2

For our initial configuration we are using the following hardware:

- 1 x IBM Storwize V7000 Gen2 model 2076-524
- 2 x four-port 8 gigabits per second (Gbps) FC host interface card (one per node)
- 1 x IBM Storwize V7000 Expansion Enclosure
- 2 x SAN Switches (for a redundant SAN fabric)

5.2.1 How to make the first connection to the Storwize V7000 Gen2

Follow these steps to connect to Storwize V7000 Gen2:

1. The first step is to connect a PC or notebook (PC) to the Technician Port on the rear of the Storwize V7000 Gen2 node. See Figure 5-3 for the location of the Technician Port. The Technician Port provides a Dynamic Host Configuration Protocol (DHCP) IP address V4, so you must ensure that your PC is configured for DHCP. The default IP address for a new node is `192.168.0.1`. You can, however, also use a static IP, which should be set to `192.168.0.2` on your PC or notebook.

   The Storwize V7000 Gen2 does not provide IPv6 IP addresses for the Technician Port.

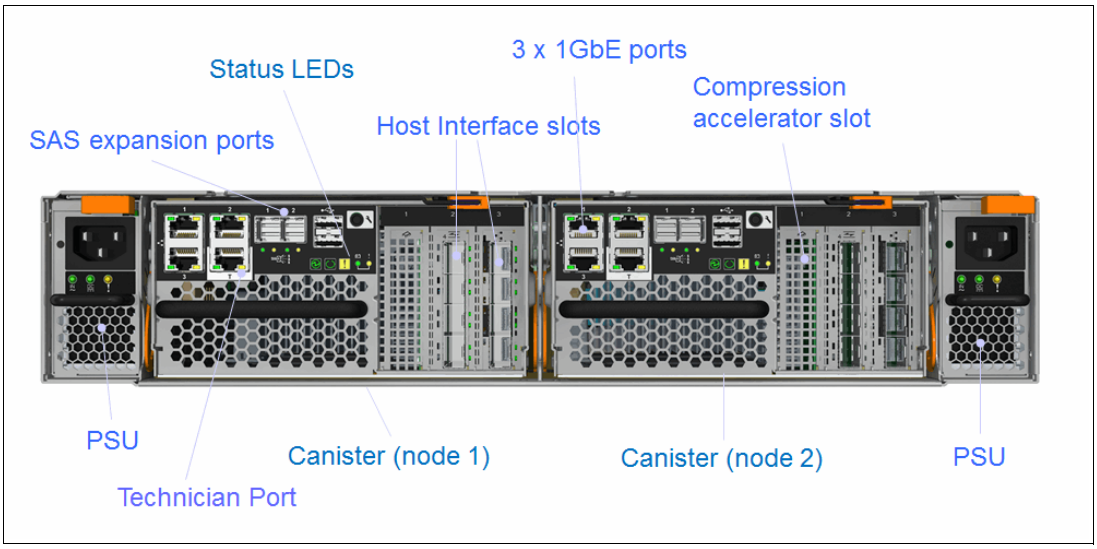


Figure 5-3 Rear of Storwize V7000 Gen2

> **Nodes:** During the initial configuration, you will probably see certificate warnings because these certificates are self-issued. You can accept these warnings because they are not harmful.

2. When your PC is connected to the Technician Port, and you have validated that you have an IPv4 DCHP address, for example `192.168.0.12` (the first IP address that the Storwize V7000 Gen2 node assigns), open a supported browser.

   This should automatically redirect you to `192.168.0.1`, and the initial configuration of the cluster can start.

3. Figure 5-4 shows the Welcome window where you start the Wizard that enables you to configure a new system, or expand an existing system.

Figure 5-4 Welcome window

4. This chapter focuses on setting up a new system, so we select **Yes** and click **Next**.

> **Remember:** If you are adding a Storwize V7000 Gen2 into an existing cluster, ensure that the existing systems are running code level 7.3 or higher, because the 2076-524 only supports code level 7.3 or higher.

5. The next window will ask you to set an IP address for the cluster. You can choose between an IPv4 or IPv6 address. In Figure 5-5, we have set an IPv4 address.

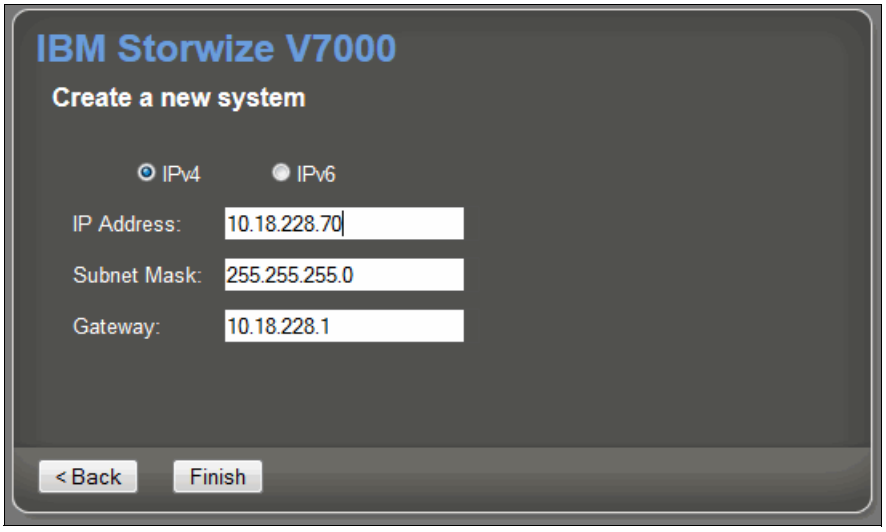


Figure 5-5 Setting the IP address

6. Click **Finish** when you have entered the IP Address, Subnet Mask, and Gateway. The system starts to initialize the system (node), and you see a window, such as the one in Figure 5-6.

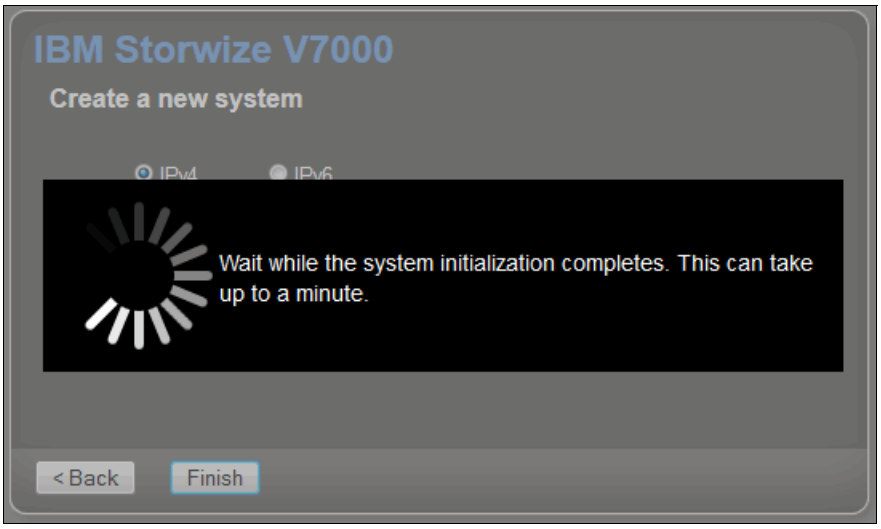


Figure 5-6 Initialization of the System

When the initialization is successfully completed, you see the message shown in Figure 5-7.

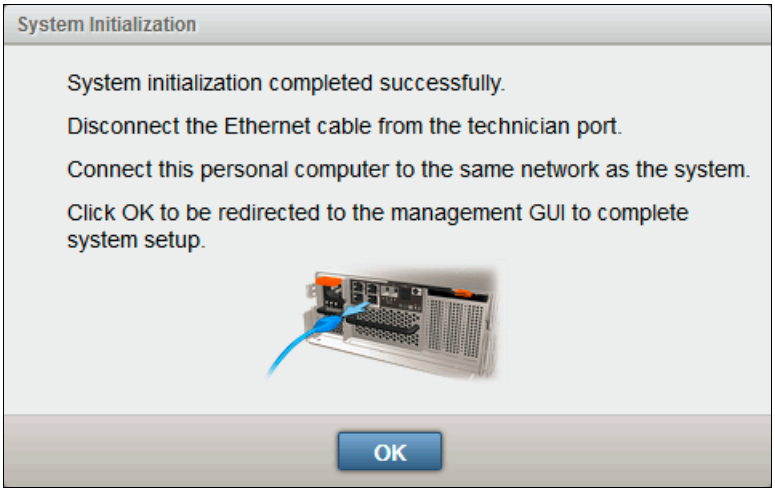


Figure 5-7 Initialization complete

7. Follow the on-screen instructions:
   a. Disconnect the Ethernet cable from the Technician Port and from your PC or notebook.
   b. Connect the same PC or notebook to the same network as the system.
   c. When you click **OK**, you are redirected to the GUI for completion of the system setup.

5.2.2 System Setup wizard

You can connect to the System IP address from any Management Console that is connected to the same network as the system:

1. Whether you are redirected from your PC or notebook, or you are connecting to the Management IP address of the system, you are taken to the Login window, as shown in Figure 5-8. You need to enter the default password (`passw0rd` with a zero).

Figure 5-8 Log in Screen

2. Click **Log in** and you are prompted to change the default password, as shown in Figure 5-9. The new password can be any combination of 6 - 63 characters.

Figure 5-9 Change default password window

3. When you have changed the password, you only need to go through a few more steps, before the initial configuration is completed.

   You get to the system setup page, as shown in Figure 5-10.

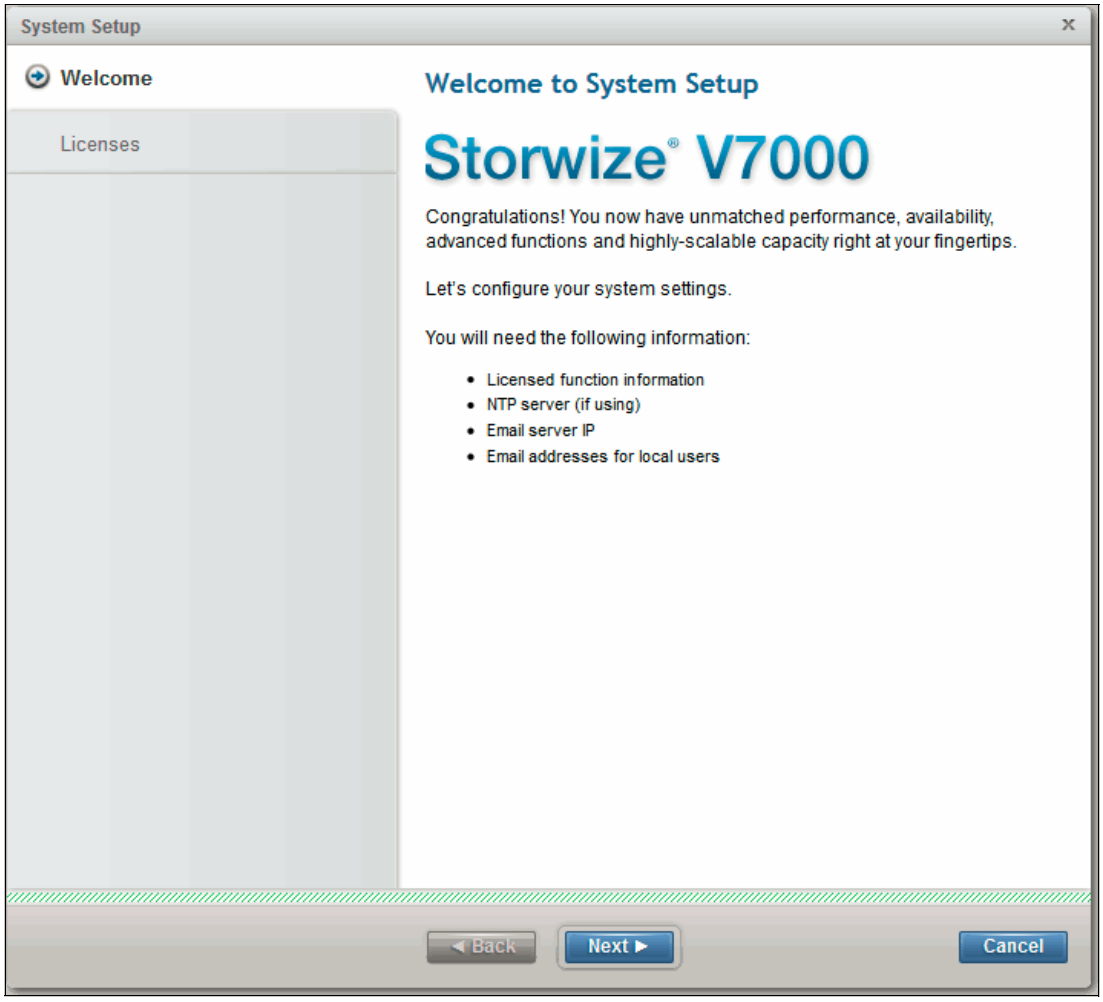


Figure 5-10 Welcome to system setup

4. Next, you must read and accept the License agreement in Figure 5-11. Click **Next**.

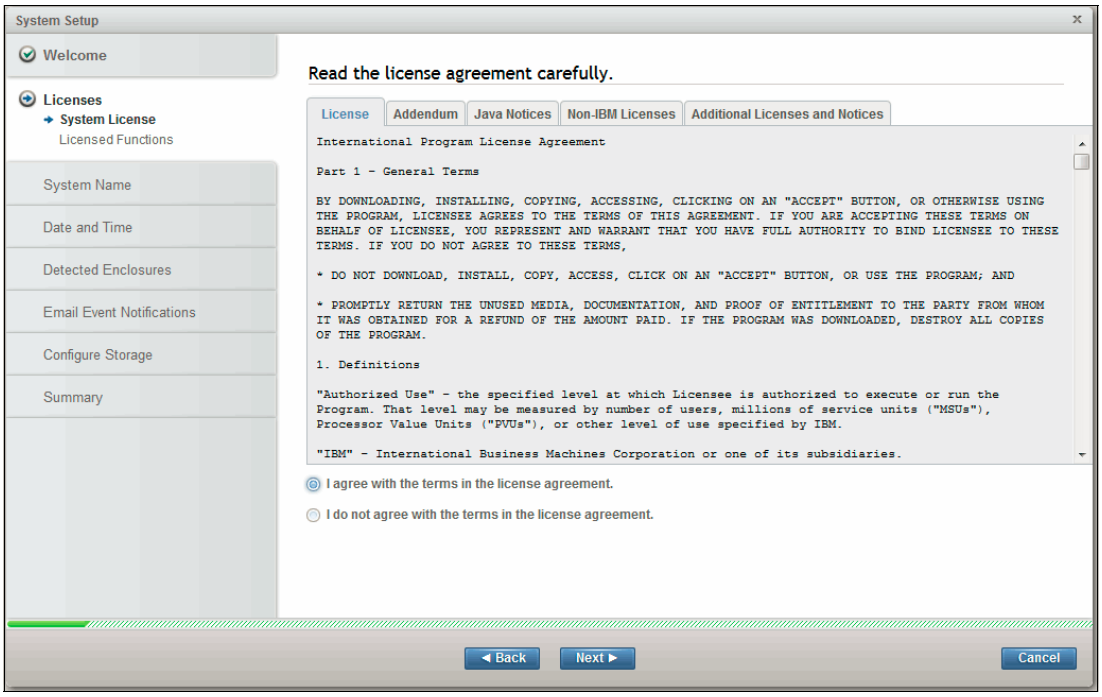


Figure 5-11 License agreement

5. You can now enter the purchased licenses for this system, as shown in Figure 5-12.

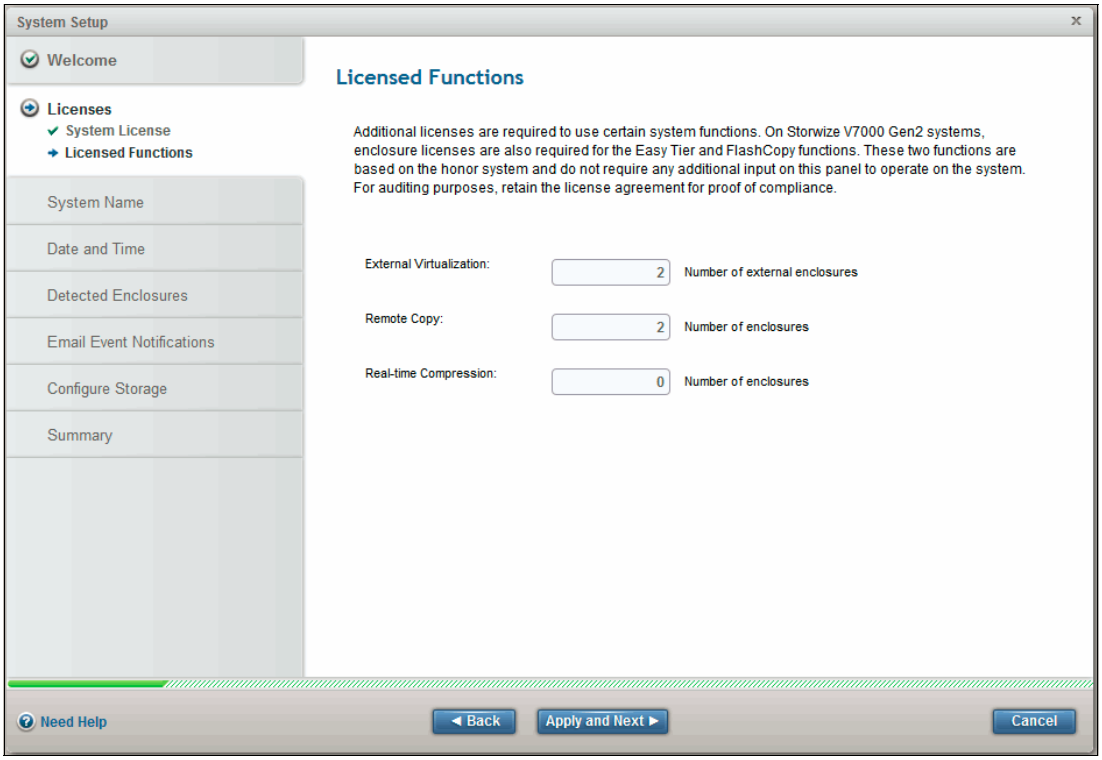


Figure 5-12 Licensed functions

6. Click **Apply and Next**. The system configures the licenses, as shown in Figure 5-13.

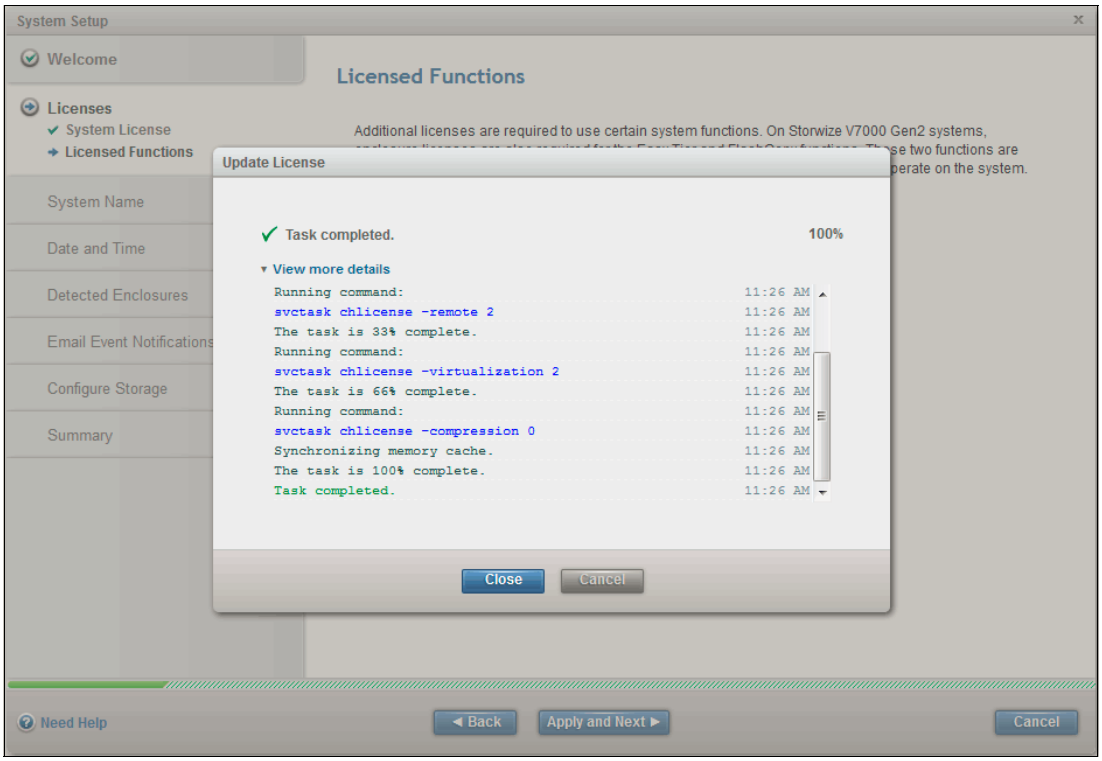


Figure 5-13 Applying licenses

On the next window, you can choose to leave the default name and change it later.

7. In this case, choose the wanted name for the system, as shown in Figure 5-14.

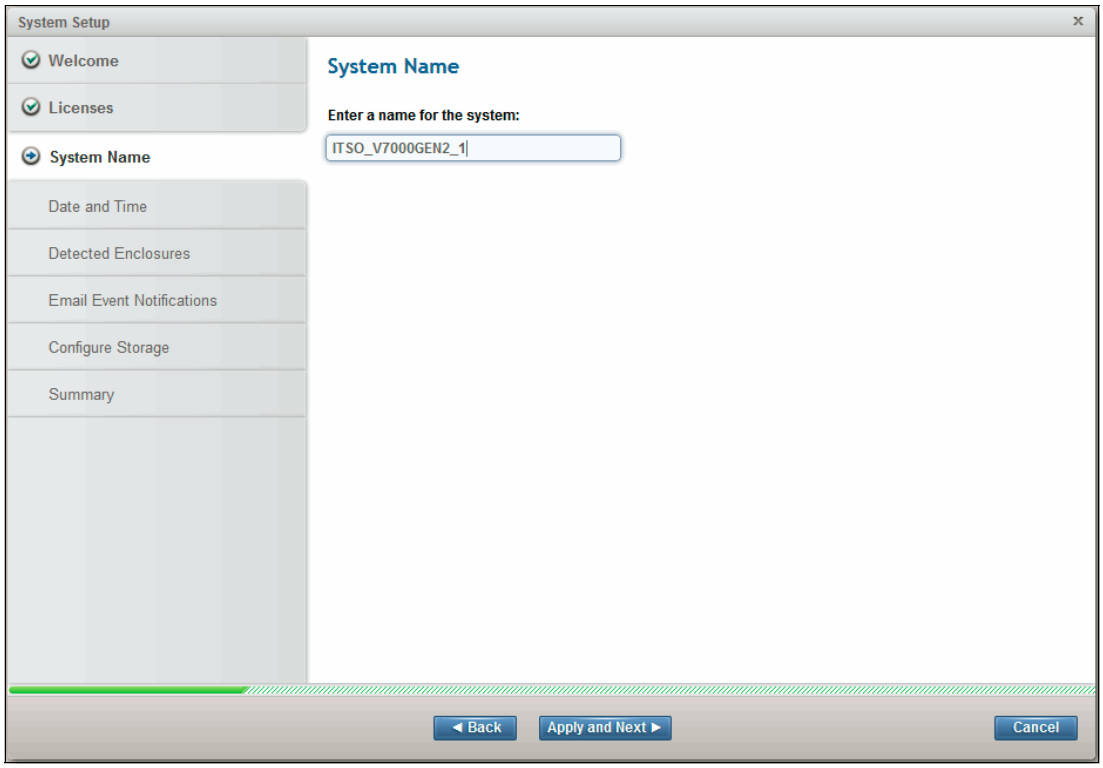


Figure 5-14 System name

8. Click **Apply and Next**. The system configures the system name, as shown in Figure 5-15.

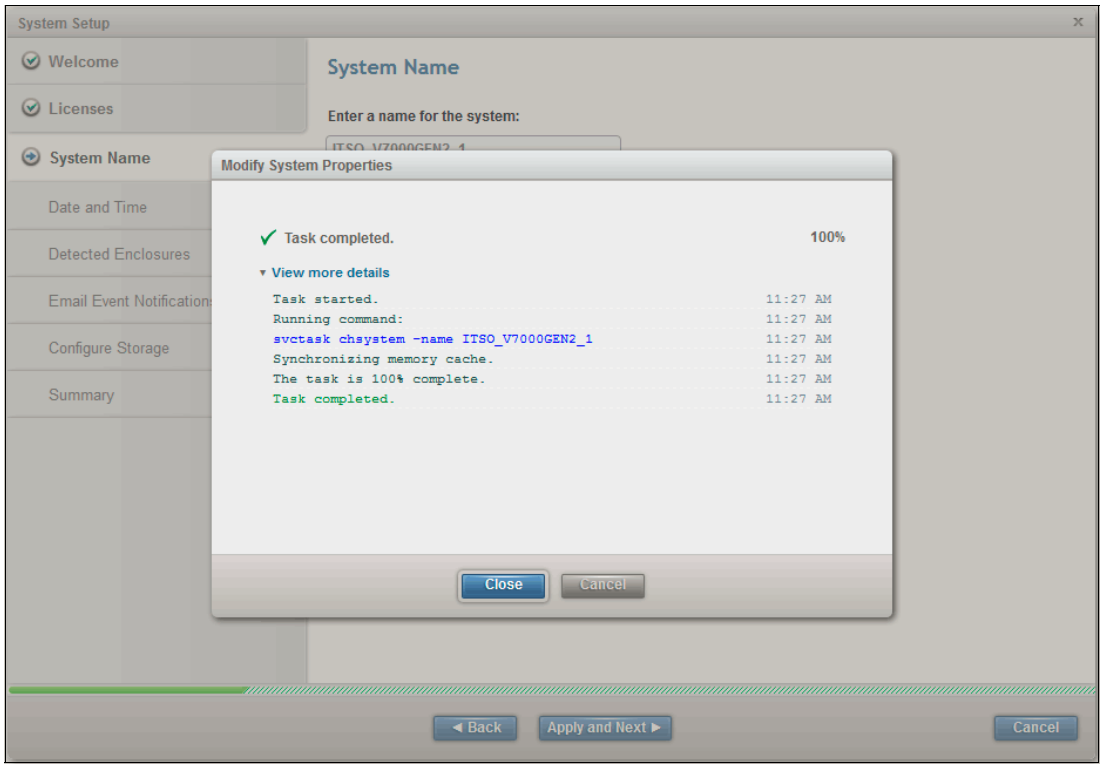


Figure 5-15 Applying system name

9. The window shown in Figure 5-16 enables you to set the date and time. You can either use manual settings, or use a Network Time Protocol (NTP) server. Using an NTP server in the environment for all devices is highly suggested, for example, to ensure that you have a common time stamp for troubleshooting, and so on.

   Note that at this stage, if you choose manual settings, you can only select 12-hour settings for a.m. or p.m. This can be changed to a 24-hour setting later.

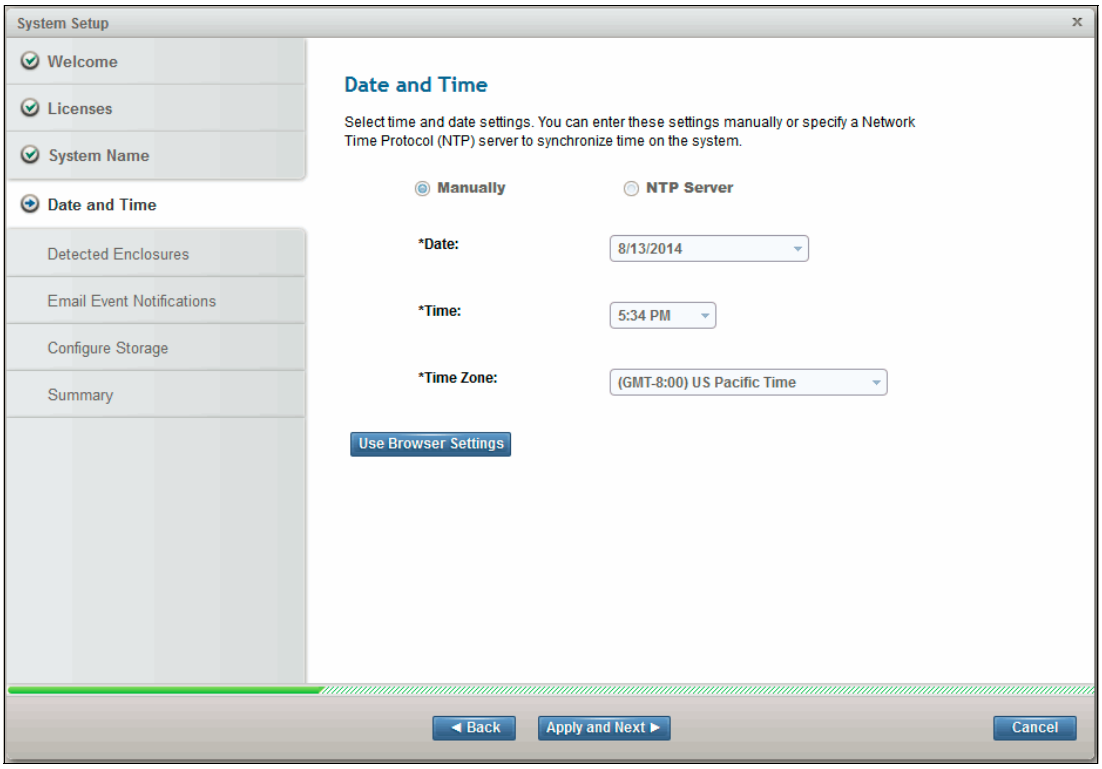


Figure 5-16 Setting Date and Time

10. Click **Apply and Next**. The system configures the Date and Time, as shown in Figure 5-17.

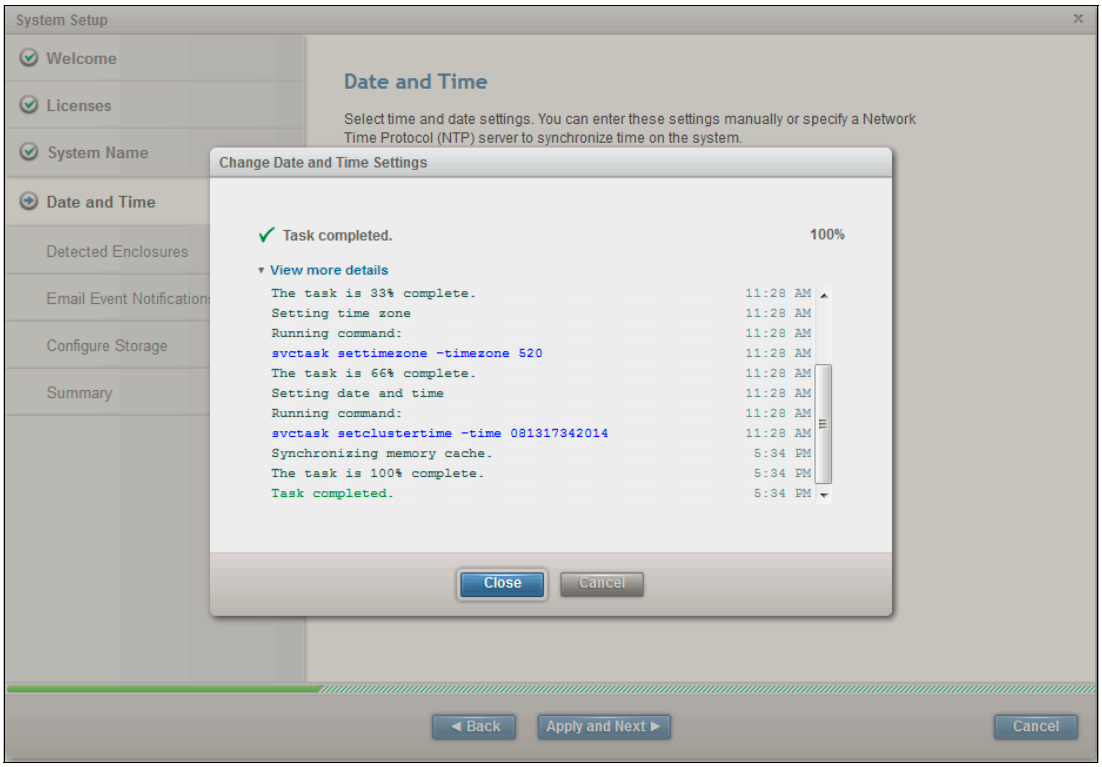


Figure 5-17 Applying Date and Time

11. The next window shows the detected enclosures. In this case, we are only using the control enclosure, so that is what we see, as shown in Figure 5-18.

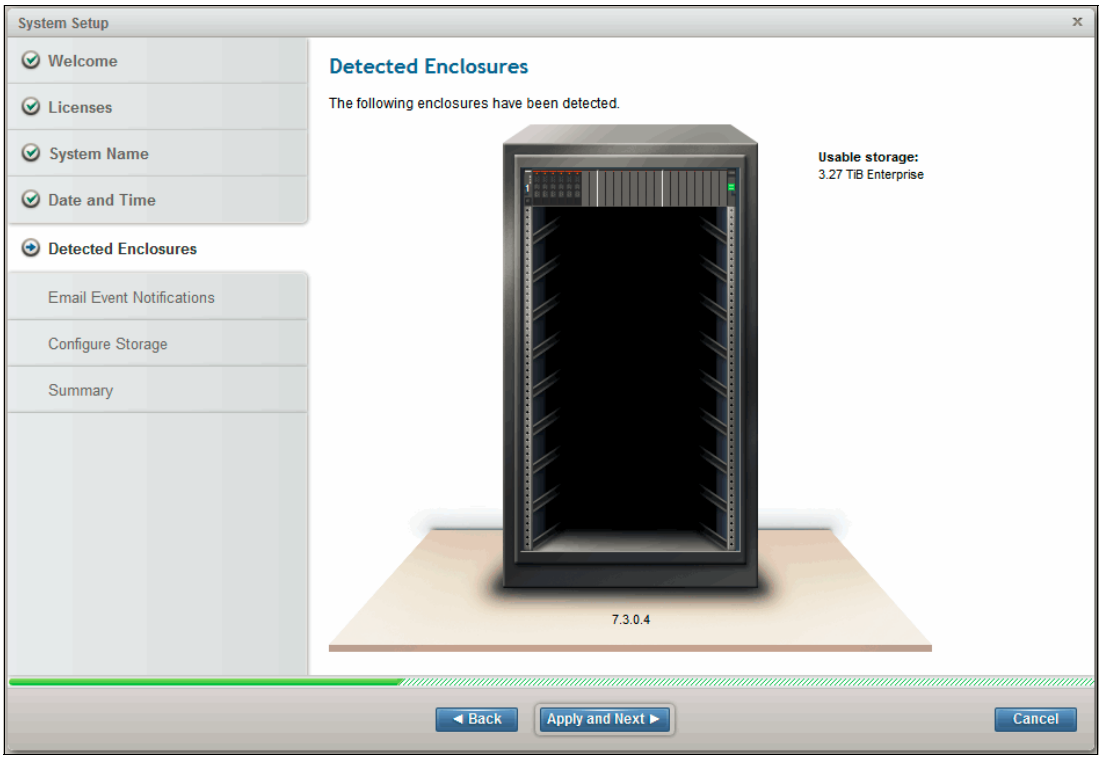


Figure 5-18 Detected enclosures

12. Click **Apply and Next**. The system configures the detected enclosures (Figure 5-19).

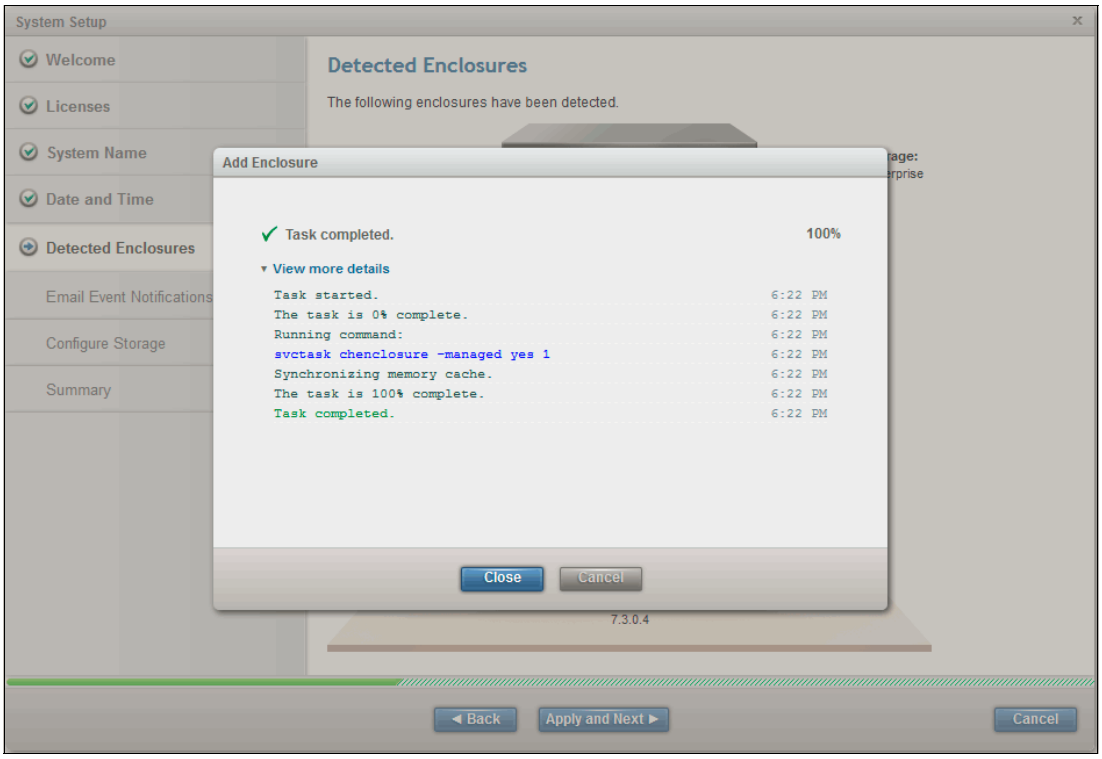


Figure 5-19 Add Enclosure

13. Click **Apply and Next**. You are asked if you want to set up Email Event Notifications, as shown in Figure 5-20.

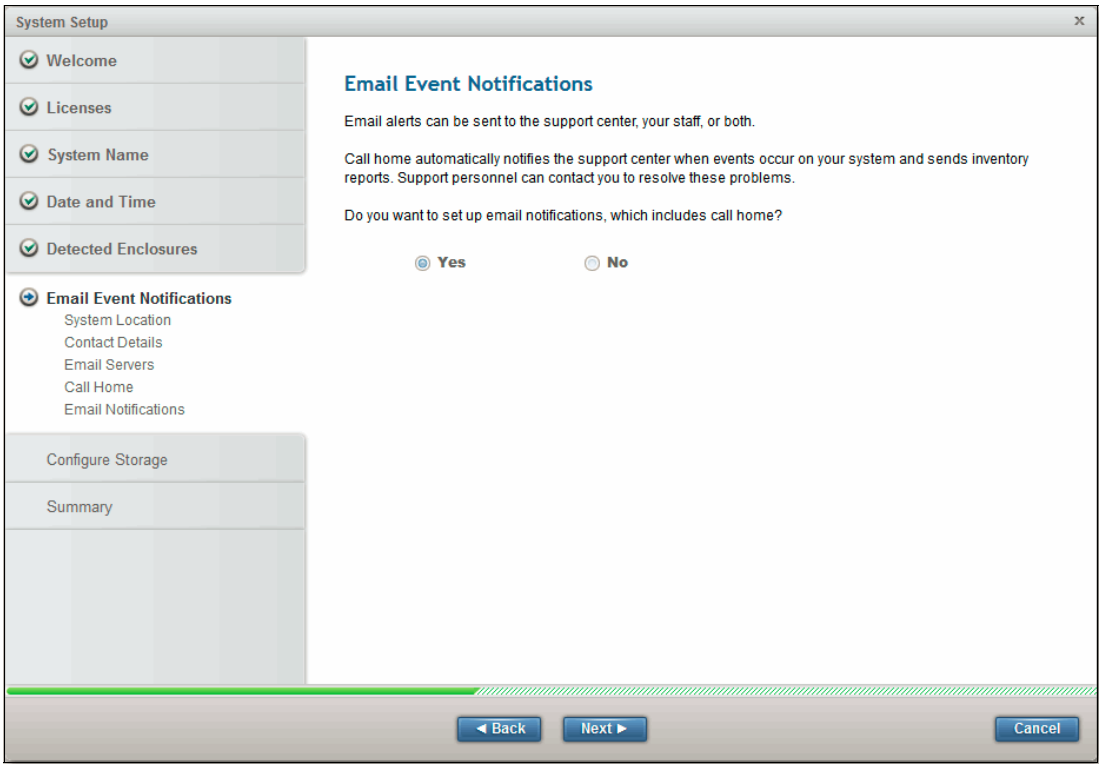


Figure 5-20 Email Event Notification

14. It is highly suggested to set up email notifications. However, they can be configured later. If you choose to say **No** to this option now, a warning displays, as shown in Figure 5-21.

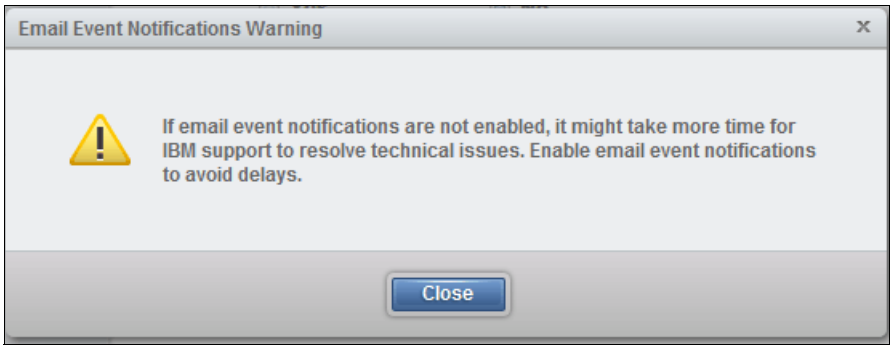


Figure 5-21 No email notifications warning

Requirement: You must have access to an Simple Mail Transfer Protocol (SMTP) server (by IP address) to be able to configure Email Event Notifications.

15. In this case, we set up Email Event Notifications starting with System Location, as shown in Figure 5-22. Complete the location details.

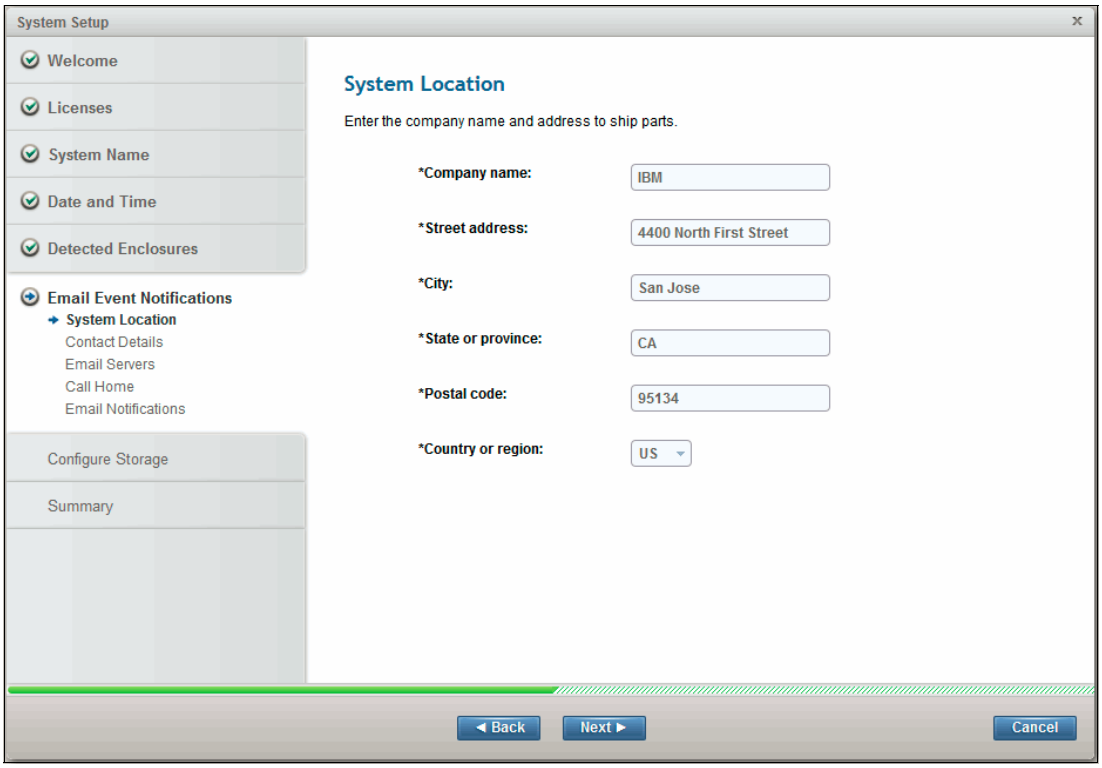


Figure 5-22 System Location

16. Click **Next** to proceed to Contact Details, as shown in Figure 5-23.

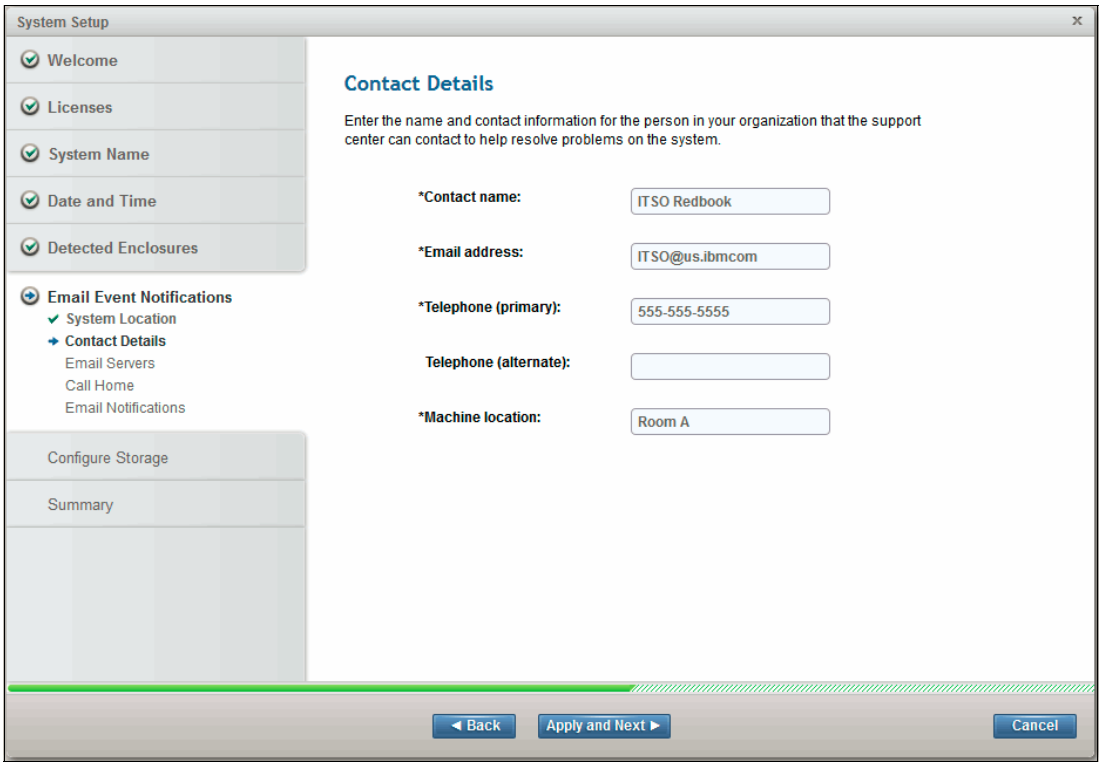


Figure 5-23 Contact Details

17. The next step is to configure Email Servers, as shown in Figure 5-24.

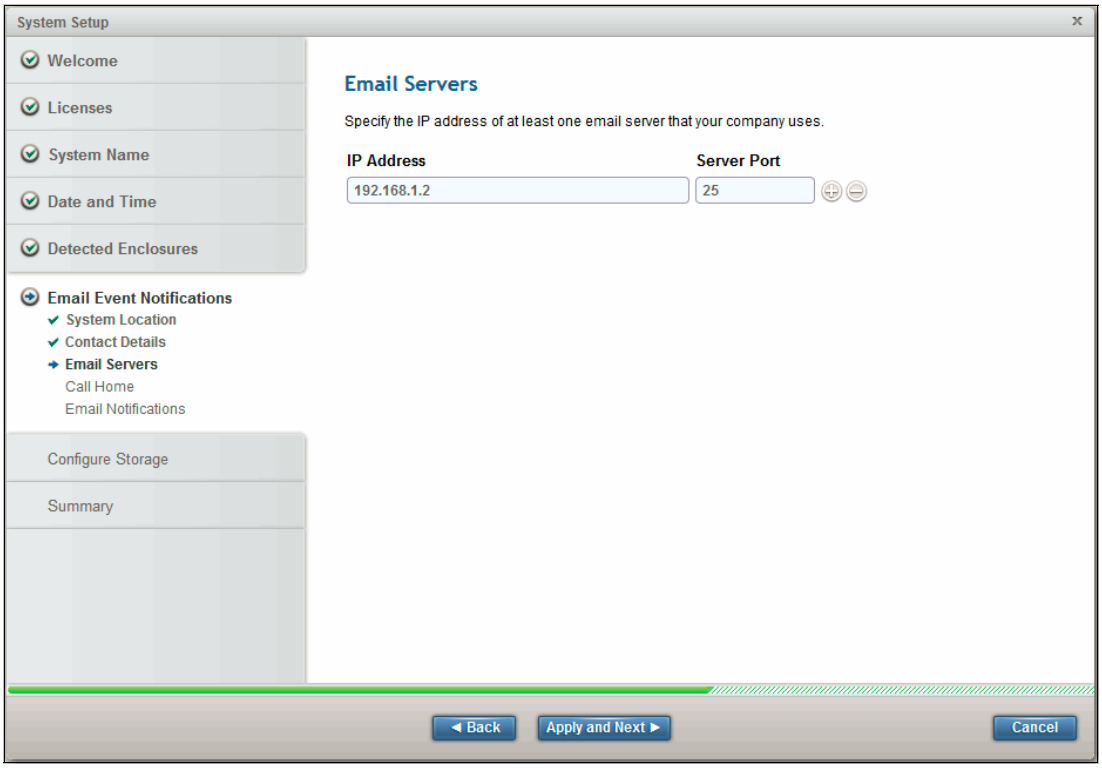


Figure 5-24 Email Servers

18. The next step is Call Home. It is automatically configured, as shown in Figure 5-25.

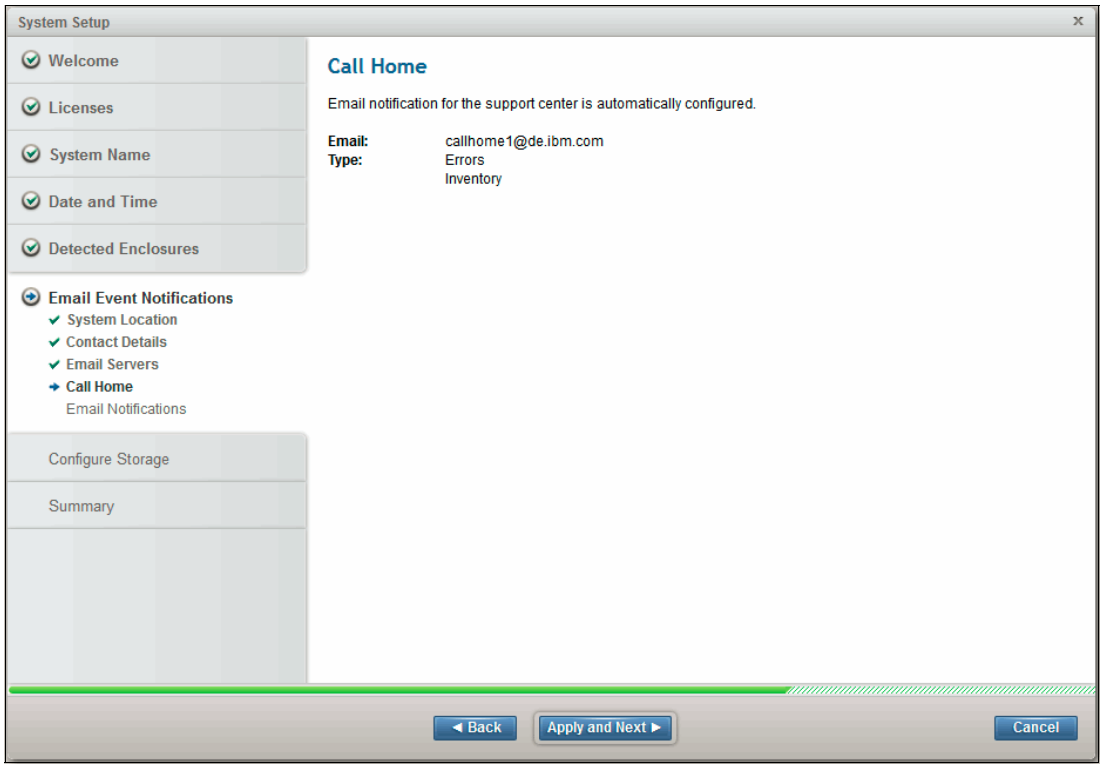


Figure 5-25 Call home

19. Click **Apply and Next**, and you are taken to the next step, which is Email Notifications, as shown in Figure 5-26 on page 113.

    There are four types of notifications:

    – Errors

      The user receives email about problems, such as hardware failures, that must be resolved immediately. To run fix procedures on these events, select **Monitoring** → **Events**.

    – Warnings

      The user receives email about problems and unexpected conditions. Investigate the cause to determinate any corrective action. To run fix procedures on these events, select **Monitoring** → **Events**.

    – Information

      The user receives email about expected events, for example when a FlashCopy has finished. No action is required for for these events.

    – Inventory

      The user receives inventory email that contains a summary of system status and configuration settings.

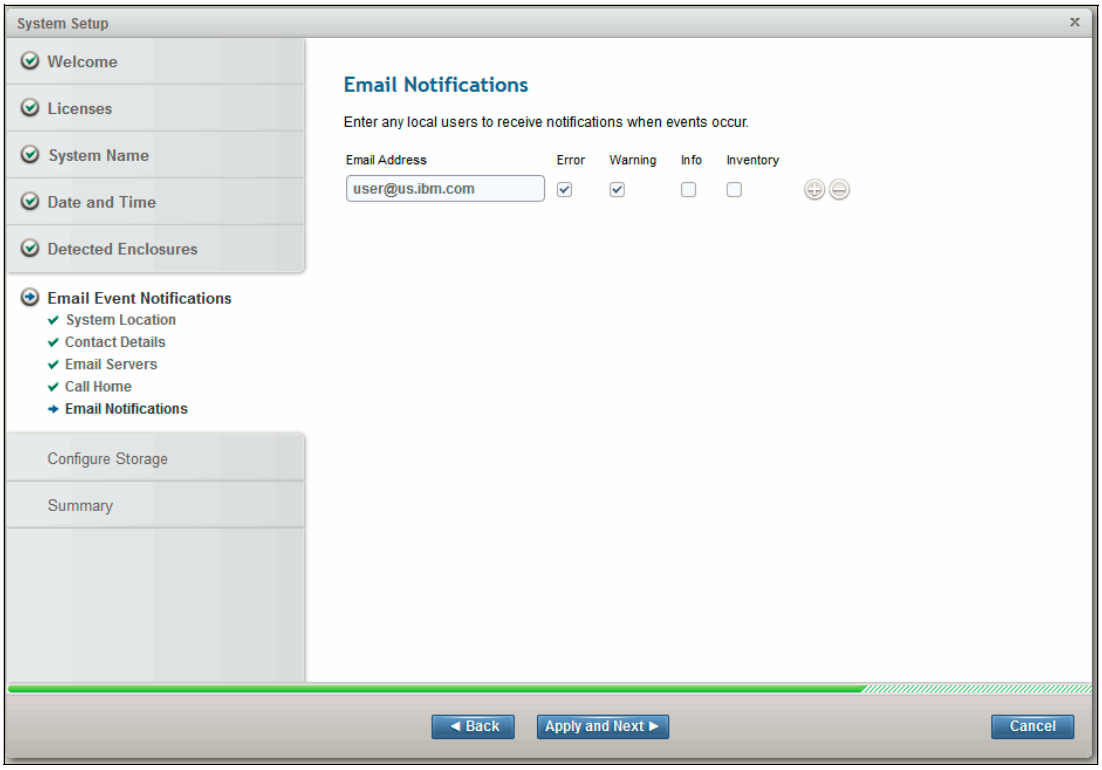


Figure 5-26 Email Notifications

20. The system then configures settings, as shown in Figure 5-27. This concludes the Event Notifications setup.

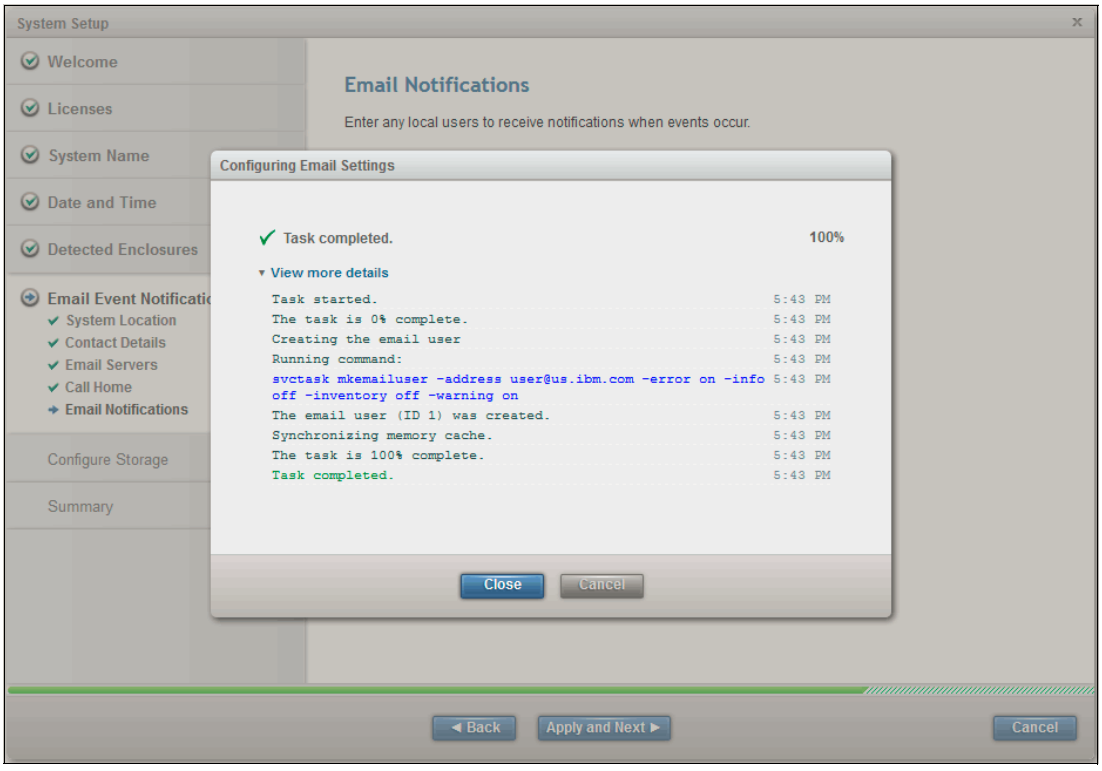


Figure 5-27 Configuring Email Settings

21. Click **Apply and Next** to be taken to the next step, which is Configure Storage.

22. You are then asked if you want the Storwize V7000 Gen2 to automatically configure storage (Figure 5-28). Select whether you want storage to be configured automatically now or later. For information about how to configure the storage manually, see 8.2.3, “Configuring internal storage” on page 157.

    To let the system configure the storage, select **Yes** and click **Next**.

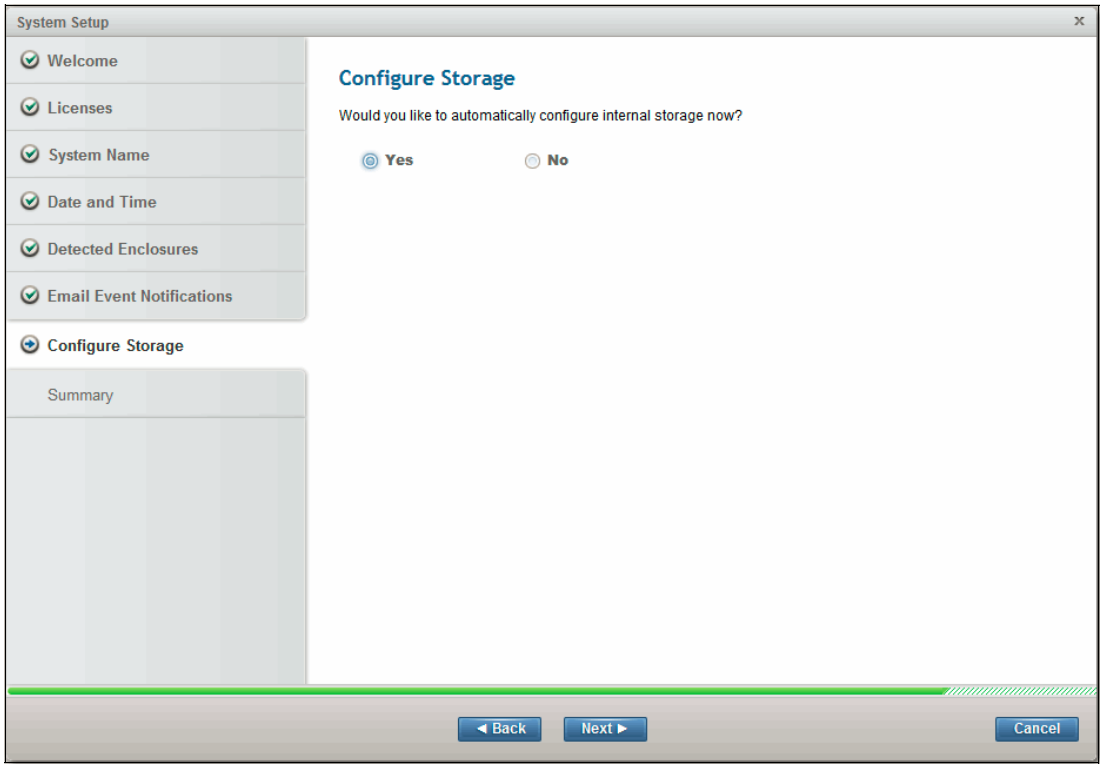


Figure 5-28 Configure Storage

23. This takes you to the final page of the Initial setup, which is the Summary, as shown in Figure 5-29. Click **Finish** to be taken to the main GUI.

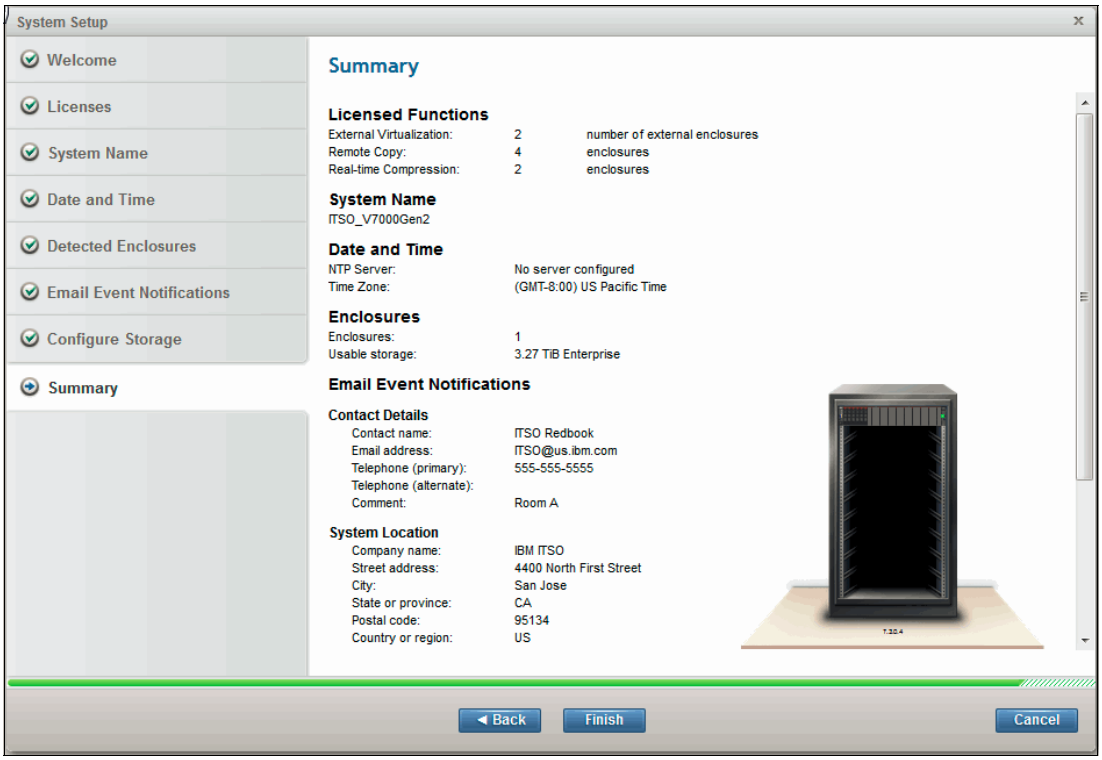


Figure 5-29 Summary

You are now ready to log in to the main GUI of the IBM Storwize V7000 Gen2, and you have full functionality, which concludes this Chapter.

For a detailed guide showing how to use the main GUI, see Chapter 9, “IBM Storwize V7000 Gen2 operations using the GUI” on page 169.

Alternatively, you can use the CLI, which is described in Chapter 8, “IBM Storwize V7000 Gen2 command-line interface” on page 143.

6

IBM Real-time Compression and the IBM Storwize V7000 Gen2

This chapter provides details about the IBM Real-time Compression feature as it pertains to the IBM Storwize V7000 Gen2.

This chapter describes the following topics:

- An overview of Real-time Compression (RtC) and its applicable use cases
- A more detailed look into RtC and the Random Access Compression Engine (RACE)
- An examination of the specific hardware options and improvements in the Storwize V7000 Gen2 that specifically benefit the use of Real-time Compression

6.1 Real-time Compression background, overview, and value proposition

Businesses and organizations around the world are challenged with tough economic conditions. The information technology (IT) environment, which was historically viewed as an expense, is now viewed as a source of innovation that must drive future revenue. However, ever-increasing data storage requirements consume the available resources and disrupt attempts to innovate in the IT environment.

The modern IT department has numerous challenges:

- Support for increasing data storage requirements

  Shrinking IT budgets are pressuring IT managers to increase the lifetime of existing storage systems. Traditional methods of cleaning up unneeded data and archiving files to auxiliary storage are time-consuming. They shift one resource constraint (physical storage) to another (the human work of storage administrators).

- Power, cooling, and floor space

  A data center provides the means to host the storage systems. However, the physical characteristics of the hard disk drive (HDD)-based systems limit the amount of data that can be stored per rack unit. Power consumption and heat dissipation are major concerns for IT managers, who must fit the storage systems into a limited data center. This conflicts with the increasing demand for computing power needed to support new types of applications.

- High availability (HA) of data

  Digital information has become the basis for any service in use today. As a result, the underlying systems that provide access to digital information are expected to be online all the time. This requirement has made it impossible to introduce data reduction solutions that impose any type of downtime. This restriction is true whether it is an actual inability to access the data, or merely a major slowdown when accessing an optimized data set.

Compression of primary storage provides an innovative approach designed to overcome these challenges.

6.1.1 The solution: IBM Real-time Compression

The Real-time Compression solution addresses the challenges listed in the previous section, because it was designed from the ground up for primary storage. Implementing Real-time Compression provides the following benefits:

- Compression for active primary data

  IBM Real-time Compression can be used with active primary data. Therefore, it supports workloads that are not candidates for compression in other solutions. A unique in-line compression mechanism enables data to be compressed before it is de-staged to the disk, which reduces disk cycles and the amount of input/output (I/O) being written to the disk.

- Compression for replicated or mirrored data

  Remote volume copies can be compressed, in addition to the volumes at the primary storage tier. This process reduces storage requirements in Metro Mirror and Global Mirror destination volumes as well.

- No changes to the existing environment are required

  IBM Real-time Compression is part of the storage system. It was designed with transparency in mind, so that it can be implemented without changes to applications, hosts, networks, fabrics, or external storage systems. The solution is not apparent to hosts, so users and applications continue to work as-is. Compression occurs within the StorwizeV7000 system itself.

- Overall savings in operational expenses

  More data is stored in a reduced rack space, so fewer storage expansion enclosures are required to store a data set. This reduced rack space has the following benefits:

  – Reduced power and cooling requirements. More data is stored in a system, therefore requiring less power and cooling per gigabyte (GB) of used capacity.

  – Reduced software licensing for additional functions in the system. More data stored per enclosure reduces the overall spending on licensing.

- Disk space savings are immediate.

  The space reduction occurs when the host writes the data. This process is unlike other compression solutions in which some or all of the reduction is realized only afterward, when a post-process compression batch job is run.

6.1.2 Common use cases

This section addresses the most common use cases for implementing compression:

- General-purpose volumes
- Databases
- Virtualized infrastructures
- Log server data stores

General-purpose volumes

Most general-purpose volumes are used for highly compressible data types, such as home directories, computer-aided design and computer-aided manufacturing (CAD/CAM), and oil and gas geo-seismic data. Storing such types of data in compressed volumes provides immediate capacity reduction to the overall used space. More space can be provided to users without any change to the environment.

There can be many file types stored in general-purpose servers. However, for practical information, the estimated compression ratios are based on actual field experience. Expected compression ratios are 50% - 60%.

File systems that contain audio, video files, and compressed files are not good candidates for compression. The overall capacity savings on these file types are minimal.

Databases

Database information is stored in table space files. It is common to observe high compression ratios in database volumes. Examples of databases that can greatly benefit from real-time compression are IBM DB2®, Oracle, and Microsoft SQL Server. Expected compression ratios are 50% - 80%.

Tip: Some databases offer optional built-in compression. Generally, do not compress already-compressed database files.

Virtualized infrastructures

The proliferation of open systems virtualization in the market has increased the use of storage space, with more virtual server images and backups kept online. The use of compression reduces the storage requirements at the source. Examples of virtualization solutions that can greatly benefit from Real-time Compression are VMWare, Microsoft Hyper-V, and Kernel-based Virtual Machine (KVM). Expected compression ratios are 45% - 75%.

Tip: Virtual machines with file systems that contain compressed files are not good candidates for compression, as described in “General-purpose volumes” on page 119.

Log server data stores

Logs are a critical element for any IT department in any organization. Log aggregates or syslog servers are a central point for administrators, and immediate access and a smooth work process are necessary. Log server data stores are very good candidates for Real-time Compression. Expected compression ratios are up to 90%.

6.2 IBM Real-time Compression technology

IBM Real-time Compression technology is based on RACE. RACE is an integral part of the software stack of Storwize V7000 version 6.4.0 and later. This integration does not alter the behavior of the system, so that previously existing features are supported for compressed volumes. RACE uses a lossless data compression algorithm along with a Real-time Compression technology, enabling it to meet the industry requirements of performance, reliability, and scalability.

RACE technology makes use of over 50 patents, many of which are not about compression. Rather, they define how to make industry standard Lempel-Ziv (L)-based compression of primary storage operate *in* real time while *enabling* random access. The primary intellectual property behind this is RACE. At a high level, the IBM RACE component compresses data written into the storage system dynamically.

This compression occurs transparently, so Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI)-connected hosts are not aware of the compression. RACE is an in-line compression technology, so each host write is compressed as it passes through RACE to the disks.

This has a clear benefit over other compression technologies that are post-processing in nature. These alternative technologies do not provide immediate capacity savings, and therefore are not a good fit for primary storage workloads, such as databases and active data set applications.

RACE is based on the Lempel-Ziv lossless data compression algorithm, and operates in real time. When a host sends a write request, it is acknowledged by the upper-level write cache of the system, and then de-staged to the storage pool.

As part of its de-staging, the request passes through the compression engine, and is then stored in compressed format onto the storage pool. Writes are therefore acknowledged immediately after being received by the upper write cache, with compression occurring as part of the destaging to internal or external physical storage. Capacity is saved when the data is written by the host, because the host writes are smaller when written to the storage pool.

IBM Real-time Compression is a self-tuning solution, similar to the Storwize V7000 system itself. It adapts to the workload that runs on the system at any particular moment.

6.2.1 Random Access Compression Engine

To understand why RACE is unique, you need to review the traditional compression techniques. This description is not about the compression algorithm itself (how the data structure is reduced in size mathematically), but rather a description about how the data is laid out within the resulting compressed output.

Compression utilities

Compression is probably most known to users because of the widespread use of compression utilities, such as Zip and Gzip. At a high level, these utilities take a file as their input, and parse the data by using a sliding window technique. Repetitions of data are detected within the sliding window history, most often 32 kilobytes (KB). Repetitions outside of the window cannot be referenced. Therefore, the file cannot be reduced in size unless data is repeated when the window “slides” to the next 32 KB slot.

Figure 6-1 shows compression that uses a sliding window, where the first two repetitions of the string “ABCDEF” fall within the same compression window, and can therefore be compressed using the same dictionary. Note that the third repetition of the string falls outside of this window, and cannot, therefore, be compressed using the same compression dictionary as the first two repetitions, reducing the overall achieved compression ratio.

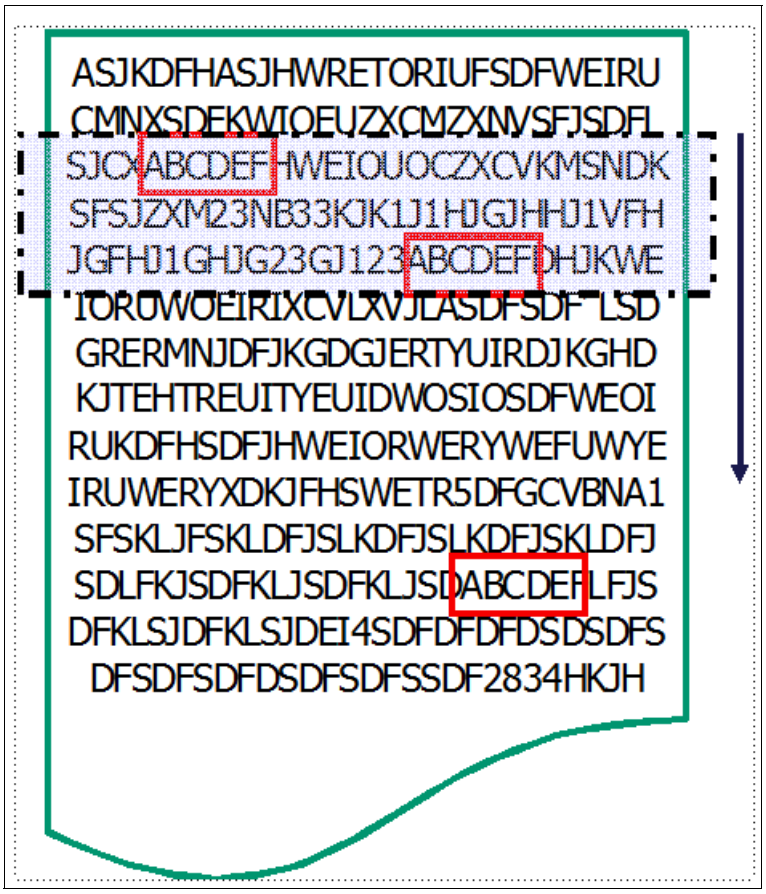


Figure 6-1 Compression that uses a sliding window

Traditional data compression in storage systems

The traditional approach taken to implement data compression in storage systems is an extension of how compression works in the compression utilities previously mentioned. Similar to compression utilities, the incoming data is split into fixed chunks, and then each chunk is compressed and extracted independently.

However, there are drawbacks to this approach. An update to a chunk requires a read of the chunk followed by a recompression of the chunk to include the update. The larger the chunk size chosen, the heavier the I/O penalty to recompress the chunk. If a small chunk size is chosen, the compression ratio is reduced, because the repetition detection potential is reduced.

Figure 6-2 shows an example of how the data is split into fixed-size chunks (in the upper-left side of the figure). It also shows how each chunk gets compressed independently into variable-length compressed chunks (in the upper-right side of the figure). The resulting compressed chunks are stored sequentially in the compressed output.

Although this approach is an evolution from compression utilities, it is limited to low-performance use cases. This limitation is mainly because it does not provide real random access to the data.

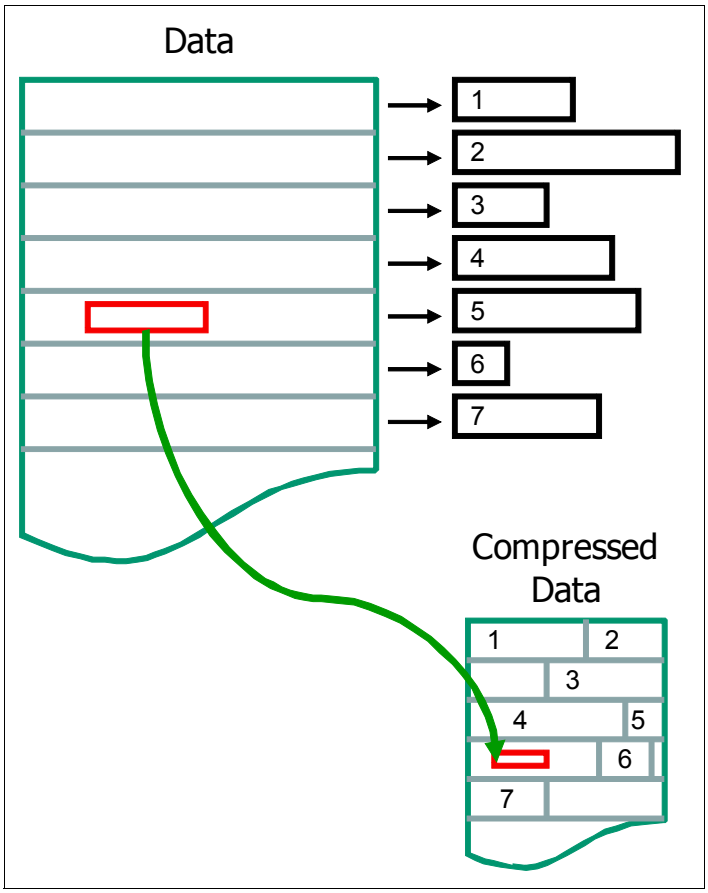


Figure 6-2 Traditional data compression in storage systems

Random Access Compression Engine approach

The IBM patented RACE implements an inverted approach when compared to traditional approaches to compression. RACE uses variable-size chunks for the input, and produces fixed-size chunks for the output.

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

Figure 6-3 shows random access compression.

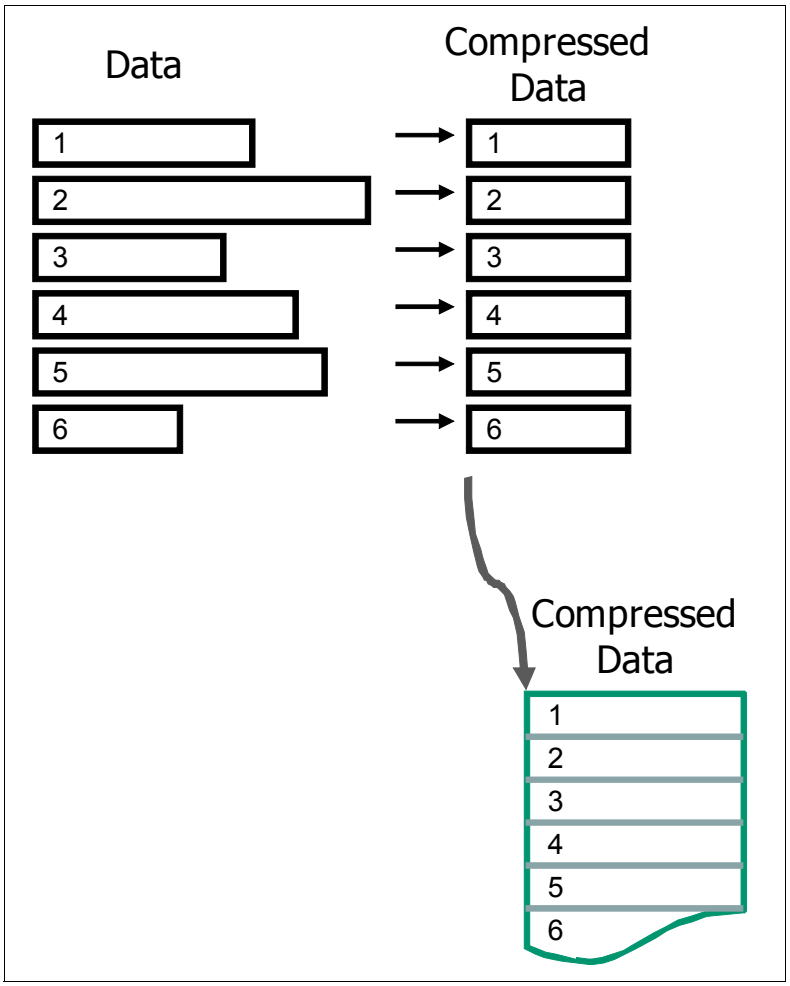


Figure 6-3 Random access compression

Location-based compression

Both compression utilities and traditional storage systems-compression approaches compress data by finding repetitions of bytes within the chunk that is being compressed. The compression ratio of this chunk depends on how many repetitions can be detected within the chunk. The number of repetitions is affected by how much the bytes stored in the chunk are related to each other.

The relation 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 this page). Because the chunking of the file is arbitrary, it has no concept of how the data is laid out within the document. Therefore, a compressed chunk can be a mixture of the textual information and part of the drawing.

This process yields a lower compression ratio, because the different data types mixed together cause a suboptimal dictionary of repetitions. 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 is also called *location-based compression*. The data repetition detection is based on the location of data within the same chunk.

This challenge was also addressed with the *predecide* mechanism that was introduced in IBM SAN Volume Controller code version 7.1.

Predecide mechanism

Some data chunks have a higher compression ratio than others. Compressing some of the chunks saves very little space but still requires resources, such as processor and memory. To avoid spending resources on incompressible data, and to provide the ability to use a different, more effective (in this particular case) compression algorithm, IBM has invented a *predecide* mechanism that was first introduced in version 7.1.

The chunks that are below a given compression ratio are skipped by the compression engine, therefore saving processor time and memory processing. Chunks that are decided not to be compressed with the main compression algorithm, but that still can be compressed well with the other algorithm, are marked and flagged accordingly. The result can vary, because predecide does not check the entire block, only a sample of it.

Figure 6-4 shows how the detection mechanism works.

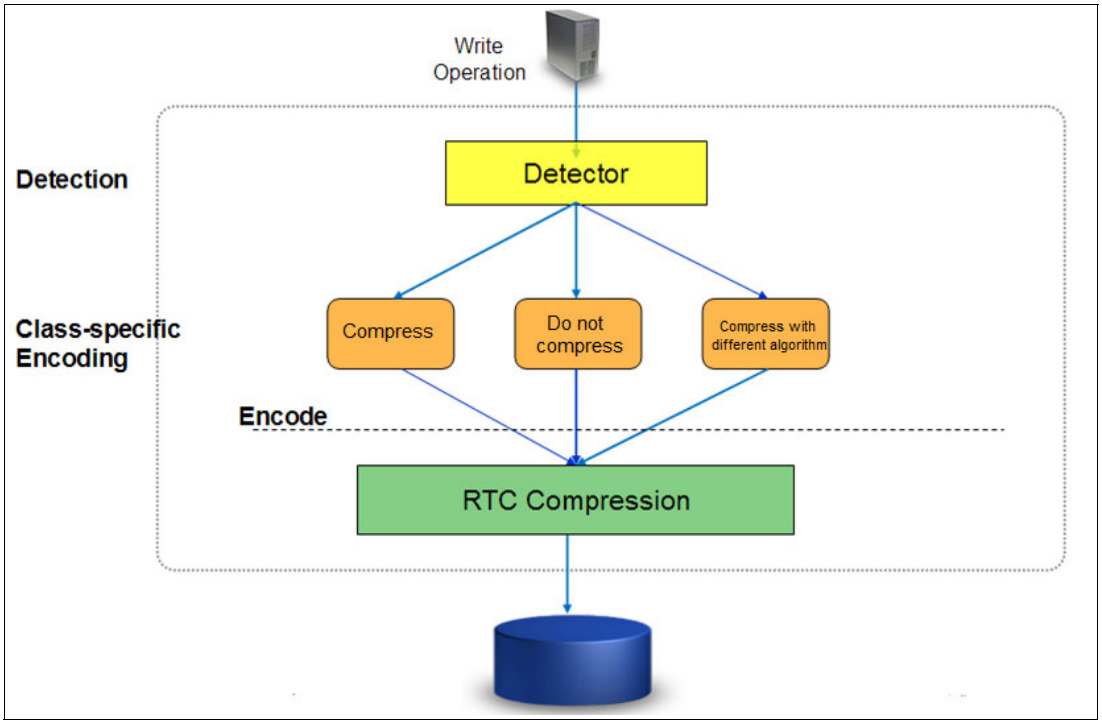


Figure 6-4 Detection mechanism

Temporal compression

RACE offers a technology leap beyond location-based compression, *temporal compression*.

When host writes arrive to RACE, they are compressed and filled up in fixed-size chunks, also called *compressed blocks*. Multiple compressed writes can be aggregated into a single compressed block.

A dictionary of the detected repetitions is stored in 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 from the same data type, more repetitions are usually detected by the compression algorithm.

This type of data compression is called *temporal compression* because the data repetition detection is based on the time the data was written into the same compressed block. 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 a more homogeneous set of input data.

The upper part of Figure 6-5 shows how three writes, sent one after the other by a host, end up in different chunks. They get compressed into different chunks because their location on the volume is not adjacent. This 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 lower part of the figure), the writes are compressed together by using a single dictionary. This yields a higher compression ratio than location-based compression.

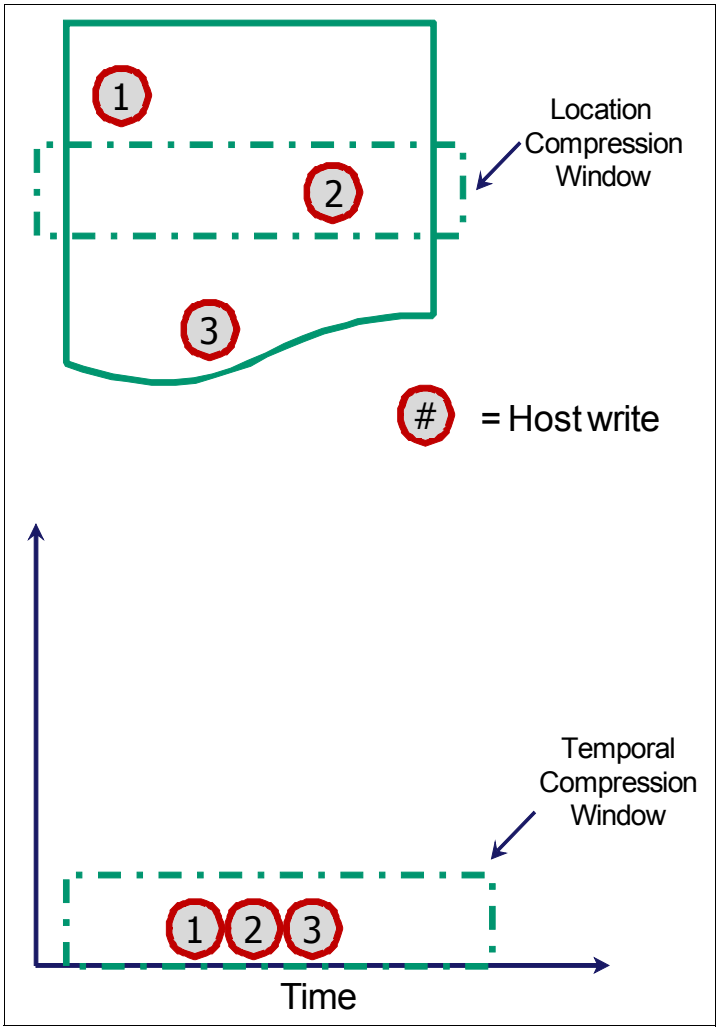


Figure 6-5 Location-based versus temporal compression

6.2.2 RACE in Storwize V7000 Gen2 software stack

It is important to understand where the RACE technology is implemented in the Storwize V7000 Gen2 software stack. RACE technology is implemented into the Storwize system thin provisioning layer, and is an organic part of the stack. The Storwize V7000 Gen2 software stack is shown in Figure 6-6. Compression is transparently integrated with existing system management design. All of the Storwize V7000 features are supported on compressed volumes.

You can create, delete, migrate, map (assign), and unmap (unassign) a compressed volume as though it were a fully allocated volume. In addition, you can use Real-time Compression with IBM Easy Tier on the same volumes. This compression method provides non-disruptive conversion between compressed and decompressed volumes. This conversion provides a uniform user experience, and eliminates the need for special procedures when dealing with compressed volumes.

Figure 6-6 shows the Storwize V7000 Gen2 software stack.

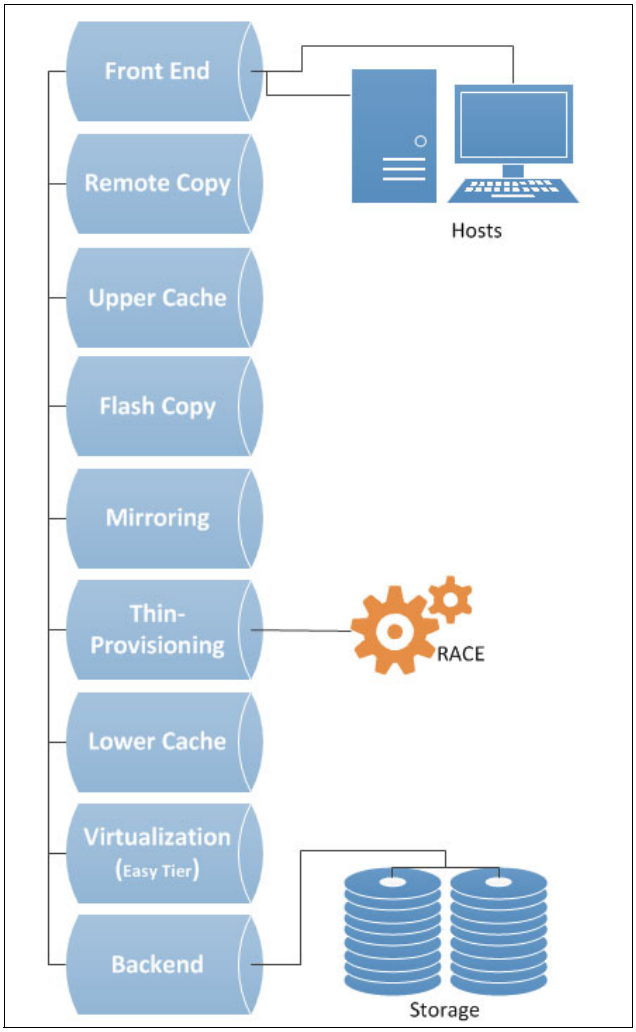


Figure 6-6 The Storwize V7000 Gen2 software stack

6.2.3 Data write flow

When a host sends a write request to the Storwize V7000 Gen2, it reaches the upper cache layer. The I/O is duplicated to the second node upper cache, and then the host is immediately sent an acknowledgment of its I/Os.

When the upper cache layer de-stages to the RACE, the I/Os are sent to the thin-provisioning layer. They are then sent to RACE and, if necessary, the original host write or writes. The metadata that holds the index of the compressed volume is updated if needed, and is compressed as well.

6.2.4 Data read flow

When a host sends a read request to the Storwize V7000 Gen2 for compressed data, it is forwarded directly to the RtC component:

1. If the RtC component contains the requested data, the Storwize V7000 Gen2 cache replies to the host with the requested data, *without having to read the data* from the lower-level cache or disk.
2. If the RtC component does not contain the requested data, the request is forwarded to the Storwize V7000 Gen2 lower-level cache.
3. If the lower-level cache contains the requested data, it is sent up the stack and returned to the host *without accessing the storage*.
4. If the lower-level cache does not contain the requested data, it sends a read request to the storage for the requested data.

6.2.5 Compression of existing data

In addition to compressing data in real time, it is also possible to compress existing data sets. This compression adds a compressed mirrored copy to an existing volume. You then delete the original copy after the synchronization of the compressed copy is complete. This process is nondisruptive, so the data remains online and accessible by applications and users.

This capability enables customers to regain space from the storage pool, which can then be reused for other applications.

6.3 Storwize V7000 Gen2 software and hardware updates that enhance Real-time Compression

The Storwize V7000 Gen2 hardware and software version 7.3 introduced significant improvements that enhance and extend the applicability of the Real-time Compression feature. In this section, we will provide an overview of these enhancements:

- Software enhancements:
  - Cache re-architecture
- Hardware enhancements:
  - Additional and enhanced processor options
  - Increased memory options
  - Intel Assist Acceleration Technology (Coletto Creek) compression acceleration cards

6.3.1 Software enhancements

Cache is the most significant software enhancement.

Cache

As mentioned in Chapter 1, “Introduction to IBM storage virtualization” on page 1, Storwize V7000 Gen2 software version 7.3 introduces an enhanced, dual-level caching model. This model differs from the single-level cache model of previous software versions.

In the previous model, the Real-time Compression software component sat below the single-level read/write cache. The benefit of this model is that the upper-level read/write cache masks from the host any latency introduced by the Real-time Compression software component. However, in this single-level caching model, the de-staging of writes for compressed I/Os to disk might not be optimal for certain workloads, due to the fact that the RACE component is interacting directly with un-cached storage.

In the new, dual-level caching model, the Real-time Compression software component sits *below* the upper-level, fast-write cache, and *above* the lower-level advanced read/write cache. There are several advantages to this dual-level model regarding Real-time Compression:

- Host writes, whether to compressed or decompressed volumes, are still serviced directly using the upper-level write cache, preserving low host write I/O latency. Response time can improve with this model, becuase the upper cache flushes less data to RACE more frequently.
- The performance of the de-staging of compressed write I/Os to storage is improved, because these I/Os are now de-staged via the advanced, lower-level cache, as opposed to directly to storage.
- The existence of a lower-level write cache below the RtC component in the software stack enables the coalescing of compressed writes and, as a result, a reduction in back-end I/Os, due to the ability to perform full-stride writes for compressed data.
- The existence of a lower-level read cache below the Real-time Compression component in the software stack enables the temporal locality nature of RtC to benefit from pre-fetching from the back-end storage.
- The main (lower-level) cache now stores compressed data for compressed volumes, increasing the effective size of the lower-level cache.
- Support for larger numbers of compressed volumes.

6.3.2 Hardware enhancements

The Storwize V7000 Gen2 introduces numerous hardware enhancements. Several of these enhancements relate directly to the Real-time Compression feature, and offer significant performance and scalability improvements over previous hardware versions.

New enhanced processor

The Storwize V7000 Gen2 offers updated eight-core processors, as compared to the four-core processor available in the previous hardware version.

Increased memory options

The Storwize V7000 Gen2 offers the option to increase the system memory from the base of 32 GB to 64 GB. This additional memory enables improved overall system performance over previous hardware models when using compression.

Intel Quick Assist Acceleration Technology (Coletto Creek) compression acceleration cards

The Storwize V7000 Gen2 includes one Intel Quick Assist compression acceleration card on-board based on the Coletto Creek chipset. The introduction of these Intel-based compression acceleration cards in the Storwize V7000 Gen2 is an industry first, providing dedicated processing power and greater throughput over previous models.

For additional details about the new hardware specification, see Chapter 2, “IBM Storwize V7000 Gen2 Hardware” on page 21.

7

Performance overview of the IBM Storwize V7000 Gen2

In this chapter, we provide a brief overview of the performance analysis capabilities of the IBM Storwize V7000 Gen2. We also describe a method that you can use to collect and process performance statistics.

However, it is beyond the intended scope of this book to provide an in-depth understanding of performance statistics, or explain how to interpret them. For a more comprehensive look at the performance of the IBM Storwize V7000 Gen2, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521, which is available at the following website:

http://www.redbooks.ibm.com/abstracts/sg247521.html?Open

For the Storwize family, as with all other IBM storage subsystems, the official IBM tool for the collection of performance statistics, and to supply performance reporting, is IBM Tivoli Storage Productivity Center.

You can obtain more information about IBM Tivoli Storage Productivity Center usage and configuration in *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364*:*

http://www.redbooks.ibm.com/abstracts/sg247364.html?Open

7.1 IBM Storwize V7000 Gen2 performance overview

The IBM Storwize V7000 Gen2 is equipped with additional processor, memory, and input/output (I/O) slots. These additional components enable enhancements, such as the number of compressed volumes that can be managed by a single I/O Group, the number of Fibre Channel (FC) Ports per I/O Group, and the amount of read cache available to an I/O Group.

In addition to the hardware upgrades, the 7.3 code level adds IBM Easy Tier 3 and storage pool balancing, which can further optimize performance. In this topic, we look at the performance basics, and best practices. Easy Tier is described in more detail in Chapter 4, “IBM Storwize V7000 Gen2 Easy Tier” on page 85.

When using Real-time Compression, ensure that you use the *Comprestimator* tool to assess the expected compression ratio for the workload that you will be delivering. With the addition of the optional acceleration cards in the Storwize V7000 Gen2, there are additional performance benefits for compressed workloads.

Note: Comprestimator is a command-line interface (CLI), host-based utility that can be used to estimate an expected compression rate for block devices. It can be downloaded from the following website:

`http://www-01.ibm.com/support/docview.wss?uid=ssg1S4001012`

For more information about RtC, see Chapter 6, “IBM Real-time Compression and the IBM Storwize V7000 Gen2” on page 117.

7.1.1 Drive Performance

In general, hard disk drive (HDD) technology has not changed significantly in the last 60 years, with the exception of flash drives. The performance of spinning drives still depends on seek time and rotational latency. The modern flash drive eliminates these factors, and has no mechanical parts.

Spinning existing drive performance

Spinning HDD performance is measured by the following three factors:

- Seek Time

  Time taken for head to current track to destination track. Performance is affected by generation and form factor (3.5 in. versus 2.5 in.).

- Rotational Latency

  Time taken for drive to spin the platter, so the destination track is under the head. Performance is affected by revolutions per minute (RPM) of the drive.

- Average Latency

  ½ seek time, plus ½ single rotation time

Flash drive performance

The performance of flash drives, also referred to as solid-state drives (SSDs), depends on only two technology factors:

- Architecture, for example single-level cell (SLC), enterprise multi-level cell (eMLC), and so on
- Chip generation, for example 34 nanometers (nm), 25 nm, 20 nm, and so on

Figure 7-1 shows a general HDD performance comparison taken from an IBM Storwize V7000, single-drive, Redundant Array of Independent Disks 0 (RAID 0) cache disabled configuration.

	NL-SAS 7,200 RPM	SAS 10,000 RPM	SAS 15,000 RPM	SSD Consumer	SSD Enterprise
Read IOPS	100 *(150)*	200 *(320)*	300 *(500)*	~5,000	25,000 - 45,000
Write IOPS	100 *(120)*	200 *(275)*	300 *(400)*	~1,000*	7,000 - 30,000*
Read MB/s	100-180	120-200	175-200	200	300-500
Write MB/s	100-180	120-200	150-200	200*	100-500*
Minimum Response Time	9ms	6ms	3.5ms	0.5ms	0.2ms

Figure 7-1 Drive performance comparison: Value in (x) are short stroked (<25% capacity used)

7.1.2 RAID performance

The Storwize V7000 Gen2 supports RAID 0, 1, 5, 6, and 10. In the following sections, we show the attributes of each:

RAID 0

- Striping
- Full capacity
- No protection against drive loss
- One host write equals one disk write (fastest performance)

RAID 1

- Mirroring between two drives
- Effective capacity of 50%
- Protects against one drive loss
- One host write equals two disk writes (fast performance)

RAID 5

- Striping with parity
- Protects against 1 drive loss
- Effective capacity of total drives minus one (n-1)
- One host write equals two disk reads and two disk writes

RAID 6

- Striping with double parity
- Protects against two drive losses
- Effective capacity of total drives minus two (n-2)
- One host write equals three disk reads and three disk writes (poor performance)

RAID 10

- Mirroring between two sets of striped drives
- Protects against up to 50% drive loss
- Effective capacity or 50%
- One write equals two disk writes (good performance)

7.1.3 Drive, RAID, and storage pool best practices

Generally speaking, the IBM Storwize V7000 Gen2 performs well, and the Storage Configuration Wizard provides an excellent setup for most designs. Even so, the following considerations should be taken into account:

- It is suggested to use the same extent size for all pools in a cluster.
- It is suggested to use only one class of array or managed disk per storage pool. The only exception is when using hybrid pools with Easy Tier.
- RAID 6 is generally advised for Nearline serial-attached SCSI (NL-SAS) drives.
- Generally, Storwize controllers have the same performance in all attached enclosures, though flash drives should be installed in or closer to the control enclosure.
- Use Easy Tier with flash drives for optimal price/performance. Typically, less than 5% flash is needed (The IBM Storage Tier Advisor Tool (STAT) tool is useful and can be used without hybrid pools).
- If virtualizing external storage, use tiering at highest level and disable at lower levels.

7.2 SAN Performance considerations

When designing a storage area network (SAN) infrastructure, or maintaining an existing infrastructure, you need to consider many factors in terms of their potential effect on performance. These factors include, but are not limited to, the following conditions:

- Dissimilar workloads competing for the same resources
- Overloaded resources
- Insufficient resources available
- Poorly performing resources
- Similar performance constraints

Remember the following high-level rules when designing your SAN and Storwize layout:

- Host-to-Storwize inter-switch link (ISL) oversubscription

  This area is the most significant I/O load across ISLs. The suggestion is to maintain a maximum of 7-to-1 oversubscription. Going higher is possible, but it could lead to I/O bottlenecks. This suggestion also assumes a core-edge design, where the hosts are on the edge and the IBM SAN Volume Controller is on the core.

- Storage-to-Storwize ISL oversubscription

  This area is the second most-significant I/O load across ISLs. The maximum oversubscription is 7-to-1. Going higher is not supported. Again, this suggestion assumes a multiple-switch SAN fabric design.

- ISL trunking/port channeling

  For the best performance and availability, we highly suggest that you use ISL trunking/port channeling. Independent ISL links can easily become overloaded and turn into performance bottlenecks. Bonded or trunked ISLs automatically share load, and provide better redundancy in the case of a failure.

- Number of paths per host multipath device

  The maximum supported number of paths per multipath device that is visible on the host is eight. Although the IBM Subsystem Device Driver Path Control Module (SDDPCM)-related products, and most vendor multipathing software, can support more paths, the suggested number of paths to a volume is four. Although the Storwize V7000 Gen2 can work with more than eight paths, this design is technically unsupported.

Rules and guidelines are no substitution for monitoring performance. Monitoring performance can both provide a validation that design expectations are met, and identify opportunities for improvement.

7.3 Performance monitoring

In this section, we highlight several performance monitoring techniques.

7.3.1 Collecting performance statistics

The IBM Storwize V7000 Gen2 is constantly collecting performance statistics. The default frequency by which files are created is at five-minute intervals. The collection interval can be changed using the **`startstats`** command. The statistics files (named VDisk, MDisk, and Node), and both canisters in the Storwize V7000 Gen2 system, keep the most recent 16 files of each type.

This design provides statistics for the most recent 80-minute period if using the default five-minute sampling interval. The Storwize V7000 Gen2 supports user-defined sampling intervals of from1 - 60 minutes. You can define the sampling interval by using the **`startstats -interval 2`** command to collect statistics at 2-minute intervals.

Collection intervals: Although more frequent collection intervals provide a more detailed view of what happens in the Storwize V7000 Gen2, they shorten the amount of time that the historical data is available. For example, rather than an 80-minute period of data with the default 5-minute interval, if you adjust to 2-minute intervals, you have a 32-minute period instead.

Statistics file naming

The files that are generated are written to the `/dumps/iostats/` directory. The file name is in the following formats:

- **`Nm`**`_stats_<`*`node_serial_number`*`>_<`*`date`*`>_<`*`time`*`>` for managed disk (MDisk) statistics
- **`Nv`**`_stats_<`*`node_serial_number`*`>_<`*`date`*`>_<`*`time`*`>` for virtual disk (VDisk) statistics
- **`Nn`**`_stats_<`*`node_serial_number`*`>_<`*`date`*`>_<`*`time`*`>` for node statistics
- **`Nd`**`_stats_<`*`node_serial_number`*`>_<`*`date`*`>_<`*`time`*`>` for disk drive statistics

The <*node_serial_number*> is of the node on which the statistics were collected. The date is in the form <*yymmdd*> and the time is in the form <*hhmmss*>. The following example shows an MDisk statistics file name:

```
Nm_stats_7836640-2_140901_164012
```

The **lsdumps -prefix /dumps/iostats** command shows typical MDisk volume, node, and disk drive statistics file names, as shown in Example 7-1. Note that the output is truncated and shows only part of the available statistics.

Example 7-1 The lsdumps command output

```
IBM_Storwize:ITSO_V7000Gen2:admin>lsdumps -prefix /dumps/iostats
id filename
0  Nm_stats_7836640-2_140901_125505
1  Nv_stats_7836640-2_140901_125505
2  Nn_stats_7836640-2_140901_125505
3  Nd_stats_7836640-2_140901_125505
4  Nn_stats_7836640-2_140901_131006
.... (truncated).....
```

Tip: The performance statistics files can be copied from the Storwize V7000 Gen2 nodes to a local drive on your workstation using the **pscp.exe** command (included with PuTTY) from an MS-DOS CLI, as shown in the following example:

```
C:\Program Files\PuTTY>pscp -unsafe -load ITSO_V7000Gen2
admin@10.18.229.81:/dumps/iostats/* c:\statsfiles
```

Use the **-load** parameter to specify the session that is defined in PuTTY.

Specify the **-unsafe** parameter when you use wildcards.

7.3.2 Real-time performance monitoring

Real-time performance statistics provide short-term status information for the Storwize V7000 Gen2. The statistics are shown as graphs in the management graphical user interface (GUI), or can be viewed from the CLI. With system-level statistics, you can quickly view the processor use and the bandwidth of volumes, interfaces, and MDisks. Each graph displays the current bandwidth in either megabytes per second (MBps) or I/O operations per second (IOPS), and a view of bandwidth over time.

Each node collects various performance statistics, mostly at five-second intervals, and the statistics that are available from the config node in a clustered environment. This information can help you determine the performance effect of a specific node. As with system statistics, node statistics help you to evaluate whether the node is operating within normal performance metrics.

Real-time performance monitoring gathers the following system-level performance statistics:

- Central processing unit (CPU) use
- Port use and I/O rates
- Volume and MDisk I/O rates
- Bandwidth
- Latency

Real-time statistics are not a configurable option, and cannot be disabled.

Real-time performance monitoring with the CLI

The following commands are available for monitoring the statistics through the CLI:

- **`lssystemstats`**

  Use the **`lssystemstats`** command to display the most recent values of all of the node statistics in a clustered system (system), or to display a history of values for a given subset of available statistics across all nodes in a system. This command also can be used to display a history of values for a given subset of available statistics.

- **`lsnodecanisterstas`**

  Use the **`lsnodecanisterstats`** command to display the most recent values of statistics for all of the nodes or node canisters, and display all statistics for a particular Node Canister. Additionally, You can use this command to display a history of values for a given subset of available statistics.

Both commands lists the same set of statistics, but either representing all node canisters in the cluster, or a particular Node Canister. The values for these statistics are calculated from the node statistics values in the following way:

- Bandwidth: Sum of bandwidth of all nodes
- Latency: Average latency for the cluster, which is calculated using data from the whole cluster, not an average of the single node values
- IOPS: Total IOPS of all nodes
- CPU percentage: Average CPU percentage of all nodes

Example 7-2 shows the resulting output of the **`lssystemstats`** command.

Example 7-2 The lssystemstats command output

```
IBM_Storwize:ITSO_V7000Gen2:admin>lssystemstats
stat_name          stat_current stat_peak stat_peak_time
compression_cpu_pc 8            8         140901175343
cpu_pc             0            0         140901175343
fc_mb              0            0         140901175343
fc_io              1350         1371      140901175208
sas_mb             0            44        140901175128
sas_io             0            285       140901174948
.....(truncated)......
drive_w_mb         0            0         140901175343
drive_w_io         0            5         140901175248
drive_w_ms         0            5         140901175248
power_w            339          344       140901175023
temp_c             26           26        140901175343
temp_f             78           78        140901175343
iplink_mb          0            0         140901175343
iplink_io          0            0         140901175343
IBM_Storwize:ITSO_V7000Gen2:admin>
```

Table 7-1 has a brief description of each of the statistics presented by the **lssystemstats** and **lsnodecanisterstats** commands.

Table 7-1 Field name descriptions for lssystemstats and lsnodecanisterstats statistics

Field name	Unit	Description
compression_cpu_pc	Percentage	Compression CPU use
cpu_pc	Percentage	Use of node CPUs
fc_mb	MBps	Fibre Channel bandwidth
fc_io	IOPS	Fibre Channel throughput
sas_mb	MBps	SAS bandwidth
sas_io	IOPS	SAS throughput
iscsi_mb	MBps	iSCSI bandwidth
iscsi_io	IOPS	iSCSI throughput
write_cache_pc	Percentage	Write cache fullness, updated every ten seconds
total_cache_pc	Percentage	Total cache fullness, updated every ten seconds
vdisk_mb	MBps	Total VDisk bandwidth
vdisk_io	IOPS	Total VDisk throughput
vdisk_ms	Milliseconds (ms)	Average VDisk latency
mdisk_mb	MBps	MDisk (SAN and RAID) bandwidth
mdisk_io	IOPS	MDisk (SAN and RAID) throughput
mdisk_ms	Milliseconds	Average MDisk latency
drive_mb	MBps	Drive bandwidth
drive_io	IOPS	Drive throughput
drive_ms	Milliseconds	Average drive latency
vdisk_w_mb	MBps	VDisk write bandwidth
vdisk_w_io	IOPS	VDisk write throughput
vdisk_w_ms	Milliseconds	Average VDisk write latency
mdisk_w_mb	MBps	MDisk (SAN and RAID) write bandwidth
mdisk_w_io	IOPS	MDisk (SAN and RAID) write throughput
mdisk_w_ms	Milliseconds	Average MDisk write latency
drive_w_mb	MBps	Drive write bandwidth
drive_w_io	IOPS	Drive write throughput
drive_w_ms	Milliseconds	Average drive write latency
vdisk_r_mb	MBps	VDisk read bandwidth
vdisk_r_io	IOPS	VDisk read throughput
vdisk_r_ms	Milliseconds	Average VDisk read latency

Field name	Unit	Description
vdisk_w_mb	MBps	VDisk write bandwidth
vdisk_w_io	IOPS	VDisk write throughput
vdisk_w_ms	Milliseconds	Average VDisk write latency
mdisk_r_mb	MBps	MDisk (SAN and RAID) read bandwidth
mdisk_r_io	IOPS	MDisk (SAN and RAID) read throughput
mdisk_r_ms	Milliseconds	Average MDisk read latency
mdisk_w_mb	MBps	MDisk (SAN and RAID) write bandwidth
mdisk_w_io	IOPS	MDisk (SAN and RAID) write throughput
mdisk_w_ms	Milliseconds	Average MDisk write latency
drive_r_mb	MBps	Drive read bandwidth
drive_r_io	IOPS	Drive read throughput
drive_r_ms	Milliseconds	Average drive read latency
drive_w_mb	MBps	Drive write bandwidth
drive_w_io	IOPS	Drive write throughput
drive_w_ms	Milliseconds	Average drive write latency
power_w	Watts	Power consumption
temp_c	Celsius	Temperature in Celsius
temp_f	Fahrenheit	Temperature in Fahrenheit
iplink_mb	MBps	Internet Protocol (IP) link Bandwidth
iplink_io	IOPS	IP link Throughput

Real-time performance monitoring with the GUI

The real-time statistics are also available from the Storwize V7000 Gen2 GUI. Select **Monitoring** → **Performance** (Figure 7-2) to open the performance monitoring window.

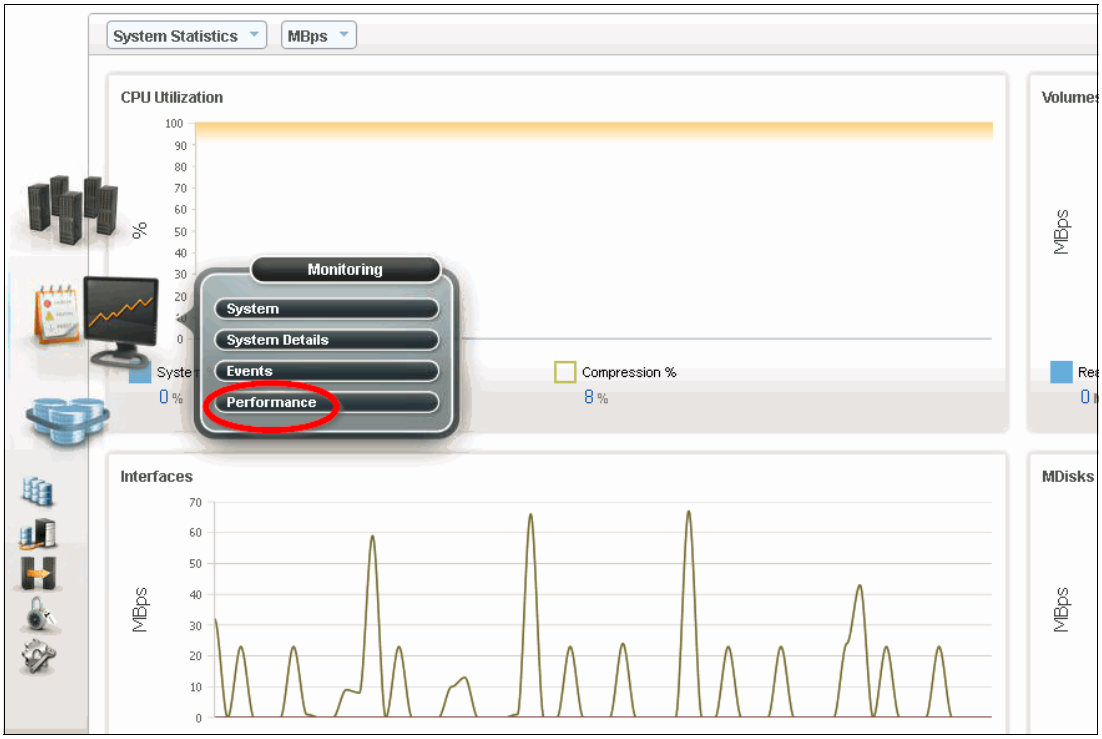


Figure 7-2 Storwize Monitoring menu

The window, as shown in Figure 7-3 on page 141, is divided into four sections that provide use views for the following resources:

- CPU Use
  - Shows the CPU usage for general tasks%
  - Shows the CPU7 usage for compression (when enabled)%
- Volumes. This shows the overall volume use with the following fields:
  - Read
  - Write
  - Read latency
  - Write latency
- Interfaces. This shows the overall statistics for each of the available interfaces:
  - Fibre Channel
  - iSCSI
  - Serial-attached SCSI (SAS)
  - IP Replication
- MDisks. This shows the following overall statistics for the MDisks:
  - Read
  - Write
  - Read latency
  - Write latency

Figure 7-3 shows real-time performance graphs.

Figure 7-3 Real-time performance graphs

You can also select to view performance statistics for each of the available canisters of the system, as shown in Figure 7-4.

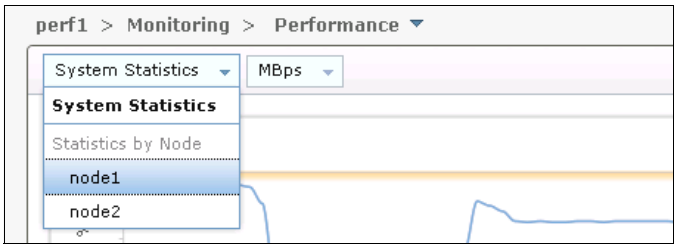


Figure 7-4 Select system node (Canister)

It is also possible to change the metric between MBps or IOPS (Figure 7-5).

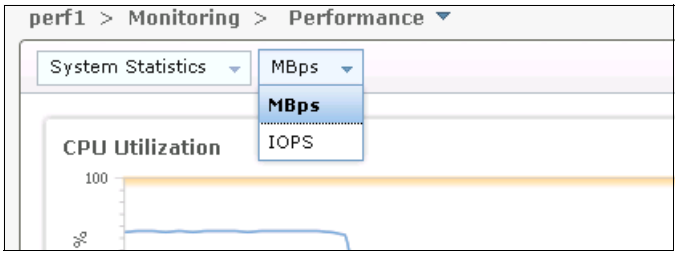


Figure 7-5 Changing to MBps

On any of these views, you can select any point in time with your cursor to know the exact value and when it occurred. As soon as you place your cursor over the timeline, it becomes a dotted line with the various values gathered (Figure 7-6).

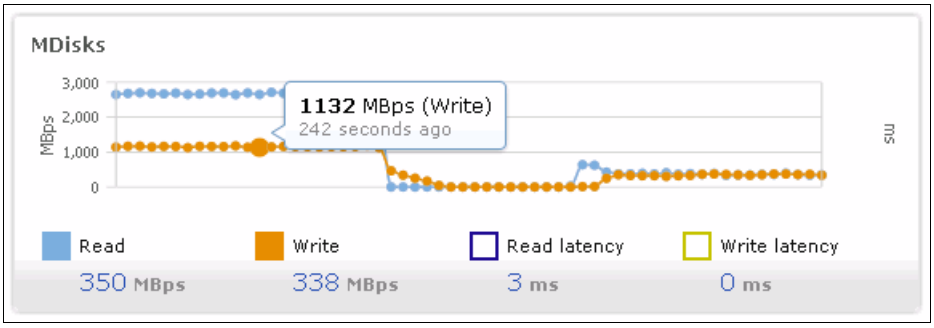


Figure 7-6 Detailed resource use

For each of the resources, there are various values that you can view by selecting the check box next to a value. For example, for the MDisks view, as shown in Figure 7-7, the four available fields are selected:

- Read
- Write
- Read latency
- Write latency

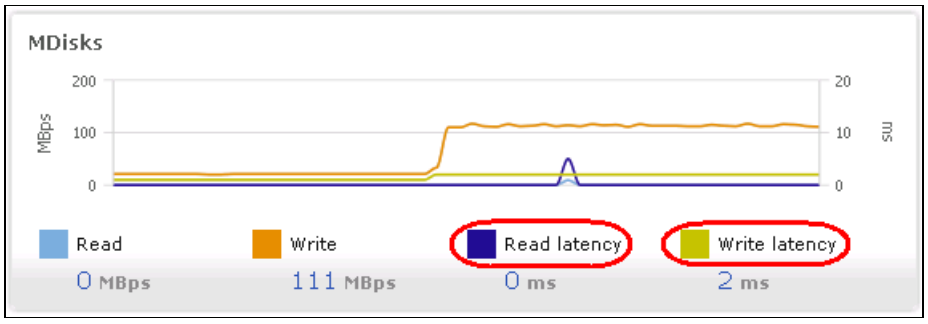


Figure 7-7 Detailed resource use

7.3.3 Performance data collection and Tivoli Storage Productivity Center for Disk

Although you can obtain performance statistics in standard .xml files, using .xml files is a less practical and less user-friendly method to analyze the Storwize V7000 Gen2 performance statistics. Tivoli Storage Productivity Center for Disk is the supported IBM tool to collect and analyze performance statistics.

For more information about using Tivoli Storage Productivity Center to monitor your storage subsystem, see the following IBM Redbooks publications:

- *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364, which is available at the following website:

  http://www.rebooks.ibm.com/abstracts/sg247364.html?Open

- *Tivoli Storage Productivity Center V5.2 Release Guide*, SG24-8204, which is available at the following website:

  http://www.redbooks.ibm.com/redpieces/abstracts/sg248204.html?Open

8

IBM Storwize V7000 Gen2 command-line interface

In this chapter, we describe the command-line interface (CLI):

- How to set up CLI
- Configuring the IBM Storwize V7000 Gen2 using the CLI:
  - Viewing the Node Canister details
  - Configuring event notification settings
  - Configuring internal storage
  - Configuring external storage
  - Configuring volumes
  - Configuring remote copy
  - Configuring hosts
- New commands introduced with the IBM Storwize V7000 Gen2

The CLI is a powerful tool offering even more functionality than the graphical user interface (GUI). We show how to set it up, and how to manage and operate your IBM Storwize V7000 Gen2. We do not delve into the advanced functionality, because it is beyond the intended scope of this book. If you want to learn about the advanced commands,see the folllowing website:

http://www.ibm.com/support/knowledgecenter/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/v7000_ichome_730.html

Furthermore, the IBM Storwize V7000 Gen2 shares the underlying platform with the IBM storage area network (SAN) Volume Controller. Therefore, we also suggest the CLI chapter in *IBM SAN Volume Controller 2145-DH8 Introduction and Implementation*, SG24-8229:

http://www.redbooks.ibm.com/Redbooks.nsf/RedpieceAbstracts/sg248229.html

8.1 How to set up the CLI

In the IBM Storwize V7000 Gen2 GUI, authentication is done by using a user name and password. The CLI uses a Secure Shell (SSH) to connect from the host to the IBM Storwize V7000 Gen2 system. There are two ways of connecting SSH to the Storwize V7000 Gen2. The first option is user name and password, and the other option is installing SSH keys. We suggest using SSH keys for maximum security. The following steps are required to enable CLI access with SSH keys:

- A public key and a private key are generated together as a pair.
- A public key is uploaded to the IBM Storwize V7000 Gen2 through the GUI.
- A client SSH tool must be configured to authenticate with the private key.
- A secure connection can be established between the client and Storwize V7000 Gen2.

SSH is the communication vehicle between the management workstation and the Storwize V7000 Gen2. The SSH client provides a secure environment from which to connect to a remote machine. It uses the principles of public and private keys for authentication.

SSH keys are generated by the SSH client software. The SSH keys include a public key, which is uploaded and maintained by the clustered system, and a private key, which is kept private on the workstation that is running the SSH client. These keys authorize specific users to access the administration and service functions on the system. Each key pair is associated with a user-defined ID string that can consist of up to 40 characters.

Up to 100 keys can be stored on the system. New IDs and keys can be added, and unwanted IDs and keys can be deleted. To use the CLI, an SSH client must be installed on that system, the SSH key pair must be generated on the client system, and the client's SSH public key must be stored on the IBM Storwize V7000 Gen2.

The SSH client used in this book is PuTTY. Also, a PuTTY key generator can be used to generate the private and public key pair. The PuTTY client can be downloaded from the following address at no initial cost:

`http://www.chiark.greenend.org.uk`

Download the following tools:

- PuTTY SSH client: `putty.exe`
- PuTTY key generator: `puttygen.exe`

Generating a public and private key pair

To generate a public and private key pair, complete the following steps:

1. Start the PuTTY key generator to generate the public and private key pair shown in Figure 8-1.

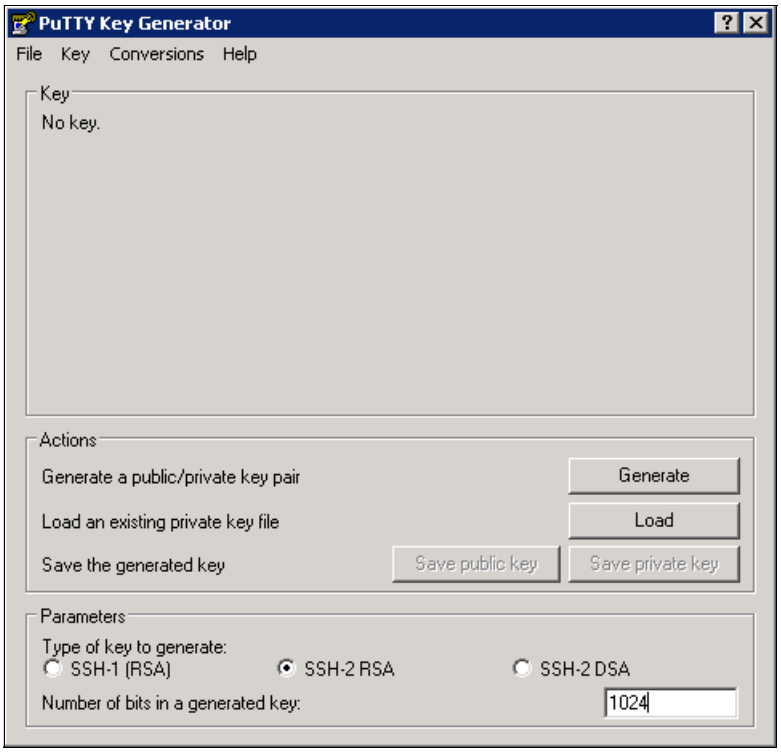


Figure 8-1 PuTTY Key Generator

Make sure that the following options are selected:

- SSH2 Rivest-Shamir-Adleman (RSA)
- Number of bits in a generated key: 1024

2. Click **Generate** and move the cursor over the blank area to generate keys (Figure 8-2).

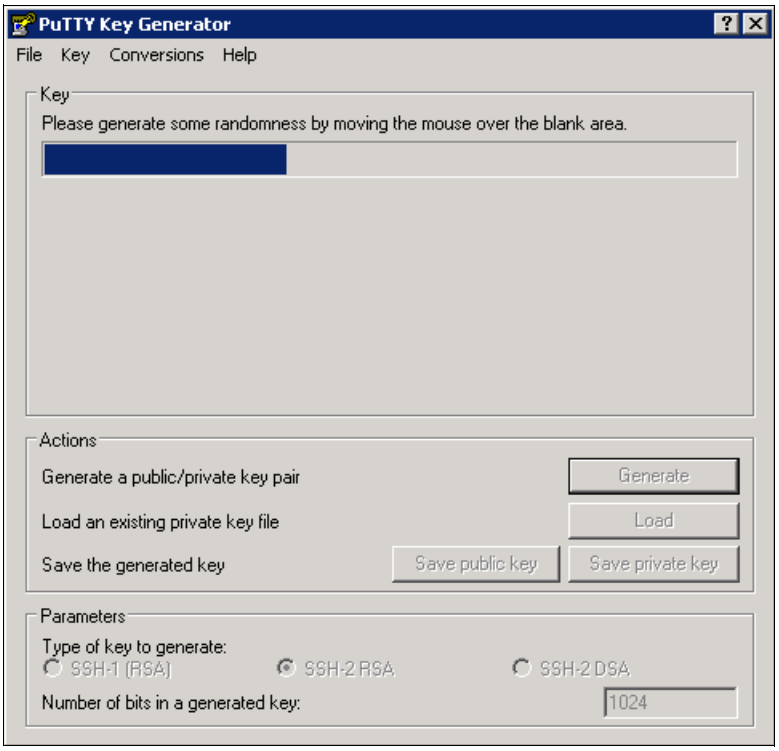


Figure 8-2 Generating Keys

To generate keys: The blank area that is indicated by the message is the large blank rectangle on the GUI inside the section of the GUI labeled *Key*. Continue to move the mouse pointer over the blank area until the progress bar reaches the far right. This action generates random characters to create a unique key pair.

3. After the keys are generated, save them for later use. Click **Save public key** (Figure 8-3).

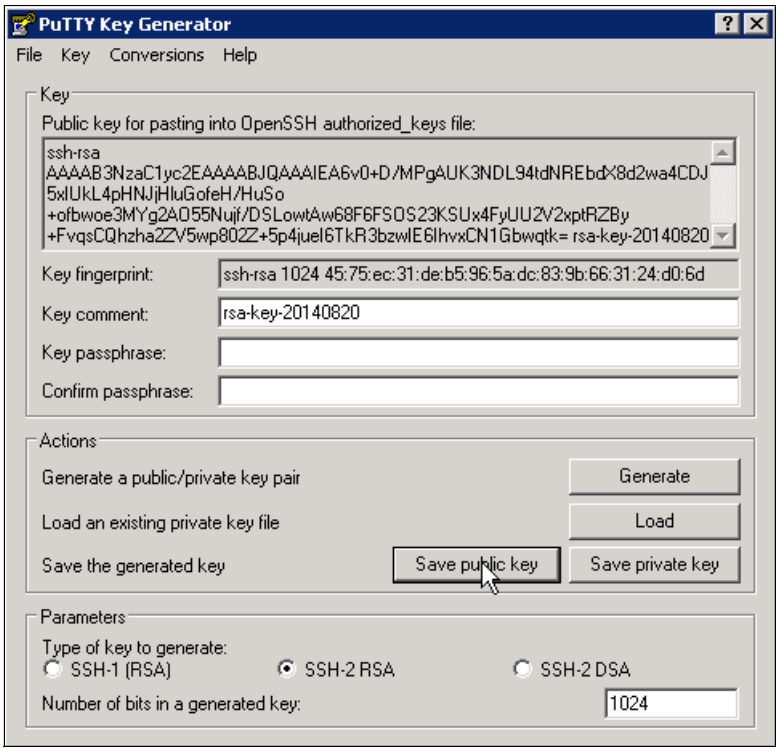


Figure 8-3 Saving public key

4. You are prompted for a name (for example, `pubkey`) and a location for the public key (for example, `C:\Support Utils\PuTTY`). Click **Save**.

   Ensure that you record the name and location, because the name and location of this SSH public key must be specified later.

> **Public key extension:** By default, the PuTTY key generator saves the public key with no extension. Use the string `pub` for naming the public key, for example, `pubkey`, to easily differentiate the SSH public key from the SSH private key.

5. Click **Save private key** (Figure 8-4).

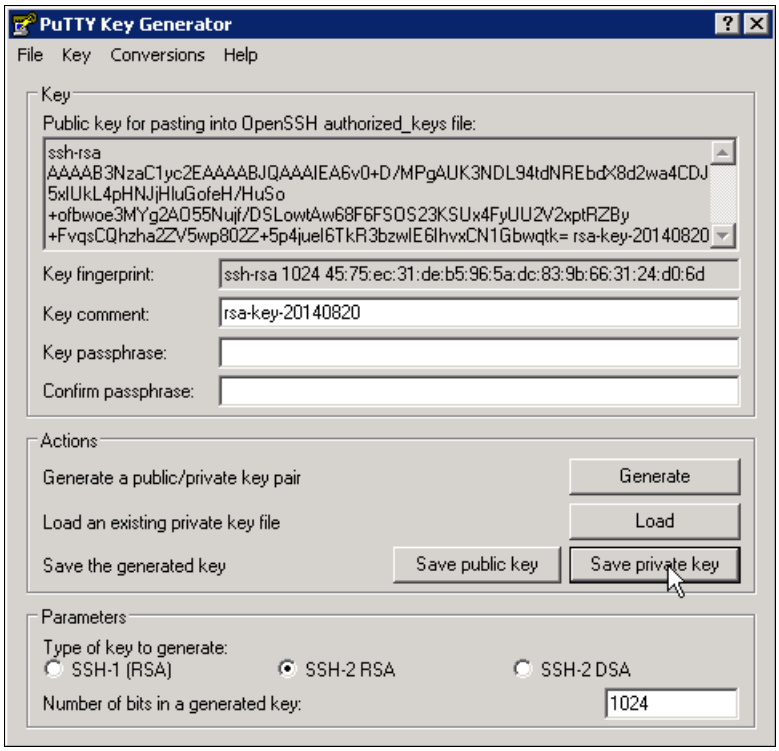


Figure 8-4 Save private key

6. You are prompted with a warning message (Figure 8-5). Click **Yes** to save the private key without a passphrase.

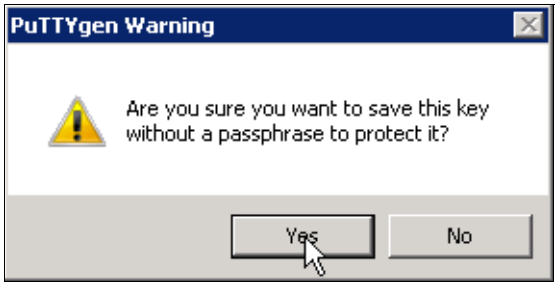


Figure 8-5 Warning message

7. When prompted, enter a name (for example, `icat`), select a secure place as the location, and click **Save**.

> **Key generator:** The PuTTY key generator saves the private key with the PPK extension.

8. Close the PuTTY key generator.

Uploading the SSH public key to the IBM Storwize V7000

After you create your SSH key pair, upload your SSH public key onto the SAN Volume Controller system. Complete the following steps:

1. Open the user section (Figure 8-6).

Figure 8-6 User section

2. Right-click the user name for which you want to upload the key and click **Properties** (Figure 8-7).

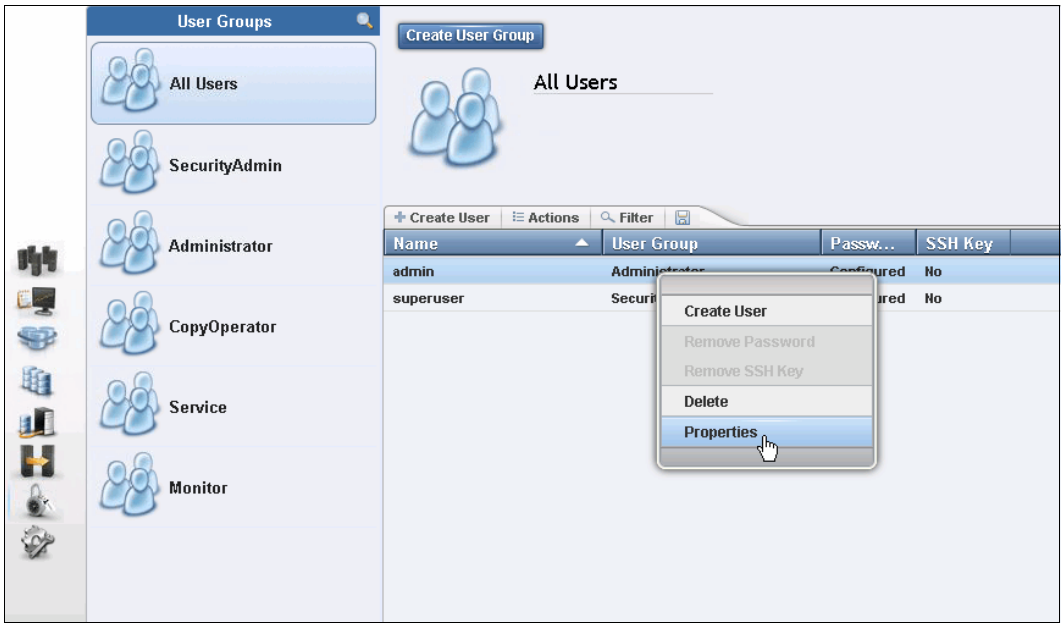


Figure 8-7 User Properties

3. To upload the public key, click **Browse**, select your public key, and click **OK** (Figure 8-8).

Figure 8-8 Upload Public Key

Configuring the SSH client

Before the CLI can be used, the SSH client must be configured:

1. Start PuTTY. The PuTTY Configuration window opens (Figure 8-9).

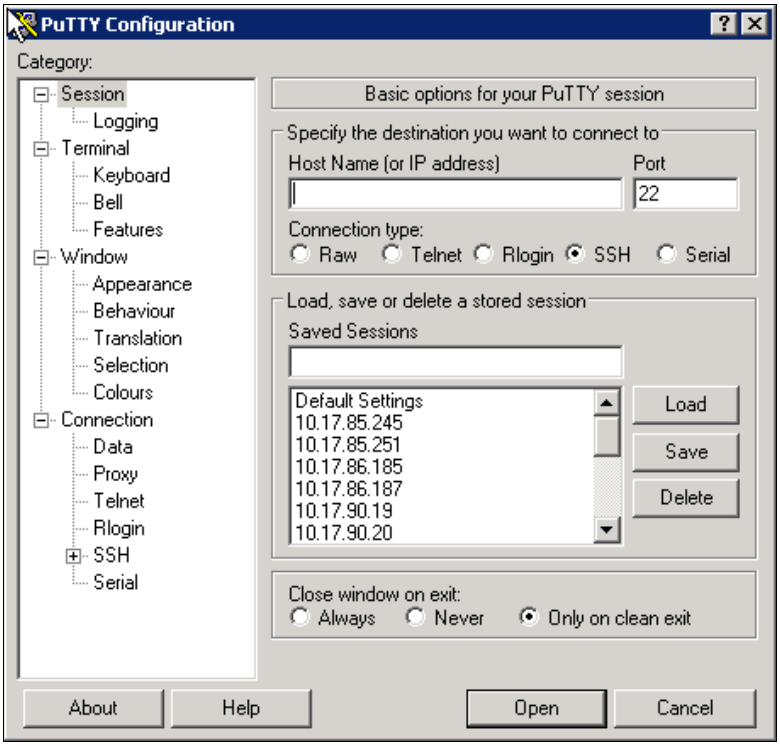


Figure 8-9 PuTTY Main window

2. In the right pane, select **SSH** as the connection type. In the **Close window on exit** section, select **Only on clean exit**, which ensures that if any connection errors occur, they are displayed on the user's window.

3. In the Category pane, on the left side of the PuTTY Configuration window (Figure 8-10), click **Connection** → **SSH** to open the PuTTY SSH Configuration window.

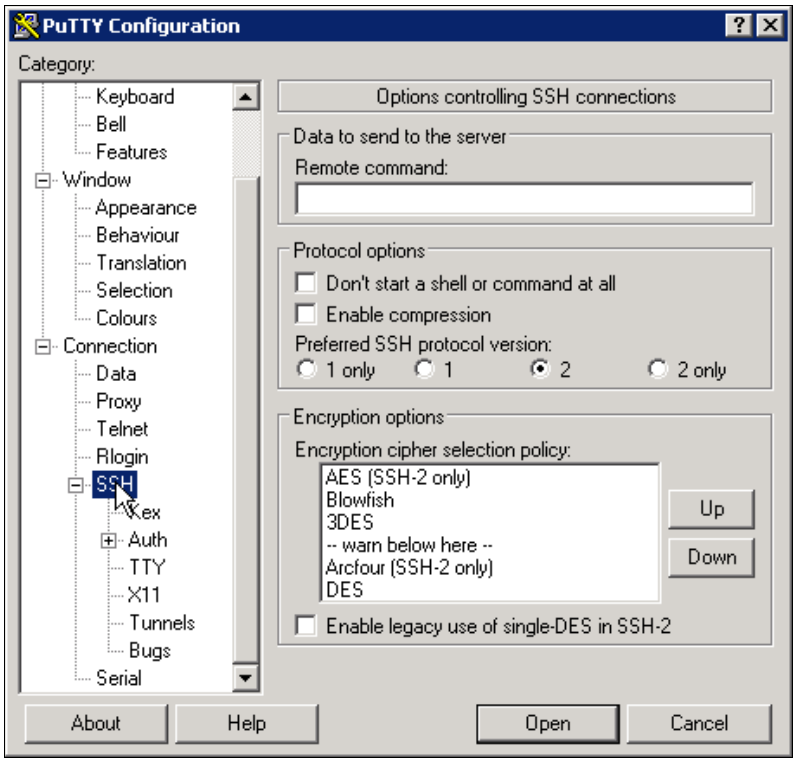


Figure 8-10 SSH subcategory

4. In the Preferred SSH protocol version section, select **2**.
5. In the Category pane on the left, click **Connection** → **SSH** → **Auth**. More options are displayed for controlling SSH authentication.

6. In the **Private key file for authentication** field (Figure 8-11), either browse to or type the fully qualified directory path and file name of the SSH client private key file, which was created previously (for example, `C:\Support Utils\PuTTY\icat.PPK`).

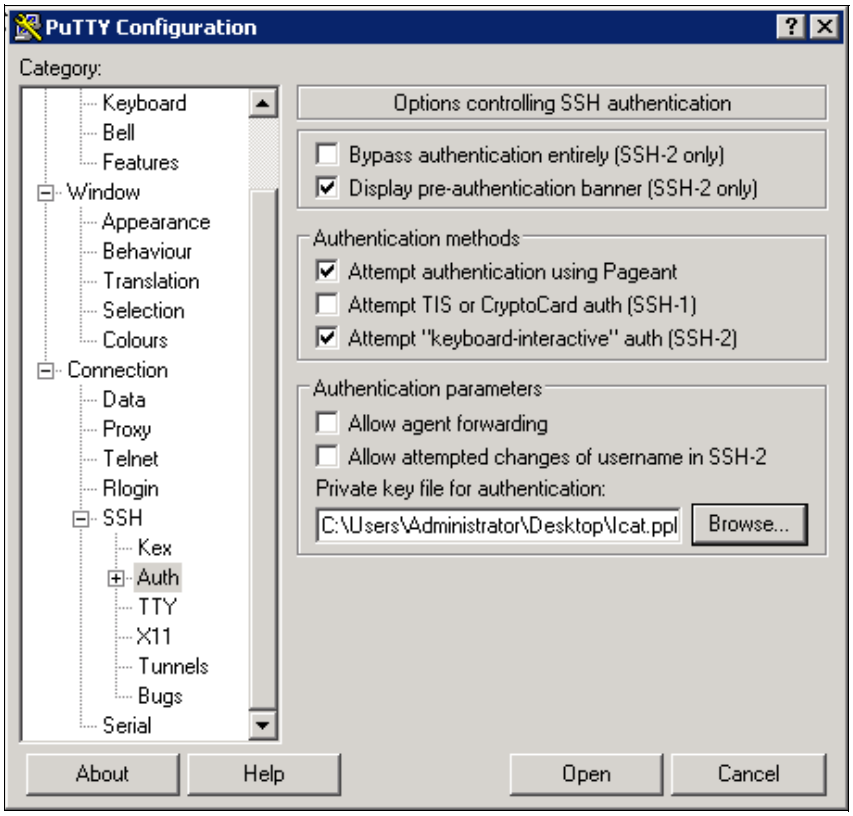


Figure 8-11 Private key for authentication

7. In the Category pane, click **Session** to return to the Basic options for your PuTTY session view (Figure 8-12 on page 154).
8. Enter the following information in the fields (Figure 8-12 on page 154) in the right pane:
   - Host Name: Specify the host name or cluster IP address of the IBM Storwize V7000 Gen2.
   - Saved Sessions: Enter a session name.

9. Click **Save** to save the new session (Figure 8-12).

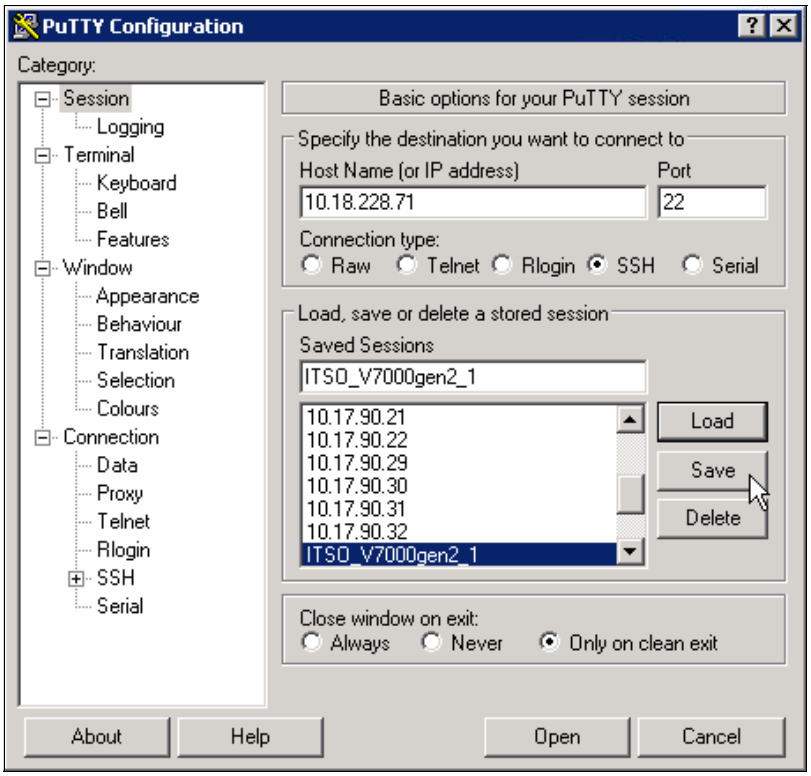


Figure 8-12 Saving a session

10. Select the new session and click **Open** to connect to the IBM Storwize V7000 system. A PuTTY Security Alert opens; confirm it by clicking **Yes** (Figure 8-13).

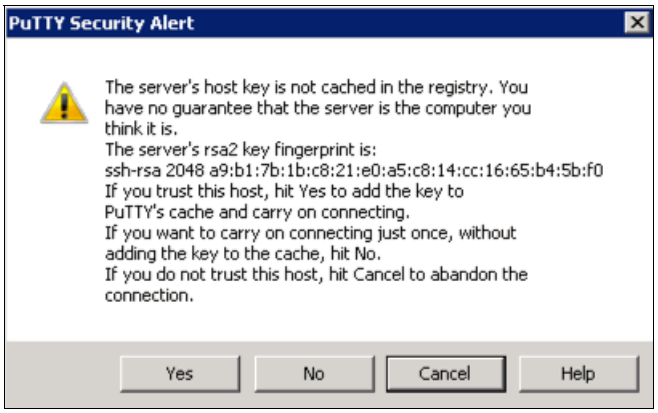


Figure 8-13 Putty Security Alert

11. PuTTY now connects to the system and prompts you for a user name to log in as. Enter `admin` as the user name, as shown in Example 8-1. Click Enter.

Example 8-1 Enter user name

```
login as: admin
Authenticating with public key "rsa-key-20140820"
IBM_2076:ITSO_V7000Gen2_1:admin>
```

12. You have now completed the tasks to configure the CLI for IBM Storwize V7000 Gen2 administration.

8.2 Configuring the IBM Storwize V7000 Gen2 using the CLI

Now we describe how to use the CLI to configure the Storwize V7000 Gen2

Tip: For a listing a full commands including syntaxes, variables, and arguments, see the IBM Knowledge Center. You can also use the help command followed by the command you want to learn more about.

Our lab setup includes the following hardware:

- 1 x IBM Storwize V7000 Gen2 model 2076-524
  - 1x Expansion Enclosure model 2076-24F
  - 12 x SAS 600 GB 10 K SFF HDDs (6 in control enclosure and 6 in Expansion)
  - 2 x 4 Port 8 Gbps Fiber Channel Host Interface Cards
- A balanced, redundant two-Fabric SAN Fibre Channel switch setup
- 2 x IBM Storwize V7000 model 2076-324, which we will use as external storage

8.2.1 Viewing the Node Canister details

Now we review how to retrieve hardware and configuration details from your Storwize V7000 Gen2. Although many prefer to use the GUI, and with good reason, there are some advantages to using the CLI. To the trained user it's more precise, quicker, and extremely lightweight.

To view the Node Canister details, follow these steps:

1. Use the **`lsnodecanister`** command to return a concise list or a detailed view of node canisters that are part of the clustered system. To view the concise list, type the command (Example 8-2).

Example 8-2 The lsnodecanister concise view

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>lsnodecanister
id name  UPS_serial_number WWNN             status IO_group_id IO_group_name
config_node UPS_unique_id hardware iscsi_name
iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number
1  node1                  500507680B00217A online 0           io_grp0
yes                       500
iqn.1986-03.com.ibm:2145.itsov7000gen22.node1            01-1      1
1           7836640
3 node2            500507680B00217B online 0       io_grp0     no
500    iqn.1986-03.com.ibm:2145.itsov7000gen22.node2         01-2     1
2           7836640
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

2. To get a detailed view, add the canister ID or name as a parameter to the command (Example 8-3).

Example 8-3 The lsnodecanister detailed view

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>lsnodecanister 1
id 1
name node1
UPS_serial_number
WWNN 500507680B00217A
status online
IO_group_id 0
IO_group_name io_grp0
partner_node_id 3
partner_node_name node2
config_node yes
UPS_unique_id
port_id 500507680B21217A
port_status active
port_speed 8Gb
port_id 500507680B22217A
port_status active
port_speed 8Gb
port_id 500507680B23217A
port_status active
port_speed 8Gb
port_id 500507680B24217A
port_status active
port_speed 8Gb
hardware 500
iscsi_name iqn.1986-03.com.ibm:2145.itsov7000gen22.node1
iscsi_alias
failover_active no
failover_name node2
failover_iscsi_name iqn.1986-03.com.ibm:2145.itsov7000gen22.node2
failover_iscsi_alias
panel_name 01-1
enclosure_id 1
canister_id 1
enclosure_serial_number 7836640
service_IP_address 10.18.228.57
service_gateway 10.18.228.1
service_subnet_mask 255.255.255.0
service_IP_address_6
service_gateway_6
service_prefix_6
service_IP_mode static
service_IP_mode_6
site_id
site_name
identify_LED off
product_mtm 2076-524
IBM_Storwize:ITSO_V7000Gen2:admin>
```

8.2.2 Configuring event notification settings

You can use the CLI to set up your system to send event notification and inventory reports to specified recipients and the IBM Support Center:

1. First, we need to set up an email server. This is done with the **mkemailserver** command (Example 8-4):

   ```
   mkemailserver -ip {ip_address -port} {port_number}
   ```

2. Then we add a recipient with the **mkemailuser** command (Example 8-4):

   ```
   mkemailuser -address {recipient email address} -error {on|off} -warning
   {on|off} -info {on|off} inventory {on|off}
   ```

3. Finally, we add the contact details for the recipient email address with the **chemail** command (Example 8-4):

   ```
   chemail -reply {recipient email address} -contact {contact name} -primary
   {primary phone number} -alternate {alternate phone number} -location {location
   eg. room3} -contact2 {alternate contac name} -primary2 {primary contact2 phone}
   -alternate2 {alternate contact2 phone} -nocontact {organization} -address
   {location address}
   -city <location city>
   ```

Example 8-4 Event notification example

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkemailserver -ip 2.2.2.2 -port 78
Email Server id [1] successfully created

IBM_Storwize:ITSO_V7000GEN2_2:admin>mkemailuser -address ibmuser@ibm.com -error
on -warning on -info off -inventory off
User, id [1], successfully created

IBM_Storwize:ITSO_V7000GEN2_2:admin>chemail -reply ibmuser@ibm.com -primary
1112225555 -location machine_room_3
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

8.2.3 Configuring internal storage

In this topic, we describe how to configure the internal drives in the Storwize V7000 Gen2. Our lab configuration consists of 12 drives spread across two enclosures and, as an example, we would like to configure it as a single storage pool.

Note: Command syntax, arguments, and variables are not shown in this topic. Use **-?** or >**help {command name}** to obtain more details.

To configure internal storage, follow these steps:

1. We start by checking available internal storage. Issue the **lsdrive** command (Example 8-5).

Example 8-5 The lsdrive command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>lsdrive
id status error_sequence_number use    tech_type capacity mdisk_id mdisk_name
member_id enclosure_id slot_id node_id node_name
0  online                       unused sas_hdd   558.4GB 1           1
1  online                       unused sas_hdd   558.4GB 1           3
2  online                       unused sas_hdd   558.4GB 1           5
3  online                       unused sas_hdd   558.4GB 1           2
4  online                       unused sas_hdd   558.4GB 1           4
5  online                       unused sas_hdd   558.4GB 1           6
6  online                       unused sas_hdd   558.4GB 2           4
7  online                       unused sas_hdd   558.4GB 2           1
8  online                       unused sas_hdd   558.4GB 2           3
9  online                       unused sas_hdd   558.4GB 2           2
10 online                       unused sas_hdd   558.4GB 2           5
11 online                       unused sas_hdd   558.4GB 2           6
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

2. At first, we make our drives candidates with the **chdrive** command (Example 8-6), then repeat for the remaining drives.

Example 8-6 The chdrive command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>chdrive -use candidate
0:1:2:3:4:5:6:7:8:9:10:11
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

3. In our case, we create one storage pool to hold our storage with the **mkmdiskgrp** command (Example 8-7).

Example 8-7 The mkmdiskgrp command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkmdiskgrp -ext 1024 -tier enterprise -name
INT_V7KGEN2
MDisk Group, id [0], successfully created
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

4. We then create one array in the storage pool with the **mkarray** command, which gives us a storage pool consisting of 10 drives. In our setup, we would like the two remaining drives to act as hot spares. Use the **chdrive** command again (Example 8-8).

Example 8-8 The mkarray and chdrive commands

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkarray -level raid5 -drive
0:1:2:3:4:6:7:8:9:10 -strip 256 -name V7KG2 0
MDisk, id [1], successfully created
IBM_Storwize:ITSO_V7000GEN2_2:admin>chdrive -use candidate 5
IBM_Storwize:ITSO_V7000GEN2_2:admin>chdrive -use candidate 11
```

5. This concludes our internal storage configuration, and we now have a managed disk, as shown in Example 8-9 on page 159.

Example 8-9 The lsdrive command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>lsdrive
id status error_sequence_number use    tech_type capacity mdisk_id mdisk_name
member_id enclosure_id slot_id node_id node_name
0  online                    member sas_hdd   558.4GB  0       V7KG2     0
1            1
1  online                    member sas_hdd   558.4GB  0       V7KG2     1
1            3
2  online                    member sas_hdd   558.4GB  0       V7KG2     2
1            5
3  online                    member sas_hdd   558.4GB  0       V7KG2     3
1            2
4  online                    member sas_hdd   558.4GB  0       V7KG2     4
1            4
5  online                     spare  sas_hdd   558.4GB
1            6
6  online                    member sas_hdd   558.4GB  0       V7KG2     5
2            4
7  online                    member sas_hdd   558.4GB  0       V7KG2     6
2            1
8  online                    member sas_hdd   558.4GB  0       V7KG2     7
2            3
9  online                    member sas_hdd   558.4GB  0       V7KG2     8
2            2
10 online                    member sas_hdd   558.4GB  0       V7KG2     9
2            5
11 online                     spare  sas_hdd   558.4GB
2            6
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

8.2.4 Configuring external storage

Now, you learn how to virtualize external storage. In our case, we are using two Storwize V7000 model 324 as back-end storage systems.

To configure external storage, follow these steps:

1. When virtualizing external storage, you must zone the external storage to be visible from the Storwize V7000 Gen2. Then, issue the **`lscontroller`** command (Example 8-10).

Example 8-10 The lscontroller command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>lscontroller
id controller_name ctrl_s/n            vendor_id           product_id_low
product_id_high
0  controller0     2076                 IBM                 2145
1  controller1     2076                 IBM                 2145
2  controller2     2076                 IBM                 2145
3  controller3     2076                 IBM                 2145
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

This reveals our two IBM Storwize V7000, which are visible on the SAN.

Because the names do not have any meaning for us, we change them.

2. Issue the **chsystem** command (Example 8-11), and repeat for the other names.

Example 8-11 The chsystem command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>chcontroller -name ITSO_EXT1A 0
IBM_Storwize:ITSO_V7000GEN2_2:admin>chcontroller -name ITSO_EXT1B 1
IBM_Storwize:ITSO_V7000GEN2_2:admin>chcontroller -name ITSO_EXT2A 2
IBM_Storwize:ITSO_V7000GEN2_2:admin>chcontroller -name ITSO_EXT2B 3
IBM_Storwize:ITSO_V7000Gen2:admin>lscontroller
id controller_name ctrl_s/n            vendor_id           product_id_low
product_id_high
0  ITSO_EXT1A      2076                IBM                   2145
1  ITSO_EXT1B      2076                IBM                   2145
2  ITSO_EXT2A      2076                IBM                   2145
3  ITSO_EXT2B      2076                IBM                   2145
IBM_Storwize:ITSO_V7000Gen2:admin>
```

Now we have set up external storage, and we are ready to start configuring it.

Remember: Create the Storwize V7000 Gen2 as a host on the external disk system and present the volumes from the external storage to the Storwize V7000 Gen2. This action is also referred to as *host mapping* and *LUN mapping*.

3. The first thing we do when configuring external storage is to issue the **detectmdisk** command, which adds the external volumes as managed disks (MDisks). After this, we issue the **lsmdisk** command to reveal all of the visible managed disks (Example 8-12).

Example 8-12 The detectmdisk and lsmdisk commands

```
IBM_Storwize:ITSO_V7000Gen2:admin>detectmdisk

IBM_Storwize:ITSO_V7000Gen2:admin>lsmdisk
id name             status mode      mdisk_grp_id mdisk_grp_name capacity
ctrl_LUN_#     controller_name UID
tier
0 mdisk0    online array  0    INT_V7KGEN2 558.4GB
ssd
1 mdisk1    online array  0    INT_V7KGEN2 558.4GB
nearline
2 mdisk2    online array  0    INT_V7KGEN2 558.4GB
enterprise
3 mdisk3    online array  0    INT_V7KGEN2 558.4GB
enterprise
4 mdisk4    online array  0    INT_V7KGEN2 558.4GB
enterprise
5  md_v7kgen1-1_003 online unmanaged                              500.0GB
0000000000000000 ITSO_EXT1A
60050768028980026800000000000000c0000000000000000000000000000000 enterprise
6  md_v7kgen1-1_002 online unmanaged                              500.0GB
0000000000000001 ITSO_EXT1A
60050768028980026800000000000000d0000000000000000000000000000000 enterprise
7  md_v7kgen1-1_005 online unmanaged                              500.0GB
0000000000000002 ITSO_EXT1A
60050768028980026800000000000000e0000000000000000000000000000000 enterprise
8  md_v7kgen1-1_004 online unmanaged
500.0GB0000000000000003 ITSO_EXT1A
```

```
600507680289800268000000000000f000000000000000000000000000000000 enterprise
9  md_v7kgen1-1_001 online unmanaged                                2.0TB
0000000000000004 ITSO_EXT1A
600507680289800268000000000001d000000000000000000000000000000000 enterprise
10 md_v7kgen1-2-001 online unmanaged                                5.0GB
0000000000000000 ITSO_EXT2A
600507680282818b300000000000000e000000000000000000000000000000000 ssd
11 md_v7kgen1-2-002 online unmanaged                                50.0GB
0000000000000001 ITSO_EXT2A
600507680282818b300000000000000f000000000000000000000000000000000 nearline
12 md_v7kgen1-2-003 online unmanaged                                50.0GB
0000000000000002 ITSO_EXT2A
600507680282818b3000000000000010000000000000000000000000000000000 enterprise
13 md_v7kgen1-2-004 online unmanaged                                50.0GB
0000000000000003 ITSO_EXT2A
600507680282818b3000000000000011000000000000000000000000000000000 enterprise
14 md_v7kgen1-2-005 online unmanaged                                50.0GB
0000000000000004 ITSO_EXT2A
600507680282818b3000000000000012000000000000000000000000000000000 enterprise
15 md_v7kgen1-2-006 online unmanaged                                50.0GB
0000000000000005 ITSO_EXT2A
600507680282818b3000000000000013000000000000000000000000000000000 enterprise
16 md_v7kgen1-2-007 online unmanaged                                50.0GB
0000000000000006 ITSO_EXT2A
600507680282818b3000000000000014000000000000000000000000000000000 enterprise
17 md_v7kgen1-2-008 online unmanaged                                50.0GB
0000000000000007 ITSO_EXT2A
600507680282818b3000000000000015000000000000000000000000000000000 enterprise
18 md_v7kgen1-2-009 online unmanaged                                50.0GB
0000000000000008 ITSO_EXT2A
600507680282818b3000000000000016000000000000000000000000000000000 enterprise
19 md_v7kgen1-2-010 online unmanaged                                50.0GB
0000000000000009 ITSO_EXT2A
600507680282818b3000000000000017000000000000000000000000000000000 enterprise
20 md_v7kgen1-2-011 online unmanaged                                50.0GB
000000000000000A ITSO_EXT2A
600507680282818b3000000000000018000000000000000000000000000000000 enterprise
21 md_v7kgen1-2-012 online unmanaged                                50.0GB
000000000000000B ITSO_EXT2A
600507680282818b3000000000000019000000000000000000000000000000000 enterprise
22 md_v7kgen1-2-013 online unmanaged                                50.0GB
000000000000000C ITSO_EXT2A
600507680282818b300000000000001a000000000000000000000000000000000 enterprise
23 md_v7kgen1-2-014 online unmanaged                                50.0GB
000000000000000D ITSO_EXT2A
600507680282818b300000000000001b000000000000000000000000000000000 enterprise
24 md_v7kgen1-2-015 online unmanaged                                50.0GB
000000000000000E ITSO_EXT2A
600507680282818b300000000000001c000000000000000000000000000000000 enterprise
25 md_v7kgen1-2-016 online unmanaged                                50.0GB
000000000000000F ITSO_EXT2A
600507680282818b300000000000001d000000000000000000000000000000000 enterprise
IBM_Storwize:ITSO_V7000Gen2:admin>
```

In our setup, we have mapped a total of 20 volumes from the two Storwize V7000 making them visible from the Storwize V7000 Gen2. As shown in Example 8-13, the external managed disks are visible, but unmanaged.

4. At this point, we create two new storage pools to hold our external storage and then add the managed disks. Use the **mkmdiskgrp** and **addmdisk**, as shown in Example 8-13.

Example 8-13 The mkmdiskgrp command

```
IBM_Storwize:ITSO_V7000Gen2:admin>mkmdiskgrp -name EXTSTG1 -ext 256 -tier
enterprise
MDisk Group, id [2], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkmdiskgrp -name EXTSTG2 -ext 256 -tier
enterprise
MDisk Group, id [1], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>
IBM_Storwize:ITSO_V7000Gen2:admin>addmdisk -mdisk 5:6:7:8:9 2
IBM_Storwize:ITSO_V7000Gen2:admin>addmdisk -mdisk 10:11:12:13:14:15 1
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

This concludes our setup of external storage.

8.2.5 Configuring volumes

In this topic we will describe how to set up and work with volumes (also known as virtual disks (VDisks), so “vdisk” in the commands). These are also referred to as logical unit numbers (LUNs). Because our lab setup consists of a Storwize V7000 Gen2 and two Storwize V7000 Gen1 virtualized behind it, we will show you how to create a volume mirror between them:

1. To start with, use the **mkvdisk** command to create the mirror volume, as seen in Example 8-14.

Example 8-14 The mkvdisk mirror volume

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkvdisk -copies 2 -iogrp io_grp0 -mdiskgrp
INT_V7KGEN2:EXTSTG1 -name test_mirror -size 5120
Virtual Disk, id [0], successfully created
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

We now have a 5 GB volume mirror between the IBM Storwize V7000 Gen2 and a virtualized external storage system. This is useful in many cases, for example for backup purposes. You can run your backup from the external storage without affecting performance on your primary storage.

2. Create two more volumes using **mkvdisk**, but this time we only want them to reside on our Storwize V7000 Gen2 (Example 8-15).

Example 8-15 The mkvdisk volumes

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkvdisk -iogrp io_grp0 -mdiskgrp
INT_V7KGEN2 -name test_vol1 -size 5120
Virtual Disk, id [1], successfully created

IBM_Storwize:ITSO_V7000GEN2_2:admin>mkvdisk -iogrp io_grp0 -mdiskgrp
INT_V7KGEN2 -name test_vol2 -size 5120
Virtual Disk, id [2], successfully created
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

3. Now, assume that one of our volumes is running out of space and we want to expand it with 5 GB more storage. To do this, use the **expandvdisksize** command (Example 8-16).

Example 8-16 The expandvdisksize command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>expandvdisksize -size 5120 test_vol1
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

Important: Be careful when expanding volumes. You should always make sure that the operating system accessing the volume supports volume expansion at the storage layer. Check compatibility with your operating system vendor.

8.2.6 Configuring remote copy

In this topic, we demonstrate the remote copy functions of the IBM Storwize V7000 Gen2. We are replicating to the first generation of Storwize V7000. There are many considerations to be taken into account when configuring remote copy. We only list some of the important ones:

- Know your current SAN / IP network use. As a rule, every write on your primary volume will eventually be written to the target volume.
- Decide which kind of mirror you want to use:
  - Metro Mirror: Synchronized mirror (highest bandwidth)
  - Global Mirror: Asynchronized mirror
  - Global Mirror with Change Volumes: Periodic asynchronized mirror (lowest bandwidth)

To learn more about replication, we suggest the following books:

IBM System Storage SAN Volume Controller and Storwize V7000 Replication Family Services, SG24-7574:

http://www.redbooks.ibm.com/redbooks/pdfs/sg247574.pdf

IBM SAN Volume Controller and Storwize Family Native IP Replication, REDP-5103

http://www.redbooks.ibm.com/redpieces/pdfs/redp5103.pdf

Remember: When replicating to or virtualizing to Storwize family products, you should be aware of the system layers. There are two types of layers: Storage and Replication, and these affect how Storwize and IBM SAN Volume Controller interact with each other. This system layer does not apply when replicating to or virtualizing other vendor storage.

Our lab setup includes a Storwize V7000 Gen2 and two Storwize V7000 and we show how to create a Metro Mirror Remote Copy relationship:

1. First, check if we have some partnership candidates using the **lspartnershipcandidate** command, as shown in Example 8-17.

Example 8-17 The lspartnershipcandidate command

```
IBM_Storwize:ITSO_V7000Gen2:admin>lspartnershipcandidate
id                  configured name
00000200A0A062CC no         EXTSTG2
00000200A260009A no         EXTSTG1
IBM_Storwize:ITSO_V7000Gen2:admin>
```

2. Now we create a partnership between our Storwize V7000 Gen2 and our partner candidate with the **mkfcpartnership** command (Example 8-18).

Example 8-18 The mkfcpartnership command

```
IBM_Storwize:ITSO_V7000Gen2:admin>mkfcpartnership -linkbandwidthmbits 1024
-backgroundcopyrate 25 EXTSTG1

IBM_Storwize:ITSO_V7000Gen2:admin>lspartnership
id               name           location partnership               type
cluster_ip event_log_sequence
0000010020A4001E0 ITSO_V7000Gen2 local
0000020A260009A EXTSTG1        remote   partially_configured_local fc
IBM_Storwize:ITSO_V7000Gen2:admin>
```

3. At this stage, our partnership is partially configured. This is because the **mkfcpartnership** command must also be run on the secondary system (Example 8-19).

Example 8-19 The mkfcpartnership command

```
IBM_Storwize:EXTSTG1:superuser>mkfcpartnership -linkbandwidthmbits 1024
-backgroundcopyrate 25 ITSO_V7000Gen2

IBM_Storwize:EXTSTG1:superuser>lspartnership
id               name           location partnership      type cluster_ip
event_log_sequence
0000020A260009A EXTSTG1        local
0000010020A4001E0 ITSO_V7000Gen2 remote   fully_configured fc
IBM_Storwize:EXTSTG1:superuser>
```

4. This gives us a fully configured partnership between our two Storwizes, and now we just need to create the volumes to be mirrored. For this operation, we use the **mkvdisk** command (Example 8-20) to create two volumes with the same characteristics, one on each system.

Example 8-20 The mkvdisk command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 5120
-name V7K2G2_RCVOL
Virtual Disk, id [0], successfully created
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

5. After the volumes are created, we issue the **mkrcrelationship** command (Example 8-21). This initiates a Metro Mirror relationship between the two Storwize.

Example 8-21 The mkrcrelationship command as Metro Mirror

```
IBM_Storwize:ITSO_V7000Gen2:admin>mkrcrelationship -aux EXTSTG1_RCVOL -cluster
0000020A260009A -master V7KG2_RCVOL
RC Relationship, id [0], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>
```

6. Here is an example of how to make a Global Mirror with Change Volumes (Example 8-22). This type of remote copy uses the least bandwidth, because writes are accumulated on the source and written to target in specified cycles, this is also the suggested setting for * IP replication.

Example 8-22 The mkrcrelationship command as Global Mirror with change volumes

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>mkrcrelationship -aux EXTSTG1_RCVOL
-cluster 00000200A260009A -master V7KG2_RCVOL -global -cyclingmode
RC Relationship, id [0], successfully created
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

8.2.7 Configuring hosts

In this topic, we show how to configure hosts to the IBM Storwize V7000 Gen2. There are some prerequisites that you should be aware of:

- Hosts must be zoned so that they are visible to the Storage on the SAN. The leading practice for this includes the following attributes:
  - Use of a dual fabric for redundancy
  - Single initiator zoning = one host port per zone
  - Modern servers and storage systems typically include many available Fibre Channel (FC) ports. Do not zone too many ports from a single host to a storage device, and vice versa. Performance can be severely affected if the multipathing software has to keep track of too many links.
- Plan for bandwidth, and use and monitor it (for example, with IBM Tivoli Storage Productivity Center for Replication for Storage)

For our lab setup, we have two hosts running VMware, each with a dual-port host bus adapter (HBA) and an IBM Storwize V7000 Gen2. Assuming that host installation and SAN zoning have taken place, follow these steps to configure the host:

1. we start by issuing the **lsfcportcandidate**, which gives us information about open FC ports on the SAN (Example 8-23).

Example 8-23 The lsfcportcandidate command

```
IBM_Storwize:ITSO_V7000Gen2:admin>lsfcportcandidate
fc_WWPN
100000051EC76B9C
100000051EC76B92
100000051EC76B9B
100000051EC76B91
IBM_Storwize:ITSO_V7000Gen2:admin>
```

2. In our case, this reveals four open FC ports, two for each of our hosts. With this information, we create our hosts using the **mkhost** command (Example 8-24).

Example 8-24 The mkhost command

```
IBM_Storwize:ITSO_V7000Gen2:admin>mkhost -name VMWare1 -fcwwpn
100000051EC76B91:100000051EC76B92 -type generic
Host, id [0], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkhost -name VMWare2 -fcwwpn
100000051EC76B9b:100000051EC76B9c -type generic
Host, id [1], successfully created
```

3. To get a concise or detailed view of the hosts, use the **lshost** command (Example 8-25).

Example 8-25 The lshost command

```
IBM_Storwize:ITSO_V7000Gen2:admin>lshost
id name    port_count iogrp_count status
0  VMWare1 2          4           online
1  VMWare2 2          4           online

IBM_Storwize:ITSO_V7000Gen2:admin>lshost 0
id 0
name VMWare1
port_count 2
type generic
mask 1111111111111111111111111111111111111111111111111111111111111111
iogrp_count 4
status online
WWPN 100000051EC76B92
node_logged_in_count 2
state active
WWPN 100000051EC76B91
node_logged_in_count 2
state active
IBM_Storwize:ITSO_V7000Gen2:admin>
```

4. This shows us our two configured hosts, which are ready for LUN mapping. Use the **lsvdisk** command to show available volumes, and continue with the **mkvdiskhostmap** command to map them (Example 8-26).

Example 8-26 The mkvdiskhostmap command

```
IBM_Storwize:ITSO_V7000Gen2:admin>mkvdiskhostmap -host VMWare1 VMWvol1
Virtual Disk to Host map, id [0], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkvdiskhostmap -host VMWare1 VMWvol2
Virtual Disk to Host map, id [1], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkvdiskhostmap -host VMWare1 temp
Virtual Disk to Host map, id [2], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkvdiskhostmap -host VMWare2 VMWvola
Virtual Disk to Host map, id [0], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkvdiskhostmap -host VMWare2 VMWvolb
Virtual Disk to Host map, id [1], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>mkvdiskhostmap -host VMWare2 temp1
Virtual Disk to Host map, id [2], successfully created
IBM_Storwize:ITSO_V7000Gen2:admin>
```

The provided storage should now be visible from the hosts, so this topic is concluded.

8.3 New Commands introduced with the IBM Storwize V7000 Gen2

In this topic, we look at the new commands available for the IBM Storwize V7000 Gen2. Several other commands were added with code release 7.3, but they are mainly reserved for the new IBM SAN Volume Controller model DH8. For that reason, we focus on new commands specific to the IBM Storwize V7000 Gen2.

The following new commands are available:

- **lsenclosurefanmodule**
- **lsenclosurebattery**

lsenclosurefanmodule

This command gives us a concise or detailed status of the new fan modules that are installed in the V7000 Storwize Gen2 (Example 8-27).

Example 8-27 The lsenclosurefanmodule command

```
IBM_Storwize:ITSO_V7000GEN2_2:admin>lsenclosurefanmodule
enclosure_id fan_module_id status
1            1             online
1            2             online
IBM_Storwize:ITSO_V7000GEN2_2:admin>lsenclosurefanmodule -fanmodule 1 1
enclosure_id 1
fan_module_id 1
status online
error_sequence_number
FRU_part_number 31P1847
FRU_identity 11S31P1846YM11BG42M08L
IBM_Storwize:ITSO_V7000GEN2_2:admin>
```

lsenclosurebattery

This command gives us a concise or detailed view of the canister batteries (Example 8-28).

Example 8-28 The lsenclosurebattery command

```
IBM_Storwize:ITSO_V7000Gen2:admin>lsenclosurebattery
enclosure_id battery_id status charging_status recondition_needed percent_charged
end_of_life_warning
1            1          online idle            no                 100
no
1            2          online idle            no                 100
no
IBM_Storwize:ITSO_V7000Gen2:admin>lsenclosurebattery -battery 2 1
enclosure_id 1
battery_id 2
status online
charging_status idle
recondition_needed no
percent_charged 100
end_of_life_warning no
FRU_part_number 31P1807
FRU_identity 11S00AR085YM30BG43K0A1
firmware_level 105:0
error_sequence_number
remaining_charge_capacity_mAh 3508
full_charge_capacity_mAh 3846
compatibility_level 1
last_recondition_timestamp 140817153801
powered_on_hours 347
cycle_count 0
IBM_Storwize:ITSO_V7000Gen2:admin>
```

This concludes the command-line interface chapter, and you should now be familiar with the basic commands. As stated in the beginning of this chapter, the CLI is a powerful and, when learned, easy-to-use tool that offers even more functionality than the GUI. If you do want to learn about the advanced commands, see the following IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/ST3FR7_7.3.0/com.ibm.storwize.v7000.730.doc/v7000_ichome_730.html

9

IBM Storwize V7000 Gen2 operations using the GUI

In this chapter, we illustrate selected IBM Storwize V7000 Gen2 operational management and system administration using the IBM Storwize V7000 graphical user interface (GUI). The IBM Storwize V7000 management GUI is an easy-to-use tool that helps you to monitor, manage, and configure your system.

The information is presented at a high level, because this book is based on the new hardware, and is not intended to be an in-depth coverage of every aspect of the software. For more detailed information about using the GUI, see *Implementing the IBM System Storage SAN Volume Controller V7.2*, SG24-7933.

Although the IBM Storage Tier Advisor Tool (STAT) is not part of the GUI, it is a strong and useful tool to determine the use of your tiered storage, as the 7.3 code level now supports three-tiered storage using the IBM Easy Tier functionality.

Important: It is possible for more than one user to be logged in to the GUI at any given time. However, no locking mechanism exists, so be aware that if two users change the same object at the same time, the last action entered from the GUI is the one that takes effect.

Data entries made through the GUI *are* case-sensitive.

9.1 Introduction to IBM Storwize V7000 GUI

The Storwize V7000 Overview pane is the first window you will see after you have logged into the system. It is an important user interface, and throughout this chapter we refer to it as the IBM Storwize V7000 Overview pane, or just the Overview pane.

The following steps illustrate how to start the Storwize V7000 GUI:

1. Initially, to log on to the management software, type the IP address that was set during the initial setup process into the address line of your web browser. You can connect from any workstation that can communicate with the system.
2. You start at the login window, as shown in Figure 9-1.

Figure 9-1 Login window

Dynamic menu

From any page inside the IBM Storwize V7000 GUI, you always have access to the dynamic menu. The IBM Storwize V7000 GUI dynamic menu is on the left side of the IBM Storwize V7000 GUI window. To navigate using this menu, move the mouse cursor over the various icons, and choose a page that you want to display.

The IBM Storwize V7000 dynamic menu consists of multiple panes. These panes group common configuration and administration objects, and presents individual administrative objects to the IBM Storwize V7000 GUI users.

9.1.1 Overview

After you have successfully logged in to the system, you start at the overview window as shown in Figure 9-2. The overview provides the user with a quick summary of the system, and a link to the IBM Storwize V7000 Knowledge Center:

http://www-01.ibm.com/support/knowledgecenter/ST3FR7/landing/V7000_welcome.html

Figure 9-2 Overview

9.1.2 Monitoring

Figure 9-3 shows the Monitoring menu where you can work with the following details:

- Information about the code level
- Hardware configuration

  See installed hardware and change memory allocation (also known as *bitmap allocation*).
- Events

  See warnings and alerts, and run the maintenance procedure.
- Real-time performance graphs

  See central processing unit (CPU) usage input/output operations per second (IOPS) for volumes, managed disks (MDisks), and so on.

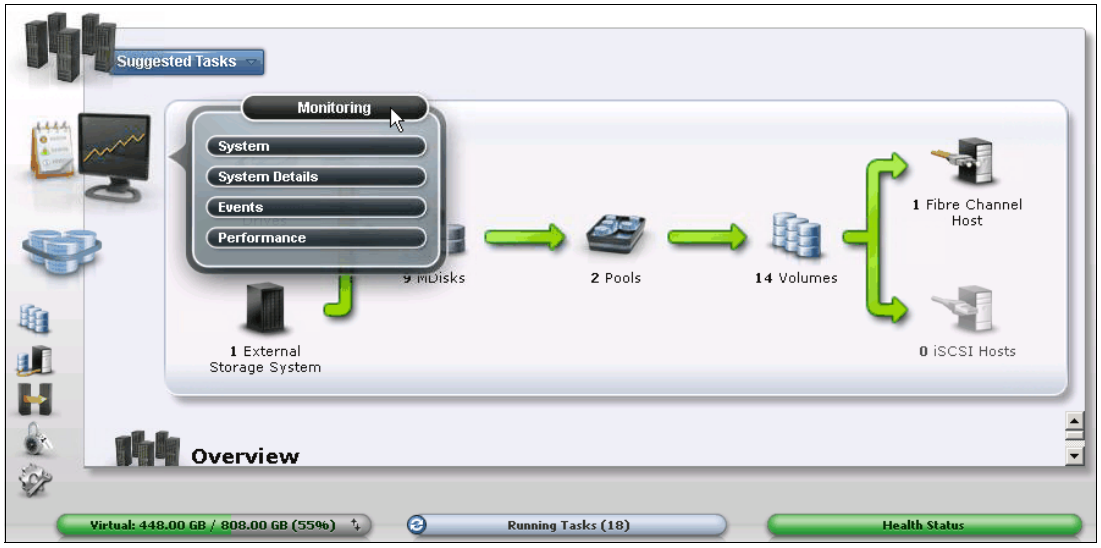


Figure 9-3 Monitoring

9.1.3 Pools

In the Pools menu, shown in Figure 9-4, you can administer everything that is related to storage pools, from creating new storage pools to working with internal or external storage.

Storage pool balancing is introduced in code level 7.3, which means that if you add new or additional MDisks to an existing pool, it balances the extents across all of the MDisks in a pool. Before release 7.3, you had to do this manually, or use a script to balance extents after adding new MDisks to an existing pool. Note that this is an automated process, and it is not configurable.

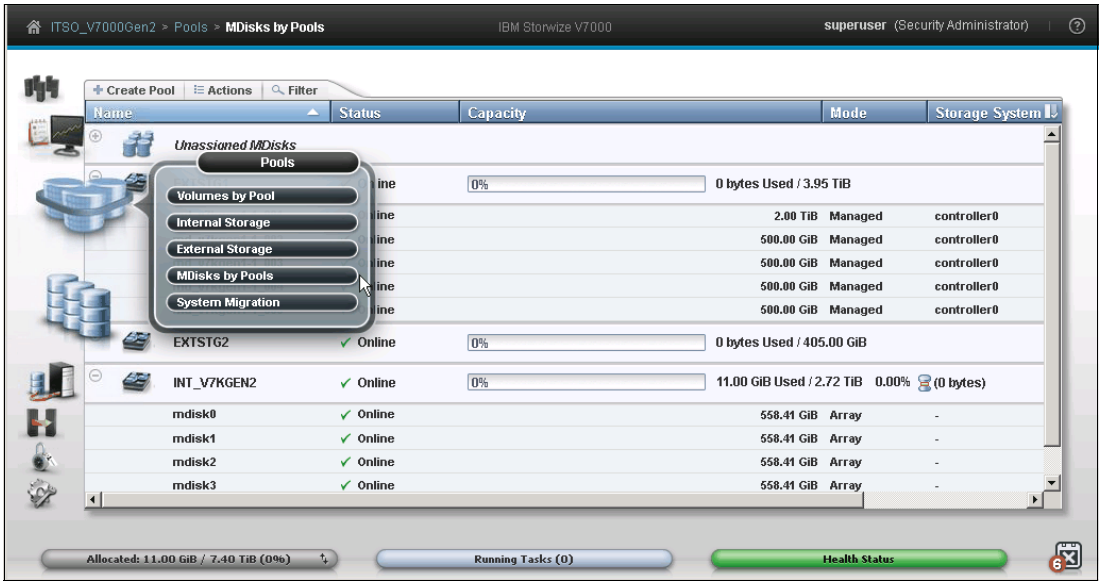


Figure 9-4 Pools

9.1.4 Volumes

The Volumes menu contains the following administrative options:

- View volumes, create volumes, and delete volumes.
- See details about volumes, if they are mapped or unmapped to a host.
- See details about volumes mapped to host.

Figure 9-5 shows the Volume menu expanded.

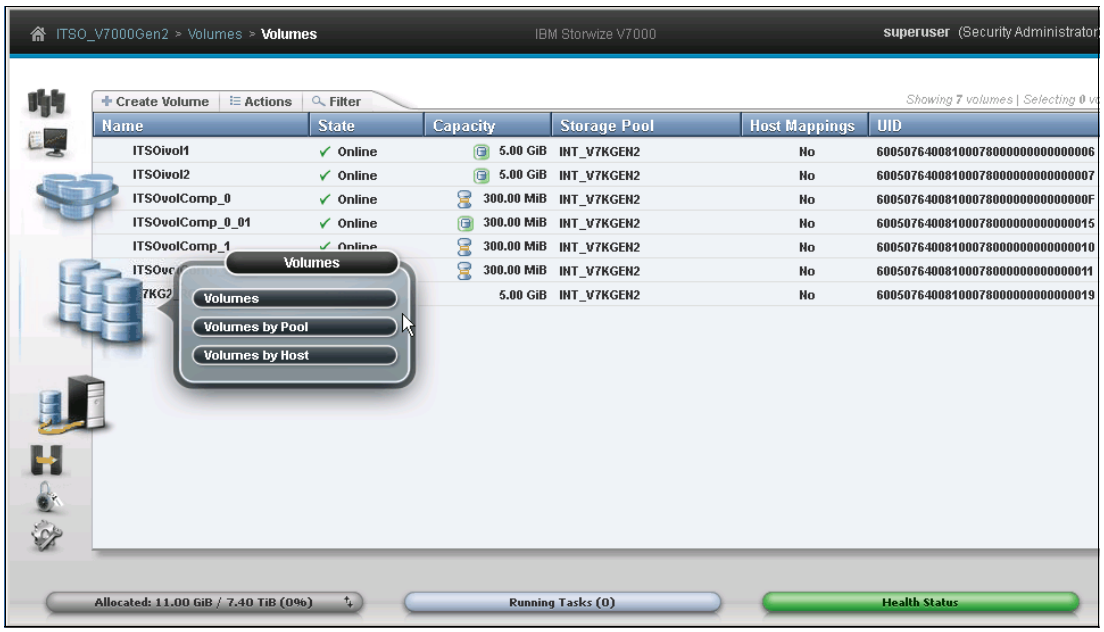


Figure 9-5 Volumes

9.1.5 Hosts

All host-related administration can be run from this menu:

- Creating and deleting hosts
- Viewing Details about each host, such as ports and worldwide port names
- Volume mappings for each host

Figure 9-6 shows the expanded view of the Hosts menu.

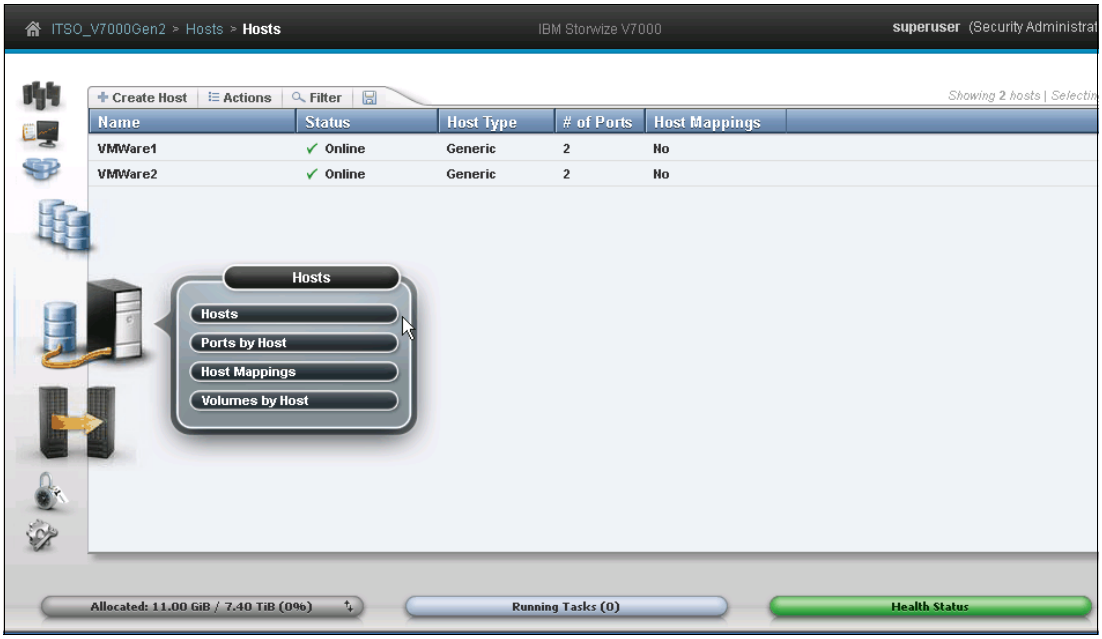


Figure 9-6 Hosts

9.1.6 Copy services

In the Copy Services menu, you can administer all copy services related activities:

- Create partnerships with other IBM SAN Volume Controller and Storwize systems.
- Create and delete Metro Mirrored volumes.
- Create and delete Global Mirrored volumes.
- Create and delete IBM FlashCopy volumes.
- View details about the copy services configured.

Figure 9-7 shows the expanded Copy Services menu.

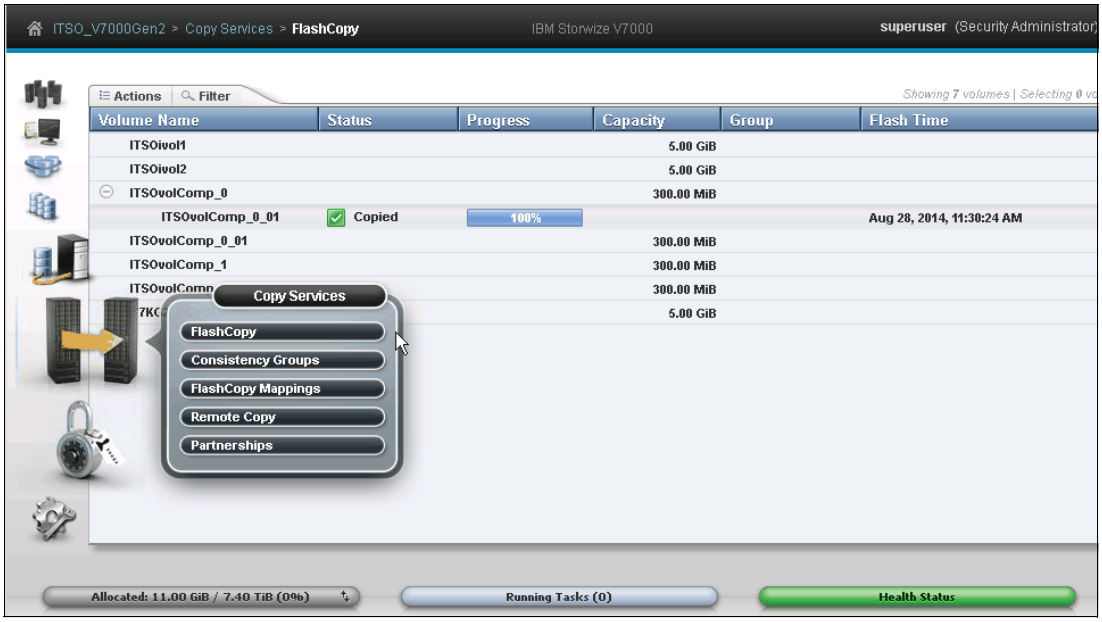


Figure 9-7 Copy Services

9.1.7 Access

In the Access menu, you can work with user administration, such as create users, delete users, or perform user group administration. Furthermore, you can see the audit log, which shows all of the activities that have been run on the system.

Figure 9-8 shows the expanded view of the Access menu.

Figure 9-8 Access

9.1.8 Settings

In the Settings menu, you have access to the following activities:

- Event notifications, such as call home (using email), Simple Network Management Protocol (SNMP), Simple Mail Transfer Protocol (SMTP), and syslog
- Directory Services, for enabling remote authentication of users
- Network, both Fibre Channel (FC) settings and Internet Protocol (IP) settings
- Support, where you can manage dumps, snaps, heatmap files, and so on
- General, where you can upgrade the system, time/date settings, and so on

Figure 9-9 shows an expanded view of the Settings menu.

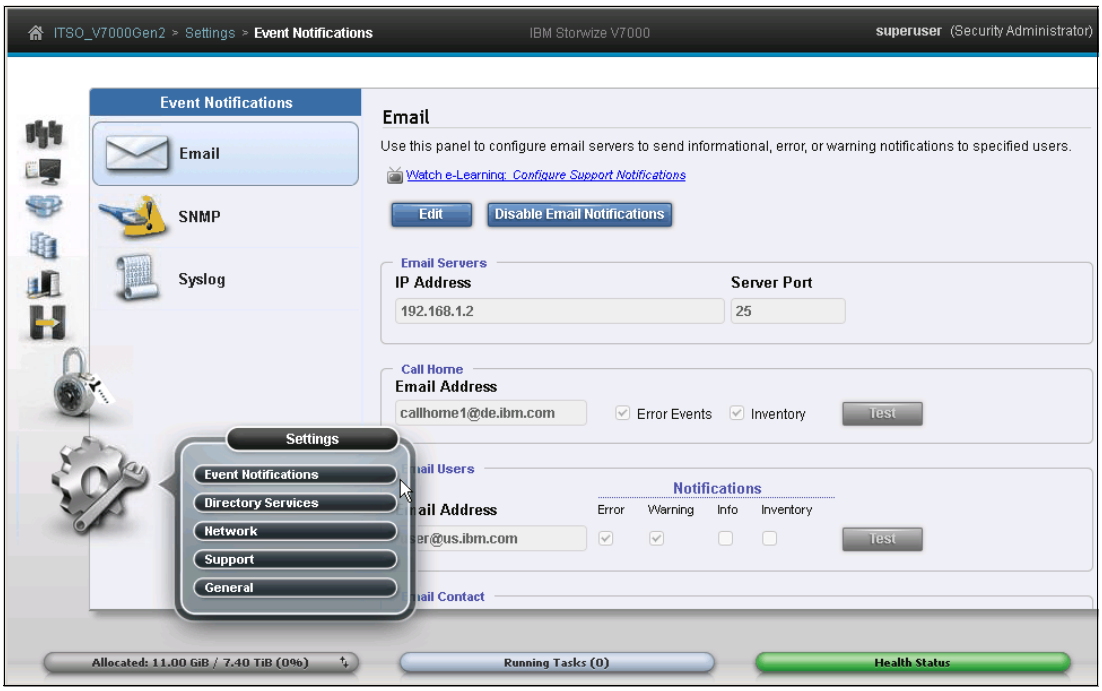


Figure 9-9 Settings

9.2 IBM Storage Tier Advisor Tool

This section shows how to use the Storage Tier Advisor Tool, also known as the STAT utility.

Note: The STAT utility is not a part of the GUI, but can be downloaded from IBM support:

`http://www.ibm.com/support/docview.wss?uid=ssg1S4000935`

STAT uses limited storage performance measurement data from a user's operational environment to model potential unbalanced workload (also known as *skew*) on disk and array resources. It is intended to supplement and support, but not replace, detailed pre-installation sizing and planning analysis.

It is most useful to obtain a broad system-wide performance projection of cumulative latency reduction on arrays and disks when a solid-state drive (SSD) configuration and the IBM Easy Tier function are used in combination to handle workload growth or skew management.

The “hot data” identification methodology in the tool is an engineering estimation based on expected cumulative latency reduction if the suggested SSD configuration is used with the measured workload and storage configuration. Care has been taken in the development of this tool, but the accuracy of any prediction of performance improvement is subject to a variety of storage system configurations, conditions, and other variables beyond the scope of this tool. Accordingly, actual results might vary.

The **STAT.exe** command creates a Hypertext Markup Language (HTML) report of the input/output (I/O) distribution. IBM Storwize V7000 input files are found under `/dumps` on the configuration node, and are named `dpa_heat.<node_name>.<time_stamp>.data`. The file must be off-loaded manually using the command-line interface (CLI) or GUI.

You can install the STAT tool on any Windows-based PC or notebook, and you don’t need to have direct access to the IBM SAN Volume Controller.

When the STAT tool is installed, it’s time to off-load or download heat files from your IBM Storwize V7000 system.

The next few screen captures show how you can download the heat files from the GUI. The heat files can also be downloaded using the CLI or PuTTY Secure Copy (PSCP). However, we show how to off-load or download these files using the GUI:

1. Log in to the GUI and select **Settings** → **Support** and click the **Show full log listing** link, as shown in Figure 9-10.

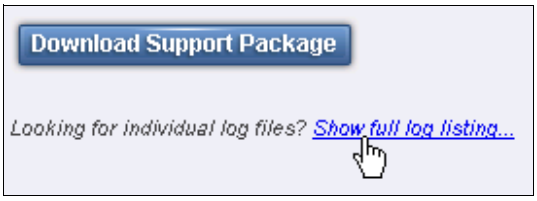


Figure 9-10 Show full log listing

2. Now you can select the heat files that you want to use for the STAT tool. Select the files, right-click, and select Download, as shown in Figure 9-11.

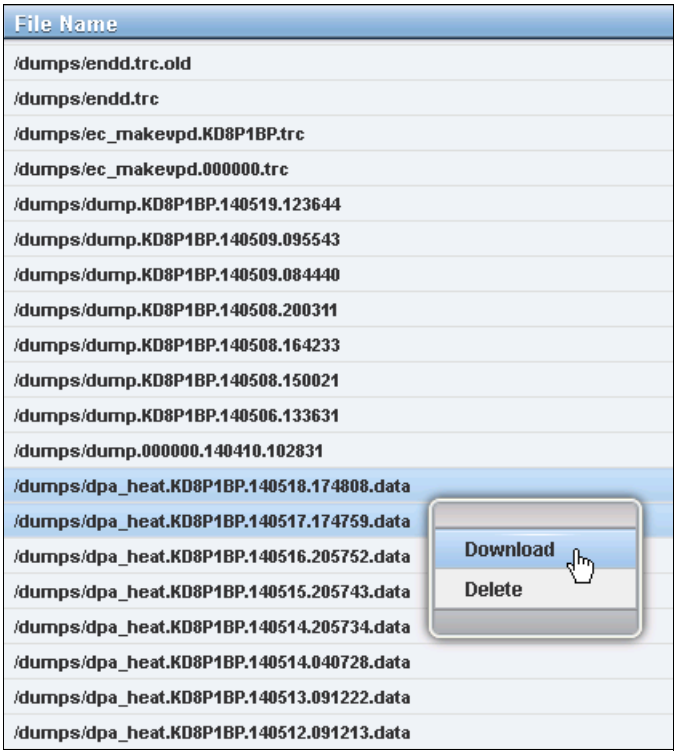


Figure 9-11 Download heat files

3. When the files have been off-loaded or downloaded, open a command prompt and go to the directory where you have installed the STAT tool (the default path on a 64-bit windows operating system is `C: Program Files (x86)\IBM\STAT`).

> **Note:** If the config node of the system reboots, asserts, and so on, note that the new config node starts the Easy Tier heatmap cycle count from `0`, which means that it takes 24 hours until you see a new heatmap file in the `/dumps` directory.

You might want to copy/move the off-loaded/downloaded files to the directory where you have installed the STAT tool for ease of the usage. Otherwise, you have to define the entire input file path every time you create a report.

4. To generate the report, (in this case we have already copied the input file to the STAT directory), run the following command (one line):

```
stat.exe -o "c:\Program Files
(x86)\IBM\STAT\ITSO_V7KGen2"dpa_heat.KD8P1BP.140518.174808.data
```

Replace the heat files with the correct file names of the ones that you have off-loaded or downloaded. For IBM Storwize V7000 Systems, you can only run one file concurrently.

Be aware that in this scenario we have used the syntax **`-o`**, which specifies an output path (a folder). This is useful if you are generating STAT files from more than one system.

> **Tip:** For detailed information about the usage of the STAT tool, see the `readme` file for the tool that is contained within the same directory where you installed it.

5. The following message is returned:

   CMUA00019I The STAT.exe command has completed.

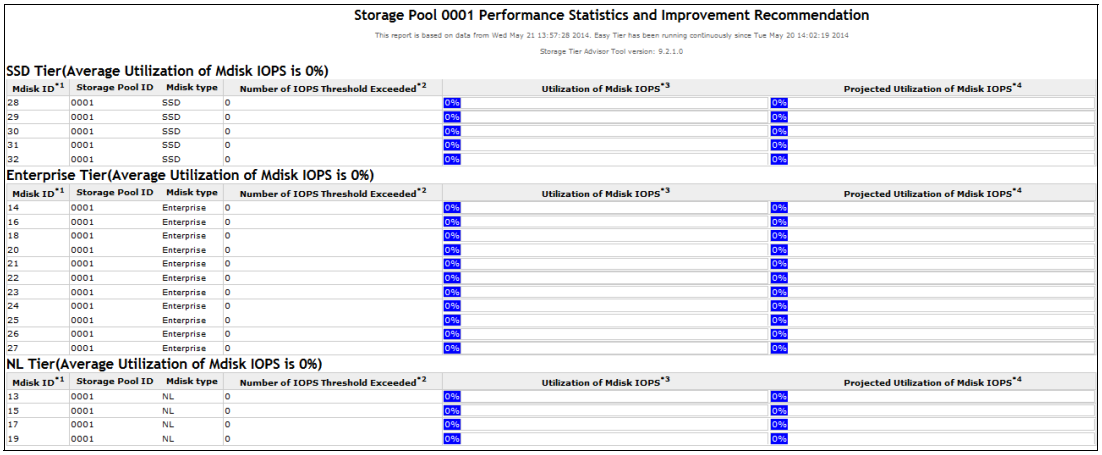
   Press any key to continue . . .

   Go to the directory where the report was generated and open the `index.html` file: In this case, the following location is the directory:

   `C:\Program Files (x86)\IBM\STAT\ITSO_V7KGen2`

   Your default browser opens the report, or you can simply open your browser and navigate to the `index.html` file.

6. You can click any of the Storage Pool IDs, which action opens the Performance Statistics and Improvement Recommendation for the selected Storage Pool ID, which we have done in Figure 9-12, with Pool ID named P1. The Pool ID **P1** is based on a three-tiered pool, consisting of flash drives (SSD), enterprise drives, and Nearline drives.



Storage Pool 0001 Performance Statistics and Improvement Recommendation

This report is based on data from Wed May 21 13:57:28 2014. Easy Tier has been running continuously since Tue May 20 14:02:19 2014

Storage Tier Advisor Tool version: 9.2.1.0

SSD Tier(Average Utilization of Mdisk IOPS is 0%)

Mdisk ID*1	Storage Pool ID	Mdisk type	Number of IOPS Threshold Exceeded*2	Utilization of Mdisk IOPS*3	Projected Utilization of Mdisk IOPS*4
28	0001	SSD	0	0%	0%
29	0001	SSD	0	0%	0%
30	0001	SSD	0	0%	0%
31	0001	SSD	0	0%	0%
32	0001	SSD	0	0%	0%

Enterprise Tier(Average Utilization of Mdisk IOPS is 0%)

Mdisk ID*1	Storage Pool ID	Mdisk type	Number of IOPS Threshold Exceeded*2	Utilization of Mdisk IOPS*3	Projected Utilization of Mdisk IOPS*4
14	0001	Enterprise	0	0%	0%
16	0001	Enterprise	0	0%	0%
18	0001	Enterprise	0	0%	0%
20	0001	Enterprise	0	0%	0%
21	0001	Enterprise	0	0%	0%
22	0001	Enterprise	0	0%	0%
23	0001	Enterprise	0	0%	0%
24	0001	Enterprise	0	0%	0%
25	0001	Enterprise	0	0%	0%
26	0001	Enterprise	0	0%	0%
27	0001	Enterprise	0	0%	0%

NL Tier(Average Utilization of Mdisk IOPS is 0%)

Mdisk ID*1	Storage Pool ID	Mdisk type	Number of IOPS Threshold Exceeded*2	Utilization of Mdisk IOPS*3	Projected Utilization of Mdisk IOPS*4
13	0001	NL	0	0%	0%
15	0001	NL	0	0%	0%
17	0001	NL	0	0%	0%
19	0001	NL	0	0%	0%

Figure 9-12 Performance Statistics and Improvement Recommendation

7. Further details can be seen in the lower section of the Performance Statistics and Improvement Recommendation page, where you can expand hyperlinks for the following information:
   - Workload Distribution Across Tiers
   - Recommended NL Configuration
   - Volume Heat Distribution

More details for planning and configuration are available in the following IBM Redbooks publications:

- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072

  http://www.redbooks.ibm.com/abstracts/tips1072.html

- *IIBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines, SG24-7521*

  http://www.redbooks.ibm.com/redpieces/pdfs/sg247521.pdf

- *IBM DS8000 Easy Tier*, REDP-4667

  http://www.redbooks.ibm.com/abstracts/redp4667.html?Open

This is described in more detail in the Chapter 4, “IBM Storwize V7000 Gen2 Easy Tier” on page 85.

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only:

- *Implementing the IBM System Storage SAN Volume Controller V7.2*, SG24-7933
- *Implementing the IBM Storwize V7000 V7.2*, SG24-7938
- *IBM b-type Gen 5 16 Gbps Switches and Network Advisor*, SG24-8186
- *Introduction to Storage Area Networks and System Networking*, SG24-5470
- *IBM SAN Volume Controller and IBM FlashSystem 820: Best Practices and Performance Capabilities*, REDP-5027
- *Implementing the IBM SAN Volume Controller and FlashSystem 820*, SG24-8172
- *Implementing IBM FlashSystem 840*, SG24-8189
- *IBM FlashSystem in IBM PureFlex System Environments*, TIPS1042
- *IBM FlashSystem 840 Product Guide*, TIPS1079
- *IBM FlashSystem 820 Running in an IBM StorwizeV7000 Environment*, TIPS1101
- *Implementing FlashSystem 840 with SAN Volume Controller*, TIPS1137
- *IBM FlashSystem V840*, TIPS1158
- *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363
- *IBM System Storage b-type Multiprotocol Routing: An Introduction and Implementation*, SG24-7544
- *IBM Tivoli Storage Area Network Manager: A Practical Introduction*, SG24-6848
- *Tivoli Storage Productivity Center for Replication for Open Systems*, SG24-8149
- *IBM Tivoli Storage Productivity Center V5.2 Release Guide*, SG24-8204
- *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116

You can search for, view, download, or order these documents and other Redbooks publications, Redpapers publications, Web Docs, drafts, and additional materials, from the following website:

ibm.com/redbooks

Other resources

These publications are also relevant as further information sources:

- *IBM System Storage Master Console: Installation and User's Guide*, GC30-4090
- *IBM System Storage Open Software Family SAN Volume Controller: CIM Agent Developers Reference*, SC26-7545
- *IBM System Storage Open Software Family SAN Volume Controller: Command-Line Interface User's Guide*, SC26-7544
- *IBM System Storage Open Software Family SAN Volume Controller: Configuration Guide*, SC26-7543
- *IBM System Storage Open Software Family SAN Volume Controller: Host Attachment Guide*, SC26-7563
- *IBM System Storage Open Software Family SAN Volume Controller: Installation Guide*, SC26-7541
- *IBM System Storage Open Software Family SAN Volume Controller: Planning Guide*, GA22-1052
- *IBM System Storage Open Software Family SAN Volume Controller: Service Guide*, SC26-7542
- *IBM System Storage SAN Volume Controller - Software Installation and Configuration Guide,* SC23-6628
- *IBM System Storage SAN Volume Controller V6.2.0 - Software Installation and Configuration Guide,* GC27-2286

  http://pic.dhe.ibm.com/infocenter/svc/ic/topic/com.ibm.storage.svc.console.doc/svc_bkmap_confguidebk.pdf
- *IBM System Storage SAN Volume Controller 6.2.0 Configuration Limits and Restrictions,* S100-3799
- *IBM TotalStorage Multipath Subsystem Device Driver User's Guide*, SC30-4096
- *IBM XIV and SVC/ Best Practices Implementation Guide*

  http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD105195
- *Considerations and Comparisons between IBM SDD for Linux and DM-MPIO*

  http://www.ibm.com/support/docview.wss?rs=540&context=ST52G7&q1=linux&uid=ssg1S7001664&loc=en_US&cs=utf-8&lang=en

Referenced websites

These websites are also relevant as further information sources:

- IBM Storage home page

  http://www.storage.ibm.com

- IBM site to download SSH for AIX

  http://oss.software.ibm.com/developerworks/projects/openssh

- IBM Tivoli Storage Area Network Manager site

  http://www-306.ibm.com/software/sysmgmt/products/support/IBMTivoliStorageAreaNetworkManager.html

- SAN Volume Controller supported platform

  http://www-1.ibm.com/servers/storage/support/software/sanvc/index.html

- SAN Volume Controller Information Center

  http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp

- Cygwin Linux-like environment for Windows

  http://www.cygwin.com

- Open source site for SSH for Windows and Mac

  http://www.openssh.com/windows.html

- Sysinternals home page

  http://www.sysinternals.com

- Subsystem Device Driver download site

  http://www-1.ibm.com/servers/storage/support/software/sdd/index.html

- Download site for Windows SSH freeware

  http://www.chiark.greenend.org.uk/~sgtatham/putty

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

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

Implementing the IBM Storwize V7000 Gen2

Learn about the latest addition to the IBM SAN Volume Controller/Storwize family

Understand the new functions and features

Benefit from an uncomplicated implementation

Data is the new currency of business, the most critical asset of the modern organization. In fact, enterprises that can gain business insights from their data are twice as likely to outperform their competitors. Nevertheless, 72% of them have not started, or are only planning, big data activities. In addition, organizations often spend too much money and time managing where their data is stored. The average firm purchases 24% more storage every year, but uses less than half of the capacity that it already has.

The IBM Storwize family, including the IBM SAN Volume Controller Data Platform, is a storage virtualization system that enables a single point of control for storage resources. This functionality helps support improved business application availability and greater resource use. The following list describes the business objectives of this system:

- To manage storage resources in your information technology (IT) infrastructure
- To make sure that those resources are used to the advantage of your business
- To do it quickly, efficiently, and in real time, while avoiding increases in administrative costs

Virtualizing storage with Storwize helps make new and existing storage more effective. Storwize includes many functions traditionally deployed separately in disk systems. By including these functions in a virtualization system, Storwize standardizes them across virtualized storage for greater flexibility and potentially lower costs.

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

Integrating these functions into Storwize also means that they are designed to operate smoothly together, reducing management effort.

This IBM Redbooks publication provides information about the latest features and functions of the Storwize V7000 Gen2 and software version 7.3 implementation, architectural improvements, and Easy Tier.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, clients, and IBM Business Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8244-00 ISBN 0738440264