

# Information Storage and Management v2 Student Guide

Education Services  
April 2012





## INFORMATION STORAGE AND MANAGEMENT V2

### Course Introduction

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Course Introduction 1

### Welcome to Information Storage and Management v2.

Copyright © 1996, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012 EMC Corporation. All Rights Reserved. EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license. EMC2, EMC, Data Domain, RSA, EMC Centera, EMC ControlCenter, EMC LifeLine, EMC OnCourse, EMC Proven, EMC Snap, EMC SourceOne, EMC Storage Administrator, Acartus, Access Logix, AdvantEdge, AlphaStor, ApplicationXtender, ArchiveXtender, Atmos, Authentica, Authentic Problems, Automated Resource Manager, AutoStart, AutoSwap, AVALONidm, Avamar, Captiva, Catalog Solution, C-Clip, Celerra, Celerra Replicator, Centera, CenterStage, CentraStar, ClaimPack, ClaimsEditor, CLARIION, ClientPak, Codebook Correlation Technology, Common Information Model, Configuration Intelligence, Configuresoft, Connectrix, CopyCross, CopyPoint, Dantz, DatabaseXtender, Direct Matrix Architecture, DiskXtender, DiskXtender 2000, Document Sciences, Documentum, elput, E-Lab, EmailXaminer, EmailXtender, Enginuity, eRoom, Event Explorer, FarPoint, FirstPass, FLARE, FormWare, Geosynchrony, Global File Virtualization, Graphic Visualization, Greenplum, HighRoad, HomeBase, InfoMover, Infoscope, Infra, InputAccel, InputAccel Express, Invista, Ionix, ISIS, Max Retriever, MediaStor, MirrorView, Navisphere, NetWorker, nLayers, OnAlert, OpenScale, PixTools, Powerlink, PowerPath, PowerSnap, QuickScan, Rainfinity, RepliCare, RepliStor, ResourcePak, Retrospect, RSA, the RSA logo, SafeLine, SAN Advisor, SAN Copy, SAN Manager, Smarts, SnapImage, SnapSure, SnapView, SRDF, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix, Symmetrix DMX, Symmetrix VMAX, TimeFinder, UltraFlex, UltraPoint, UltraScale, Unisphere, VMAX, Vblock, Viewlets, Virtual Matrix, Virtual Matrix Architecture, Virtual Provisioning, VisualSAN, VisualSRM, Voyence, VPLEX, VSAM-Assist, WebXtender, xPression, xPresso, YottaYotta, the EMC logo, and where information lives, are registered trademarks or trademarks of EMC Corporation in the United States and other countries.

All other trademarks used herein are the property of their respective owners.

© Copyright 2012 EMC Corporation. All rights reserved. Published in the USA.

Revision Date: 04/13/2012

Revision Number: MR-1CP-ISMv2\_2.0

# Course Overview

Description	Information Storage and Management (ISM) is the only course of its kind to provide a comprehensive understanding the varied storage infrastructure components in classic and virtual environments. It enables participants to make informed decisions in an increasingly complex IT environment. It provides a strong understanding of underlying storage technologies and prepares participants for advanced concepts, technologies, and products. Participants will learn the architectures, features, and benefits of intelligent storage systems; storage networking technologies such as FC SAN, IP SAN, NAS, and object-based and unified storage; business continuity solutions such as backup and replication; the increasingly critical area of information security and management, and the emerging field of Cloud computing. This unique, open course focuses on concepts and principles which are further illustrated and reinforced with EMC product examples.
Audience	EMC Customers, Partners, Internals, and Industry audience (including students)
Prerequisites	<ul style="list-style-type: none"><li>• Basic understanding of computer architecture, operating systems, networking, and databases.</li><li>• Experience in specific segments of IT infrastructure will be an advantage.</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Course Introduction

2

## Course Description:

Information Storage and Management (ISM) is the only course of its kind to provide a comprehensive understanding the varied storage infrastructure components in classic and virtual environments. It enables participants to make informed decisions in an increasingly complex IT environment. It provides a strong understanding of underlying storage technologies and prepares participants for advanced concepts, technologies, and products. Participants will learn the architectures, features, and benefits of intelligent storage systems; storage networking technologies such as FC SAN, IP SAN, NAS, and object-based and unified storage; business continuity solutions such as backup and replication; the increasingly critical area of information security and management, and the emerging field of Cloud computing. This unique, open course focuses on concepts and principles which are further illustrated and reinforced with EMC product examples.

## Audience:

- Experienced storage professionals who may not have exposure to all of the segments of modern storage infrastructure
- Experienced IT professionals managing storage infrastructure in both classic and virtualized environments
- Students and IT professionals who want to build their career in the storage industry
- Organization-wide IT teams directly or indirectly responsible for planning, designing, deploying, managing, or leveraging information infrastructure
- Individuals seeking EMC Proven Professional Information Storage Associate v2 (EMCISA) certification

## Prerequisites:

To understand the content and successfully complete this course, a participant must have basic understanding of computer architecture, operating systems, networking, and databases. Participants with experience in specific segments of storage infrastructure would also be able to assimilate the course material.

## Course Objectives

Upon completion of this course, you should be able to:

- Evaluate storage architectures and key data center elements in classic, virtualized, and cloud environments
- Explain physical and logical components of a storage infrastructure including storage subsystems, RAID, and intelligent storage systems
- Describe storage networking technologies such as FC SAN, IP SAN, FCoE, NAS, and object-based and unified storage
- Articulate business continuity solutions—backup and replication, and archive for managing fixed content
- Describe information security requirements and solutions, and identify parameters for managing and monitoring storage infrastructure in classic, virtualized, and cloud environments

Upon completion of this course, you should be able to:

- Evaluate storage architectures and key data center elements in classic, virtualized, and cloud environments
- Explain physical and logical components of a storage infrastructure including storage subsystems, RAID, and intelligent storage systems
- Describe storage networking technologies such as FC SAN, IP SAN, FCoE, NAS, and object-based and unified storage
- Articulate business continuity solutions—backup and replication, and archive for managing fixed content
- Describe information security requirements and solutions, and identify parameters for managing and monitoring storage infrastructure in classic, virtualized, and cloud environments

## Course Organization

- **Module 1** : Introduction to Information Storage
- **Module 2** : Data Center Environment
- **Module 3** : Data Protection – RAID
- **Module 4** : Intelligent Storage System
- **Module 5** : Fibre Channel Storage Area Network (FC SAN)
- **Module 6** : IP SAN and FCoE
- **Module 7** : Network-Attached Storage (NAS)
- **Module 8** : Object-based and Unified Storage
- **Module 9** : Introduction to Business Continuity
- **Module 10** : Backup and Archive
- **Module 11** : Local Replication
- **Module 12** : Remote Replication
- **Module 13** : Cloud Computing
- **Module 14** : Securing the Storage Infrastructure
- **Module 15** : Managing the Storage Infrastructure

The organization of the course is shown above. This course has 15 Modules.

# Agenda

## Modules/Lessons

<b>Day 1</b>	<ul style="list-style-type: none"><li>• Introduction to Information Storage</li><li>• Data Center Environment</li><li>• RAID</li><li>• Intelligent Storage System</li></ul>
<b>Day 2</b>	<ul style="list-style-type: none"><li>• FC SAN</li><li>• IP SAN AND FCoE</li><li>• NAS</li></ul>
<b>Day 3</b>	<ul style="list-style-type: none"><li>• Object-based and Unified Storage</li><li>• Introduction to Business Continuity</li><li>• Backup and Archive</li></ul>
<b>Day 4</b>	<ul style="list-style-type: none"><li>• Local Replication</li><li>• Remote Replication</li><li>• Cloud Computing</li></ul>
<b>Day 5</b>	<ul style="list-style-type: none"><li>• Cloud Computing</li><li>• Securing Storage Infrastructure</li><li>• Managing Storage Infrastructure</li></ul>

# Become a Certified EMC Proven Professional!

Information Storage and Management v2 course supports EMC Proven Professional certification.

The #1 certification program in the information storage and management industry.



## EMC Proven Professional certification details

Certification Level	Associate
Exam Name	E10-001 Information Storage and Management Exam Version 2
Credential Name	EMC Proven Professional Information Storage Associate v2 (EMCISA)

## Next Steps:

### 1. Assess your level of expertise

Take a Free online practice test to identify areas you may need to review.

### 2. Validate your expertise

Register for your proctored exam at a Pearson VUE testing center near you.

For more information, please visit <http://education.emc.com/certification>.

## Why Get Proven?

Being Proven means investing in yourself and formally validating your knowledge, skills, and expertise by the industry's most comprehensive education and certification program.

**Get Proven. Join a community of dedicated professionals, share exclusive benefits:**

**Keep Current** - Ensure that your knowledge grows as fast as the pace of technology changes

- Receive no-cost Knowledge Maintenance updates on the latest EMC products and technologies.
- Learn from in-depth technical papers from our Knowledge Sharing library.

**Lead and Advocate** – Establish yourself as an industry thought-leader and mentor

- Become an advocate for the Information Storage and Management industry.
- Get published. Share your expertise and best practices in our Knowledge Sharing Competition.

**Connect and Collaborate** – Join a community of the most trusted professionals in the industry

- Discuss, share, find answers, or simply connect in our EMC Proven Professional online community.



Like Us:  
[EMC Proven Professional](#)

<http://education.EMC.com/ProvenCommunity>

Follow Us:  
[@EMCEducation](#), [@EMCProven](#)

This slide intentionally left blank.

# Module – 1

# Introduction to Information Storage



## Module 1: Introduction to Information Storage

Upon completion of this module, you should be able to:

- Define data and information
- Describe types of data
- Describe the evolution of storage architecture
- Describe the core elements of a data center
- List the key characteristics of data center
- Provide an overview of virtualization and cloud computing

This module focuses on the definition of data and information, types of data, and evolution of storage architecture. It lists the five core elements of a data center and describes the key characteristics of a data center. This module also provides an overview of virtualization and cloud computing.

## Why Information Storage and Management?

- Information is the knowledge derived from data
- Growth of digital information has resulted in information explosion
- We live in an on-command, on-demand world
  - ▶ We need information when and where required
- Increasing dependency on fast and reliable access to information
- Businesses seek to store, protect, optimize, and leverage the information
  - ▶ To gain competitive advantage
  - ▶ To derive new business opportunity

Information is increasingly important in our daily lives. We have become information-dependent in the 21st century, living in an on-command, on-demand world, which means, we need information when and where it is required. We access the Internet every day to perform searches, participate in social networking, send and receive e-mails, share pictures and videos, and use scores of other applications. Equipped with a growing number of content-generating devices, more information is created by individuals than by organizations (including business, governments, non-profits and so on). Information created by individuals gains value when shared with others. When created, information resides locally on devices, such as cell phones, smartphones, tablets, cameras, and laptops. To be shared, this information needs to be uploaded to central data repository (data centers) via networks. Although the majority of information is created by individuals, it is stored and managed by a relatively small number of organizations.

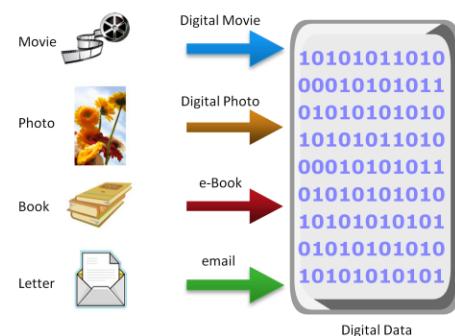
The importance, dependency, and volume of information for the business world also continue to grow at astounding rates. Businesses depend on fast and reliable access to information critical to their success. Examples of business processes or systems that rely on digital information include airline reservations, telecommunications billing, internet commerce, electronic banking, credit card transaction processing, capital/stock trading, health care claims processing, life science research and so on. The increasing dependence of businesses on information has amplified the challenges in storing, protecting, and managing data. Legal, regulatory, and contractual obligations regarding the availability and protection of data further add to these challenges.

## What is Data?

### Data

It is a collection of raw facts from which conclusions may be drawn.

- Data is converted into more convenient form – digital data
- Factors for digital data growth are:
  - ▶ Increase in data-processing capabilities
  - ▶ Lower cost of digital storage
  - ▶ Affordable and faster communication technology
  - ▶ Proliferation of applications and smart devices



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 1: Introduction to Information Storage 4

Data is a collection of raw facts from which conclusions may be drawn. Handwritten letters, a printed book, a family photograph, printed and duly signed copies of mortgage papers, a bank's ledgers, and an airline ticket are examples that contain data.

Before the advent of computers, the methods adopted for data creation and sharing were limited to fewer forms, such as paper and film. Today, the same data can be converted into more convenient forms, such as an e-mail message, an e-book, a digital image, or a digital movie. This data can be generated using a computer and stored as strings of binary numbers (0s and 1s). Data in this form is called digital data and is accessible by the user only after a computer processes it.

Businesses analyze raw data to identify meaningful trends. On the basis of these trends, a company can plan or modify its strategy. For example, a retailer identifies customers' preferred products and brand names by analyzing their purchase patterns and maintaining an inventory of those products. Effective data analysis not only extends its benefits to existing businesses, but also creates the potential for new business opportunities by using the information in creative ways.

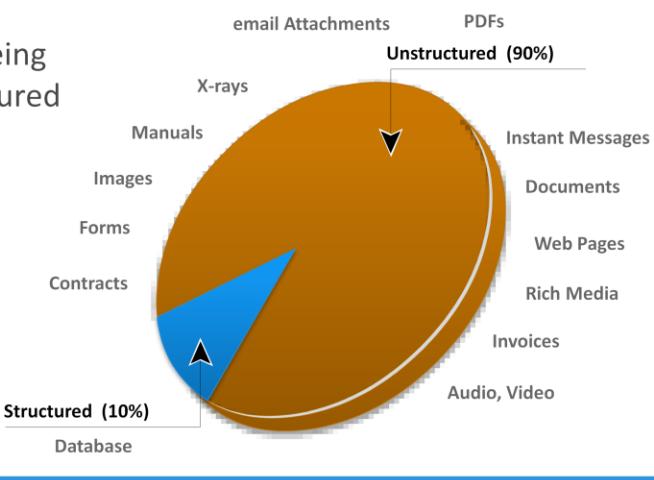
Cont..

With the advancement of computer and communication technologies, the rate of data generation and sharing has increased exponentially. The following is a list of some of the factors that have contributed to the growth of digital data:

- **Increase in data-processing capabilities:** Modern computers provide a significant increase in processing and storage capabilities. This enables the conversion of various types of content and media from conventional forms to digital formats.
- **Lower cost of digital storage:** Technological advances and the decrease in the cost of storage devices have provided low-cost storage solutions. This cost benefit has increased the rate at which digital data is generated and stored.
- **Affordable and faster communication technology:** The rate of sharing digital data is now much faster than traditional approaches. A handwritten letter might take a week to reach its destination, whereas it typically takes only a few seconds for an e-mail message to reach its recipient.
- **Proliferation of applications and smart devices:** Smartphones, tablets, and newer digital devices, along with smart applications, have significantly contributed to the generation of digital content.

## Types of Data

- Data can be classified as:
  - ▶ Structured
  - ▶ Unstructured
- Majority of data being created is unstructured



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 1: Introduction to Information Storage 6

Data can be classified as structured or unstructured based on how it is stored and managed. Structured data is organized in rows and columns in a rigidly defined format so that applications can retrieve and process it efficiently. Structured data is typically stored using a database management system (DBMS).

Data is unstructured if its elements cannot be stored in rows and columns, which makes it difficult to query and retrieve by applications. For example, customer contacts that are stored in various forms such as sticky notes, e-mail messages, business cards, or even digital format files, such as .doc, .txt, and .pdf. Due to its unstructured nature, it is difficult to retrieve this data using a traditional customer relationship management application. A vast majority of new data being created today is unstructured. The industry is challenged with new architectures, technologies, techniques, and skills to store, manage, analyze, and derive value from unstructured data from numerous sources.

# Big Data

## Big Data

It refers to data sets whose sizes are beyond the ability of commonly used software tools to capture, store, manage, and process within acceptable time limits.

- Includes both structured and unstructured data generated by variety of sources
- Big data analysis in real time requires new techniques and tools that provide:
  - ▶ High performance
  - ▶ Massively parallel processing (MPP) data platforms
  - ▶ Advanced analytics
- Big data analytics provide an opportunity to translate large volumes of data into right decisions

Big data is a new and evolving concept, which refers to data sets whose sizes are beyond the capability of commonly used software tools to capture, store, manage, and process within acceptable time limits. It includes both structured and unstructured data generated by a variety of sources, including business application transactions, web pages, videos, images, e-mails, social media, and so on. These data sets typically require real-time capture or updates for analysis, predictive modeling, and decision making.

Traditional IT infrastructure and data processing tools and methodologies are inadequate to handle the volume, variety, dynamism, and complexity of big data. Analyzing big data in real time requires new techniques, architectures, and tools that provide high performance, massively parallel processing (MPP) data platforms, and advanced analytics on the data sets.

Data Science is an emerging discipline, which enables organizations to derive business value from big data. Data Science represents the synthesis of several existing disciplines, such as statistics, math, data visualization and computer science to enable data scientists to develop advanced algorithms for the purpose of analyzing vast amounts of information to drive new value and make more data-driven decisions. Several industries and markets currently looking to employ data science techniques include medical and scientific research, healthcare, public administration, fraud detection, social media, banks, insurance companies, and other digital information-based entities that benefit from the analytics of big data. The storage architecture required for big data should be simple, efficient, and inexpensive to manage, yet provide access to multiple platforms and data sources simultaneously.

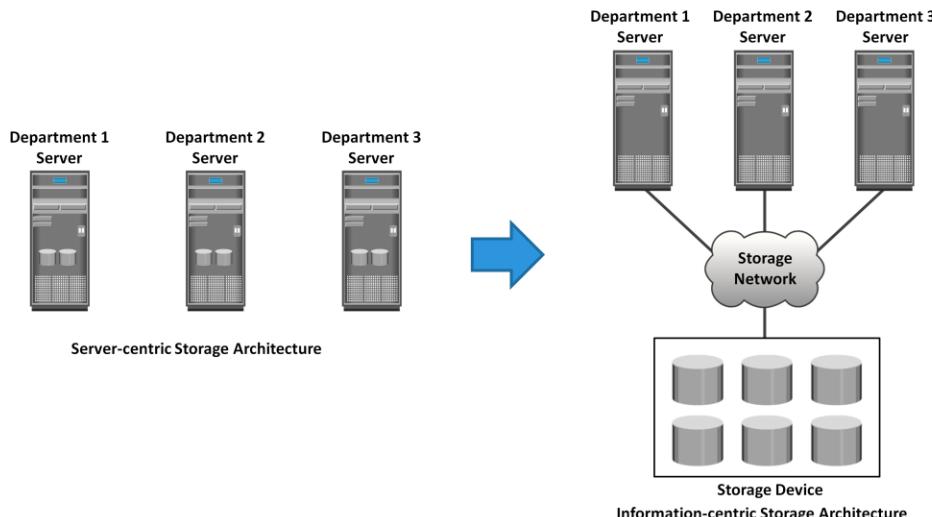
## Storage

- Stores data created by individuals and organizations
  - ▶ Provides access to data for further processing
- Examples of storage devices are:
  - ▶ Media card in a cell phone or digital camera
  - ▶ DVDs, CD-ROMs
  - ▶ Disk drives
  - ▶ Disk arrays
  - ▶ Tapes

Data created by individuals or businesses must be stored so that it is easily accessible for further processing. In a computing environment, devices designed for storing data are termed storage devices or simply storage. The type of storage used varies based on the type of data and the rate at which it is created and used. Devices, such as a media card in a cell phone or digital camera, DVDs, CD-ROMs, and disk drives in personal computers are examples of storage devices.

Businesses have several options available for storing data, including internal hard disks, external disk arrays, and tapes.

## Evolution of Storage Architecture



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 1: Introduction to Information Storage 9

Historically, organizations had centralized computers (mainframes) and information storage devices (tape reels and disk packs) in their data center. The evolution of open systems, their affordability, and ease of deployment made it possible for business units/departments to have their own servers and storage. In earlier implementations of open systems, the storage was typically internal to the server. These storage devices could not be shared with any other servers. This approach is referred to server-centric storage architecture. In this architecture, each server has a limited number of storage devices, and any administrative tasks, such as maintenance of the server or increasing storage capacity, might result in unavailability of information. The proliferation of departmental servers in an enterprise resulted in unprotected, unmanaged, fragmented islands of information and increased capital and operating expenses.

To overcome these challenges, storage evolved from server-centric to information-centric architecture. In this architecture, storage devices are managed centrally and independent of servers. These centrally-managed storage devices are shared with multiple servers. When a new server is deployed in the environment, storage is assigned from the same shared storage devices to that server. The capacity of shared storage can be increased dynamically by adding more storage devices without impacting information availability. In this architecture, information management is easier and cost-effective.

Storage technology and architecture continue to evolve, which enables organizations to consolidate, protect, optimize, and leverage their data to achieve the highest return on information assets.

## Data Center

### Data Center

It is a facility that contains storage, compute, network, and other IT resources to provide centralized data-processing capabilities.

- Core elements of a data center
  - ▶ Application
  - ▶ Database management system (DBMS)
  - ▶ Host or Compute
  - ▶ Network
  - ▶ Storage
- These core elements work together to address data-processing requirements

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 1: Introduction to Information Storage 10

Organizations maintain data centers to provide centralized data-processing capabilities across the enterprise. Data centers house and manage large amounts of data. The data center infrastructure includes hardware components, such as computers, storage systems, network devices, and power backups; and software components, such as applications, operating systems, and management software. It also includes environmental controls, such as air conditioning, fire suppression, and ventilation.

Large organizations often maintain more than one data center to distribute data processing workloads and provide backup if a disaster occurs.

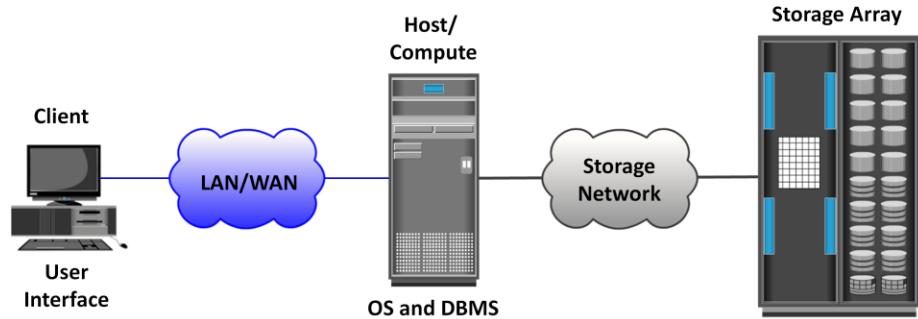
Five core elements are essential for the functionality of a data center:

- **Application:** A computer program that provides the logic for computing operations
- **Database management system (DBMS):** Provides a structured way to store data in logically organized tables that are interrelated
- **Host or compute:** A computing platform (hardware, firmware and software) that runs applications and databases
- **Network:** A data path that facilitates communication among various networked devices
- **Storage:** A device that stores data persistently for subsequent use

These core elements are typically viewed and managed as separate entities, but all the elements must work together to address data-processing requirements.

*Note:* In this course host, compute, and server are used interchangeably to represent the element that runs applications.

## Data Center: Online Order Transaction System Example



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

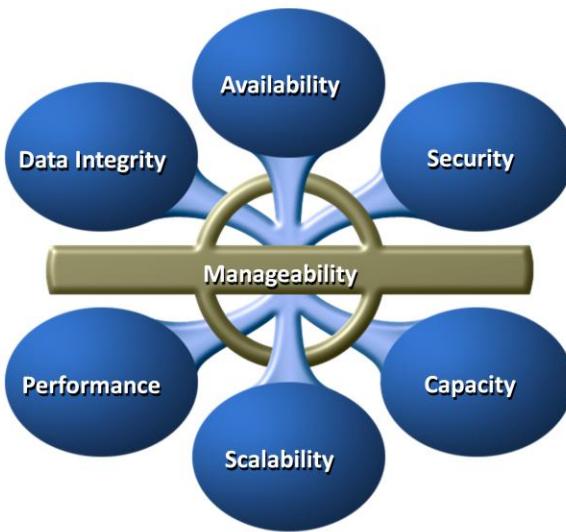
Module 1: Introduction to Information Storage 11

Figure in the slide shows an example of an online order transaction system that involves the five core elements of a data center and illustrates their functionality in a business process.

A customer places an order through a client machine connected over a LAN/WAN to a host running an order-processing application. The client accesses the DBMS on the host through the application to provide order-related information, such as the customer name, address, payment method, products ordered, and quantity ordered.

The DBMS uses the host operating system to write this data to the physical disks in the storage array. The storage networks provide the communication link between the host and the storage array and transports the request to read or write data between them. The storage array, after receiving the read or write request from the host, performs the necessary operations to store the data on physical disks.

## Key Characteristics of a Data Center



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 1: Introduction to Information Storage 12

Uninterrupted operation of data centers is critical to the survival and success of a business. Although the characteristics shown in the slide are applicable to all elements of the data center infrastructure, the focus here is on storage systems.

- **Availability:** A data center should ensure the availability of information when required. Unavailability of information could cost millions of dollars per hour to businesses, such as financial services, telecommunications, and e-commerce.
- **Security:** Data centers must establish policies, procedures, and core element integration to prevent unauthorized access to information.
- **Scalability:** Business growth often requires deploying more servers, new applications, and additional databases. Data center resources should scale based on requirements, without interrupting business operations.
- **Performance:** All the elements of the data center should provide optimal performance based on the required service levels.
- **Data integrity:** Data integrity refers to mechanisms, such as error correction codes or parity bits, which ensure that data is stored and retrieved exactly as it was received.

Cont..

- **Capacity:** Data center operations require adequate resources to store and process large amounts of data, efficiently. When capacity requirements increase, the data center must provide additional capacity without interrupting availability or with minimal disruption. Capacity may be managed by reallocating the existing resources or by adding new resources.
- **Manageability:** A data center should provide easy and integrated management of all its elements. Manageability can be achieved through automation and reduction of human (manual) intervention in common tasks.

## Managing Data Center

- Key management activities include
  - ▶ Monitoring
    - ▶ Continuous process of gathering information on various elements and services running in a data center
  - ▶ Reporting
    - ▶ Details on resource performance, capacity, and utilization
  - ▶ Provisioning
    - ▶ Configuration and allocation of resources to meet the capacity, availability, performance, and security requirements
- Virtualization and cloud computing have changed the way data center infrastructure resources are provisioned and managed

Managing a data center involves many tasks. The key management activities include the following:

- **Monitoring:** It is a continuous process of gathering information on various elements and services running in a data center. The aspects of a data center that are monitored include security, performance, availability, and capacity.
- **Reporting:** It is done periodically on resource performance, capacity, and utilization. Reporting tasks help to establish business justifications and chargeback of costs associated with data center operations.
- **Provisioning:** It is a process of providing the hardware, software, and other resources required to run a data center. Provisioning activities primarily include resources management to meet capacity, availability, performance, and security requirements.

Virtualization and cloud computing have dramatically changed the way data center infrastructure resources are provisioned and managed. Organizations are rapidly deploying virtualization on various elements of data centers to optimize their utilization. Further, continuous cost pressure on IT and on-demand data processing requirements have resulted in the adoption of cloud computing.

## Virtualization: An Overview

- Virtualization is a technique of abstracting physical resources and making them appear as logical resources
  - ▶ For example partitioning of raw disks
- Pools physical resources and provides an aggregated view of physical resource capabilities
- Virtual resources can be created from pooled physical resources
  - ▶ Improves utilization of physical IT resources

Virtualization is a technique of abstracting physical resources, such as compute, storage, and network, and making them appear as logical resources. Virtualization existed in the IT industry for several years and in different forms. Common examples of virtualization are virtual memory used on compute systems and partitioning of raw disks.

Virtualization enables pooling of physical resources and providing an aggregated view of the physical resource capabilities. For example, storage virtualization enables multiple pooled storage devices to appear as a single large storage entity. Similarly, by using compute virtualization, the CPU capacity of the pooled physical servers can be viewed as aggregation of the power of all CPUs (in megahertz). Virtualization also enables centralized management of pooled resources.

Virtual resources can be created and provisioned from the pooled physical resources. For example, a virtual disk of a given capacity can be created from a storage pool or a virtual server with specific CPU power and memory can be configured from a compute pool. These virtual resources share pooled physical resources, which improves the utilization of physical IT resources. Based on business requirements, capacity can be added to or removed from the virtual resources without any disruption to applications or users. With improved utilization of IT assets, organizations save the costs associated with procurement and management of new physical resources. Moreover, fewer physical resources means less space and energy, which leads to better economics and green computing.

## Cloud Computing: An Overview

- Enables individuals and organizations to use IT resources as a service over network
- Enables self-service requesting and automates request-fulfillment process
  - ▶ Enables users to scale up or scale down the usage of computing resources quickly
- Enables consumption-based metering
  - ▶ Consumers pay only for the resources they use
    - ▶ Example: CPU hours used, amount of data transferred, and Gigabytes of data stored

In today's fast-paced and competitive environment, organizations must be agile and flexible to meet changing market requirements. This leads to rapid expansion and upgrade of resources while meeting stagnant IT budgets. Cloud computing addresses these challenges efficiently. Cloud computing enables individuals or businesses to use IT resources as a service over the network. It provides highly scalable and flexible computing that enables provisioning of resources on demand. Users can scale up or scale down the demand of computing resources, including storage capacity, with minimal management effort or service provider interaction. Cloud computing empowers self-service requesting through a fully automated request-fulfillment process. Cloud computing enables consumption-based metering; therefore, consumers pay only for the resources they use, such as CPU hours used, amount of data transferred, and gigabytes of data stored.

Cloud infrastructure is usually built upon virtualized data centers, which provide resource pooling and rapid provisioning of resources. Information storage in virtualized and cloud environments is detailed later in this course.

## Module 1: Summary

Key points covered in this module:

- Data and information
- Types of data
- Big data
- Evolution of storage architecture
- Core elements of data center
- Key characteristics of data center
- Virtualization and cloud computing

This module covered the definition of data and information. Data is a collection of raw facts from which conclusions may be drawn and information is the intelligence and knowledge derived from data. Businesses analyze raw data to identify meaningful trends. On the basis of these trends, a company can plan or modify its strategy.

Data can be classified as structured and unstructured. Big data refers to data sets whose sizes are beyond the ability of commonly used software tools to capture, store, manage, and process within acceptable time limits.

Information-centric architecture is commonly deployed in today's data center. It helps to overcome the challenges of server-centric storage architecture.

A data center has five core elements such as application, database management system (DBMS), host, network, and storage.

The key characteristics of data are availability, security, scalability, performance, data integrity, capacity, and manageability.

Virtualization is a technique of abstracting physical resources, such as compute, storage, and network, and making them appear as logical resources.

Cloud computing enables individuals or businesses to use IT resources as a service over the network.

## Check Your Knowledge – 1

- Which is an example of structured data?
  - A. Image
  - B. PDF document
  - C. Database
  - D. Web page
- Which is true about big data?
  - A. Includes only unstructured data
  - B. Includes data from a single source
  - C. Captured efficiently using traditional software tools
  - D. Data size is beyond the capability of traditional software to process

## Check Your Knowledge – 2

- Which is a feature of information-centric architecture?
  - A. Storage is internal to the servers
  - B. Prevents sharing of storage among servers
  - C. Consists of server, network, and storage in a single system
  - D. Storage is managed centrally and independent of servers
- What accurately describes virtualization?
  - A. Provides on-demand, metered services
  - B. Abstracts physical resources into logical resources
  - C. Pools logical resources to provide data integrity
  - D. Enables decentralized management across data centers

## Check Your Knowledge – 3

- Which requirement refers to the ability of a storage solution to grow with the business?
  - A. Availability
  - B. Manageability
  - C. Integrity
  - D. Scalability

# Module – 2

# Data Center Environment



## Module 2: Data Center Environment

Upon completion of this module, you should be able to:

- Describe the core elements of a data center
- Describe virtualization at application and host layer
- Describe disk drive components and performance
- Describe host access to storage through DAS
- Describe working and benefits of flash drives

This module focuses on the key components of a data center. It also includes virtualization at compute, memory, desktop, and application. Storage and network virtualization are discussed later in the course. This module also focuses on storage subsystems and provides details on components, geometry, and performance parameters of a disk drive. The connectivity between the host and storage facilitated by various technologies is also explained.

## Module 2: Data Center Environment

### Lesson 1: Application, DBMS, and Host (Compute)

During this lesson the following topics are covered:

- Application and application virtualization
- DBMS
- Components of host system
- Compute and memory virtualization

This lesson covers three key components of a data center – application, DBMS and compute. Hardware and software components of a compute system including OS, logical volume manager, file system, and device driver are also explained. Virtualization at application and compute is also discussed in the lesson.

## Application

- A software program that provides logic for computing operations
- Commonly deployed applications in a data center
  - ▶ Business applications – email, enterprise resource planning (ERP), decision support system (DSS)
  - ▶ Management applications – resource management, performance tuning, virtualization
  - ▶ Data protection applications – backup, replication
  - ▶ Security applications – authentication, antivirus
- Key I/O characteristics of an application
  - ▶ Read intensive vs. write intensive
  - ▶ Sequential vs. random
  - ▶ I/O size

An *application* is a computer program that provides the logic for computing operations. The application sends requests to the underlying operating system to perform read/write (R/W) operations on the storage devices. Applications can be layered on the database, which in turn uses the OS services to perform R/W operations on the storage devices.

Applications deployed in a data center environment are commonly categorized as business applications, infrastructure management applications, data protection applications, and security applications. Some examples of these applications are e-mail, enterprise resource planning (ERP), decision support system (DSS), resource management, backup, authentication and antivirus applications, and so on.

The characteristics of I/Os (Input/Output) generated by the application influence the overall performance of storage system and storage solution designs. Common I/O characteristics of an application are:

- Read versus Write intensive
- Sequential versus Random
- I/O size

## Application Virtualization

### Application Virtualization

It is the technique of presenting an application to an end user without any installation, integration, or dependencies on the underlying computing platform.

- Allows application to be delivered in an isolated environment
  - ▶ Aggregates Operating System (OS) resources and the application into a virtualized container
  - ▶ Ensures integrity of Operating System (OS) and applications
  - ▶ Avoids conflicts between different applications or different versions of the same application

*Application virtualization* breaks the dependency between the application and the underlying platform (OS and hardware). Application virtualization encapsulates the application and the required OS resources within a virtualized container. This technology provides the ability to deploy applications without making any change to the underlying OS, file system, or registry of the computing platform on which they are deployed. Because virtualized applications run in an isolated environment, the underlying OS and other applications are protected from potential corruptions. There are many scenarios in which conflicts might arise if multiple applications or multiple versions of the same application are installed on the same computing platform. Application virtualization eliminates this conflict by isolating different versions of an application and the associated O/S resources.

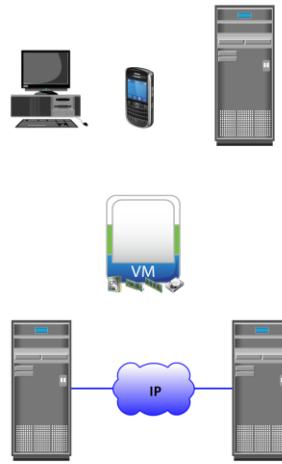
## Database Management System (DBMS)

- Database is a structured way to store data in logically organized tables that are interrelated
  - ▶ Helps to optimize the storage and retrieval of data
- DBMS controls the creation, maintenance, and use of databases
  - ▶ Processes an application's request for data
  - ▶ Instructs the OS to retrieve the appropriate data from storage
- Popular DBMS examples are MySQL, Oracle RDBMS, SQL Server, etc.

A database is a structured way to store data in logically organized tables that are interrelated. A database helps to optimize the storage and retrieval of data. A DBMS controls the creation, maintenance, and use of a database. The DBMS processes an application's request for data and instructs the operating system to transfer the appropriate data from the storage.

## Host (Compute)

- Resource that runs applications with the help of underlying computing components
  - ▶ Example: Servers, mainframes, laptop, desktops, tablets, server clusters, etc.
- Consists of hardware and software components
- Hardware components
  - ▶ Include CPU, memory, and input/output (I/O) devices
- Software components
  - ▶ Include OS, device driver, file system, volume manager, and so on



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 7

Users store and retrieve data through applications. The computers on which these applications run are referred to as *hosts* or *compute systems*. Hosts can be physical or virtual machines. A compute virtualization software enables creating virtual machines on top of physical compute infrastructure. Compute virtualization and virtual machines are discussed later in this module.

Examples of physical hosts include desktop computers, servers or a cluster of servers, virtual servers, laptops, and mobile devices. A host consists of CPU, memory, I/O devices, and a collection of software to perform computing operations. This software includes the operating system, file system, logical volume manager, device drivers, and so on. These software can be installed individually or may be part of the operating system.

## Operating Systems and Device Driver

- In a traditional environment OS resides between the applications and the hardware
  - ▶ Responsible for controlling the environment
- In a virtualized environment virtualization layer works between OS and hardware
  - ▶ Virtualization layer controls the environment
  - ▶ OS works as a guest and only controls the application environment
  - ▶ In some implementation OS is modified to communicate with virtualization layer
- Device driver is a software that enables the OS to recognize the specific device

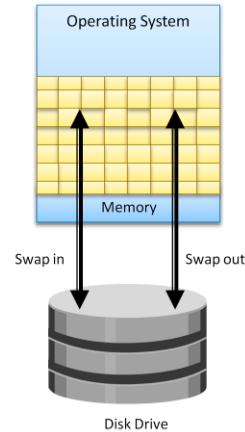
In a traditional computing environment, an *operating system* controls all the aspects of computing. It works between the application and physical components of a compute system. One of the services it provides to the application is data access. The operating system also monitors and responds to user actions and the environment. It organizes and controls hardware components and manages the allocation of hardware resources. It provides basic security for the access and usage of all managed resources. An operating system also performs basic storage management tasks while managing other underlying components, such as the file system, volume manager, and device drivers.

In a virtualized compute environment, the virtualization layer works between the operating system and the hardware resources. Here the OS might work differently based on the type of the compute virtualization implemented. In a typical implementation, the OS works as a guest and performs only the activities related to application interaction. In this case, hardware management functions are handled by the virtualization layer.

A *device driver* is special software that permits the operating system to interact with a specific device, such as a printer, a mouse, or a disk drive. A device driver enables the operating system to recognize the device and to access and control devices. Device drivers are hardware-dependent and operating-system-specific.

## Memory Virtualization

- An OS feature that presents larger memory to the application than physically available
  - ▶ Additional memory space comes from disk storage
  - ▶ Space used on the disk for virtual memory is called 'swap space/swap file or page file'
  - ▶ Inactive memory pages are moved from physical memory to the swap file
  - ▶ Provides efficient use of available physical memory
  - ▶ Data access from swap file is slower – use of flash drives for swap space gives best performance



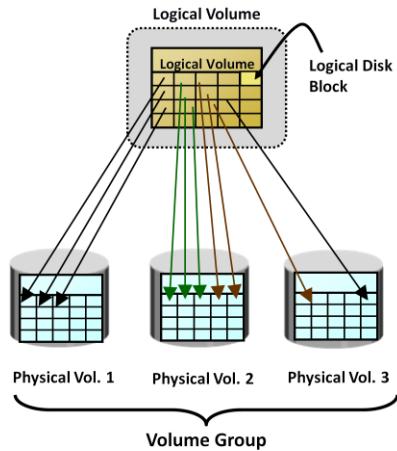
Memory has been, and continues to be, an expensive component of a host. It determines both the size and number of applications that can run on a host. *Memory virtualization* enables multiple applications and processes, whose aggregate memory requirement is greater than the available physical memory, to run on a host without impacting each other.

Memory virtualization is an operating system feature that virtualizes the physical memory (RAM) of a host. It creates a virtual memory with an address space larger than the physical memory space present in the compute system. The virtual memory encompasses the address space of the physical memory and part of the disk storage. The operating system utility that manages the virtual memory is known as the *virtual memory manager* (VMM). The VMM manages the virtual-to-physical memory mapping and fetches data from the disk storage when a process references a virtual address that points to data at the disk storage. The space used by the VMM on the disk is known as a swap space. A *swap space* (also known as *page file* or *swap file*) is a portion of the disk drive that appears like physical memory to the operating system.

In a virtual memory implementation, the memory of a system is divided into contiguous blocks of fixed-size pages. A process known as *paging* moves inactive physical memory pages onto the swap file and brings them back to the physical memory when required. This enables efficient use of the available physical memory among different applications. The operating system typically moves the least used pages into the swap file so that enough RAM is available for processes that are more active. Access to swap file pages is slower than physical memory pages because swap file pages are allocated on the disk drive which is slower than physical memory.

## Logical Volume Manager (LVM)

- Responsible for creating and controlling host level logical storage
  - ▶ Physical view of storage is converted to a logical view
  - ▶ Logical data blocks are mapped to physical data blocks
- One or more Physical Volumes form a Volume Group
  - ▶ LVM manages Volume Groups as a single entity
- Logical volumes are created from the volume group



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 10

In the early days, entire disk drive would be allocated to the file system or other data entity used by the operating system or application. The disadvantage was lack of flexibility. When a disk drive ran out of space, there was no easy way to extend the file system's size. Also, as the storage capacity of the disk drive increased, allocating the entire disk drive for the file system often resulted in underutilization of storage capacity.

The evolution of *Logical Volume Managers* (LVMs) enabled dynamic extension of file system capacity and efficient storage management. LVM is software that runs on the compute system and manages logical and physical storage. LVM is an intermediate layer between the file system and the physical disk. It can partition a larger-capacity disk into virtual, smaller-capacity volumes (the process is called *partitioning*) or aggregate several smaller disks to form a larger virtual volume. (The process is called *concatenation*).

The LVM provides optimized storage access and simplifies storage resource management. It hides details about the physical disk and the location of data on the disk. It enables administrators to change the storage allocation even when the application is running.

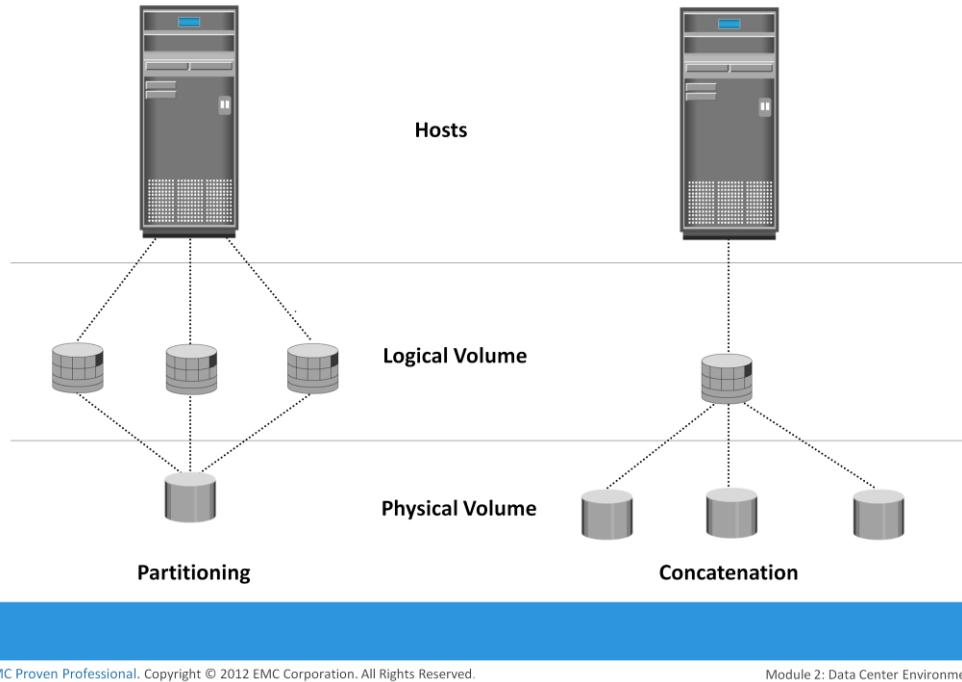
Cont...

The basic LVM components are physical volumes, volume groups, and logical volumes. In LVM terminology, each physical disk connected to the host system is a *physical volume* (PV). A *volume group* is created by grouping together one or more physical volumes. A unique *physical volume identifier* (PVID) is assigned to each physical volume when it is initialized for use by the LVM. Physical volumes can be added or removed from a volume group dynamically. They cannot be shared between different volume groups; which means, the entire physical volume becomes part of a volume group. Each physical volume is divided into equal-sized data blocks called *physical extents* when the volume group is created.

*Logical volumes* (LV) are created within a given volume group. A LV can be thought of as a disk partition, whereas the volume group itself can be thought of as a disk. The size of a LV is based on a multiple of the physical extents. The LV appears as a physical device to the operating system. A LV is made up of noncontiguous physical extents and may span over multiple physical volumes. A file system is created on a logical volume. These LVs are then assigned to the application. A logical volume can also be mirrored to provide enhanced data availability.

Today, logical volume managers are mostly offered as part of the operating system.

## LVM Example: Partitioning and Concatenation



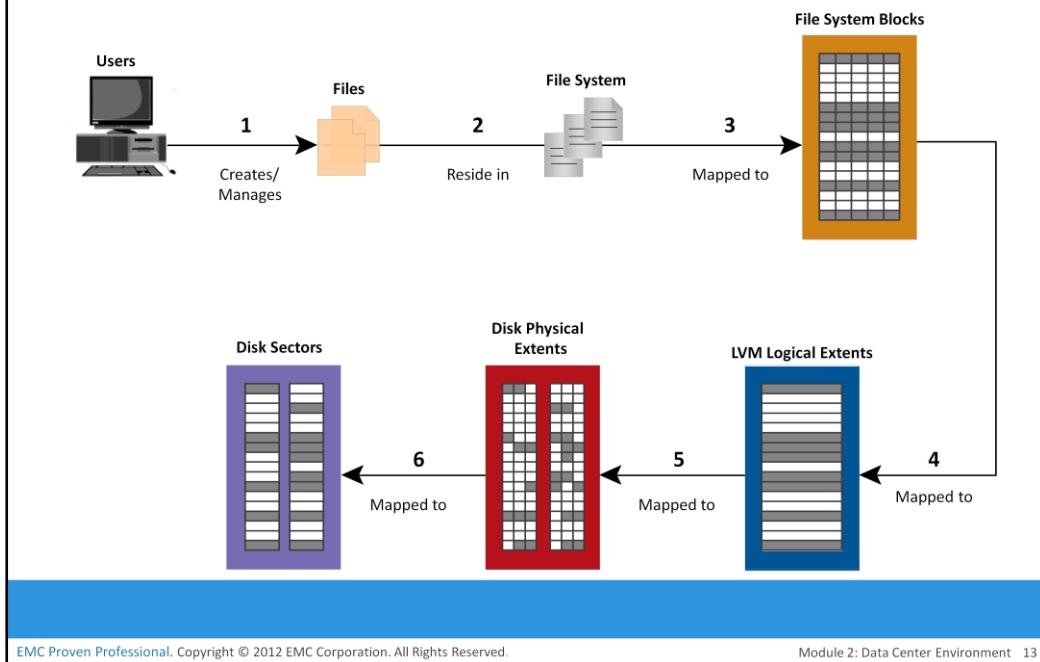
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 12

*Disk partitioning* was introduced to improve the flexibility and utilization of disk drives. In partitioning, a disk drive is divided into logical containers called *logical volumes* (LVs). For example, a large physical drive can be partitioned into multiple LVs to maintain data according to the file system and application requirements. The partitions are created from groups of contiguous cylinders when the hard disk is initially set up on the host. The host's file system accesses the logical volumes without any knowledge of partitioning and physical structure of the disk.

*Concatenation* is the process of grouping several physical drives and presenting them to the host as one big logical volume.

## File System



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 13

A *file* is a collection of related records or data stored as a unit with a name. A *file system* is a hierarchical structure of files. A file system enables easy access to data files residing within a disk drive, a disk partition, or a logical volume. A file system consists of logical structures and software routines that control access to files. It provides users with the functionality to create, modify, delete, and access files. Access to files on the disks is controlled by the permissions assigned to the file by the owner, which are also maintained by the file system.

A file system organizes data in a structured hierarchical manner via the use of directories, which are containers for storing pointers to multiple files. All file systems maintain a pointer map to the directories, subdirectories, and files that are part of the file system. The following list shows the process of mapping user files to the disk storage that uses an LVM:

1. Files are created and managed by users and applications.
2. These files reside in the file systems.
3. The file systems are mapped to file system blocks.
4. The file system blocks are mapped to logical extents of a logical volume.
5. These logical extents in turn are mapped to the disk physical extents either by the operating system or by the LVM.
6. These physical extents are mapped to the disk sectors in a storage subsystem.

If there is no LVM, then there are no logical extents. Without LVM, file system blocks are directly mapped to disk sectors.

Cont..

A file system *block* is the smallest unit allocated for storing data. Each file system block is a contiguous area on the physical disk. The block size of a file system is fixed at the time of its creation. The file system size depends on the block size and the total number of files system blocks. A file can span multiple file system blocks because most files are larger than the predefined block size of the file system. File system blocks cease to be contiguous and become fragmented when new blocks are added or deleted. Over time, as files grow larger, the file system becomes increasingly fragmented.

Apart from the files and directories, the file system also includes a number of other related records, which are collectively called the *metadata*. The metadata of a file system must be consistent for the file system to be considered healthy.

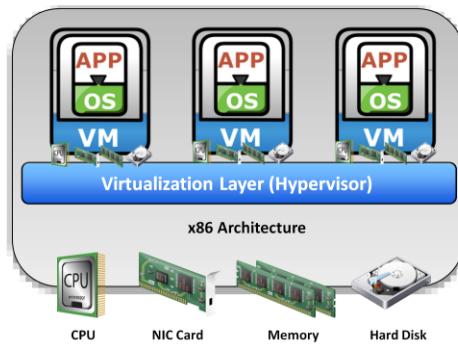
Examples of some common file systems are, FAT 32 (File Allocation Table) and NT File System (NTFS) for Microsoft Windows, UNIX File System (UFS) and Extended File System (EXT2/3) for Linux.

## Compute Virtualization

### Compute Virtualization

It is a technique of masking or abstracting the physical compute hardware and enabling multiple operating systems (OSs) to run concurrently on a single or clustered physical machine(s).

- Enables creation of multiple virtual machines (VMs), each running an OS and application
  - ▶ VM is a logical entity that looks and behaves like physical machine
- Virtualization layer resides between hardware and VMs
  - ▶ Also known as hypervisor
- VMs are provided with standardized hardware resources



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

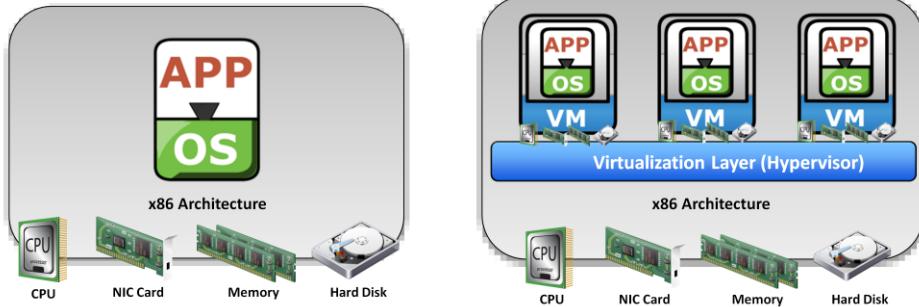
Module 2: Data Center Environment 15

Compute virtualization is a technique of masking or abstracting the physical hardware from the operating system. It enables multiple operating systems to run concurrently on a single or clustered physical machine(s). This technique enables creating portable virtual compute systems called *virtual machines (VMs)*. Each VM runs an operating system and application instance in an isolated manner.

Compute virtualization is achieved by a virtualization layer that resides between the hardware and virtual machines. This layer is also called the *hypervisor*. The hypervisor provides hardware resources, such as CPU, memory, and network to all the virtual machines. Within a physical server, a large number of virtual machines can be created depending on the hardware capabilities of the physical server.

A virtual machine is a logical entity but appears like a physical host to the operating system, with its own CPU, memory, network controller, and disks. However, all VMs share the same underlying physical hardware in an isolated manner. From a hypervisor perspective, virtual machines are discrete sets of files that include VM configuration file, data files, and so on.

## Need for Compute Virtualization



Before Virtualization	After Virtualization
<ul style="list-style-type: none"><li>Runs single operating system (OS) per machine at a time</li><li>Couples s/w and h/w tightly</li><li>May create conflicts when multiple applications run on the same machine</li><li>Underutilizes resources</li><li>Is inflexible and expensive</li></ul>	<ul style="list-style-type: none"><li>Runs multiple operating systems (OSs) per physical machine concurrently</li><li>Makes OS and applications h/w independent</li><li>Isolates VM from each other, hence, no conflict</li><li>Improves resource utilization</li><li>Offers flexible infrastructure at low cost</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 16

Typically, a physical server often faces resource-conflict issues when two or more applications running on the server have conflicting requirements. For example, applications might need different values in the same registry entry, different versions of the same DLL, and so on. These issues are further compounded with an application's high-availability requirements. As a result, the servers are limited to serve only one application at a time. This causes organizations to purchase new physical machines for every application they deploy, resulting in expensive and inflexible infrastructure. On the other hand, many applications do not take full advantage of the hardware capabilities available to them. Consequently, resources such as processors, memory, and storage remain underutilized.

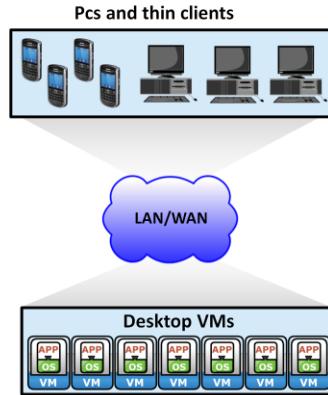
Compute virtualization enables to overcome these challenges by allowing multiple operating systems and applications to run on a single physical machine. This technique significantly improves server utilization and provides server consolidation. Server consolidation enables organizations to run their data center with fewer servers. This, in turn, cuts down the cost of new server acquisition, reduces operational cost, and saves data center floor and rack space. Creation of VMs takes less time compared to a physical server setup; organizations can provision servers faster and with ease. Individual VMs can be restarted, upgraded, or even crashed, without affecting the other VMs on the same physical machine. Moreover, VMs can be copied or moved from one physical machine to another without causing application downtime.

## Desktop Virtualization

### Desktop Virtualization

It is a technology which enables detachment of the user state, the Operating System (OS), and the applications from endpoint devices.

- Enables organizations to host and centrally manage desktops
  - ▶ Desktops run as virtual machines within the data center and accessed over a network
- Desktop virtualization benefits
  - ▶ Flexibility of access due to enablement of thin clients
  - ▶ Improved data security
  - ▶ Simplified data backup and PC maintenance



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 17

With the traditional desktop, the OS, applications, and user profiles are all tied to a specific piece of hardware. With legacy desktops, business productivity is impacted greatly when a client device is broken or lost. *Desktop virtualization* breaks the dependency between the hardware and its OS, applications, user profiles, and settings. This enables the IT staff to change, update, and deploy these elements independently. Desktops hosted at the data center and runs on virtual machines, whereas users remotely access these desktops from a variety of client devices, such as laptop, desktop, and mobile devices (also called Thin devices). Application execution and data storage are performed centrally at the data center instead of at the client devices. Because desktops run as virtual machines within an organization's data center, it mitigates the risk of data leakage and theft. It also helps to perform centralized backup and simplifies compliance procedures. Virtual desktops are easy to maintain because it is simple to apply patches, deploy new applications and OS, and provision or remove users centrally.

## Module 2: Data Center Environment

### Lesson 2: Connectivity

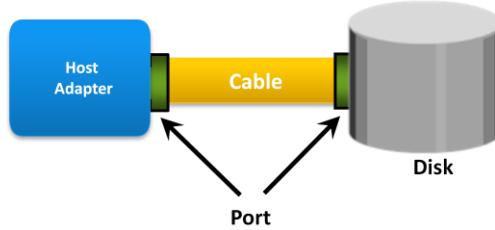
During this lesson the following topics are covered:

- Physical components of connectivity
- Storage connectivity protocols

This lesson covers physical components of connectivity and storage connectivity protocols. These protocols include IDE/ATA, SCSI, Fibre Channel and IP.

## Connectivity

- Interconnection between hosts or between a host and peripheral devices, such as storage
- Physical Components of Connectivity are:
  - ▶ Host interface card, port, and cable
- Protocol = a defined format for communication between sending and receiving devices
  - ▶ Popular storage interface protocols: IDE/ATA and SCSI



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 19

Connectivity refers to the interconnection between hosts or between a host and peripheral devices, such as printers or storage devices. The discussion here focuses only on the connectivity between the host and the storage device. Connectivity and communication between host and storage are enabled using physical components and interface protocols.

The *physical components* of connectivity are the hardware elements that connect the host to storage. Three physical components of connectivity between the host and storage are host interface device, port, and cable

A *host interface device* or *host adapter* connects a host to other hosts and storage devices. Examples of host interface devices are host bus adapter (HBA) and network interface card (NIC). *Host bus adaptor (HBA)* is an application-specific integrated circuit (ASIC) board that performs I/O interface functions between the host and storage, relieving the CPU from additional I/O processing workload. A host typically contains multiple HBAs.

A *port* is a specialized outlet that enables connectivity between the host and external devices. An HBA may contain one or more ports to connect the host to the storage device. *Cables* connect hosts to internal or external devices using copper or fiber optic media.

A *protocol* enables communication between the host and storage. Protocols are implemented using interface devices (or controllers) at both source and destination. The popular interface protocols used for host to storage communications are *Integrated Device Electronics/Advanced Technology Attachment (IDE/ATA)*, *Small Computer System Interface (SCSI)*, *Fibre Channel (FC)* and *Internet Protocol (IP)*.

## IDE/ATA and Serial ATA

- Integrated Device Electronics (IDE)/Advanced Technology Attachment (ATA)
  - ▶ Popular interface used to connect hard disks or CD-ROM drives
  - ▶ Available with variety of standards and names
- Serial Advanced Technology Attachment (SATA)
  - ▶ Serial version of the IDE/ATA specification that has replaced the parallel ATA
  - ▶ Inexpensive storage interconnect, typically used for internal connectivity
  - ▶ Provides data transfer rate up to 6 Gb/s (standard 3.0)

IDE/ATA is a popular interface protocol standard used for connecting storage devices, such as disk drives and CD-ROM drives. This protocol supports parallel transmission and therefore is also known as Parallel ATA (PATA) or simply ATA. IDE/ATA has a variety of standards and names. The Ultra DMA/133 version of ATA supports a throughput of 133 MB per second. In a master-slave configuration, an ATA interface supports two storage devices per connector. However, if the performance of the drive is important, sharing a port between two devices is not recommended.

The serial version of this protocol supports single bit serial transmission and is known as Serial ATA (SATA). High performance and low cost SATA has largely replaced PATA in the newer systems. SATA revision 3.0 provides a data transfer rate up to 6 Gb/s.

## SCSI and SAS

- Parallel Small computer system interface (SCSI)
  - ▶ Popular standard for connecting host and peripheral devices
    - ▶ Commonly used for storage connectivity in servers
  - ▶ Higher cost than IDE/ATA, therefore not popular in PC environments
  - ▶ Available in wide variety of related technologies and standards
  - ▶ Support up to 16 devices on a single bus
  - ▶ Ultra-640 version provides data transfer speed up to 640 MB/s
- Serial Attached SCSI (SAS)
  - ▶ Point-to-point serial protocol replacing parallel SCSI
  - ▶ Supports data transfer rate up to 6 Gb/s (SAS 2.0)

SCSI has emerged as a preferred connectivity protocol in high-end computers. This protocol supports parallel transmission and offers improved performance, scalability, and compatibility compared to ATA. However, the high cost associated with SCSI limits its popularity among home or personal desktop users. Over the years, SCSI has been enhanced and now includes a wide variety of related technologies and standards. SCSI supports up to 16 devices on a single bus and provides data transfer rates up to 640 MB/s (for the Ultra-640 version).

Serial attached SCSI (SAS) is a point-to-point serial protocol that provides an alternative to parallel SCSI. A newer version (SAS 2.0) of serial SCSI supports a data transfer rate up to 6 Gb/s.

## Fibre Channel and IP

- Fibre Channel (FC)
  - ▶ Widely used protocol for high speed communication to the storage device
  - ▶ Provides a serial data transmission that operates over copper wire and/or optical fiber
  - ▶ Latest version of the FC interface '16FC' allows transmission of data up to 16 Gb/s
- Internet Protocol (IP)
  - ▶ Traditionally used to transfer host-to-host traffic
  - ▶ Provide opportunity to leverage existing IP based network for storage communication
    - ▶ Examples: iSCSI and FCIP protocols

Fibre Channel is a widely used protocol for high-speed communication to the storage device. The Fibre Channel interface provides gigabit network speed. It provides a serial data transmission that operates over copper wire and optical fiber. The latest version of the FC interface '16FC' allows transmission of data up to 16 Gb/s. The FC protocol and its features are covered in more detail in Module 5.

IP is a network protocol that has been traditionally used for host-to-host traffic. With the emergence of new technologies, an IP network has become a viable option for host-to-storage communication. IP offers several advantages in terms of cost and maturity and enables organizations to leverage their existing IP-based network. iSCSI and FCIP protocols are common examples that leverage IP for host-to-storage communication. These protocols are detailed in module 6.

## Module 2: Data Center Environment

### Lesson 3: Storage

During this lesson the following topics are covered:

- Various storage options
- Disk drive components, addressing, and performance
- Enterprise Flash drives
- Host access to storage and direct-attached storage

This lesson covers the most important element of a data center – Storage. Various storage medias and options are discussed with focus on disk drives. Components, structure, addressing and factors that impacts disk drives performance are detailed in the lesson. Further it covers new generation flash drives and their benefits. Finally it introduces various methods of accessing storage from the host with details of direct-attached storage options.

## Storage Options

- Magnetic Tape
  - ▶ Low cost solution for long term data storage
    - ▶ Preferred option for backup destination in the past
  - ▶ Limitations
    - ▶ Sequential data access
    - ▶ Single application access at a time
    - ▶ Physical wear and tear
    - ▶ Storage/retrieval overheads

The storage is a core component in a data center. A storage device uses magnetic, optic, or solid state media. Disks, tapes, and diskettes use magnetic media, whereas CD/DVD uses optical media for storage. Removable Flash memory or Flash drives are examples of solid state media.

In the past *tapes* were the most popular storage option for backups because of their low cost. However, tapes have various limitations in terms of performance and management as listed here:

- Data is stored on the tape linearly along the length of the tape. Search and retrieval of data are done sequentially, and it invariably takes several seconds to access the data. As a result, random data access is slow and time-consuming. This limits tapes as a viable option for applications that require real-time, rapid access to data.
- In a shared computing environment, data stored on tape cannot be accessed by multiple applications simultaneously, restricting its use to one application at a time.
- On a tape drive, the read/write head touches the tape surface, so the tape degrades or wears out after repeated use.
- The storage and retrieval requirements of data from the tape and the overhead associated with managing the tape media are significant.

Due to these limitations and availability of low-cost disk drives, tapes are no longer a preferred choice as a backup destination for enterprise-class data centers.

## Storage Options (contd.)

- Optical discs
  - ▶ Popularly used as distribution medium in small, single-user computing environments
  - ▶ Limited in capacity and speed
  - ▶ Write once and read many (WORM): CD-ROM, DVD-ROM
  - ▶ Other variations: CD-RW, Blu-ray discs
- Disk drive
  - ▶ Most popular storage medium
  - ▶ Large storage capacity
  - ▶ Random read/write access
- Flash drives
  - ▶ Uses semiconductor media
  - ▶ Provide high performance and low power consumption

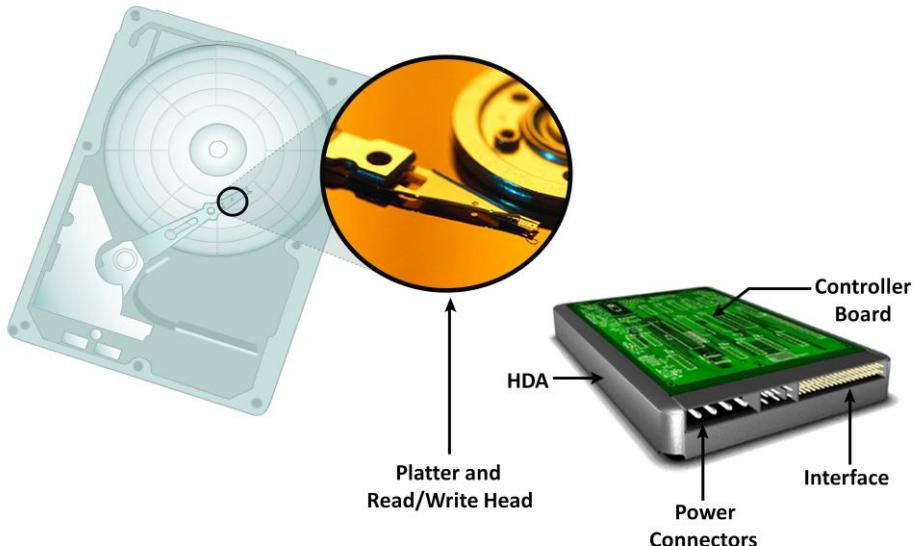
*Optical disc storage* is popular in small, single-user computing environments. It is frequently used by individuals to store photos or as a backup medium on personal or laptop computers. It is also used as a distribution medium for small applications, such as games, or as a means to transfer small amounts of data from one computer to another.

Optical discs have limited capacity and speed, which limit the use of optical media as a business data storage solution. The capability to write once and read many (WORM) is one advantage of optical disc storage. A CD-ROM is an example of a WORM device. Optical discs, to some degree, guarantee that the content has not been altered. Therefore, it can be used as a low-cost alternative for long-term storage of relatively small amounts of fixed content that do not change after it is created. Collections of optical discs in an array, called a *jukebox*, are still used as a fixed-content storage solution. Other forms of optical discs include CD-RW, Blu-ray disc, and other variations of DVD.

*Disk drives* are the most popular storage medium used in modern computers for storing and accessing data for performance-intensive, online applications. Disks support rapid access to random data locations. This means that data can be written or retrieved quickly for a large number of simultaneous users or applications. In addition, disks have a large capacity. Disk storage arrays are configured with multiple disks to provide increased capacity and enhanced performance.

Flash drives (or solid state drives - SSDs) uses semiconductor media and provides high performance and low power consumption. Flash drives are discussed in detail later in this module.

## Disk Drive Components



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 26

The key components of a hard disk drive are platter, spindle, read-write head, actuator arm assembly, and controller board. I/O operations in a HDD is performed by rapidly moving the arm across the rotating flat platters coated with magnetic particles. Data is transferred between the disk controller and magnetic platters through the read-write (R/W) head which is attached to the arm. Data can be recorded and erased on magnetic platters any number of times.

**Platter:** A typical HDD consists of one or more flat circular disks called *platters*. The data is recorded on these platters in binary codes (0s and 1s). The set of rotating platters is sealed in a case, called *Head Disk Assembly* (HDA). A platter is a rigid, round disk coated with magnetic material on both surfaces (top and bottom). The data is encoded by polarizing the magnetic area, or domains, of the disk surface. Data can be written to or read from both surfaces of the platter. The number of platters and the storage capacity of each platter determine the total capacity of the drive.

**Spindle:** A spindle connects all the platters and is connected to a motor. The motor of the spindle rotates with a constant speed. The disk platter spins at a speed of several thousands of revolutions per minute (rpm). Common spindle speeds are 5,400 rpm, 7,200 rpm, 10,000 rpm, and 15,000 rpm. The speed of the platter is increasing with improvements in technology; although, the extent to which it can be improved is limited.

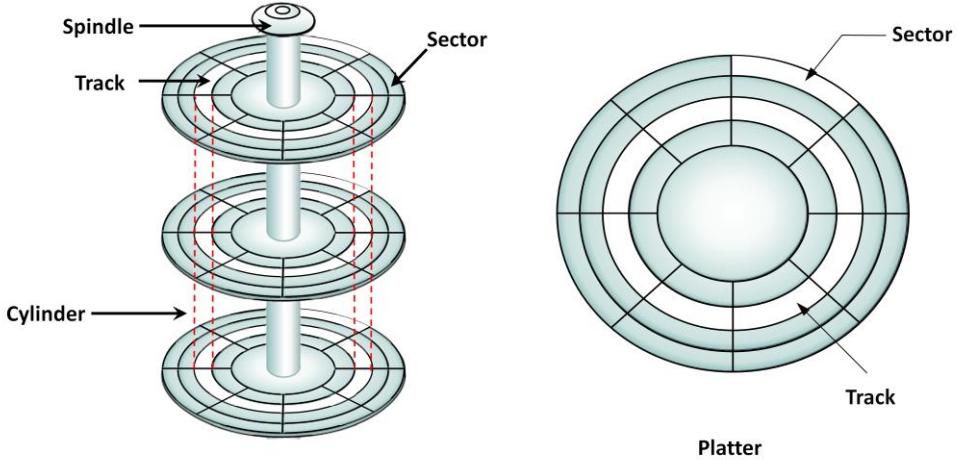
Cont...

**Read/Write Head:** Read/Write (R/W) heads, read and write data from or to platters. Drives have two R/W heads per platter, one for each surface of the platter. The R/W head changes the magnetic polarization on the surface of the platter when writing data. While reading data, the head detects the magnetic polarization on the surface of the platter. During reads and writes, the R/W head senses the magnetic polarization and never touches the surface of the platter. When the spindle is rotating, there is a microscopic air gap maintained between the R/W heads and the platters, known as the *head flying height*. This air gap is removed when the spindle stops rotating and the R/W head rests on a special area on the platter near the spindle. This area is called the *landing zone*. The landing zone is coated with a lubricant to reduce friction between the head and the platter. The logic on the disk drive ensures that heads are moved to the landing zone before they touch the surface. If the drive malfunctions and the R/W head accidentally touches the surface of the platter outside the landing zone, a *head crash* occurs. In a head crash, the magnetic coating on the platter is scratched and may cause damage to the R/W head. A head crash generally results in data loss.

**Actuator Arm Assembly:** R/W heads are mounted on the *actuator arm assembly*, which positions the R/W head at the location on the platter where the data needs to be written or read. The R/W heads for all platters on a drive are attached to one actuator arm assembly and move across the platters simultaneously.

**Drive Controller Board:** The controller is a printed circuit board, mounted at the bottom of a disk drive. It consists of a microprocessor, internal memory, circuitry, and firmware. The firmware controls the power to the spindle motor and the speed of the motor. It also manages the communication between the drive and the host. In addition, it controls the R/W operations by moving the actuator arm and switching between different R/W heads, and performs the optimization of data access.

## Physical Disk Structure



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 28

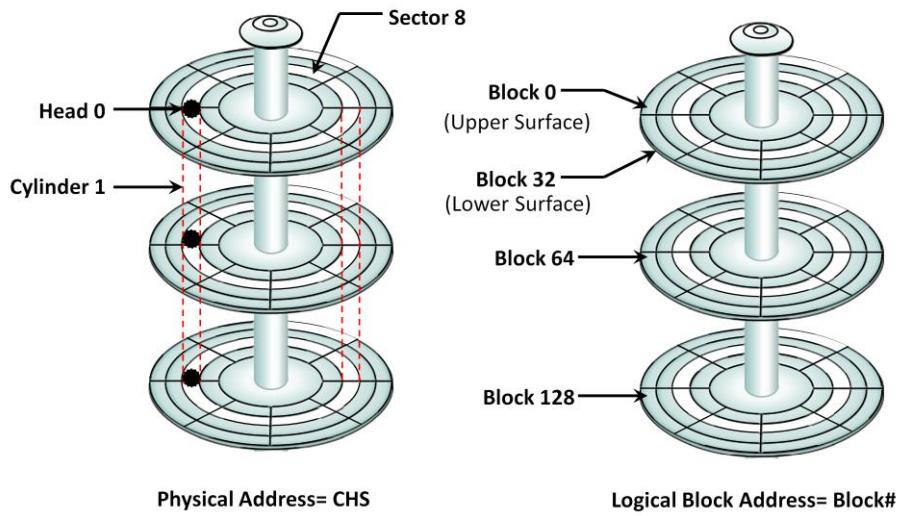
Data on the disk is recorded on *tracks*, which are concentric rings on the platter around the spindle. The tracks are numbered, starting from zero, from the outer edge of the platter. The number of *tracks per inch* (TPI) on the platter (or the *track density*) measures how tightly the tracks are packed on a platter.

Each track is divided into smaller units called *sectors*. A sector is the smallest, individually addressable unit of storage. The track and sector structure is written on the platter by the drive manufacturer using a low-level formatting operation. The number of sectors per track varies according to the drive type. The first personal computer disks had 17 sectors per track. Recent disks have a much larger number of sectors on a single track. There can be thousands of tracks on a platter, depending on the physical dimensions and recording density of the platter.

Typically, a sector holds 512 bytes of user data; although, some disks can be formatted with larger sector sizes. In addition to user data, a sector also stores other information, such as the sector number, head number or platter number, and track number. This information helps the controller to locate the data on the drive.

A cylinder is a set of identical tracks on both surfaces of each drive platter. The location of R/W heads is referred to by the cylinder number, not by the track number.

## Logical Block Addressing



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 29

Earlier drives used physical addresses consisting of the *cylinder, head, & sector (CHS)* number to refer to specific locations on the disk, and the host operating system had to be aware of the geometry of each disk used. *Logical block addressing (LBA)* has simplified the addressing by using a linear address to access physical blocks of data. The disk controller translates LBA to a CHS address, and the host needs to know only the size of the disk drive in terms of the number of blocks. The logical blocks are mapped to physical sectors on a 1:1 basis.

In the slide, the drive shows eight sectors per track, six heads, and four cylinders. This means a total of  $8 \times 6 \times 4 = 192$  blocks, so the block number ranges from 0 to 191. Each block has its own unique address.

Assuming that the sector holds 512 bytes, a 500-GB drive with a formatted capacity of 465.7 GB has in excess of 976,000,000 blocks.

## Disk Drive Performance

- Electromechanical device
  - ▶ Impacts the overall performance of the storage system
- Disk service time
  - ▶ Time taken by a disk to complete an I/O request, depends on:
    - ▶ Seek time
    - ▶ Rotational latency
    - ▶ Data transfer rate

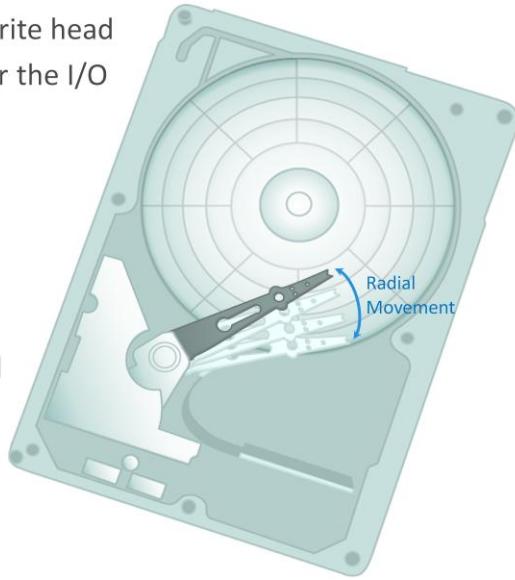
Disk service time = seek time + rotational latency + data transfer time

A disk drive is an electromechanical device that governs the overall performance of the storage system environment. The various factors that affect the performance of disk drives are:

- Seek time
- Rotational latency
- Data transfer rate

## Seek Time

- Time taken to position the read/write head
- The lower the seek time, the faster the I/O operation
- Seek time specifications include
  - ▶ Full stroke
  - ▶ Average
  - ▶ Track-to-track
- The seek time of a disk is specified by the drive manufacturer



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 31

The *seek time* (also called *access time*) describes the time taken to position the R/W heads across the platter with a radial movement (moving along the radius of the platter). In other words, it is the time taken to position and settle the arm and the head over the correct track. Therefore, the lower the seek time, the faster the I/O operation. Disk vendors publish the following seek time specifications:

**Full Stroke:** The time taken by the R/W head to move across the entire width of the disk, from the innermost track to the outermost track.

**Average:** The average time taken by the R/W head to move from one random track to another, normally listed as the time for one-third of a full stroke.

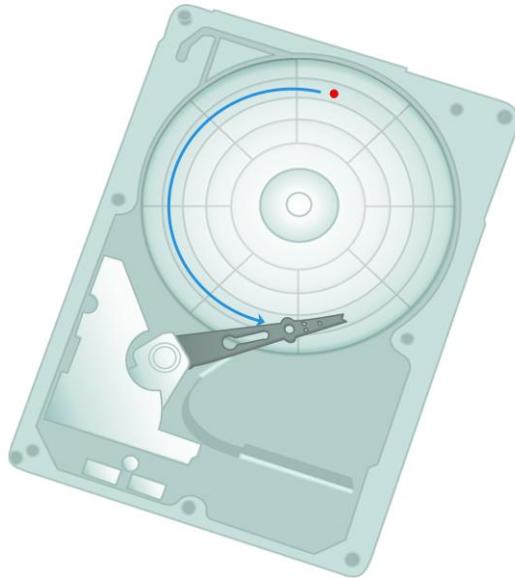
**Track-to-Track:** The time taken by the R/W head to move between adjacent tracks.

Each of these specifications is measured in milliseconds. The seek time of a disk is typically specified by the drive manufacturer. The average seek time on a modern disk is typically in the range of 3 to 15 milliseconds. Seek time has more impact on the I/O operation of random tracks rather than the adjacent tracks. To minimize the seek time, data can be written to only a subset of the available cylinders. This results in lower usable capacity than the actual capacity of the drive. For example, a 500-GB disk drive is set up to use only the first 40 percent of the cylinders and is effectively treated as a 200-GB drive. This is known as *short-stroking* the drive.

## Rotational Latency

- The time taken by the platter to rotate and position the data under the R/W head
- Depends on the rotation speed of the spindle
- Average rotational latency
  - ▶ One-half of the time taken for a full rotation
  - ▶ For 'X' rpm, drive latency is calculated in milliseconds as:

$$= \frac{1/2}{(X/60)}$$



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 32

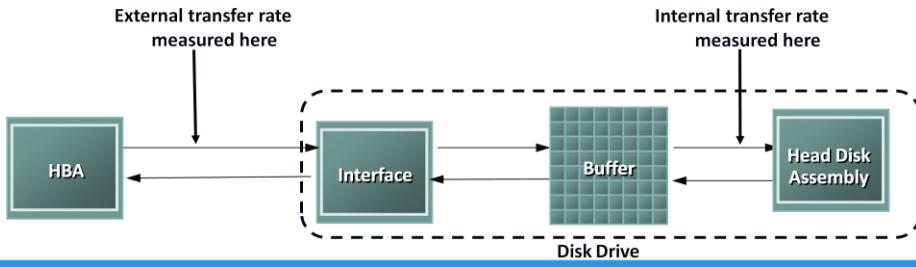
To access data, the actuator arm moves the R/W head over the platter to a particular track while the platter spins to position the requested sector under the R/W head. The time taken by the platter to rotate and position the data under the R/W head is called *rotational latency*. This latency depends on the rotation speed of the spindle and is measured in milliseconds. The average rotational latency is one-half of the time taken for a full rotation. Similar to the seek time, rotational latency has more impact on the reading/writing of random sectors on the disk than on the same operations on adjacent sectors.

Average rotational latency is approximately 5.5 ms for a 5,400-rpm drive, and around 2.0 ms for a 15,000-rpm (or 250-rps revolution per second) drive as shown here.

Av. rotational latency for 15K rpm or 250 rps (15000/60) drive is =  $(1/2)/250=2$  milliseconds

## Data Transfer Rate

- Average amount of data per unit time that the drive can deliver to the HBA
  - ▶ Internal transfer rate : Speed at which data moves from a platter's surface to the internal buffer of the disk
  - ▶ External transfer rate: Rate at which data move through the interface to the HBA



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 33

The *data transfer rate* (also called *transfer rate*) refers to the average amount of data per unit time that the drive can deliver to the HBA. In a *read operation*, the data first moves from disk platters to R/W heads; then it moves to the drive's internal *buffer*. Finally, data moves from the buffer through the interface to the host HBA. In a *write operation*, the data moves from the HBA to the internal buffer of the disk drive through the drive's interface. The data then moves from the buffer to the R/W heads. Finally, it moves from the R/W heads to the platters. The data transfer rates during the R/W operations are measured in terms of internal and external transfer rates, as shown in the slide.

*Internal transfer rate* is the speed at which data moves from a platter's surface to the internal buffer (cache) of the disk. The internal transfer rate takes into account factors such as the seek time and rotational latency. *External transfer rate* is the rate at which data can move through the interface to the HBA. The external transfer rate is generally the advertised speed of the interface, such as 133 MB/s for ATA. The sustained external transfer rate is lower than the interface speed.

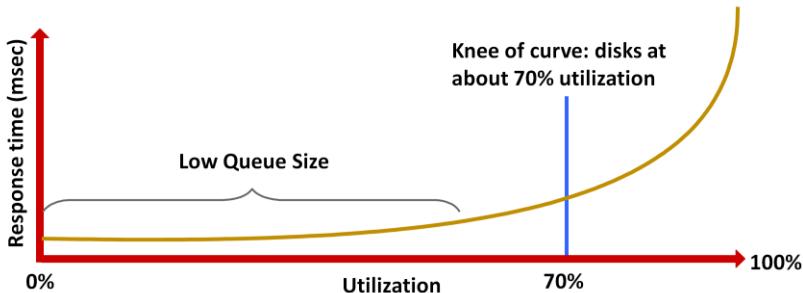
## I/O Controller Utilization Vs. Response Time

- Based on fundamental laws of disk drive performance:

$$\text{Av. Response Time} = \frac{\text{Service Time}}{(1 - \text{Utilization})}$$

Service time is time taken by the controller to serve an I/O

- For performance-sensitive applications disks are commonly utilized below 70% of their I/O serving capability



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 34

Utilization of a disk I/O controller has a significant impact on the I/O response time. Consider that a disk is viewed as a black box consisting of two elements queue and disk I/O controller. *Queue* is the location where an I/O request waits before it is processed by the I/O controller and *disk I/O controller* processes I/Os waiting in the queue one by one.

The I/O requests arrive at the controller at the rate generated by the application. The I/O arrival rate, the queue length, and the time taken by the I/O controller to process each request determines the I/O response time. If the controller is busy or heavily utilized, the queue size will be large and the response time will be high. Based on the fundamental laws of disk drive performance, the relationship between controller utilization and average response time is given as:

$$\text{Average response time} = \text{Service time} / (1 - \text{Utilization})$$

where, service time is the time taken by the controller to serve an I/O.

As the utilization reaches 100 percent that is, as the I/O controller saturates, the response time is closer to infinity. In essence, the saturated component, or the bottleneck, forces the serialization of I/O requests; meaning, each I/O request must wait for the completion of the I/O requests that preceded it. Figure in the slide shows a graph plotted between utilization and response time. The graph indicates that the response time changes are nonlinear as the utilization increases. When the average queue sizes are low, the response time remains low. The response time increases slowly with added load on the queue and increases exponentially when the utilization exceeds 70 percent. Therefore, for performance-sensitive applications, it is common to utilize disks below their 70 percent of I/O serving capability.

## Storage Design Based on Application Requirements and Disk Drive Performance

- Disks required to meet an application's capacity need ( $D_C$ ):

$$D_C = \frac{\text{Total capacity required}}{\text{Capacity of a single disk}}$$

- Disks required to meet application's performance need ( $D_P$ ):

$$D_P = \frac{\text{IOPS generated by an application at peak workload}}{\text{IOPS serviced by single disk}}$$

- IOPS serviced by a disk (S) depends upon disk service time ( $T_S$ ):

$$T_S = \text{Seek time} + \frac{0.5}{(\text{Disk rpm}/60)} + \frac{\text{Data block size}}{\text{Data transfer rate}}$$

►  $T_S$  is time taken for an I/O to complete, therefore IOPS serviced by a disk (S) is equal to  $(1/T_S)$

► For performance sensitive application  $(S) = 0.7 \times \frac{1}{T_S}$

Disk required for an application =  $\max(D_C, D_P)$

Determining storage requirements for an application begins with determining the required storage capacity and I/O performance. Capacity can be easily estimated by the size and number of file systems and database components used by applications. The I/O size, I/O characteristics, and the number of I/Os generated by the application at peak workload are other factors that affect performance, I/O response time and design of storage system.

The disk service time ( $T_S$ ) for an I/O is a key measure of disk performance;  $T_S$ , along with disk utilization rate (U), determines the I/O response time for an application. As discussed earlier the total disk service time is the sum of the seek time, rotational latency, and transfer time.

Note that transfer time is calculated based on the block size of the I/O and given data transfer rate of a disk drive—for example, an I/O with a block size of 32 KB and given disk data transfer rate 40MB/s; the transfer time will be 32 KB / 40 MB.

$T_S$  determines the time taken by the I/O controller to serve an I/O, therefore, the maximum number of I/Os serviced per second or IOPS is  $(1/T_S)$ .

The IOPS calculated above represents the IOPS that can be achieved at potentially high levels of I/O controller utilization (close to 100 percent). If the application demands a faster response time, then the utilization for the disks should be maintained below 70 percent.

Based on this discussion, the total number of disks required for an application is computed as :

= Max (Disks required for meeting capacity, Disks required for meeting performance)

Cont..

Consider an example in which the capacity requirement for an application is 1.46 TB. The number of IOPS generated by the application at peak workload is estimated at 9,000 IOPS. The vendor specifies that a 146-GB, 15,000-rpm drive is capable of doing a maximum 180 IOPS.

In this example, the number of disks required to meet the capacity requirements will be  $1.46 \text{ TB} / 146 \text{ GB} = 10$  disks.

To meet the application IOPS requirements, the number of disks required is  $9,000 / 180 = 50$ . However, if the application is response-time sensitive, the number of IOPS a disk drive can perform should be calculated based on 70-percent disk utilization. Considering this, the number of IOPS a disk can perform at 70 percent utilization is  $180 \times 0.7 = 126$  IOPS. Therefore, the number of disks required to meet the application IOPS requirement will be  $9,000 / 126 = 72$ .

As a result, the number of disks required to meet the application requirements will be  $\text{Max}(10, 72) = 72$  disks.

The preceding example indicates that from a capacity-perspective, 10 disks are sufficient; however, the number of disks required to meet application performance is 72. To optimize disk requirements from a performance perspective, various solutions are deployed in a real-time environment. Examples of these solutions are disk native command queuing, use of flash drives, RAID, and the use of cache memory. RAID and cache are detailed in module 3 and 4 respectively.

## Enterprise Flash Drives

Conventional Hard Drives	Flash Drives
Mechanical delay due to seek time and rotational latency	Highest possible throughput per drive due to no mechanical movement
Limited performance and I/O serving capability	Very low latency per I/O and consistent I/O performance
More power consumption due to mechanical operations	High Energy efficiency <ul style="list-style-type: none"><li>• Lower power requirement per GB</li><li>• Lower power requirement per IOPS</li></ul>
Low mean time between failure (MTBF)	High reliability due to no moving parts
Higher TCO due to more number of disks, power, cooling, and management cost	Overall less TCO

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

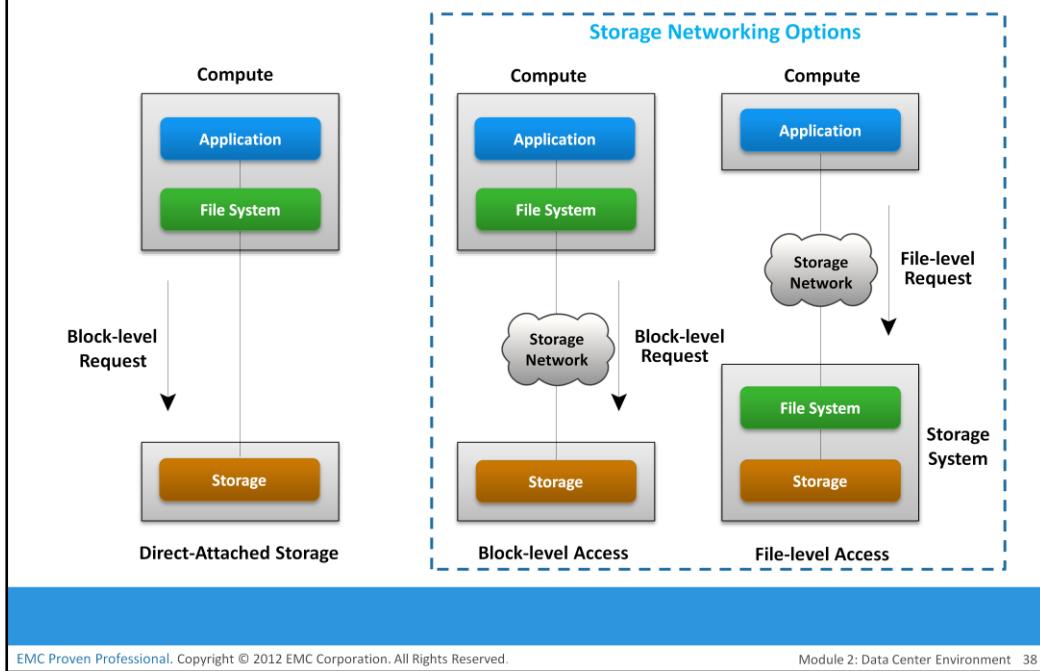
Module 2: Data Center Environment 37

Traditionally, high I/O requirements of an application were met by simply using more disks. Availability of enterprise class flash drives (EFD) has changed the scenario.

Flash drives, also referred as solid state drives (SSDs), are new generation drives that deliver ultra-high performance required by performance-sensitive applications. Flash drives use semiconductor-based solid state memory (flash memory) to store and retrieve data. Unlike conventional mechanical disk drives, flash drives contain no moving parts; therefore, they do not have seek and rotational latencies. Flash drives deliver a high number of IOPS with very low response times. Also, being a semiconductor-based device, flash drives consume less power, compared to mechanical drives. Flash drives are especially suited for applications with small block size and random-read workloads that require consistently low (less than 1 ms) response times. Applications that need to process massive amounts of information quickly, such as currency exchange, electronic trading systems, and real-time data feed processing, benefit from flash drives.

Overall, flash drives provide better total cost of ownership (TCO) even though they cost more on \$/GB basis. By implementing flash drives, businesses can meet application performance requirements with far fewer drives (approximately 20 to 30 times less number of drives compared to conventional mechanical drives). This reduction not only provides savings in terms of drive cost, but also translates to savings for power, cooling, and space consumption. Fewer numbers of drives in the environment also means less cost for managing the storage.

## Host Access to Storage



Data is accessed and stored by applications using the underlying infrastructure. The key components of this infrastructure are the operating system (or file system), connectivity, and storage. The storage device can be internal and (or) external to the host. In either case, the host controller card accesses the storage devices using predefined protocols, such as IDE/ATA, SCSI, or Fibre Channel (FC). IDE/ATA and SCSI are popularly used in small and personal computing environments for accessing internal storage. FC and iSCSI protocols are used for accessing data from an external storage device (or subsystems). External storage devices can be connected to the host directly or through the storage network. When the storage is connected directly to the host, it is referred as Direct-Attached Storage (DAS).

Data can be accessed over a network in one of the following ways: block level, file level, or object level. In general, the application requests data from the file system (or operating system) by specifying the filename and location. The file system maps the file attributes to the logical block address of the data and sends the request to the storage device. The storage device converts the logical block address (LBA) to a cylinder-head-sector (CHS) address and fetches the data.

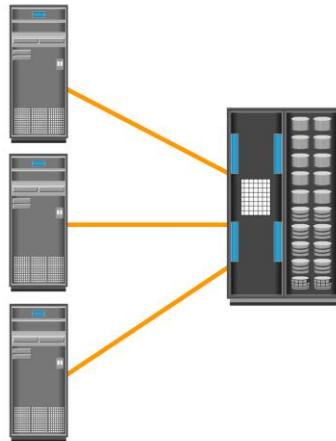
In a block-level access, the file system is created on a host, and data is accessed on a network at the block level. In this case, raw disks or logical volumes are assigned to the host for creating the file system.

In a file-level access, the file system is created on a separate file server or at the storage side, and the file-level request is sent over a network. Because data is accessed at the file level, this method has higher overhead, as compared to the data accessed at the block level. Object-level access is an intelligent evolution, whereby data is accessed over a network in terms of self-contained objects with a unique object identifier. Details of storage networking technologies and deployments are covered in later modules of this course.

## Direct-Attached Storage (DAS)



Internal Direct Connect



External Direct Connect

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 2: Data Center Environment 39

DAS is an architecture in which storage is connected directly to the hosts. The internal disk drive of a host and the directly connected external storage array are examples of DAS. Although the implementation of storage networking technologies is gaining popularity, DAS has remained suitable for localized data access in a small environment, such as personal computing and workgroups. DAS is classified as internal or external, based on the location of the storage device with respect to the host.

In *internal DAS* architectures, the storage device is internally connected to the host by a serial or parallel bus. The physical bus has distance limitations and can be sustained only over a shorter distance for high-speed connectivity. In addition, most internal buses can support only a limited number of devices, and they occupy a large amount of space inside the host, making maintenance of other components difficult. In *external DAS* architectures, the host connects directly to the external storage device, and data is accessed at the block level. In most cases, communication between the host and the storage device takes place over a SCSI or FC protocol. Compared to internal DAS, an external DAS overcomes the distance limitations and provides centralized management of storage devices.

**DAS Benefits and limitations:** DAS requires a relatively lower initial investment than storage networking architectures. The DAS configuration is simple and can be deployed easily and rapidly. It requires fewer management tasks and less hardware and software elements to set up and operate. However, DAS does not scale well. A storage array has a limited number of ports, which restricts the number of hosts that can directly connect to the storage. Therefore, DAS does not make optimal use of resources and, moreover unused resources cannot be easily re-allocated, resulting in islands of over-utilized and under-utilized storage pools.

## Module 2: Data Center Environment

### Concept in Practice

- VMware ESXi

The concept in practice covers the product example of compute virtualization. It covers industry's leading hypervisor software VMware ESXi.

## VMware ESXi

- Industry's leading hypervisor
  - ▶ Enable virtualization of x86 hardware platforms
- Physical machine that houses ESXi is called ESXi host
  - ▶ ESXi host abstracts physical compute resources to run multiple VMs concurrently on same physical server
- Two Components
  - ▶ VMKernel
    - ▶ Work similar to OS – responsible for process creation, resource scheduling, and so on
  - ▶ Virtual machine monitor
    - ▶ Performs binary translation for privileged OS instructions that can not be virtualized

VMware is the leading provider for server virtualization solution. VMware ESXi provides a platform called hypervisor. The hypervisor abstracts CPU, memory, and storage resources to run multiple virtual machines concurrently on the same physical server.

VMware ESXi is a hypervisor that installs on x86 hardware to enable server virtualization. It enables creating multiple virtual machines (VMs) that can run simultaneously on the same physical machine. A VM is a discrete set of files that can be moved, copied, and used as a template. All the files that make up a VM are typically stored in a single directory on a cluster file system called Virtual Machine File System (VMFS). The physical machine that houses ESXi is called ESXi host. The ESXi hosts provide physical resources used to run virtual machines. ESXi has two key components: VMkernel and Virtual Machine Monitor.

VMkernel provides functionality similar to that found in other operating systems, such as process creation, file system management, and process scheduling. It is designed to specifically support running multiple VMs and provide core functionality such as resource scheduling, I/O stacks, and so on.

The virtual machine monitor is responsible for executing commands on the CPUs and performing Binary Translation (BT). A virtual machine monitor performs hardware abstraction to appear as a physical machine with its own CPU, memory, and I/O devices. Each VM is assigned a virtual machine monitor that has a share of the CPU, memory, and I/O devices to successfully run the VM.

## Module 2: Summary

Key points covered in this module:

- Key data center elements
- Application and compute virtualization
- Disk drive components and performance
- Enterprise flash drives
- Host access to storage

This module covered the key elements of a data center – application, DBMS, compute, network, and storage.

It also covered virtualization at application and compute that enable better utilization of resources and ease of management. This module also elaborated on disk drive components and factors governing disk drive performance. It also covered enterprise flash drives that are superior to mechanical disk drives in many ways. This module also covered various options of host to storage access with focus on DAS.

## Check Your Knowledge – 1

- Which is a benefit of compute virtualization?
  - A. Enables compute memory swapping
  - B. Improves compute utilization
  - C. Isolates compute memory from the applications
  - D. Isolates OS from the applications
- What best describes virtual machines (VMs)?
  - A. All VMs on a physical server must run same OS
  - B. VM files are deleted when VM is powered off
  - C. VMs are discrete sets of files
  - D. All VMs share available resources equally

## Check Your Knowledge – 2

- What is concatenation?
  - A. Grouping multiple physical drives into a logical drive
  - B. Dividing a physical drive into multiple logical drives
  - C. Process of writing disk metadata on a logical drive
  - D. Adding more capacity to a physical drive through de-fragmentation
- Which factors contribute to the overall service time of a mechanical disk?
  - A. Disk buffer time, full stroke, and rotation latency
  - B. Internal transfer rate, external transfer rate, and buffer time
  - C. Full stroke, average seek time, and track-to-track seek time
  - D. Average seek time, rotational latency, and data transfer rate

## Check Your Knowledge – 3

- Which is a challenge of DAS environment?
  - A. Low performance
  - B. Limited scalability
  - C. Deployment complexity
  - D. High initial investment

## Exercise: Design Storage Solution for New Application

- Scenario
  - ▶ Characteristics of new application:
    - ▶ Require 1TB of storage capacity
    - ▶ Peak I/O workload 4900 IOPS
    - ▶ Typical I/O size is 4KB
  - ▶ Specifications of the available disk drives:
    - ▶ 15K rpm drive with storage capacity = 100 GB
    - ▶ Average seek time = 5ms
    - ▶ Data transfer rate = 40 MB/sec
  - ▶ As it is business critical application, response time must be within acceptable range
- Task
  - ▶ Calculate the number of disks required for the application

An organization is deploying a new business application in their environment. The new application requires 1TB of storage space for business and application data. During peak workload, application is expected to generate 4900 IOPS (I/O per second) with typical I/O size of 4KB.

The available disk drive option is 15,000 rpm drive with 100 GB capacity. Other specification of the drives are:

Av. Seek time = 5 millisecond and data transfer rate = 40MB/sec.

You are asked to calculate the required number of disk drives that can meet both capacity and performance requirements of an application.

# Module – 3

# Data Protection – RAID



## Module 3: Data Protection – RAID

Upon completion of this module, you should be able to:

- Describe RAID implementation methods
- Describe the three RAID techniques
- Describe commonly used RAID levels
- Describe the impact of RAID on performance
- Compare RAID levels based on their cost, performance, and protection

This module focuses on RAID and its use to improve performance and protection. It details on various RAID implementations, techniques, and levels commonly used. This module also describes the impact of RAID on performance and compares the commonly used RAID levels.

## Module 3: Data Protection – RAID

### Lesson 1: RAID Overview

During this lesson the following topics are covered:

- RAID Implementation methods
- RAID array components
- RAID techniques

This lesson focuses on RAID implementation methods and RAID array components. This lesson also focuses on various RAID techniques.

## Why RAID?

### RAID

It is a technique that combines multiple disk drives into a logical unit (RAID set) and provides protection, performance, or both.

- Due to mechanical components in a disk drive it offers limited performance
- An individual drive has a certain life expectancy and is measured in MTBF:
  - ▶ For example: If the MTBF of a drive is 750,000 hours, and there are 1000 drives in the array, then the MTBF of the array is 750 hours ( $750,000/1000$ )
- RAID was introduced to mitigate these problems

Today's data centers house hundreds of disk drives in their storage infrastructure. Disk drives are inherently susceptible to failures due to mechanical wear and tear and other environmental factors, which could result in data loss. The greater the number of disk drives in a storage array, the greater the probability of a disk failure in the array. For example, consider a storage array of 100 disk drives, each with an average life expectancy of 750,000 hours. The average life expectancy of this collection in the array, therefore, is  $750,000/100$  or 7,500 hours. This means that a disk drive in this array is likely to fail at least once in 7,500 hours.

RAID is an enabling technology that leverages multiple drives as part of a set that provides data protection against drive failures. In general, RAID implementations also improve the storage system performance by serving I/Os from multiple disks simultaneously. Modern arrays with flash drives also benefit in terms of protection and performance by using RAID.

In 1987, Patterson, Gibson, and Katz at the University of California, Berkeley, published a paper titled "A Case for Redundant Arrays of Inexpensive Disks (RAID)." This paper described the use of small-capacity, inexpensive disk drives as an alternative to large-capacity drives common on mainframe computers. The term *RAID* has been redefined to refer to *independent* disks to reflect advances in the storage technology. RAID technology has now grown from an academic concept to an industry standard and is common implementation in today's storage arrays.

## RAID Implementation Methods

- Software RAID implementation
  - ▶ Uses host-based software to provide RAID functionality
  - ▶ Limitations
    - ▶ Use host CPU cycles to perform RAID calculations, hence impact overall system performance
    - ▶ Support limited RAID levels
    - ▶ RAID software and OS can be upgraded only if they are compatible
- Hardware RAID Implementation
  - ▶ Uses a specialized hardware controller installed either on a host or on an array

There are two methods of RAID implementation, hardware and software. Both have their advantages and disadvantages.

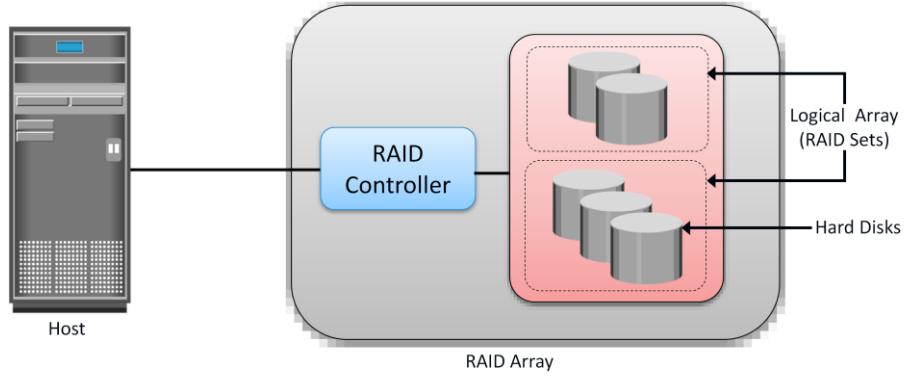
*Software RAID* uses host-based software to provide RAID functions and is implemented at the operating-system level. Software RAID implementations offer cost and simplicity benefits when compared with hardware RAID. However, they have the following limitations:

- Performance: Software RAID affects overall system performance. This is due to additional CPU cycles required to perform RAID calculations.
- Supported features: Software RAID does not support all RAID levels.
- Operating system compatibility: Software RAID is tied to the host operating system; hence, upgrades to software RAID or to the operating system should be validated for compatibility. This leads to inflexibility in the data-processing environment.

In hardware RAID implementations, a specialized hardware controller is implemented either on the host or on the array. Controller card RAID is a host-based hardware RAID implementation in which a specialized RAID controller is installed in the host, and disk drives are connected to it. Manufacturers also integrate RAID controllers on motherboards. A host-based RAID controller is not an efficient solution in a data center environment with a large number of hosts. The external RAID controller is an array-based hardware RAID. It acts as an interface between the host and disks. It presents storage volumes to the host, and the host manages these volumes as physical drives. The key functions of the RAID controllers are as follows:

- Management and control of disk aggregations
- Translation of I/O requests between logical disks and physical disks
- Data regeneration in the event of disk failures

## RAID Array Components



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 6

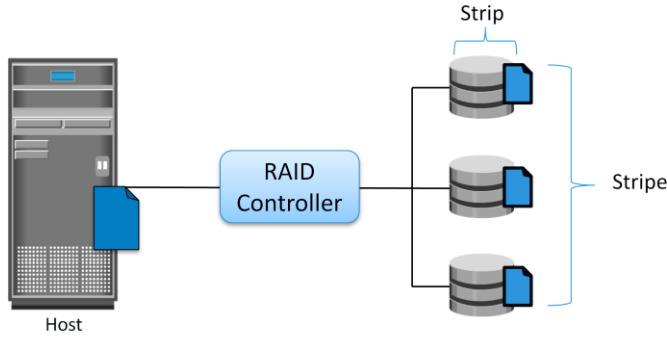
A *RAID array* is an enclosure that contains a number of disk drives and supporting hardware to implement RAID. A subset of disks within a RAID array can be grouped to form logical associations called logical arrays, also known as a *RAID set* or a *RAID group*.

## RAID Techniques

- Three key techniques used for RAID are:
  - ▶ Striping
  - ▶ Mirroring
  - ▶ Parity

RAID techniques – striping, mirroring, and parity – form the basis for defining various RAID levels. These techniques determine the data availability and performance characteristics of a RAID set.

## RAID Technique – Striping



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 8

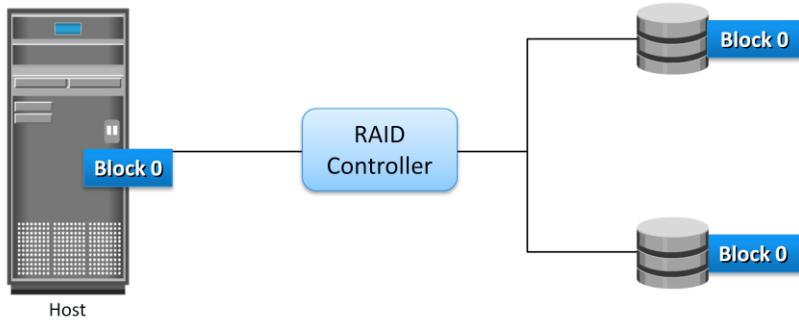
Striping is a technique of spreading data across multiple drives (more than one) in order to use the drives in parallel. All the read-write heads work simultaneously, allowing more data to be processed in a shorter time and increasing performance, compared to reading and writing from a single disk.

Within each disk in a RAID set, a predefined number of contiguously addressable disk blocks are defined as strip. The set of aligned strips that spans across all the disks within the RAID set is called a stripe. Figure on the slide shows physical and logical representations of a striped RAID set.

*Strip size* (also called *stripe depth*) describes the number of blocks in a *strip*, and is the maximum amount of data that can be written to or read from a single disk in the set, assuming that the accessed data starts at the beginning of the strip. All strips in a stripe have the same number of blocks. Having a smaller strip size means that the data is broken into smaller pieces while spread across the disks.

Stripe size is a multiple of strip size by the number of data disks in the RAID set. For example, in a five disk striped RAID set with a strip size of 64KB, the stripe size is 320 KB (64KB x 5). *Stripe width* refers to the number of data strips in a stripe. Striped RAID does not provide any data protection unless parity or mirroring is used.

## RAID Technique – Mirroring



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 9

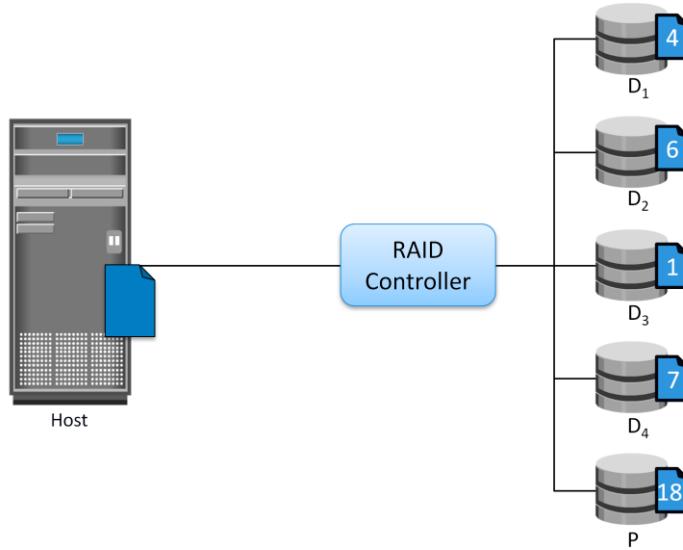
Mirroring is a technique whereby the same data is stored on two different disk drives, yielding two copies of the data. If one disk drive failure occurs, the data is intact on the surviving disk drive and the controller continues to service the host's data requests from the surviving disk of a mirrored pair.

When the failed disk is replaced with a new disk, the controller copies the data from the surviving disk of the mirrored pair. This activity is transparent to the host.

In addition to providing complete data redundancy, mirroring enables fast recovery from disk failure. However, disk mirroring provides only data protection and is not a substitute for data backup. Mirroring constantly captures changes in the data, whereas a backup captures point-in-time images of the data.

Mirroring involves duplication of data—the amount of storage capacity needed is twice the amount of data being stored. Therefore, mirroring is considered expensive and is preferred for mission-critical applications that cannot afford the risk of any data loss. Mirroring improves read performance because read requests can be serviced by both disks. However, write performance is slightly lower than that in a single disk because each write request manifests as two writes on the disk drives. Mirroring does not deliver the same levels of write performance as a striped RAID.

## RAID Technique – Parity



*Actual parity calculation is a bitwise XOR operation*

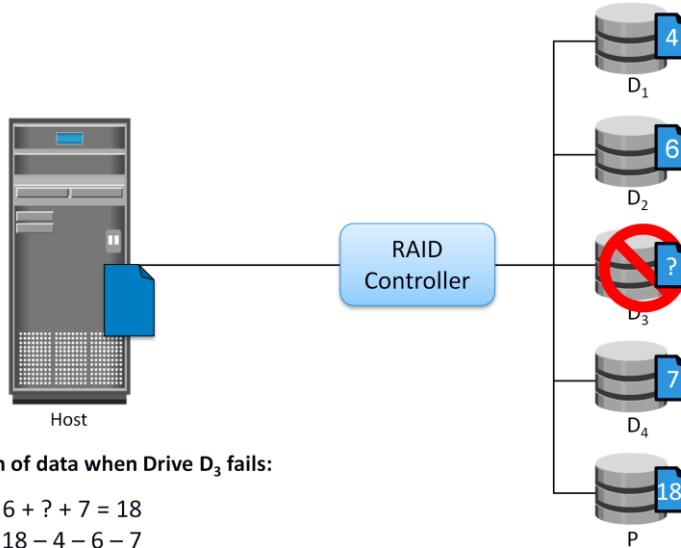
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 10

Parity is a method to protect striped data from disk drive failure without the cost of mirroring. An additional disk drive is added to hold parity, a mathematical construct that allows re-creation of the missing data. Parity is a redundancy technique that ensures protection of data without maintaining a full set of duplicate data. Calculation of parity is a function of the RAID controller.

Parity information can be stored on separate, dedicated disk drives or distributed across all the drives in a RAID set. The first four disks in the figure, labeled  $D_1$  to  $D_4$ , contain the data. The fifth disk, labeled  $P$ , stores the parity information, which, in this case, is the sum of the elements in each row.

## Data Recovery in Parity Technique



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 11

Now, if one of the data disks fails, the missing value can be calculated by subtracting the sum of the rest of the elements from the parity value. Here, for simplicity, the computation of parity is represented as an arithmetic sum of the data. However, parity calculation is a *bitwise XOR* operation.

Compared to mirroring, parity implementation considerably reduces the cost associated with data protection. Consider an example of a parity RAID configuration with five disks where four disks hold data, and the fifth holds the parity information. In this example, parity requires only 25 percent extra disk space compared to mirroring, which requires 100 percent extra disk space. However, there are some disadvantages of using parity. Parity information is generated from data on the data disk. Therefore, parity is recalculated every time there is a change in data. This recalculation is time-consuming and affects the performance of the RAID array.

For parity RAID, the stripe size calculation does not include the parity strip. For example in a five (4 + 1) disk parity RAID set with a strip size of 64 KB, the stripe size will be 256 KB (64 KB x 4).

## Module 3: Data Protection – RAID

### Lesson 2: RAID Levels

During this lesson the following topics are covered:

- Commonly used RAID levels
- RAID impacts on performance
- RAID comparison
- Hot spare

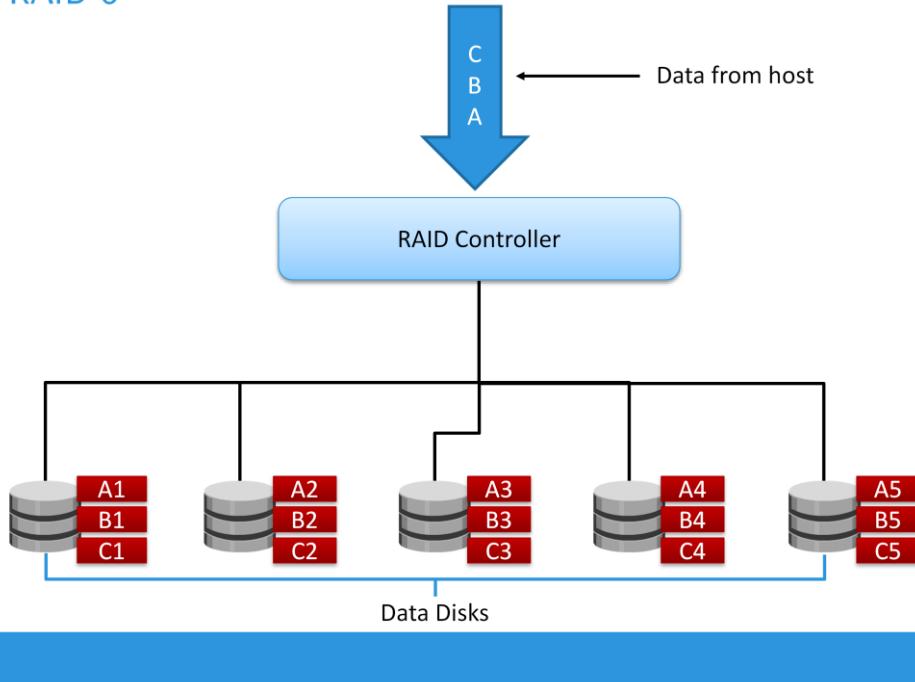
This lesson focuses on commonly used RAID levels and their comparisons. This lesson also focuses on Hot spare.

## RAID Levels

- Commonly used RAID levels are:
  - ▶ RAID 0 – Striped set with no fault tolerance
  - ▶ RAID 1 – Disk mirroring
  - ▶ RAID 1 + 0 – Nested RAID
  - ▶ RAID 3 – Striped set with parallel access and dedicated parity disk
  - ▶ RAID 5 – Striped set with independent disk access and a distributed parity
  - ▶ RAID 6 – Striped set with independent disk access and dual distributed parity

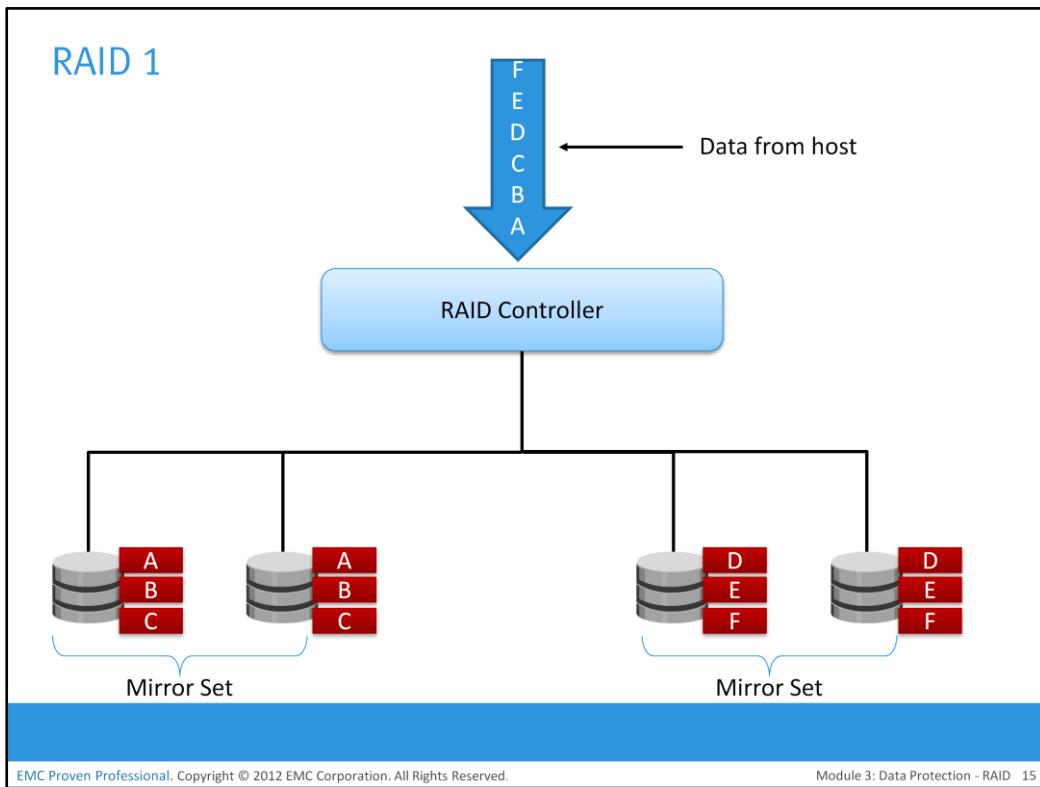
Application performance, data availability requirements, and cost determine the RAID level selection. These RAID levels are defined on the basis of striping, mirroring, and parity techniques. Some RAID levels use a single technique, whereas others use a combination of techniques. The commonly used RAID levels are listed on the slide.

## RAID 0



RAID 0 configuration uses data striping techniques, where data is striped across all the disks within a RAID set. Therefore it utilizes the full storage capacity of a RAID set. To read data, all the strips are put back together by the controller. When the number of drives in the RAID set increases, performance improves because more data can be read or written simultaneously. RAID 0 is a good option for applications that need high I/O throughput. However, if these applications require high availability during drive failures, RAID 0 does not provide data protection and availability.

## RAID 1

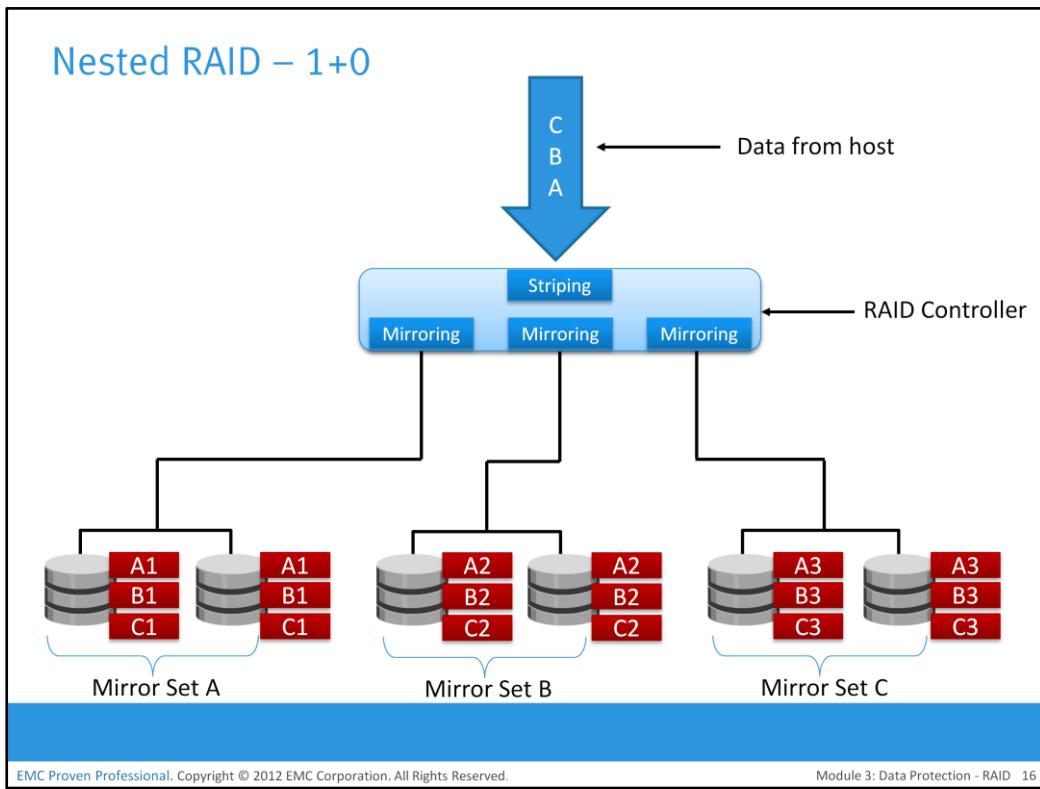


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 15

RAID 1 is based on the mirroring technique. In this RAID configuration, data is mirrored to provide fault tolerance. A RAID 1 set consists of two disk drives and every write is written to both disks. The mirroring is transparent to the host. During disk failure, the impact on data recovery in RAID 1 is the least among all RAID implementations. This is because the RAID controller uses the mirror drive for data recovery. RAID 1 is suitable for applications that require high availability and cost is no constraint.

## Nested RAID – 1+0



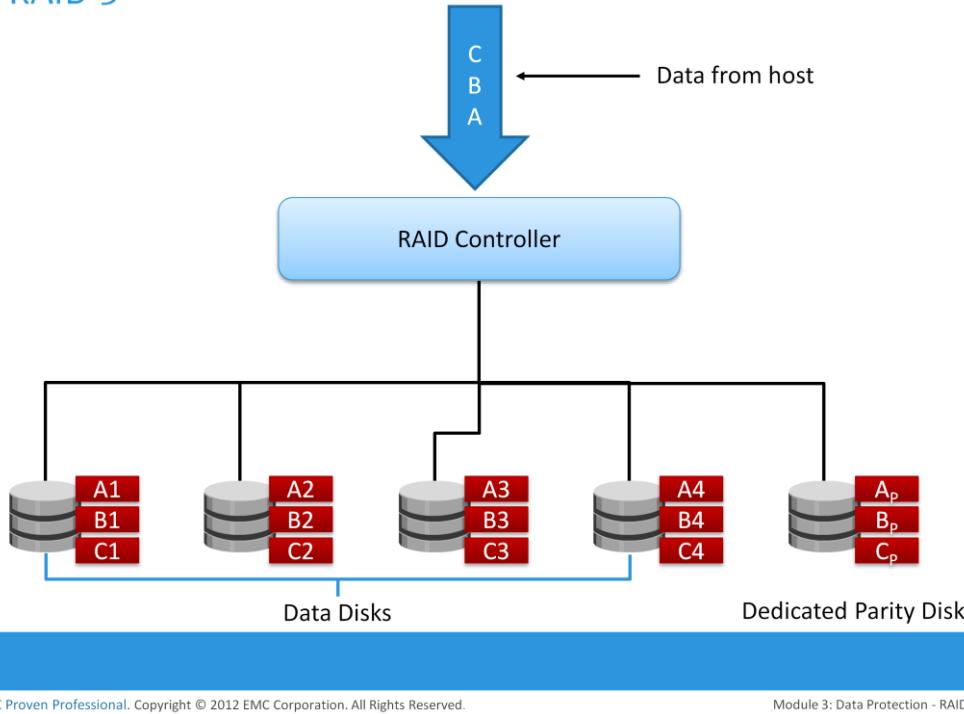
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 16

Most data centers require data redundancy and performance from their RAID arrays. RAID 1+0 combines the performance benefits of RAID 0 with the redundancy benefits of RAID 1. It uses mirroring and striping techniques and combine their benefits. This RAID type requires an even number of disks, the minimum being four.

RAID 1+0 is also known as RAID 10 (Ten) or RAID 1/0. RAID 1+0 is also called striped mirror. The basic element of RAID 1+0 is a mirrored pair, which means that data is first mirrored and then both copies of the data are striped across multiple disk drive pairs in a RAID set. When replacing a failed drive, only the mirror is rebuilt. In other words, the disk array controller uses the surviving drive in the mirrored pair for data recovery and continuous operation. Data from the surviving disk is copied to the replacement disk.

## RAID 3



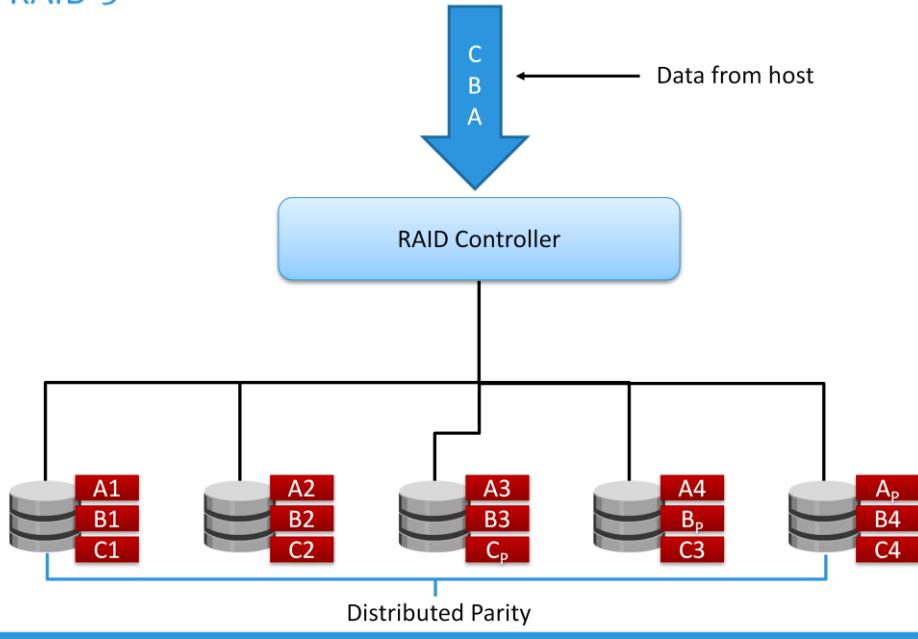
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 17

RAID 3 stripes data for performance and uses parity for fault tolerance. Parity information is stored on a dedicated drive so that the data can be reconstructed if a drive fails in a RAID set. For example, in a set of five disks, four are used for data and one for parity. Therefore, the **total disk space required is 1.25 times the size of the data disks**. RAID 3 always reads and writes complete stripes of data across all disks because the drives operate in parallel. There are no partial writes that update one out of many strips in a stripe.

Similar to RAID 3, RAID 4 stripes data for high performance and uses parity for improved fault tolerance. Data is striped across all disks except the parity disk in the array. Parity information is stored on a dedicated disk so that the data can be rebuilt if a drive fails. Unlike RAID 3, data disks in RAID 4 can be accessed independently so that specific data elements can be read or written on a single disk without reading or writing an entire stripe. RAID 4 provides good read throughput and reasonable write throughput.

## RAID 5

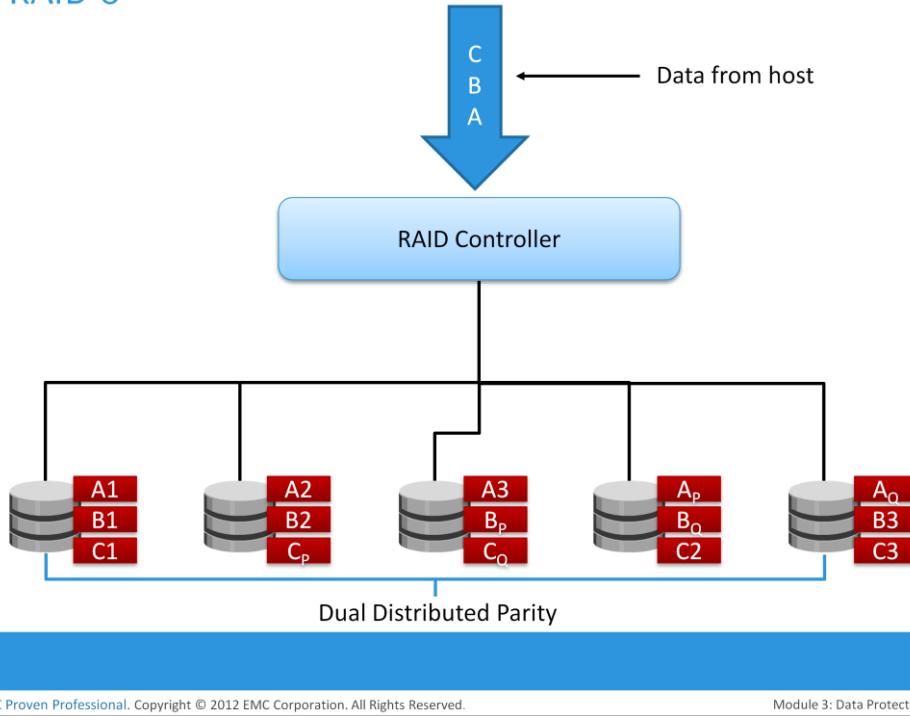


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 18

RAID 5 is a versatile RAID implementation. It is similar to RAID 4 because it uses striping. The drives (strips) are also independently accessible. The difference between RAID 4 and RAID 5 is the parity location. In RAID 4, parity is written to a dedicated drive, creating a write bottleneck for the parity disk. In RAID 5, parity is distributed across all disks to overcome the write bottleneck of a dedicated parity disk.

## RAID 6

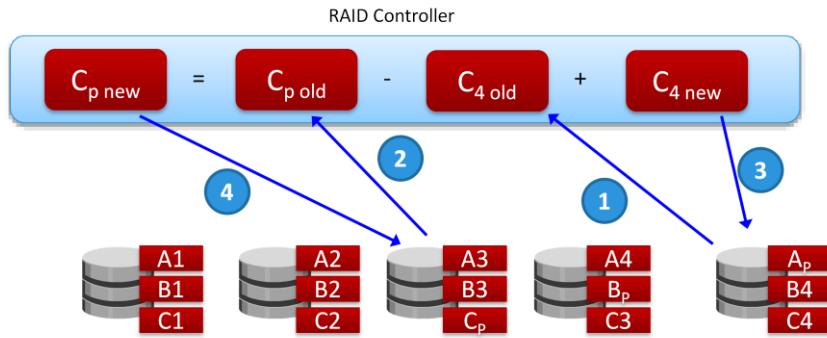


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 19

RAID 6 works the same way as RAID 5, except that RAID 6 includes a second parity element to enable survival if two disk failures occur in a RAID set. Therefore, a RAID 6 implementation requires at least four disks. RAID 6 distributes the parity across all the disks. The write penalty (explained later in this module) in RAID 6 is more than that in RAID 5; therefore, RAID 5 writes perform better than RAID 6. The rebuild operation in RAID 6 may take longer than that in RAID 5 due to the presence of two parity sets.

## RAID Impacts on Performance



- In RAID 5, every write (update) to a disk manifests as four I/O operations (2 disk reads and 2 disk writes)
- In RAID 6, every write (update) to a disk manifests as six I/O operations (3 disk reads and 3 disk writes)
- In RAID 1, every write manifests as two I/O operations (2 disk writes)

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 20

When choosing a RAID type, it is imperative to consider its impact on disk performance and application IOPS. In both mirrored and parity RAID configurations, every write operation translates into more I/O overhead for the disks, which is referred to as a *write penalty*. In a RAID 1 implementation, every write operation must be performed on two disks configured as a mirrored pair, whereas in a RAID 5 implementation, a write operation may manifest as four I/O operations. When performing I/Os to a disk configured with RAID 5, the controller has to read, recalculate, and write a parity segment for every data write operation.

This slide illustrates a single write operation on RAID 5 that contains a group of five disks. The parity (P) at the controller is calculated as follows:

$$C_p = C_1 + C_2 + C_3 + C_4 \text{ (XOR operations)}$$

Whenever the controller performs a write I/O, parity must be computed by reading the old parity ( $C_p$  old) and the old data ( $C_4$  old) from the disk, which means two read I/Os. Then, the new parity ( $C_p$  new) is computed as follows:

$$C_p \text{ new} = C_p \text{ old} - C_4 \text{ old} + C_4 \text{ new} \text{ (XOR operations)}$$

After computing the new parity, the controller completes the write I/O by writing the new data and the new parity onto the disks, amounting to two write I/Os. Therefore, the controller performs two disk reads and two disk writes for every write operation, and the write penalty is 4.

In RAID 6, which maintains dual parity, a disk write requires three read operations: two parity and one data. After calculating both new parities, the controller performs three write operations: two parity and an I/O. Therefore, in a RAID 6 implementation, the controller performs six I/O operations for each write I/O, and the write penalty is 6.

## RAID Penalty Calculation Example

- Total IOPS at peak workload is 1200
- Read/Write ratio 2:1
- Calculate disk load at peak activity for:
  - ▶ RAID 1/0
  - ▶ RAID 5

Consider an application that generates 1200 IOPS at peak workload, with read/write ratio of 2:1. Calculate disk load at peak activity for RAID 1/0 and RAID 5 configuration.

## Solution: RAID Penalty

- For RAID 1/0, the disk load (read + write)  
$$\begin{aligned} &= (1200 \times 2/3) + (1200 \times (1/3) \times 2) \\ &= 800 + 800 \\ &= 1600 \text{ IOPS} \end{aligned}$$
- For RAID 5, the disk load (read + write)  
$$\begin{aligned} &= (1200 \times 2/3) + (1200 \times (1/3) \times 4) \\ &= 800 + 1600 \\ &= 2400 \text{ IOPS} \end{aligned}$$

## RAID Comparison

RAID level	Min disks	Available storage capacity (%)	Read performance	Write performance	Write penalty	Protection
1	2	50	Better than single disk	Slower than single disk, because every write must be committed to all disks	Moderate	Mirror
1+0	4	50	Good	Good	Moderate	Mirror
3	3	$[(n-1)/n]*100$	Fair for random reads and good for sequential reads	Poor to fair for small random writes fair for large, sequential writes	High	Parity (Supports single disk failure)
5	3	$[(n-1)/n]*100$	Good for random and sequential reads	Fair for random and sequential writes	High	Parity (Supports single disk failure)
6	4	$[(n-2)/n]*100$	Good for random and sequential reads	Poor to fair for random and sequential writes	Very High	Parity (Supports two disk failures)

where n = number of disks

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 23

The table on the slide compare different RAID levels.

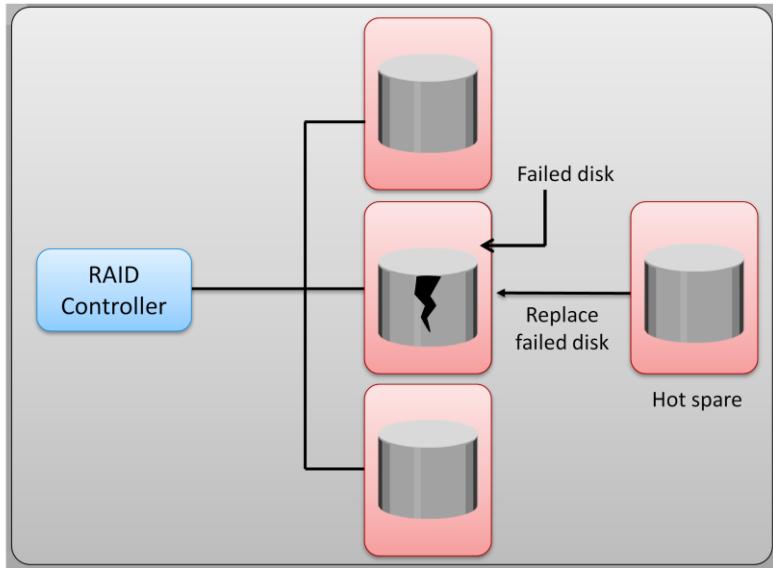
## Suitable RAID Levels for Different Applications

- RAID 1+0
  - ▶ Suitable for applications with small, random, and write intensive (writes typically greater than 30%) I/O profile
  - ▶ Example: OLTP, RDBMS – Temp space
- RAID 3
  - ▶ Large, sequential read and write
  - ▶ Example: data backup and multimedia streaming
- RAID 5 and 6
  - ▶ Small, random workload (writes typically less than 30%)
  - ▶ Example: email, RDBMS – Data entry

Common applications that benefit from different RAID levels.

- RAID 1+0 performs well for workloads that use small, random, write-intensive I/Os. Some applications that benefit from RAID 1+0 are high transaction rate online transaction processing (OLTP), RDBMS temp space and so on.
- RAID 3 provides good performance for applications that involve large sequential data access, such as data backup or video streaming.
- RAID 5 is good for random, read intensive I/O applications and preferred for messaging, medium-performance media serving, and relational database management system (RDBMS) implementations, in which database administrators (DBAs) optimize data access.

## Hot Spare



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 25

A *hot spare* refers to a spare drive in a RAID array that temporarily replaces a failed disk drive by taking the identity of the failed disk drive. With the hot spare, one of the following methods of data recovery is performed depending on the RAID implementation:

- If parity RAID is used, the data is rebuilt onto the hot spare from the parity and the data on the surviving disk drives in the RAID set.
- If mirroring is used, the data from the surviving mirror is used to copy the data onto the hot spare.

When a new disk drive is added to the system, data from the hot spare is copied to it. The hot spare returns to its idle state, ready to replace the next failed drive. Alternatively, the hot spare replaces the failed disk drive permanently. This means that it is no longer a hot spare, and a new hot spare must be configured on the array.

A hot spare should be large enough to accommodate data from a failed drive. Some systems implement multiple hot spares to improve data availability.

A hot spare can be configured as automatic or user initiated, which specifies how it will be used in the event of disk failure. In an automatic configuration, when the recoverable error rates for a disk exceed a predetermined threshold, the disk subsystem tries to copy data from the failing disk to the hot spare automatically. If this task is completed before the damaged disk fails, the subsystem switches to the hot spare and marks the failing disk as unusable. Otherwise, it uses parity or the mirrored disk to recover the data. In the case of a user-initiated configuration, the administrator has control of the rebuild process. For example, the rebuild could occur overnight to prevent any degradation of system performance. However, the system is at risk of data loss if another disk failure occurs.

## Module 3: Summary

Key points covered in this module:

- RAID implementation methods and techniques
- Common RAID levels
- RAID write penalty
- Compare RAID levels based on their cost and performance

This module covered the two methods of RAID implementation, hardware and software.

The three techniques on which the RAID levels are built are striping, mirroring, and parity.

The commonly used RAID levels are 0, 1, 1+0, 3, 5, and 6.

When choosing a RAID type, it is imperative to consider its impact on disk performance and application IOPS. In both mirrored and parity RAID configurations, every write operation translates into more I/O overhead for the disks, which is referred to as a *write penalty*.

Finally, this module compared different RAID levels based on their cost, performance, and write penalty.

## Check Your Knowledge – 1

- Which statement is true about software RAID implementation?
  - A. Upgrades to operating system do not require compatibility validation with RAID software
  - B. It is expensive than hardware RAID implementation
  - C. Supports all RAID levels
  - D. Uses host CPU cycles to perform RAID calculations
- An application generates 400 small random IOPS with a read/write ratio of 3:1. What is the RAID-corrected IOPS on the disk for RAID 5 ?
  - A. 400
  - B. 500
  - C. 700
  - D. 900

## Check Your Knowledge – 2

- What is write penalty in a RAID 6 configuration for small random I/Os?
  - A. 2
  - B. 3
  - C. 4
  - D. 6
- Which application is most benefited by using RAID 3?
  - A. Backup
  - B. OLTP
  - C. e-commerce
  - D. email

## Check Your Knowledge – 3

- What is the stripe size of a five disk parity RAID 5 set that has a strip size of 64 KB?
  - A. 64 KB
  - B. 128 KB
  - C. 256 KB
  - D. 320 KB

## Exercise 1: RAID

- A company is planning to reconfigure storage for their accounting application for high availability
  - ▶ Current configuration and challenges
    - ▶ Application performs 15% random writes and 85% random reads
    - ▶ Currently deployed with five disk RAID 0 configuration
    - ▶ Each disk has an advertised formatted capacity of 200 GB
    - ▶ Total size of accounting application's data is 730 GB which is unlikely to change over 6 months
    - ▶ Approaching end of financial year, buying even one disk is not possible
  - Task
    - ▶ Recommend a RAID level that the company can use to restructure their environment fulfilling their needs
    - ▶ Justify your choice based on cost, performance, and availability

### Business Profile:

A company, involved in mobile wireless services across the country, has about 5000 employees worldwide. This company has 7 regional offices across the country. Although the company is financially doing well, they continue to feel the competitive pressure. As a result, the company needs to ensure that the IT infrastructure takes advantage of fault tolerant features.

### Current Configuration and Challenges:

The company uses different applications for communication, accounting, and management. All the applications are hosted on individual servers with disks configured as RAID 0. All financial activity is managed and tracked by a single accounting application. It is very important for the accounting data to be highly available. The application performs around 15% random write operations and the remaining 85% are random reads. The accounting data is currently stored on a 5-disk RAID 0 set. Each disk has an advertised formatted capacity of 200 GB and the total size of their files is 730 GB. The company performs nightly backups and removes old information — so the amount of data is unlikely to change much over the next 6 months. The company is approaching the end of the financial year and the IT budget is depleted. It won't be possible to buy even one new disk drive.

### Tasks:

Recommend a RAID level that the company can use to restructure their environment fulfilling their needs.

Justify your choice based on cost, performance, and availability of the new solution.

## Exercise 2: RAID

- A company (same as discussed in exercise 1) is now planning to reconfigure storage for their database application for HA
  - ▶ Current configuration and challenges
    - ▶ The application performs 40% writes and 60% reads
    - ▶ Currently deployed on six disk RAID 0 configuration with advertised capacity of each disk being 200 GB
    - ▶ Size of the database is 900 GB and amount of data is likely to change by 30% over the next 6 months
    - ▶ It is a new financial year and the company has an increased budget
  - Task
    - ▶ Recommend a suitable RAID level to fulfill company's needs
    - ▶ Estimate the cost of the new solution (200GB disk costs \$1000)
    - ▶ Justify your choice based on cost, performance, and availability

### Business Profile:

A company, involved in mobile wireless services across the country, has about 5000 employees worldwide. This company has 7 regional offices across the country. Although the company is financially doing well, they continue to feel the competitive pressure. As a result, the company needs to ensure that the IT infrastructure takes advantage of fault tolerant features.

### Current Configuration and Challenges:

The company uses an accounting application that is hosted on an individual server with disks configured as RAID 0. It is now the beginning of a new financial year and the IT department has an increased budget. You are called in to recommend changes to their database environment. You investigate their database environment closely and observe that the data is stored on a 6-disk RAID 0 set. Each disk has an advertised formatted capacity of 200 GB and the total size of their files is 900 GB. The amount of data is likely to change by 30 % over the next 6 months and your solution must accommodate this growth. The application performs around 40% write operations and the remaining 60 % are reads.

### Tasks:

Recommend a RAID level that the company can use to restructure their environment and fulfill their needs. What is the cost of the new solution?

Justify your choice based on cost, performance, and data availability of the new solution.

**Note:** A new 200 GB disk drive costs \$1000. The controller can handle all commonly used RAID levels, so will not need to be replaced.

This slide intentionally left blank.

**EMC<sup>2</sup> PROVEN PROFESSIONAL**

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 3: Data Protection - RAID 32

# Module – 4

# Intelligent Storage System



## Module 4: Intelligent Storage System

Upon completion of this module, you should be able to:

- Describe the key components of intelligent storage system
- Describe cache management and protection techniques
- Describe two storage provisioning methods
- Describe two types of intelligent storage systems

This module focuses on the key components of an intelligent storage system. It details the function of each component, including cache management and protection techniques. The module also focuses on the two storage provisioning methods. Finally, it describes the two types of intelligent storage systems.

# Module 4: Intelligent Storage System

## Lesson 1: Key Components of an Intelligent Storage System

During this lesson the following topics are covered:

- Intelligent storage system overview
- Key components of an intelligent storage system
- Cache management

This lesson focuses on intelligent storage system overview and key components of an intelligent storage system. This lesson also focuses on cache management.

## What is an Intelligent Storage System (ISS) ?

### Intelligent Storage System

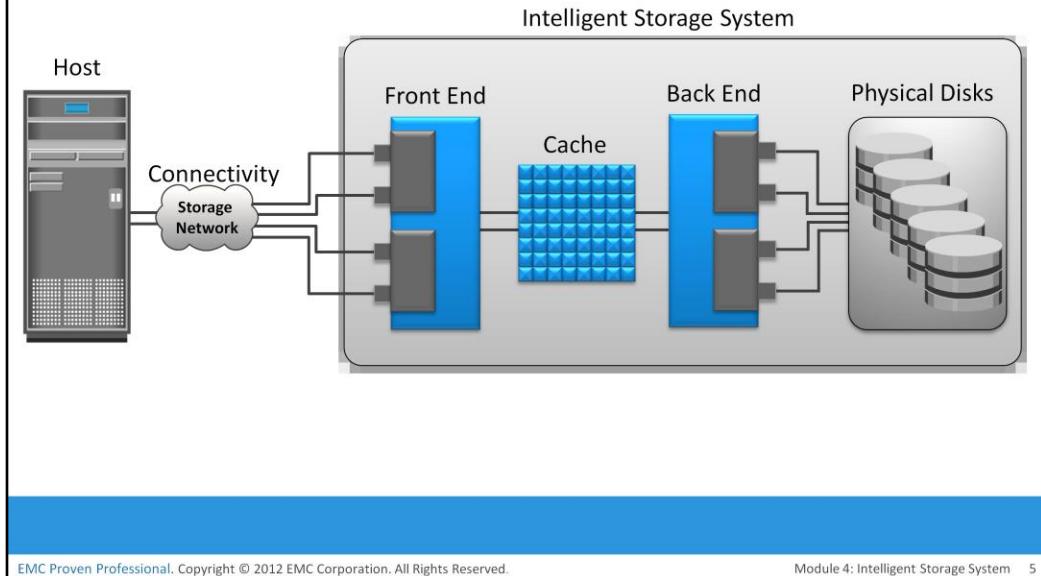
It is a feature-rich RAID array that provides highly optimized I/O processing capabilities.

- Provides large amount of cache and multiple I/O paths that enhances the performance
- Has an operating environment that provides
  - ▶ Intelligent cache management
  - ▶ Array resource management
  - ▶ Connectivity to heterogeneous hosts
- Supports flash drive, virtual provisioning, and automated storage tiering

Business-critical applications require high levels of performance, availability, security, and scalability. A disk drive is a core element of storage that governs the performance of any storage system. Some of the older disk-array technologies could not overcome performance constraints due to the limitations of disk drives and their mechanical components. RAID technology made an important contribution to enhancing storage performance and reliability, but disk drives, even with a RAID implementation, could not meet the performance requirements of today's applications.

With advancements in technology, a new breed of storage solutions, known as *intelligent storage systems*, has evolved. These intelligent storage systems are feature-rich RAID arrays that provide highly optimized I/O processing capabilities. These storage systems are configured with a large amount of memory (called *cache*) and multiple I/O paths and use sophisticated algorithms to meet the requirements of performance-sensitive applications. These arrays have an operating environment that intelligently and optimally handles the management, allocation, and utilization of storage resources. Support for flash drives and other modern-day technologies, such as virtual storage provisioning and automated storage tiering, has added a new dimension to storage system performance, scalability, and availability.

## Key Components of an ISS

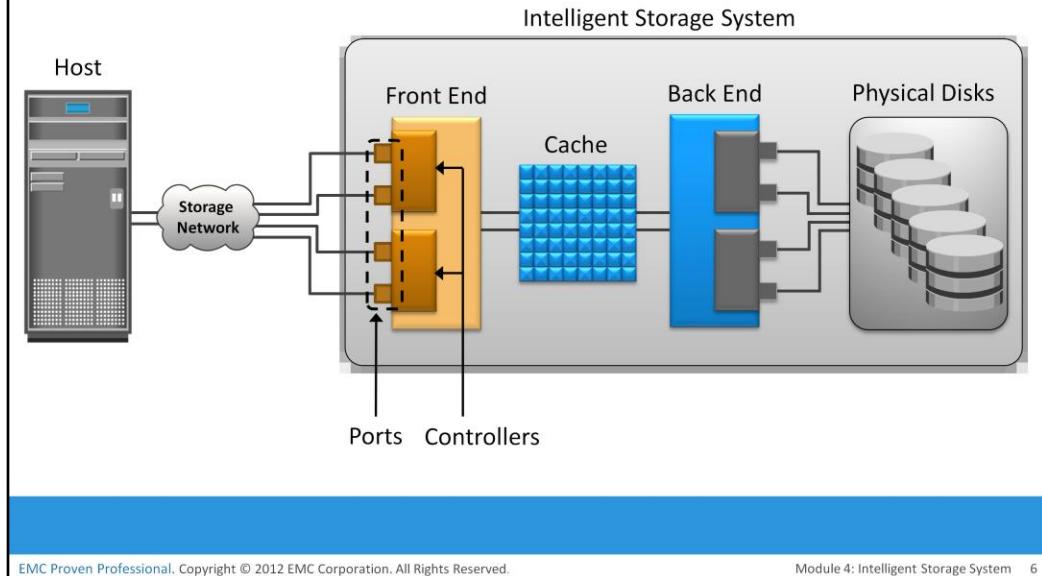


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 5

An intelligent storage system consists of four key components: *front end*, *cache*, *back end*, and *physical disks*. An I/O request received from the host at the front-end port is processed through cache and back end, to enable storage and retrieval of data from the physical disk. A read request can be serviced directly from cache if the requested data is found in the cache. In modern intelligent storage systems, front end, cache, and back end are typically integrated on a single board (referred as a *storage processor* or *storage controller*).

## Key Components of ISS: Front End

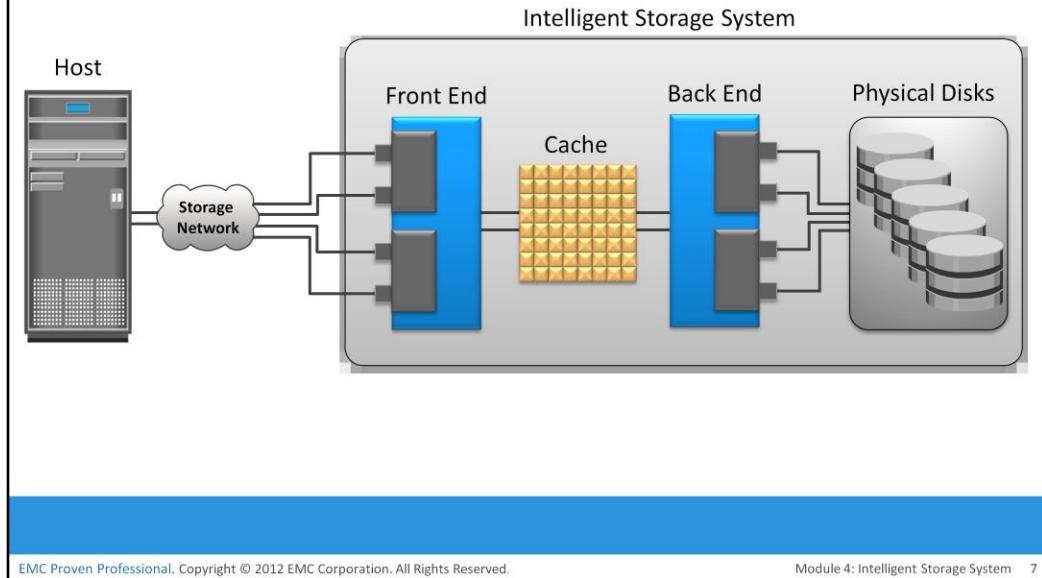


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 6

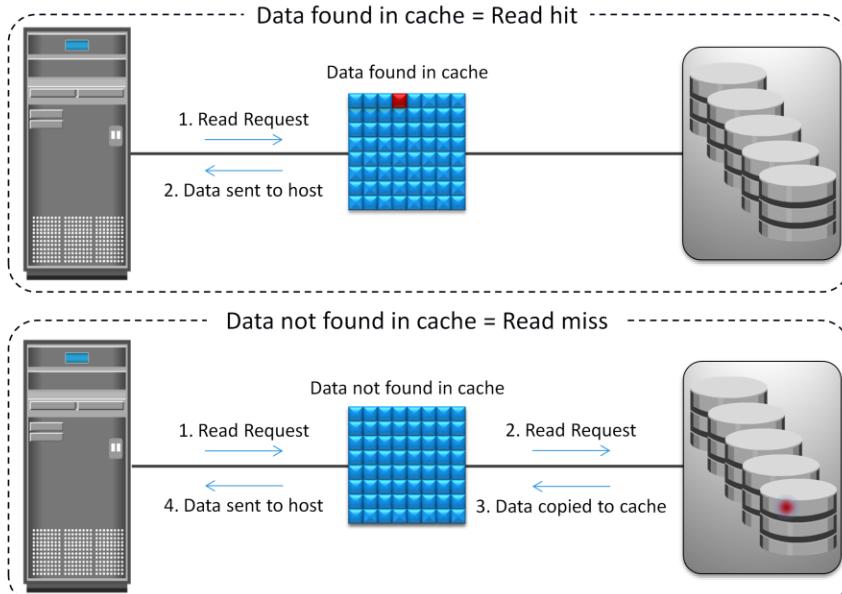
The front end provides the interface between the storage system and the host. It consists of two components: front-end ports and front-end controllers. Typically, a front end has redundant controllers for high availability, and each controller contains multiple ports that enable large numbers of hosts to connect to the intelligent storage system. Each front-end controller has processing logic that executes the appropriate transport protocol, such as Fibre Channel, iSCSI, FICON, or FCoE for storage connections. *Front-end controllers* route data to and from cache via the internal data bus. When the cache receives the write data, the controller sends an acknowledgment message back to the host.

## Key Components of ISS: Cache



**Cache** is semiconductor memory where data is placed temporarily to reduce the time required to service I/O requests from the host. Cache improves storage system performance by isolating hosts from the mechanical delays associated with rotating disks or hard disk drive (HDD). Rotating disks are the slowest component of an intelligent storage system. Data access on rotating disks usually takes several millisecond because of seek time and rotational latency. Accessing data from cache is fast and typically takes less than a millisecond. On intelligent arrays, write data is first placed in cache and then written to disk.

## Read Operation with Cache



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

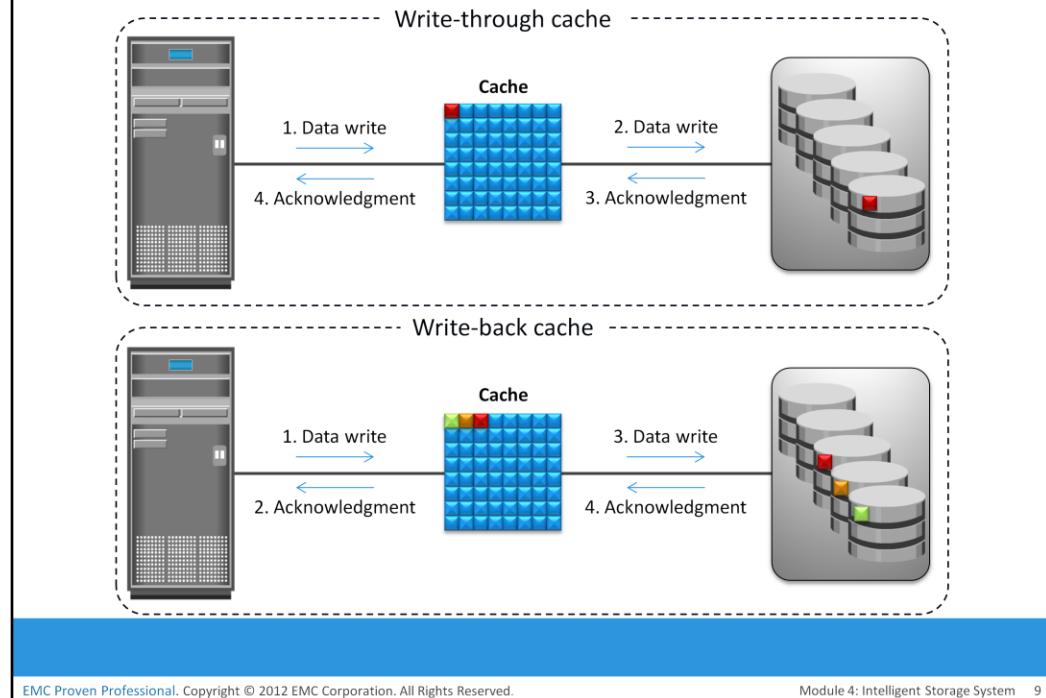
Module 4: Intelligent Storage System 8

When a host issues a read request, the storage controller reads the tag RAM to determine whether the required data is available in cache. If the requested data is found in the cache, it is called a *read cache hit* or *read hit* and data is sent directly to the host, without any disk operation. This provides a fast response time to the host (about a millisecond). If the requested data is not found in cache, it is called a *cache miss* and the data must be read from the disk. The back end accesses the appropriate disk and retrieves the requested data. Data is then placed in cache and finally sent to the host through the front end. Cache misses increase the I/O response time.

A *prefetch* or *read-ahead* algorithm is used when read requests are sequential. In a sequential read request, a contiguous set of associated blocks is retrieved. Several other blocks that have not yet been requested by the host can be read from the disk and placed into cache in advance. When the host subsequently requests these blocks, the read operations will be read hits. This process significantly improves the response time experienced by the host. The intelligent storage system offers fixed and variable prefetch sizes. In *fixed prefetch*, the intelligent storage system prefetches a fixed amount of data. It is most suitable when host I/O sizes are uniform. In *variable prefetch*, the storage system prefetches an amount of data in multiples of the size of the host request. *Maximum prefetch* limits the number of data blocks that can be prefetched to prevent the disks from being rendered busy with prefetch at the expense of other I/Os.

Read performance is measured in terms of the *read hit ratio*, or the *hit rate*, usually expressed as a percentage. This ratio is the number of read hits with respect to the total number of read requests. A higher read hit ratio improves the read performance.

## Write Operation with Cache



Write operations with cache provide performance advantages over writing directly to disks. When an I/O is written to cache and acknowledged, it is completed in far less time (from the host's perspective) than it would take to write directly to disk. Sequential writes also offer opportunities for optimization because many smaller writes can be coalesced for larger transfers to disk drives with the use of cache.

A write operation with cache is implemented in the following ways:

- **Write-back cache:** Data is placed in cache and an acknowledgment is sent to the host immediately. Later, data from several writes are committed (de-staged) to the disk. Write response times are much faster because the write operations are isolated from the mechanical delays of the disk. However, uncommitted data is at risk of loss if cache failures occur.
- **Write-through cache:** Data is placed in the cache and immediately written to the disk, and an acknowledgment is sent to the host. Because data is committed to disk as it arrives, the risks of data loss are low, but the write-response time is longer because of the disk operations.

Cache can be bypassed under certain conditions, such as large size write I/O. In this implementation, if the size of an I/O request exceeds the predefined size, called *write aside size*, writes are sent to the disk directly to reduce the impact of large writes consuming a large cache space. This is particularly useful in an environment where cache resources are constrained and cache is required for small random I/Os.

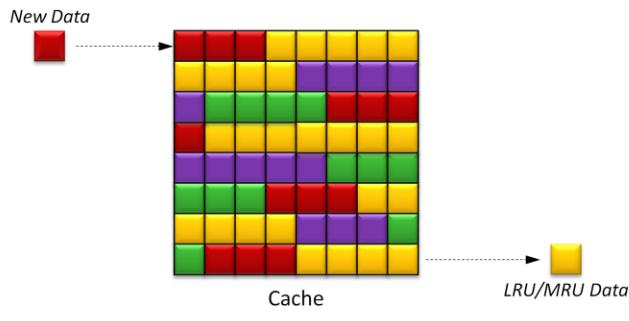
Cont..

Cache can be implemented as either dedicated cache or global cache. With dedicated cache, separate sets of memory locations are reserved for reads and writes. In global cache, both reads and writes can use any of the available memory addresses. Cache management is more efficient in a global cache implementation because only one global set of addresses has to be managed.

Global cache allows users to specify the percentages of cache available for reads and writes for cache management. Typically, the read cache is small, but it should be increased if the application being used is read-intensive. In other global cache implementations, the ratio of cache available for reads versus writes is dynamically adjusted based on the workloads.

## Cache Management: Algorithms

- Least recently used (LRU)
  - ▶ Discards data that have not been accessed for a long time
- Most recently used (MRU)
  - ▶ Discards data that have been most recently accessed



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 11

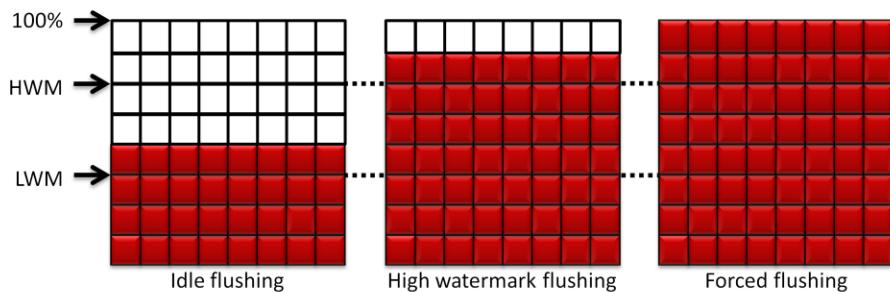
Cache is a finite and expensive resource that needs proper management. Even though modern intelligent storage systems come with a large amount of cache, when all cache pages are filled, some pages have to be freed up to accommodate new data and avoid performance degradation. Various cache management algorithms are implemented in intelligent storage systems to proactively maintain a set of free pages and a list of pages that can be potentially freed up whenever required.

The most commonly used algorithms are discussed in the following list:

- Least Recently Used (LRU): An algorithm that continuously monitors data access in cache and identifies the cache pages that have not been accessed for a long time. LRU either frees up these pages or marks them for reuse. This algorithm is based on the assumption that data that has not been accessed for a while will not be requested by the host. However, if a page contains write data that has not yet been committed to disk, the data is first written to disk before the page is reused.
- Most Recently Used (MRU): This algorithm is the opposite of LRU, where the pages that have been accessed most recently are freed up or marked for reuse. This algorithm is based on the assumption that recently accessed data may not be required for a while.

## Cache Management: Watermarking

- Manages I/O burst through flushing process
  - ▶ Flushing is the process of committing data from cache to the disk
- Three modes of flushing to manage cache utilization are:
  - ▶ Idle flushing
  - ▶ High watermark flushing
  - ▶ Forced flushing



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 12

As cache fills, the storage system must take action to flush dirty pages (data written into the cache but not yet written to the disk) to manage space availability. *Flushing* is the process that commits data from cache to the disk. On the basis of the I/O access rate and pattern, high and low levels called *watermarks* are set in cache to manage the flushing process. *High watermark (HWM)* is the cache utilization level at which the storage system starts high-speed flushing of cache data. *Low watermark (LWM)* is the point at which the storage system stops flushing data to the disks. The cache utilization level drives the mode of flushing to be used:

- **Idle flushing:** Occurs continuously, at a modest rate, when the cache utilization level is between the high and low watermark.
- **High watermark flushing:** Activated when cache utilization hits the high watermark. The storage system dedicates some additional resources for flushing. This type of flushing has some impact on I/O processing.
- **Forced flushing:** Occurs in the event of a large I/O burst when cache reaches 100 percent of its capacity, which significantly affects the I/O response time. In forced flushing, system flushes the cache on priority by allocating more resources.

## Cache Data Protection

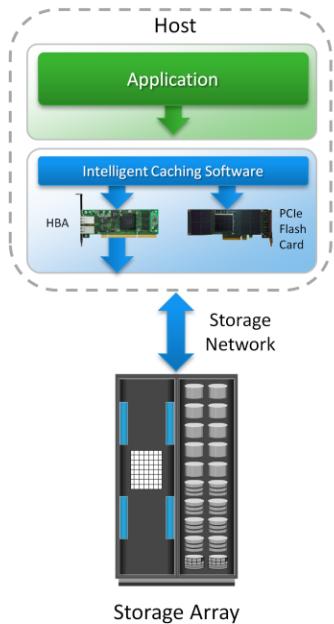
- Protects data in the cache against power or cache failures:
  - ▶ Cache mirroring
    - ▶ Provides protection to data against cache failure
    - ▶ Each write to the cache is held in two different memory locations on two independent memory cards
  - ▶ Cache vaulting
    - ▶ Provides protection to data against power failure
    - ▶ In the event of power failure, uncommitted data is dumped to a dedicated set of drives called vault drives

Cache is volatile memory, so a power failure or any kind of cache failure will cause loss of the data that is not yet committed to the disk. This risk of losing uncommitted data held in cache can be mitigated using *cache mirroring* and *cache vaulting*:

- Cache mirroring: Each write to cache is held in two different memory locations on two independent memory cards. If a cache failure occurs, the write data will still be safe in the mirrored location and can be committed to the disk. Reads are staged from the disk to the cache; therefore, if a cache failure occurs, the data can still be accessed from the disk. Because only writes are mirrored, this method results in better utilization of the available cache. In cache mirroring approaches, the problem of maintaining *cache coherency* is introduced. Cache coherency means that data in two different cache locations must be identical at all times. It is the responsibility of the array operating environment to ensure coherency.
- Cache vaulting: The risk of data loss due to power failure can be addressed in various ways: powering the memory with a battery until the AC power is restored or using battery power to write the cache content to the disk. If an extended power failure occurs, using batteries is not a viable option. This is because in intelligent storage systems, large amounts of data might need to be committed to numerous disks, and batteries might not provide power for sufficient time to write each piece of data to its intended disk. Therefore, storage vendors use a set of physical disks to dump the contents of cache during power failure. This is called *cache vaulting* and the disks are called *vault drives*. When power is restored, data from these disks is written back to write cache and then written to the intended disks.

## Server Flash-caching Technology

- Uses intelligent caching software and PCIe flash card on host
- Dramatically improves application performance
  - ▶ Provides performance acceleration for read-intensive workloads
  - ▶ Avoids network latencies associated with I/O access to the storage array
- Intelligently determines data that would benefit by sitting in server on PCIe flash
- Uses minimal CPU and memory resources
  - ▶ Flash management is offloaded onto PCIe card



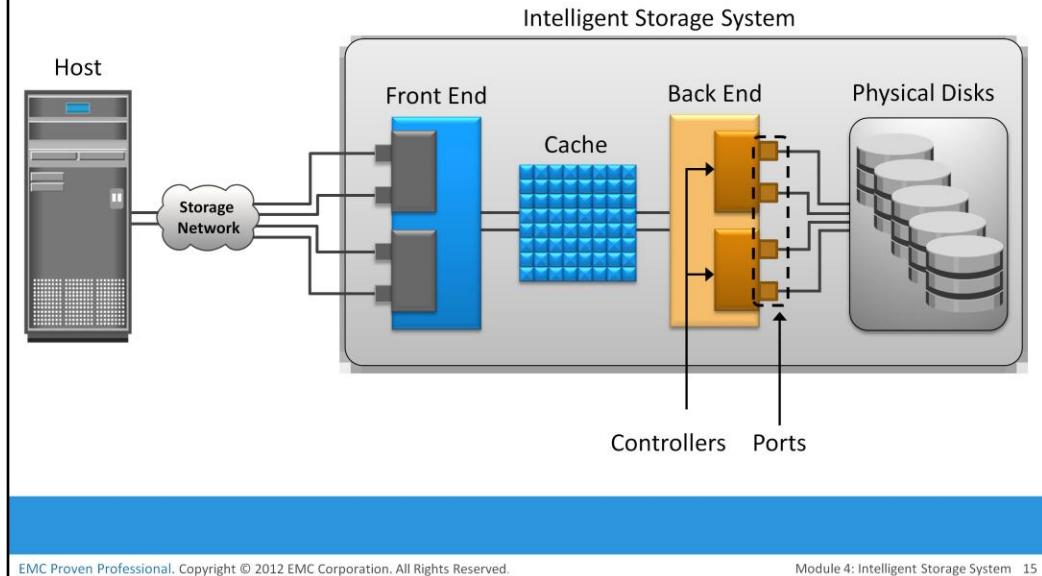
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 14

Server flash-caching technology uses intelligent caching software and PCI Express (PCIe) flash card on the host. This dramatically improves application performance by reducing latency and accelerates throughput. Server flash-caching technology works in both physical and virtual environments and provides performance acceleration for read-intensive workloads. This technology uses minimal CPU and memory resources from the server by offloading flash management onto the PCIe card.

It intelligently determines which data would benefit by sitting in the server on PCIe flash and closer to the application. This avoids the latencies associated with I/O access over the network to the storage array. With this, the processing power required for an application's most frequently referenced data is offloaded from the back-end storage to the PCIe card. Therefore, the storage array can allocate greater processing power to other applications.

## Key Components of ISS: Back End



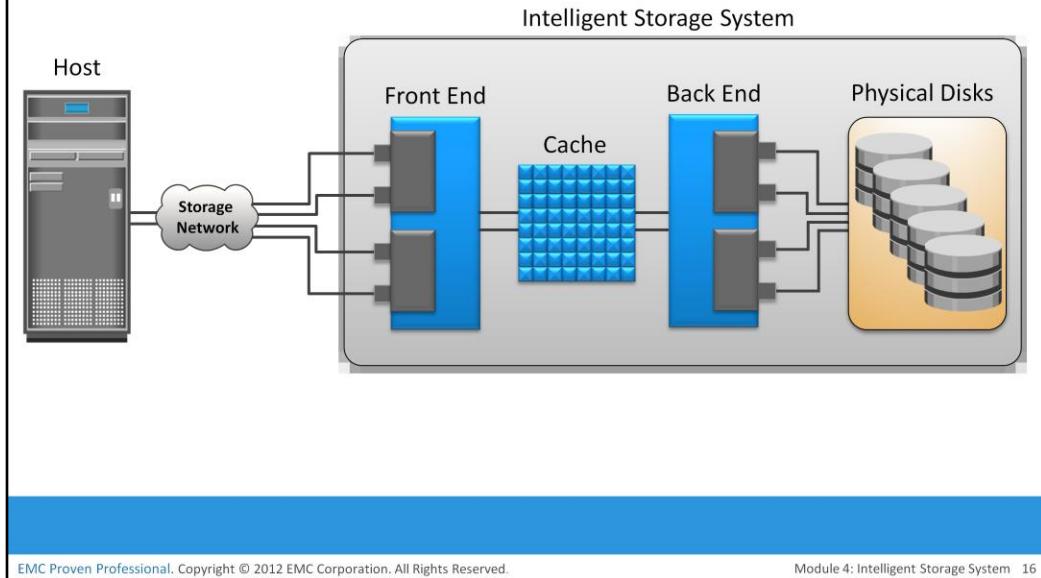
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 15

The *back end* provides an interface between cache and the physical disks. It consists of two components: back-end ports and back-end controllers. The back-end controls data transfers between cache and the physical disks. From cache, data is sent to the back end and then routed to the destination disk. Physical disks are connected to ports on the back end. The back-end controller communicates with the disks when performing reads and writes and also provides additional, but limited, temporary data storage. The algorithms implemented on back-end controllers provide error detection and correction, along with RAID functionality.

For high data protection and high availability, storage systems are configured with dual controllers with multiple ports. Such configurations provide an alternative path to physical disks if a controller or port failure occurs. This reliability is further enhanced if the disks are also dual-ported. In that case, each disk port can connect to a separate controller. Multiple controllers also facilitate load balancing.

## Key Components of ISS: Physical Disks



Physical disks are connected to the back-end storage controller and provide persistent data storage. Modern intelligent storage systems provide support to a variety of disk drives with different speeds and types, such as FC, SATA, SAS, and flash drives. They also support the use of a mix of flash, FC, or SATA within the same array.

## Module 4: Intelligent Storage System

### Lesson 2: Storage provisioning and ISS implementation

During this lesson the following topics are covered:

- Traditional storage provisioning
- Virtual storage provisioning
- ISS implementation

This lesson focuses on traditional and virtual storage provisioning. This lesson also focuses on ISS implementation.

## Assigning Storage to Host

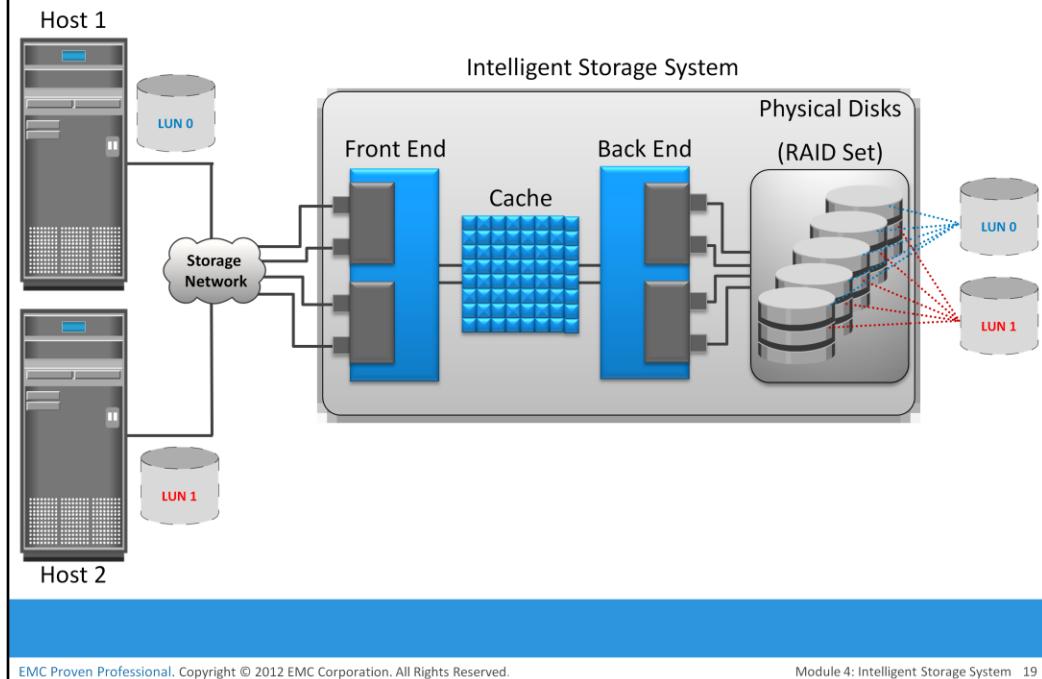
### Storage Provisioning

It is the process of assigning storage resources to hosts based on capacity, availability, and performance requirements of applications running on the hosts.

- Can be performed in two ways:
  - ▶ Traditional storage provisioning
  - ▶ Virtual storage provisioning

*Storage provisioning* is the process of assigning storage resources to hosts based on capacity, availability, and performance requirements of applications running on the hosts. Storage provisioning can be performed in two ways: traditional and virtual. *Virtual provisioning* leverages virtualization technology for provisioning storage for applications.

## Traditional Storage Provisioning



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 19

In traditional storage provisioning, physical disks are logically grouped together on which a required RAID level is applied to form a set, called RAID set. The number of drives in the RAID set and the RAID level determine the availability, capacity, and performance of the RAID set. It is highly recommended that the RAID set be created from drives of the same type, speed, and capacity to ensure maximum usable capacity, reliability, and consistency in performance. For example, if drives of different capacities are mixed in a RAID set, the capacity of the smallest drive is used from each disk in the set to make up the RAID set's overall capacity. The remaining capacity of the larger drives remains unused. Likewise, mixing higher revolutions per minute (RPM) drives with lower RPM drives lowers the overall performance of the RAID set.

RAID sets usually have a large capacity because they combine the total capacity of individual drives in the set. *Logical units* are created from the RAID sets by partitioning (seen as slices of the RAID set) the available capacity into smaller units. These units are then assigned to the host based on their storage requirements. Logical units are spread across all the physical disks that belong to that set. Each logical unit created from the RAID set is assigned a unique ID, called a *logical unit number* (LUN). LUNs hide the organization and composition of the RAID set from the hosts. LUNs created by traditional storage provisioning methods are also referred to as *thick LUNs* to distinguish them from the LUNs created by virtual provisioning methods.

Figure on the slide shows a RAID set consisting of five disks that have been sliced, or partitioned, into two LUNs: LUN 0 and LUN 1. These LUNs are then assigned to Host1 and Host 2 for their storage requirements.

Cont..

When a LUN is configured and assigned to a non-virtualized host, a bus scan is required to identify the LUN. This LUN appears as a raw disk to the operating system. To make this disk usable, it is formatted with a file system and then the file system is mounted.

In a virtualized host environment, the LUN is assigned to the hypervisor, which recognizes it as a raw disk. This disk is configured with the hypervisor file system, and then virtual disks are created on it. *Virtual disks* are files on the hypervisor file system. The virtual disks are then assigned to virtual machines and appear as raw disks to them. To make the virtual disk usable to the virtual machine, similar steps are followed as in a non-virtualized environment. Here, the LUN space may be shared and accessed simultaneously by multiple virtual machines.

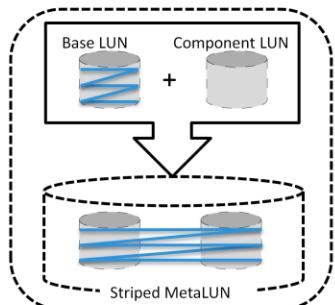
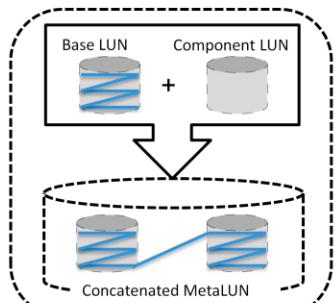
Virtual machines can also access a LUN directly on the storage system. In this method the entire LUN is allocated to a single virtual machine. Storing data in this way is recommended when the applications running on the virtual machine are response-time sensitive, and sharing storage with other virtual machines may impact their response time. The direct access method is also used when a virtual machine is clustered with a physical machine. In this case, the virtual machine is required to access the LUN that is being accessed by the physical machine.

## LUN Expansion

### MetalUN

It is a method to expand LUNs that require additional capacity or performance.

- Created by combining two or more LUNs
- MetaLUNs can either be concatenated or striped
- Concatenated metaLUN
  - ▶ Provides only additional capacity but no performance
  - ▶ Expansion is quick as data is not restriped
- Striped metaLUN
  - ▶ Provides capacity and performance
  - ▶ Expansion is slow as data is restriped



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 21

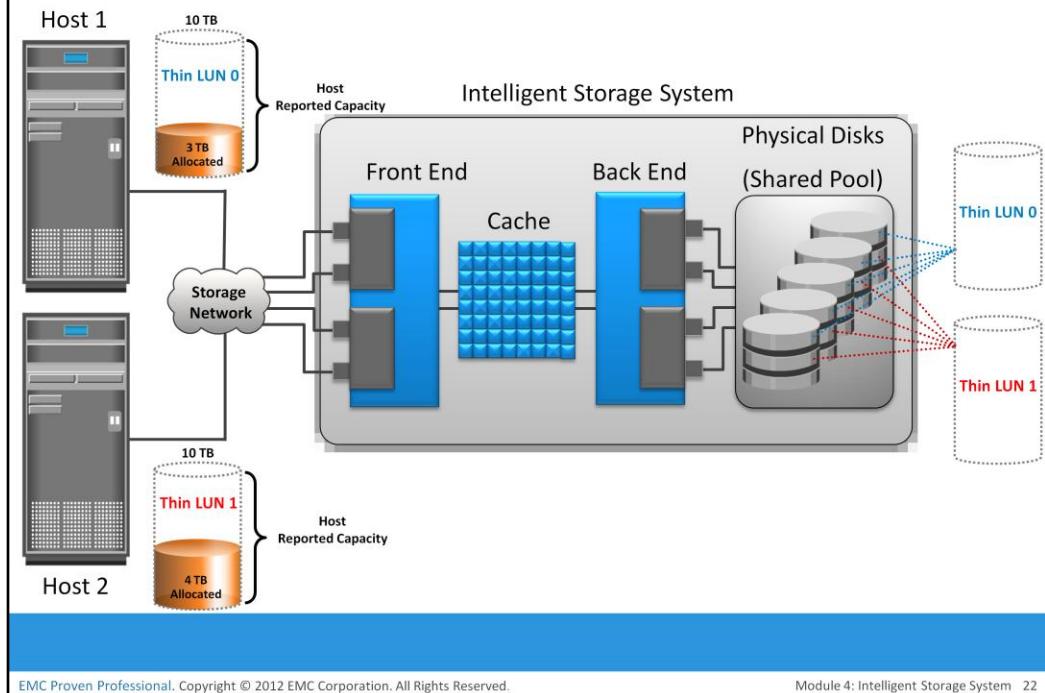
*MetaLUN* is a method to expand LUNs that require additional capacity or performance. A metaLUN can be created by combining two or more LUNs. A metaLUN consists of a base LUN and one or more component LUNs. MetaLUNs can be either *concatenated* or *striped*.

Concatenated expansion simply adds additional capacity to the base LUN. In this expansion, the component LUNs are not required to be of the same capacity as the base LUN. All LUNs in a concatenated metaLUN must be either protected (parity or mirrored) or unprotected (RAID 0). RAID types within a metaLUN can be mixed. For example, a RAID 1/0 LUN can be concatenated with a RAID 5 LUN. However, a RAID 0 LUN can be concatenated only with another RAID 0 LUN. Concatenated expansion is quick but does not provide any performance benefit.

Striped expansion restripes the base LUN's data across the base LUN and component LUNs. In striped expansion, all LUNs must be of the same capacity and RAID level. Striped expansion provides improved performance due to the increased number of drives being striped.

All LUNs in both concatenated and striped expansion must reside on the same disk-drive type: either all Fibre Channel or all ATA.

## Virtual Storage Provisioning



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

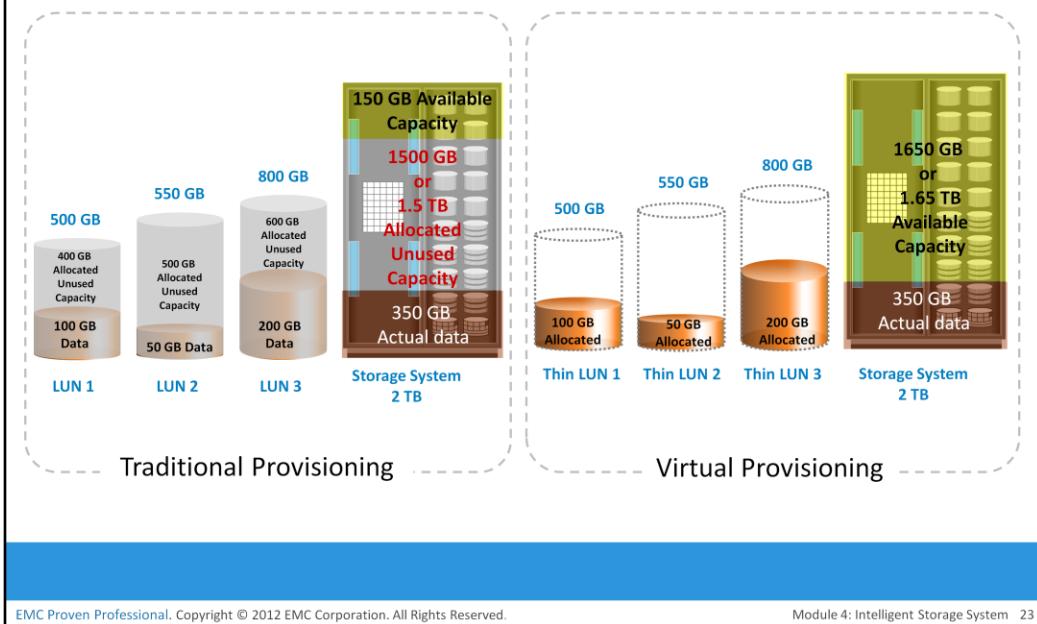
Module 4: Intelligent Storage System 22

*Virtual provisioning* enables creating and presenting a LUN with more capacity than is physically allocated to it on the storage array. The LUN created using virtual provisioning is called a *thin LUN* to distinguish it from the traditional LUN.

Thin LUNs do not require physical storage to be completely allocated to them at the time they are created and presented to a host. Physical storage is allocated to the host “on-demand” from a shared pool of physical capacity. A *shared pool* consists of physical disks. A shared pool in virtual provisioning is analogous to a RAID group, which is a collection of drives on which LUNs are created. Similar to a RAID group, a shared pool supports a single RAID protection level. However, unlike a RAID group, a shared pool might contain large numbers of drives. Shared pools can be homogeneous (containing a single drive type) or heterogeneous (containing mixed drive types, such as flash, FC, SAS, and SATA drives).

Virtual provisioning enables more efficient allocation of storage to hosts. Virtual provisioning also enables oversubscription, where more capacity is presented to the hosts than is actually available on the storage array. Both shared pool and thin LUN can be expanded nondisruptively as the storage requirements of the hosts grow. Multiple shared pools can be created within a storage array, and a shared pool may be shared by multiple thin LUNs.

## Traditional Provisioning vs. Virtual Provisioning



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 23

Administrators typically allocate storage capacity based on anticipated storage requirements. This generally results in the over provisioning of storage capacity, which then leads to higher costs and lower capacity utilization. Administrators often over-provision storage to an application for various reasons such as, to avoid frequent provisioning of storage if the LUN capacity is exhausted, and to reduce disruption to application availability. Over provisioning of storage often leads to additional storage acquisition and operational costs.

Virtual provisioning addresses these challenges. Virtual provisioning improves storage capacity utilization and simplifies storage management. Figure on the slide illustrates an example, comparing virtual provisioning with traditional storage provisioning.

With traditional provisioning, three LUNs are created and presented to one or more hosts. The total storage capacity of the storage system is 2 TB. The allocated capacity of LUN 1 is 500 GB, of which only 100 GB is consumed, and the remaining 400 GB is unused. The size of LUN 2 is 550 GB, of which 50 GB is consumed, and 500 GB is unused. The size of LUN 3 is 800 GB, of which 200 GB is consumed, and 600 GB is unused. In total, the storage system has 350 GB of data, 1.5 TB of allocated but unused capacity, and only 150 GB of remaining capacity available for other applications.

Now consider the same 2 TB storage system with virtual provisioning. Here, three thin LUNs of the same sizes are created. However, there is no allocated unused capacity. In total, the storage system with virtual provisioning has the same 350 GB of data, but 1.65 TB of capacity is available for other applications, whereas only 150 GB is available in traditional storage provisioning.

Cont..

Virtual provisioning and thin LUN offer many benefits, although in some cases traditional LUN is better suited for an application. Thin LUNs are appropriate for applications that can tolerate performance variations. In some cases, performance improvement is perceived when using a thin LUN, due to striping across a large number of drives in the pool. However, when multiple thin LUNs contend for shared storage resources in a given pool, and when utilization reaches higher levels, the performance can degrade. Thin LUNs provide the best storage space efficiency and are suitable for applications where space consumption is difficult to forecast. Using thin LUNs benefits organizations in reducing power and acquisition costs and in simplifying their storage management.

Traditional LUNs are suited for applications that require predictable performance. Traditional LUNs provide full control for precise data placement and allow an administrator to create LUNs on different RAID groups if there is any workload contention. Organizations that are not highly concerned about storage space efficiency may still use traditional LUNs.

Both traditional and thin LUNs can coexist in the same storage array. Based on the requirement, an administrator may migrate data between thin and traditional LUNs.

## LUN Masking

### LUN Masking

It is a process that provides data access control by defining which LUNs a host can access.

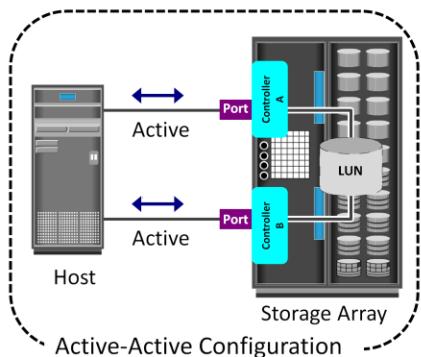
- Implemented on storage array
- Prevents unauthorized or accidental use of LUNs in a shared environment

*LUN masking* is a process that provides data access control by defining which LUNs a host can access. The LUN masking function is implemented on the storage array. This ensures that volume access by hosts is controlled appropriately, preventing unauthorized or accidental use in a shared environment.

For example, consider a storage array with two LUNs that store data of the sales and finance departments. Without LUN masking, both departments can easily see and modify each other's data, posing a high risk to data integrity and security. With LUN masking, LUNs are accessible only to the designated hosts.

## Types of ISS: High-end Storage Systems

- Referred as active-active arrays, and generally aimed at large enterprise applications
  - ▶ Performs I/Os to LUNs through all the available paths
- These arrays provide the following capabilities:
  - ▶ Large storage capacity and cache
  - ▶ Fault tolerant architecture
  - ▶ Connectivity to mainframe and open systems
  - ▶ Multiple front-end ports and interface protocols
  - ▶ Ability to handle large amount of concurrent I/Os
  - ▶ Support local and remote data replication



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 26

Intelligent storage systems generally fall into one of the following two categories: high-end storage systems, and midrange storage systems.

High-end storage systems, referred to as *active-active arrays*, are generally aimed at large enterprise applications. These systems are designed with a large number of controllers and cache memory. An active-active array implies that the host can perform I/Os to its LUNs through any of the available controllers .

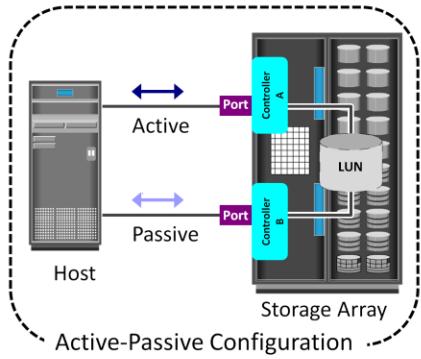
To address enterprise storage needs, these arrays provide the following capabilities:

- Large storage capacity
- Large amounts of cache to service host I/Os optimally
- Fault tolerance architecture to improve data availability
- Connectivity to mainframe computers and open systems hosts
- Availability of multiple front-end ports and interface protocols to serve a large number of hosts
- Availability of multiple back-end controllers to manage disk processing
- Scalability to support increased connectivity, performance, and storage capacity requirements
- Ability to handle large amounts of concurrent I/Os from a number of hosts and applications
- Support for array-based local and remote data replication

In addition to these features, high-end systems possess some unique features that are required for mission-critical applications.

## Types of ISS: Midrange Storage Systems

- Referred as active-passive arrays, and generally aimed at small and medium-sized enterprise applications
  - ▶ Performs I/Os to LUNs only through active paths
- These arrays typically have two controllers, each with cache, RAID controllers, and disks drive interfaces
- Less front-end ports, storage capacity, and cache as compared to high-end arrays
- Support local and remote data replication



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 27

Midrange storage systems are also referred to as *active-passive arrays* and are best suited for small- and medium-sized enterprise applications. They also provide optimal storage solutions at a lower cost. In an active-passive array, a host can perform I/Os to a LUN only through the controller that owns the LUN. The host can perform reads or writes to the LUN only through the path to controller A because controller A is the owner of that LUN. The path to controller B remains passive and no I/O activity is performed through this path.

Midrange storage systems are typically designed with two controllers, each of which contains host interfaces, cache, RAID controllers, and interface to disk drives.

Midrange arrays are designed to meet the requirements of small and medium enterprise application; therefore, they host less storage capacity and cache than high-end storage arrays. There are also fewer front-end ports for connection to hosts. However, they ensure high redundancy and high performance for applications with predictable workloads. They also support array-based local and remote replication.

## Module 4: Intelligent Storage System

### Concept in Practice

- EMC VNX
- EMC Symmetrix VMAX

The Concept in Practice covers the product example of intelligent storage system. It covers two products: EMC Symmetrix and EMC VNX.

## EMC VNX

- EMC's midrange storage offering
- Unified storage offering that provides storage for block, file, and object data
- Ideally suited for applications with predictable workloads



EMC VNX

The EMC VNX storage array is EMC's midrange storage offering that delivers enterprise-quality features and functionalities. EMC VNX is a unified storage platform that offers storage for block, file, and object-based data within the same array. It is ideally suited for applications with predictable workloads that require moderate-to-high throughput.

## EMC Symmetrix VMAX

- EMC's high-end storage offering
- Key features supported by Symmetrix VMAX are:
  - ▶ Incrementally scalable to 2,400 disks
  - ▶ Supports up to 8 VMAX engines
  - ▶ Supports flash drives, fully automated storage tiering (FAST), virtual provisioning, and cloud computing
  - ▶ Supports up to 1 TB of global cache memory
  - ▶ Supports FC, iSCSI, GigE, and FICON for host connectivity
  - ▶ Supports RAID levels 1, 1+0, 5, and 6
  - ▶ Supports storage-based replication via EMC TimeFinder and SRDF



EMC Symmetrix VMAX

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 4: Intelligent Storage System 30

EMC Symmetrix establishes the highest standards for performance and capacity for an enterprise information storage solution and is recognized as the industry's most trusted storage platform. Symmetrix offers the highest level of scalability and performance to meet even unpredictable I/O workload requirements. The EMC Symmetrix offering includes Symmetrix Virtual Matrix (VMAX) series.

The EMC Symmetrix VMAX series is an innovative platform built around a scalable Virtual Matrix architecture to support the future storage growth demands of virtual IT environments. The key features supported by Symmetrix VMAX follows:

- Incrementally scalable to 2,400 disks
- Supports up to 8 VMAX engines (Each VMAX engine contains a pair of directors)
- Supports flash drives, fully automated storage tiering (FAST), virtual provisioning, and Cloud computing
- Supports up to 1 TB of global cache memory
- Supports FC, iSCSI, GigE, and FICON for host connectivity
- Supports RAID levels 1, 1+0, 5, and 6
- Supports storage-based replication through EMC TimeFinder and EMC SRDF
- Highly fault-tolerant design that allows nondisruptive upgrades and full component-level redundancy with hot-swappable replacements

## Module 4: Summary

Key points covered in this module:

- Key components of intelligent storage system
- Cache management and protection techniques
- Storage provisioning methods
- Types of intelligent storage systems

This module covered the four key components of intelligent storage systems, front end, cache, back end, and physical disks. Further, this module detailed the cache management and protection techniques such as flushing, least recently used, and most recently used algorithm. It also covered the two storage provisioning techniques, traditional and virtual storage provisioning. Finally, the module described the two types of intelligent storage systems such as high-end and midrange storage systems.

## Check Your Knowledge – 1

- Which component of an intelligent storage system isolates host from the mechanical delays associated with rotating disks?
  - A. Front-end controller
  - B. Back-end controller
  - C. Cache
  - D. Storage network
- Which mode of flushing is activated when the cache reaches 100% of its capacity?
  - A. Idle
  - B. High watermark
  - C. Forced
  - D. Low watermark

## Check Your Knowledge – 2

- In traditional storage provisioning, which LUN expansion technique provides improved performance?
  - A. Concatenated metaLUN
  - B. Striped metaLUN
  - C. Base LUN
  - D. Component LUN
- Which process provides data access control by restricting host access to specific LUN(s)?
  - A. LUN masking
  - B. Zoning
  - C. Trespassing
  - D. VSAN

## Check Your Knowledge – 3

- Which mechanism provides protection to ‘uncommitted data in cache’ against power failure?
  - A. Mirroring
  - B. Vaulting
  - C. Watermarking
  - D. Tiering

# Module – 5

# Fibre Channel Storage Area Network (FC SAN)



## Module 5: Fibre Channel Storage Area Network (FC SAN)

Upon completion of this module, you should be able to:

- Describe FC SAN and its components
- Describe FC architecture
- Describe FC SAN topologies and zoning
- Describe virtualization in SAN environment

This module focuses on FC SAN components, FC interconnectivity options, and FC architecture. This module also focuses on virtualization in SAN environment.

# Module 5: Fibre Channel Storage Area Network (FC SAN)

## Lesson 1: Overview of FC SAN

During this lesson the following topics are covered:

- Evolution of FC SAN
- Components of FC SAN
- FC interconnectivity options
- FC port types

This lesson covers evolution of FC SAN, its components, and three FC interconnectivity options. This lesson also covers various FC port types.

## Business Needs and Technology Challenges

- An effective information management solution must provide:
  - ▶ Just-in-time information to business users
  - ▶ Flexible and resilient storage infrastructure
- Information management challenges in DAS environment:
  - ▶ Explosive growth of information storage that remains isolated and underutilized
  - ▶ Proliferation of new servers and applications
  - ▶ Complexity in sharing storage resources across multiple servers
  - ▶ High cost of managing information
- Storage area network (SAN) addresses these challenges

Organizations are experiencing an explosive growth in information. This information needs to be stored, protected, optimized, and managed efficiently. Data center managers are burdened with the challenging task of providing low-cost, high-performance information management solutions. An effective information management solution must provide the following:

**Just-in-time information to business users:** Information must be available to business users when they need it. 24 x 7 data availability is becoming one of the key requirements of today's storage infrastructure. The explosive growth in storage, proliferation of new servers and applications, and the spread of mission-critical data throughout enterprises are some of the challenges that need to be addressed to provide information availability in real time.

**Flexible and resilient storage infrastructure:** The storage infrastructure must provide flexibility and resilience that aligns with changing business requirements. Storage should scale without compromising the performance requirements of applications and, at the same time, the total cost of managing information must be low.

**Direct-attached storage (DAS) is often referred to as a stovepiped storage environment.** Hosts "own" the storage, and it is difficult to manage and share resources on these isolated storage devices. Efforts to organize this dispersed data led to the emergence of the storage area network (SAN).

## What is a SAN?

SAN

It is a high-speed, dedicated network of servers and shared storage devices.

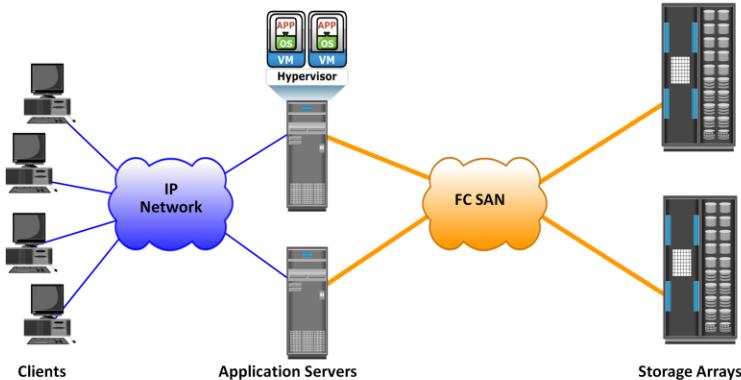
- Centralizes storage and management
- Enables sharing of storage resources across multiple servers at block level
- Meets increasing storage demands efficiently with better economies of scale
- Common SAN deployments are:
  - ▶ Fibre Channel (FC) SAN: uses FC protocol for communication
  - ▶ IP SAN: uses IP-based protocols for communication

SAN is a high-speed, dedicated network of servers and shared storage devices. It enables storage consolidation and enables storage to be shared across multiple servers. This improves the utilization of storage resources compared to direct-attached storage architecture and reduces the total amount of storage an organization needs to purchase and manage. With consolidation, storage management becomes centralized and less complex, which further reduces the cost of managing information. SAN also enables organizations to connect geographically dispersed servers and storage. Further, it meets the storage demands efficiently with better economies of scale and also provides effective maintenance and protection of data.

Common SAN deployments are Fibre Channel (FC) SAN and IP SAN. Fibre Channel SAN uses Fibre Channel protocol for the transport of data, commands, and status information between servers (or hosts) and storage devices. IP SAN uses IP-based protocols for communication.

## Understanding Fibre Channel

- High-speed network technology
  - ▶ Latest FC implementation supports speed up to 16 Gb/s
- Highly scalable
  - ▶ Theoretically, accommodate approximately 15 million devices



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 6

The FC architecture forms the fundamental construct of the FC SAN infrastructure. Fibre Channel is a high-speed network technology that runs on high-speed optical fiber cables and serial copper cables. The FC technology was developed to meet the demand for increased speeds of data transfer between servers and mass storage systems. Technical Committee T11, which is the committee within International Committee for Information Technology Standards (INCITS), is responsible for Fibre Channel interface standards.

High data transmission speed is an important feature of the FC networking technology. In comparison with Ultra-SCSI that is commonly used in DAS environments, FC is a significant leap in storage networking technology. **The latest FC implementations of 16 GFC (Fibre Channel) offers a throughput of 3200 MB/s (raw bit rates of 16 Gb/s), whereas Ultra640 SCSI is available with a throughput of 640 MB/s.** Credit-based flow control mechanism in FC delivers data as fast as the destination buffer is able to receive it, without dropping frames. Also FC has very little transmission overhead. The FC architecture is highly scalable, and theoretically, a single FC network can accommodate approximately 15 million devices.

Note: FibRE refers to the protocol, whereas fibER refers to a media.

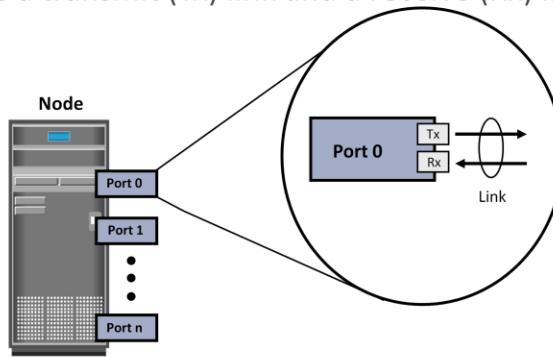
## Components of FC SAN

- Node (server and storage) ports
- Cables
- Connectors
- Interconnecting devices such as FC switches and hubs
- SAN management software

FC SAN is a network of servers and shared storage devices. Servers and storage are the end points or devices in the SAN (called 'nodes'). FC SAN infrastructure consists of node ports, cables, connectors, interconnecting devices (such as FC switches or hubs), along with SAN management software.

## Node Ports

- Provide physical interface for communicating with other nodes
- Exist on
  - ▶ HBA in server
  - ▶ Front-end adapters in storage
- Each port has a transmit (Tx) link and a receive (Rx) link



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

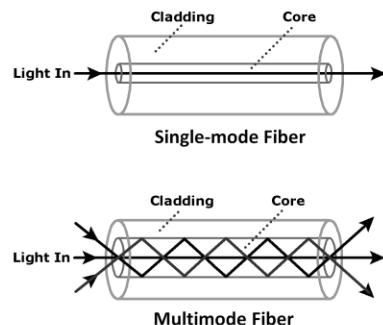
Module 5: Fibre Channel Storage Area Network 8

In a Fibre Channel network, the end devices, such as hosts, storage arrays, and tape libraries, are all referred to as nodes. Each node is a source or destination of information. Each node requires one or more ports to provide a physical interface for communicating with other nodes. These ports are integral components of host adapters, such as HBA, and storage front-end controllers or adapters. In an FC environment a port operates in full-duplex data transmission mode with a transmit (Tx) link and a receive (Rx) link

## Cables

- SAN implementation uses
  - ▶ Copper cables for short distance
  - ▶ Optical fiber cables for long distance
- Two types of optical cables: single-mode and multimode

Single-mode	Multimode
Carries single beam of light	Can carry multiple beams of light simultaneously
Distance up to 10km	Used for short distance (Modal dispersion weakens signal strength after certain distance)



SAN implementations use optical fiber cabling. Copper can be used for shorter distances for back-end connectivity because it provides acceptable signal-to-noise ratio for distances up to 30 meters. Optical fiber cables carry data in the form of light. There are two types of optical cables: multimode and single-mode. Multimode fiber (MMF) cable carries multiple beams of light projected at different angles simultaneously onto the core of the cable. Based on the bandwidth, multimode fibers are classified as OM1 (62.5µm core), OM2 (50µm core), and laser-optimized OM3 (50µm core). In an MMF transmission, multiple light beams traveling inside the cable tend to disperse and collide. This collision weakens the signal strength after it travels a certain distance—a process known as modal dispersion. An MMF cable is typically used for short distances because of signal degradation (attenuation) due to modal dispersion.

Single-mode fiber (SMF) carries a single ray of light projected at the center of the core. These cables are available in core diameters of 7 to 11 microns; the most common size is 9 microns. In an SMF transmission, a single light beam travels in a straight line through the core of the fiber. The small core and the single light wave help to limit modal dispersion. Among all types of fiber cables, single-mode provides minimum signal attenuation over maximum distance (up to 10 km). A single-mode cable is used for long-distance cable runs, and distance usually depends on the power of the laser at the transmitter and sensitivity of the receiver.

MMFs are generally used within data centers for shorter distance runs, whereas SMFs are used for longer distances.

## Connectors

- Attached at the end of a cable
- Enable swift connection and disconnection of the cable to and from a port
- Commonly used connectors for fiber optic cables are:
  - ▶ Standard Connector (SC)
    - ▶ Duplex connectors
  - ▶ Lucent Connector (LC)
    - ▶ Duplex connectors
  - ▶ Straight Tip (ST)
    - ▶ Patch panel connectors
    - ▶ Simplex connectors



Standard Connector



Lucent Connector



Straight Tip Connector

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 10

A connector is attached at the end of a cable to enable swift connection and disconnection of the cable to and from a port. A Standard connector (SC) and a Lucent connector (LC) are two commonly used connectors for fiber optic cables. Straight Tip (ST) is another fiber-optic connector, which is often used with fiber patch panels.

## Interconnecting Devices

- Commonly used interconnecting devices in FC SAN are:
  - ▶ Hubs, switches, and directors
- Hubs provide limited connectivity and scalability
- Switches and directors are intelligent devices
  - ▶ Switches are available with fixed port count or modular design
  - ▶ Directors are always modular, and its port count can be increased by inserting additional 'line cards' or 'blades'
  - ▶ High-end switches and directors contain redundant components

FC hubs, switches, and directors are the interconnect devices commonly used in FC SAN.

Hubs are used as communication devices in FC-AL implementations. Hubs physically connect nodes in a logical loop or a physical star topology. All the nodes must share the loop because data travels through all the connection points. Because of the availability of low-cost and high-performance switches, hubs are no longer used in FC SANs.

Switches are more intelligent than hubs and directly route data from one physical port to another. Therefore, nodes do not share the data path. Instead, each node has a dedicated communication path.

Directors are high-end switches with a higher port count and better fault-tolerance capabilities.

Switches are available with a fixed port count or with modular design. In a modular switch, the port count is increased by installing additional port cards to open slots. The architecture of a director is always modular, and its port count is increased by inserting additional line cards or blades to the director's chassis. High-end switches and directors contain redundant components to provide high availability. Both switches and directors have management ports (Ethernet or serial) for connectivity to SAN management servers.

## SAN Management Software

- A suite of tools used in a SAN to manage interfaces between host and storage arrays
- Provides integrated management of SAN environment
- Enables web-based management using GUI or CLI



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 12

SAN management software manages the interfaces between hosts, interconnect devices, and storage arrays. The software provides a view of the SAN environment and enables management of various resources from one central console.

It provides key management functions, including mapping of storage devices, monitoring and generating alerts for discovered devices, and zoning (discussed later in the module).

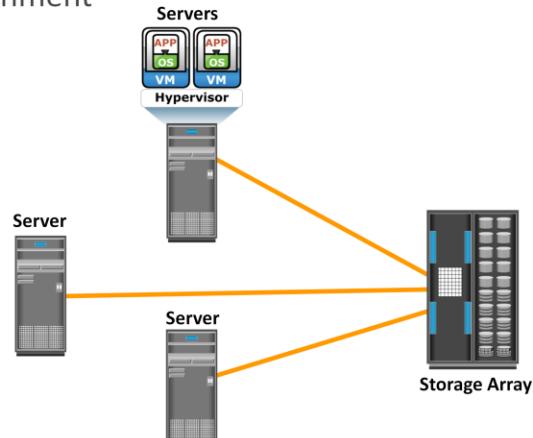
## FC Interconnectivity Options

- Point-to-Point
- Fibre Channel Arbitrated Loop (FC-AL)
- Fibre Channel Switched Fabric (FC-SW)

The FC architecture supports three basic interconnectivity options: point-to-point, fibre channel arbitrated loop (FC-AL), and fibre channel switched fabric (FC-SW).

## Point-to-Point Connectivity

- Enables direct connection between nodes
- Offers limited connectivity and scalability
- Used in DAS environment



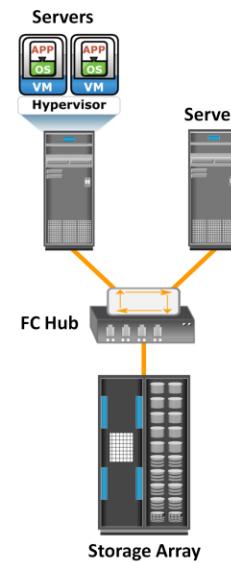
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 14

Point-to-point is the simplest FC configuration—two devices are connected directly to each other, as shown in the slide. This configuration provides a dedicated connection for data transmission between nodes. However, the point-to-point configuration offers limited connectivity, because only two devices can communicate with each other at a given time. Moreover, it cannot be scaled to accommodate a large number of nodes. Standard DAS uses point-to-point connectivity.

## FC-AL Connectivity

- Provides shared loop to attached nodes
  - ▶ Nodes must arbitrate to gain control
- Implemented using ring or star topology
- Limitations of FC-AL
  - ▶ Only one device can perform I/O operation at a time
  - ▶ Supports up to 126 nodes
  - ▶ Addition or removal of a node causes momentary pause in loop traffic



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 15

In the FC-AL configuration, devices are attached to a shared loop. FC-AL has the characteristics of a token ring topology and a physical star topology. In FC-AL, each device contends with other devices to perform I/O operations. Devices on the loop must “arbitrate” to gain control of the loop. At any given time, only one device can perform I/O operations on the loop.

As a loop configuration, FC-AL can be implemented without any interconnecting devices by directly connecting one device to another two devices in a ring through cables.

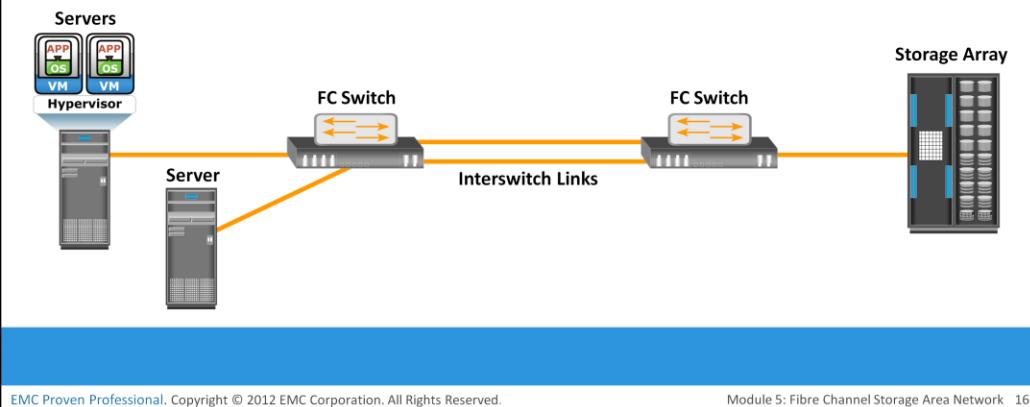
However, FC-AL implementations may also use hubs whereby the arbitrated loop is physically connected in a star topology.

The FC-AL configuration has the following limitations in terms of scalability:

- FC-AL shares the loop and only one device can perform I/O operations at a time. Because each device in a loop must wait for its turn to process an I/O request, overall performance in FC-AL environment is low.
- FC-AL uses only 8-bits of 24-bit Fibre Channel addressing (the remaining 16-bits are masked) and enables the assignment of 127 valid addresses to the ports. Hence, it can support up to 127 devices on a loop. One address is reserved for optionally connecting the loop to an FC switch port. Therefore, up to 126 nodes can be connected to the loop.
- Adding or removing a device results in loop re-initialization, which can cause a momentary pause in loop traffic.

## FC-SW Connectivity

- Creates a logical space (called fabric) in which all nodes communicate with one another using switches
  - ▶ Interswitch links (ISLs) enable switches to be connected together
- Provides dedicated path between nodes
- Addition/removal of node does not affect traffic of other nodes



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 16

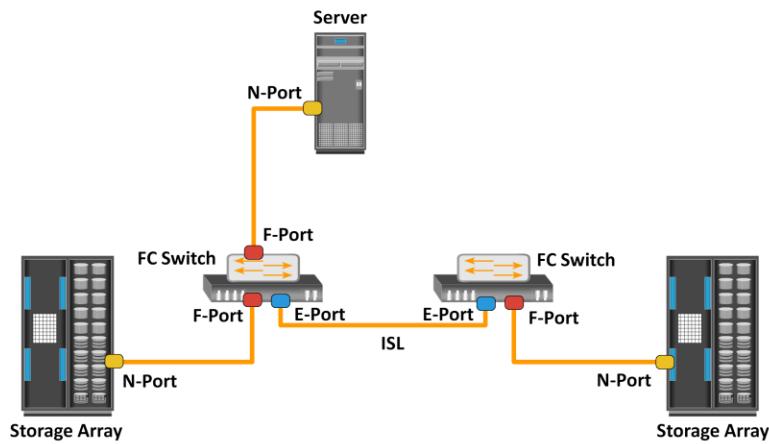
FC-SW is also referred to as fabric connect. A fabric is a logical space in which all nodes communicate with one another in a network. This virtual space can be created with a switch or a network of switches. Each switch in a fabric contains a unique domain identifier, which is part of the fabric's addressing scheme. In FC-SW, nodes do not share a loop; instead, data is transferred through a dedicated path between the nodes. Each port in a fabric has a unique 24-bit Fibre Channel address for communication.

In a switched fabric, the link between any two switches is called an interswitch link (ISL). ISLs enable switches to be connected together to form a single, larger fabric. ISLs are used to transfer host-to-storage data and fabric management traffic from one switch to another. By using ISLs, a switched fabric can be expanded to connect a large number of nodes.

FC-SW uses switches that are intelligent devices. They can switch data traffic between nodes directly through switch ports. Frames are routed between source and destination by the fabric.

Unlike a loop configuration, a FC-SW network provides dedicated path and scalability. The addition or removal of a device in a switched fabric is minimally disruptive; it does not affect the ongoing traffic between other devices.

## Port Types in Switched Fabric



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 17

Ports in a switched fabric can be one of the following types:

- **N\_Port:** An end point in the fabric. This port is also known as the node port. Typically, it is a host port (HBA) or a storage array port that is connected to a switch in a switched fabric.
- **E\_Port:** A port that forms the connection between two FC switches. This port is also known as the expansion port. The E\_Port on an FC switch connects to the E\_Port of another FC switch in the fabric ISLs.
- **F\_Port:** A port on a switch that connects an N\_Port. It is also known as a fabric port.
- **G\_Port:** A generic port on a switch that can operate as an E\_Port or an F\_Port and determines its functionality automatically during initialization.

## Module 5: Fibre Channel Storage Area Network (FC SAN)

### Lesson 2: Fibre Channel (FC) Architecture

During this lesson the following topics are covered:

- FC protocol stack
- FC addressing
- WWN addressing
- Structure and organization of FC data
- Fabric services
- Fabric login types

This lesson covers FC protocol stack, FC and WWN addressing, and structure and organization of FC data. This lesson also covers fabric services and login types.

## FC Architecture Overview

- Provides benefits of both channel and network technologies
  - ▶ Provides high performance with low protocol overheads
  - ▶ Provides high scalability with long distance capability
- Implements SCSI over FC network
  - ▶ Transports SCSI data through FC network
- Storage devices, attached to SAN, appear as local storage devices to host operating system

Traditionally, host computer operating systems have communicated with peripheral devices over channel connections, such as ESCON and SCSI. Channel technologies provide high levels of performance with low protocol overheads. Such performance is achievable due to the static nature of channels and the high level of hardware and software integration provided by the channel technologies. However, these technologies suffer from inherent limitations in terms of the number of devices that can be connected and the distance between these devices.

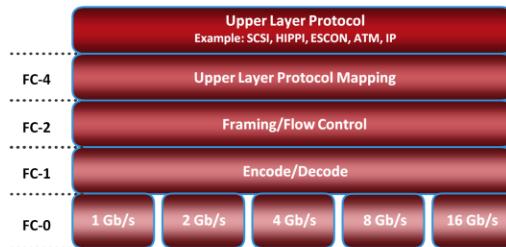
In contrast to channel technology, network technologies are more flexible and provide greater distance capabilities. Network connectivity provides greater scalability and uses shared bandwidth for communication. This flexibility results in greater protocol overhead and reduced performance.

The FC architecture represents true channel/network integration and captures some of the benefits of both channel and network technology. FC SAN uses the Fibre Channel Protocol (FCP) that provides both channel speed for data transfer with low protocol overhead and scalability of network technology.

FCP forms the fundamental construct of the FC SAN infrastructure. Fibre Channel provides a serial data transfer interface that operates over copper wire and optical fiber. FCP is the implementation of SCSI over an FC network. In FCP architecture, all external and remote storage devices attached to the SAN appear as local devices to the host operating system. The key advantages of FCP are as follows:

- Sustained transmission bandwidth over long distances.
- Support for a larger number of addressable devices over a network. Theoretically, FC can support more than 15 million device addresses on a network.
- Support speeds up to 16 Gbps (16 GFC).

## Fibre Channel Protocol Stack



FC Layer	Function	Features Specified by FC Layer
FC-4	Mapping interface	Mapping upper layer protocol (e.g. SCSI) to lower FC layers
FC-3	Common services	Not implemented
FC-2	Routing, flow control	Frame structure, FC addressing, flow control
FC-1	Encode/decode	8b/10b or 64b/66b encoding, bit and frame synchronization
FC-0	Physical layer	Media, cables, connector

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 20

It is easier to understand a communication protocol by viewing it as a structure of independent layers. FCP defines the communication protocol in five layers: FC-0 through FC-4 (except FC-3 layer, which is not implemented).

**FC-4 Layer:** It is the uppermost layer in the FCP stack. This layer defines the application interfaces and the way Upper Layer Protocols (ULPs) are mapped to the lower FC layers. The FC standard defines several protocols that can operate on the FC-4 layer. Some of the protocols include SCSI, High Performance Parallel Interface (HIPPI) Framing Protocol, Enterprise Storage Connectivity (ESCON), Asynchronous Transfer Mode (ATM), and IP.

**FC-2 Layer:** It provides Fibre Channel addressing, structure, and organization of data (frames, sequences, and exchanges). It also defines fabric services, classes of service, flow control, and routing.

**FC-1 Layer:** It defines how data is encoded prior to transmission and decoded upon receipt. At the transmitter node, an 8-bit character is encoded into a 10-bit transmission character. This character is then transmitted to the receiver node. At the receiver node, the 10-bit character is passed to the FC-1 layer, which decodes the 10-bit character into the original 8-bit character. FC links with speed 10 Gbps and above use 64-bit to 66-bit encoding algorithm. This layer also defines the transmission words such as FC frame delimiters, which identify the start and end of a frame and primitive signals that indicate events at a transmitting port. In addition to these, the FC-1 layer performs link initialization and error recovery.

**FC-0 Layer :** It is the lowest layer in the FCP stack. This layer defines the physical interface, media, and transmission of bits. The FC-0 specification includes cables, connectors, and optical and electrical parameters for a variety of data rates. The FC transmission can use both electrical and optical media.

## FC Addressing in Switched Fabric

- FC Address is assigned to nodes during fabric login
  - ▶ Used for communication between nodes within FC SAN
- Address format



- Domain ID is a unique number provided to each switch in the fabric
  - ▶ 239 addresses are available for domain ID
- Maximum possible number of node ports in a switched fabric:
  - ▶ 239 domains X 256 areas X 256 ports = 15,663,104

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 21

An FC address is dynamically assigned when a node port logs on to the fabric. The FC address has a distinct format, as shown in the slide. The first field of the FC address contains the domain ID of the switch. A Domain ID is a unique number provided to each switch in the fabric. Although this is an 8-bit field, there are only 239 available addresses for domain ID because some addresses are deemed special and reserved for fabric management services. For example, FFFF0C is reserved for the name server, and FFFFFE is reserved for the fabric login service. The area ID is used to identify a group of switch ports used for connecting nodes. An example of a group of ports with common area ID is a port card on the switch. The last field, the port ID, identifies the port within the group.

Therefore, the maximum possible number of node ports in a switched fabric is calculated as:  
239 domains X 256 areas X 256 ports = 15,663,104

## World Wide Name (WWN)

- Unique 64 bit identifier
- Static to node ports on an FC network
  - ▶ Similar to MAC address of NIC
  - ▶ WWNN and WWPN are used to uniquely identify nodes and ports respectively

World Wide Name - Array															
5	0	0	6	0	1	6	0	0	0	6	0	0	1	B	2
0101	0000	0000	0110	0000	0001	0110	0000	0000	0000	0110	0000	0000	0001	1011	0010

World Wide Name - HBA															
1	0	0	0	0	0	0	0	c	9	2	0	d	c	4	0
Format Type	Reserved 12 bits				Company ID 24 bits				Company Specific 24 bits						

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 22

Each device in the FC environment is assigned a 64-bit unique identifier called the World Wide Name (WWN). The Fibre Channel environment uses two types of WWNs: World Wide Node Name (WWNN) and World Wide Port Name (WWPN). Unlike an FC address, which is assigned dynamically, a WWN is a static name for each node on an FC network. WWNs are similar to the Media Access Control (MAC) addresses used in IP networking. WWNs are burned into the hardware or assigned through software. Several configuration definitions in a SAN use WWN for identifying storage devices and HBAs. The name server in an FC environment keeps the association of WWNs to the dynamically created FC addresses for nodes. Figure in the slide illustrates the WWN structure examples for an array and an HBA.

## Structure and Organization of FC Data

- FC data is organized as Exchange, Sequence, and Frame

FC Data Structure	Description
Exchange	<ul style="list-style-type: none"><li>Enables two N_Ports to identify and manage a set of information units<ul style="list-style-type: none"><li>Information unit: upper layer protocol-specific information that is sent to another port to perform certain operation</li><li>Each information unit maps to a sequence</li></ul></li><li>Includes one or more sequences</li></ul>
Sequence	<ul style="list-style-type: none"><li>Contiguous set of frames that correspond to an information unit</li></ul>
Frame	<ul style="list-style-type: none"><li>Fundamental unit of data transfer</li><li>Each frame consists of five parts: SOF, frame header, data field, CRC, and EOF</li></ul>



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 23

In an FC network, data transport is analogous to a conversation between two people, whereby a frame represents a word, a sequence represents a sentence, and an exchange represents a conversation.

**Exchange:** An exchange operation enables two node ports to identify and manage a set of information units. Each upper layer protocol has its protocol-specific information that must be sent to another port to perform certain operations. This protocol-specific information is called an information unit. The structure of these information units is defined in the FC-4 layer. This unit maps to a sequence. An exchange is composed of one or more sequences.

**Sequence:** A sequence refers to a contiguous set of frames that are sent from one port to another. A sequence corresponds to an information unit, as defined by the ULP.

**Frame:** A frame is the fundamental unit of data transfer at Layer 2. An FC frame consists of five parts: start of frame (SOF), frame header, data field, cyclic redundancy check (CRC), and end of frame (EOF). The SOF and EOF act as delimiters. The frame header is 24 bytes long and contains addressing information for the frame. The data field in an FC frame contains the data payload, up to 2,112 bytes of actual data—in most cases, the SCSI data. The CRC checksum facilitates error detection for the content of the frame. This checksum verifies data integrity by checking whether the content of the frames was received correctly. The CRC checksum is calculated by the sender before encoding at the FC-1 layer. Similarly, it is calculated by the receiver after decoding at the FC-1 layer.

## Fabric Services

- FC switches provide fabric services as defined in FC standards

Fabric Services	Description
Fabric Login Server	<ul style="list-style-type: none"><li>• Used during the initial part of the node's fabric login process</li><li>• Located at pre-defined address of FFFFFE</li></ul>
Name Server	<ul style="list-style-type: none"><li>• Responsible for name registration and management of node ports</li><li>• Located at pre-defined address FFFFFC</li></ul>
Fabric Controller	<ul style="list-style-type: none"><li>• Responsible for managing and distributing Registered State Change Notifications (RSCNs) to attached node ports</li><li>• Responsible for distributing SW-RSCNs to every other switch<ul style="list-style-type: none"><li>– SW-RSCNs keep the name server up-to-date on all switches</li></ul></li><li>• Located at pre-defined address FFFFFD</li></ul>
Management Server	<ul style="list-style-type: none"><li>• Enables FC SAN management using fabric management software</li><li>• Located at pre-defined address FFFFFA</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 24

All FC switches, regardless of the manufacturer, provide a common set of services as defined in the Fibre Channel standards. These services are available at certain predefined addresses. Some of these services are Fabric Login Server, Fabric Controller, Name Server, and Management Server.

The Fabric Login Server is located at the predefined address of FFFFFE and is used during the initial part of the node's fabric login process.

The Name Server (formally known as Distributed Name Server) is located at the predefined address FFFFFC and is responsible for name registration and management of node ports. Each switch exchanges its Name Server information with other switches in the fabric to maintain a synchronized, distributed name service.

Each switch has a Fabric Controller located at the predefined address FFFFFD. The Fabric Controller provides services to both node ports and other switches. The Fabric Controller is responsible for managing and distributing Registered State Change Notifications (RSCNs) to the node ports registered with the Fabric Controller. If there is a change in the fabric, RSCNs are sent out by a switch to the attached node ports. The Fabric Controller also generates Switch Registered State Change Notifications (SW-RSCNs) to every other domain (switch) in the fabric. These RSCNs keep the name server up-to-date on all switches in the fabric.

FFFFFA is the Fibre Channel address for the Management Server. The Management Server is distributed to every switch within the fabric. The Management Server enables the FC SAN management software to retrieve information and administer the fabric.

## Login Types in Switched Fabric

- Fabric login (FLOGI)
  - ▶ Occurs between an N\_Port and an F\_Port
  - ▶ Node sends a FLOGI frame with WWN to Fabric Login Server on switch
  - ▶ Node obtains FC address from switch
  - ▶ Immediately after FLOGI, N\_Port registers with Name Server on switch, indicating its WWN, port type, assigned FC address, etc.
  - ▶ N\_Port queries name server about all other logged in ports
- Port login (PLOGI)
  - ▶ Occurs between two N\_Ports to establish a session
  - ▶ Exchange service parameters relevant to the session
- Process login (PRLI)
  - ▶ Occurs between two N\_Ports to exchange ULP related parameters

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 25

Fabric services define three login types:

- Fabric login (FLOGI): Performed between an N\_Port and an F\_Port. To log on to the fabric, a node sends a FLOGI frame with the WWNN and WWPN parameters to the login service at the predefined FC address FFFFFE (Fabric Login Server). In turn, the switch accepts the login and returns an Accept (ACC) frame with the assigned FC address for the node. Immediately after the FLOGI, the N\_Port registers itself with the local Name Server on the switch, indicating its WWNN, WWPN, port type, class of service, assigned FC address and so on. After the N\_Port has logged in, it can query the name server database for information about all other logged in ports.
- Port login (PLOGI): Performed between two N\_Ports to establish a session. The initiator N\_Port sends a PLOGI request frame to the target N\_Port, which accepts it. The target N\_Port returns an ACC to the initiator N\_Port. Next, the N\_Ports exchange service parameters relevant to the session.
- Process login (PRLI): Also performed between two N\_Ports. This login relates to the FC-4 ULPs, such as SCSI. If the ULP is SCSI, N\_Ports exchange SCSI-related service parameters.

## Module 5: Fibre Channel Storage Area Network (FC SAN)

### Lesson 3: FC SAN Topologies and Zoning

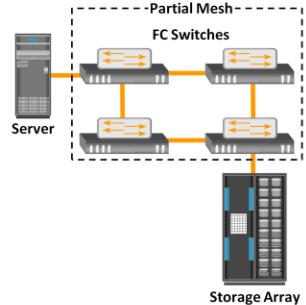
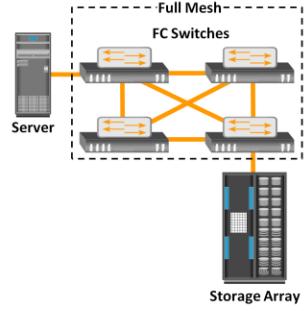
During this lesson the following topics are covered:

- Mesh and core-edge topologies
- Benefits of zoning
- Types of zoning

This lesson covers FC SAN topologies such as mesh and core-edge. This lesson also covers zoning and its benefits, components, and types.

## Mesh Topology

- Full mesh
  - ▶ Each switch is connected to every other switch
  - ▶ Maximum of one ISL or hop is required for host-to-storage traffic
  - ▶ Host and storage can be connected to any switch
- Partial mesh
  - ▶ Not all the switches are connected to every other switch



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

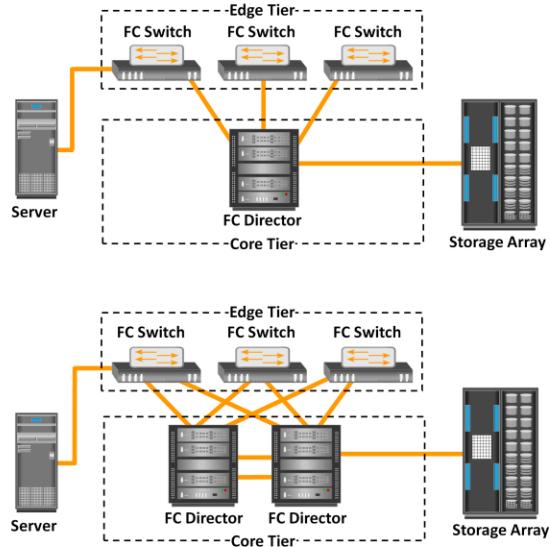
Module 5: Fibre Channel Storage Area Network 27

A mesh topology may be one of the two types: full mesh or partial mesh. In a full mesh, every switch is connected to every other switch in the topology. A full mesh topology may be appropriate when the number of switches involved is small. A typical deployment would involve up to four switches or directors, with each of them servicing highly localized host-to-storage traffic. In a full mesh topology, a maximum of one ISL or hop is required for host-to-storage traffic. However, with the increase in the number of switches, the number of switch ports used for ISL also increases. This reduces the available switch ports for node connectivity.

In a partial mesh topology, several hops or ISLs may be required for the traffic to reach its destination. Partial mesh offers more scalability than full mesh topology. However, without proper placement of host and storage devices, traffic management in a partial mesh fabric might be complicated and ISLs could become overloaded due to excessive traffic aggregation.

## Core-edge Topology

- Consists of edge and core switch tiers
- Network traffic traverses core tier or terminate at core tier
- Storage is usually connected to the core tier
- Benefits
  - ▶ High availability
  - ▶ Medium scalability
  - ▶ Medium to maximum connectivity



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 28

The core-edge fabric topology has two types of switch tiers. The edge tier is usually composed of switches and offers an inexpensive approach to adding more hosts in a fabric. Each switch at the edge tier is attached to a switch at the core tier through ISLs.

The core tier is usually composed of directors that ensure high fabric availability. In addition, typically all traffic must either traverse this tier or terminate at this tier. In this configuration, all storage devices are connected to the core tier, enabling host-to-storage traffic to traverse only one ISL. Hosts that require high performance may be connected directly to the core tier and consequently avoid ISL delays.

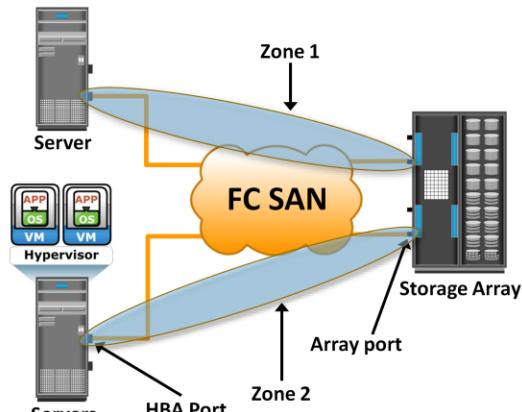
In core-edge topology, the edge-tier switches are not connected to each other. The core-edge fabric topology increases connectivity within the SAN while conserving the overall port utilization. If fabric expansion is required, additional edge switches are connected to the core. The core of the fabric is also extended by adding more switches or directors at the core tier. Based on the number of core-tier switches, this topology has different variations, such as, single-core topology and dual-core topology. To transform a single-core topology to dual-core, new ISLs are created to connect each edge switch to the new core switch in the fabric.

# Zoning

## Zoning

It is an FC switch function that enables node ports within the fabric to be logically segmented into groups, and communicate with each other within the group.

- Zone set comprises zones
- Each zone comprises zone members (HBA and array ports)
- Benefits
  - ▶ Restricts RSCN traffic
  - ▶ Provides access control



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 29

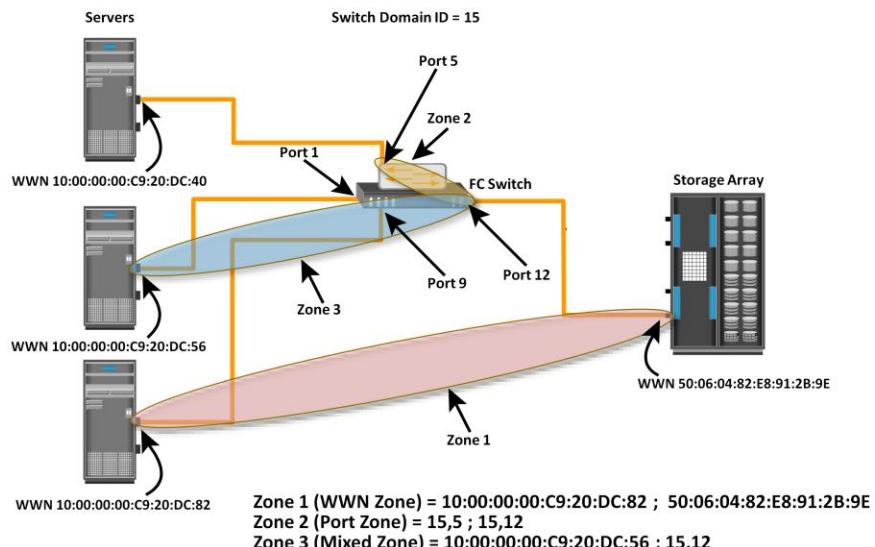
Zoning is an FC switch function that enables node ports within the fabric to be logically segmented into groups and communicate with each other within the group.

Whenever a change takes place in the name server database, the fabric controller sends a Registered State Change Notification (RSCN) to all the nodes impacted by the change. If zoning is not configured, the fabric controller sends an RSCN to all the nodes in the fabric. Involving the nodes that are not impacted by the change results in increased fabric-management traffic. For a large fabric, the amount of FC traffic generated due to this process can be significant and might impact the host-to-storage data traffic. Zoning helps to limit the number of RSCNs in a fabric. In the presence of zoning, a fabric sends the RSCN to only those nodes in a zone where the change has occurred.

Zoning also provides access control, along with other access control mechanisms, such as LUN masking. Zoning provides control by allowing only the members in the same zone to establish communication with each other.

Zone members, zones, and zone sets form the hierarchy defined in the zoning process. A zone set is composed of a group of zones that can be activated or deactivated as a single entity in a fabric. Multiple zone sets may be defined in a fabric, but only one zone set can be active at a time. Members are nodes within the SAN that can be included in a zone. Switch ports, HBA ports, and storage device ports can be members of a zone. A port or node can be a member of multiple zones. Nodes distributed across multiple switches in a switched fabric may also be grouped into the same zone. Zone sets are also referred to as zone configurations.

## Types of Zoning



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 30

Zoning can be categorized into three types:

- **Port zoning:** Uses the physical address of switch ports to define zones. In port zoning, access to node is determined by the physical switch port to which a node is connected. The zone members are the port identifier (switch domain ID and port number) to which HBA and its targets (storage devices) are connected. If a node is moved to another switch port in the fabric, port zoning must be modified to allow the node, in its new port, to participate in its original zone. However, if an HBA or storage device port fails, an administrator just has to replace the failed device without changing the zoning configuration.
- **WWN zoning:** Uses World Wide Names to define zones. The zone members are the unique WWN addresses of the HBA and its targets (storage devices). A major advantage of WWN zoning is its flexibility. WWN zoning allows nodes to be moved to another switch port in the fabric and maintain connectivity to its zone partners without having to modify the zone configuration. This is possible because the WWN is static to the node port.
- **Mixed zoning:** Combines the qualities of both WWN zoning and port zoning. Using mixed zoning enables a specific node port to be tied to the WWN of another node.

Figure in the slide shows the three types of zoning on an FC network.

## Module 5: Fibre Channel Storage Area Network (FC SAN)

### Lesson 4: Virtualization in SAN

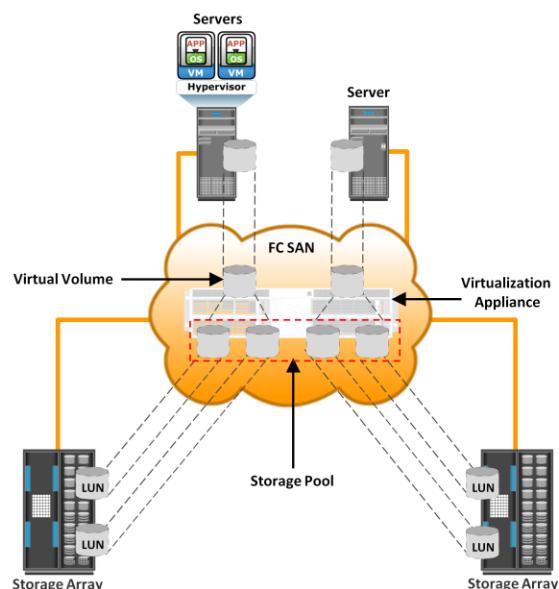
During this lesson the following topics are covered:

- Block-level storage virtualization
- Virtual SAN

This lesson covers block-level storage virtualization and virtual SAN.

## Block-level Storage Virtualization

- Provides a virtualization layer in SAN
- Abstracts block storage devices and creates a storage pool by aggregating LUNs
- Virtual volumes are created from storage pool and assigned to hosts
  - ▶ Virtualization layer maps virtual volumes to LUNs
- Benefits
  - ▶ Online volume expansion
  - ▶ Nondisruptive migration



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 32

Block-level storage virtualization aggregates block storage devices (LUNs) and enables provisioning of virtual storage volumes, independent of the underlying physical storage. A virtualization layer, which exists at the SAN, abstracts the identity of physical storage devices and creates a storage pool from heterogeneous storage devices. Virtual volumes are created from the storage pool and assigned to the hosts. Instead of being directed to the LUNs on the individual storage arrays, the hosts are directed to the virtual volumes provided by the virtualization layer. For hosts and storage arrays, the virtualization layer appears as the target and initiator devices, respectively. The virtualization layer maps the virtual volumes to the LUNs on the individual arrays. The hosts remain unaware of the mapping operation and access the virtual volumes as if they were accessing the physical storage attached to them. Typically, the virtualization layer is managed via a dedicated virtualization appliance to which the hosts and the storage arrays are connected.

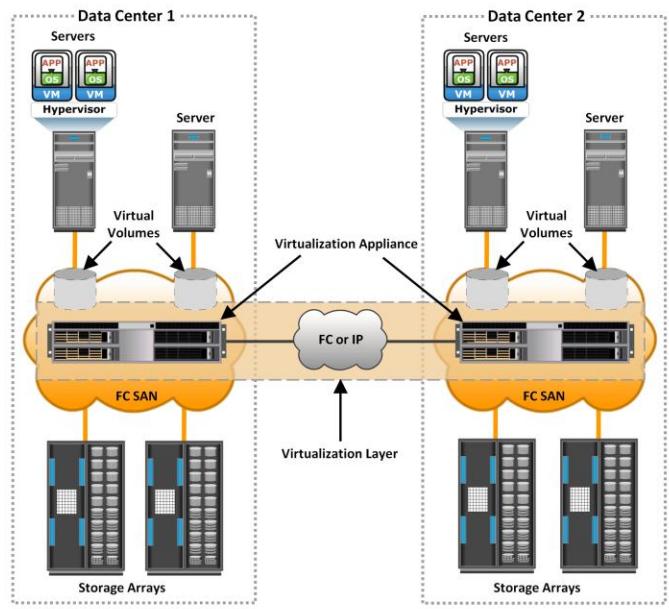
Figure in the slide illustrates a virtualized environment. It shows two physical servers, each of which has one virtual volume assigned. These virtual volumes are used by the servers. These virtual volumes are mapped to the LUNs in the storage arrays. When an I/O is sent to a virtual volume, it is redirected through the virtualization layer at the storage network to the mapped LUNs. Depending on the capabilities of the virtualization appliance, the architecture may allow for more complex mapping between array LUNs and virtual volumes.

cont..

Block-level storage virtualization enables extending the storage volumes online to meet application growth requirements. It consolidates heterogeneous storage arrays and enables transparent volume access.

Block-level storage virtualization also provides the advantage of nondisruptive data migration. In a traditional SAN environment, LUN migration from one array to another is an offline event because the hosts needed to be updated to reflect the new array configuration. In other instances, host CPU cycles were required to migrate data from one array to the other, especially in a multivendor environment. With a block-level virtualization solution in place, the virtualization layer handles the back-end migration of data, which enables LUNs to remain online and accessible while data is migrating. No physical changes are required because the host still points to the same virtual targets on the virtualization layer. However, the mappings information on the virtualization layer should be changed. These changes can be executed dynamically and are transparent to the end user.

## Use Case: Block-level Storage Virtualization across Data Centers



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 34

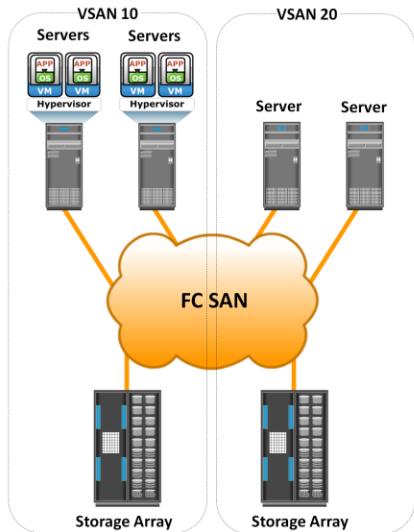
Previously, block-level storage virtualization provided nondisruptive data migration only within a data center. The new generation of block-level storage virtualization enables nondisruptive data migration both within and between data centers. It provides the capability to connect the virtualization layers at multiple data centers. The connected virtualization layers are managed centrally and work as a single virtualization layer stretched across data centers. This enables the federation of block-storage resources both within and across data centers. The virtual volumes are created from the federated storage resources.

## Virtual SAN (VSAN)/Virtual Fabric

### VSAN

It is a logical fabric on an FC SAN, enabling communication among a group of nodes, regardless of their physical location in the fabric.

- Each VSAN has its own fabric services (name server, zoning), configuration, and set of FC addresses
- VSANs improve SAN security, scalability, availability, and manageability



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 35

Virtual SAN (also called virtual fabric) is a logical fabric on an FC SAN, which enables communication among a group of nodes regardless of their physical location in the fabric. In a VSAN, a group of hosts or storage ports communicate with each other using a virtual topology defined on the physical SAN. Multiple VSANs may be created on a single physical SAN. Each VSAN acts as an independent fabric with its own set of fabric services, such as name server, and zoning. Fabric-related configurations in one VSAN do not affect the traffic in another.

VSANs improve SAN security, scalability, availability, and manageability. VSANs provide enhanced security by isolating the sensitive data in a VSAN and by restricting access to the resources located within that VSAN. The same Fibre Channel address can be assigned to nodes in different VSANs, thus increasing the fabric scalability. Events causing traffic disruptions in one VSAN are contained within that VSAN and are not propagated to other VSANs. VSANs facilitate an easy, flexible, and less expensive way to manage networks. Configuring VSANs is easier and quicker compared to building separate physical FC SANs for various node groups. To regroup nodes, an administrator simply changes the VSAN configurations without moving nodes and recabling.

## Module 5: Fibre Channel Storage Area Network (FC SAN)

### Concept in Practice:

- EMC Connectrix
- EMC VPLEX

The Concept in Practice section covers EMC Connectrix and VPLEX.

## EMC Connectrix

- Connectrix family includes networked storage connectivity products
  - ▶ Offers high-speed FC connectivity, highly resilient switching technology, intelligent IP storage networking, and I/O consolidation with Fibre Channel over Ethernet
- Connectrix family consist of enterprise directors, departmental switches, and multi-purpose switches



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 37

The EMC Connectrix family represents the industry's most extensive selection of networked storage connectivity products. Connectrix integrates high-speed FC connectivity, highly resilient switching technology, options for intelligent IP storage networking, and I/O consolidation with products that support Fibre Channel over Ethernet (FCoE). The connectivity products offered under the Connectrix brand are: Enterprise directors, departmental switches, and multi-purpose switches .

Enterprise directors offer high port density and high component redundancy. They are deployed in high-availability or large-scale environments. Connectrix directors offer several hundred ports per domain. Departmental switches are best suited for workgroup, mid-tier environments. Multi-purpose switches support various protocols such as iSCSI, FCIP, FCoE, FICON, in addition to FC protocol. In addition to FC ports, Connectrix switches and directors have Ethernet ports and serial ports for communication and switch management functions..

EMC ControlCenter SAN Manager provides a single interface for managing a SAN. With SAN Manager, an administrator can discover, monitor, manage, and configure complex heterogeneous SAN environments. It streamlines and centralizes SAN management operations across multivendor storage networks and storage devices. It enables storage administrators to manage SAN zones and LUN masking consistently across multivendor SAN arrays and switches. EMC ControlCenter SAN Manager also supports virtual environments, including VMware, and virtual SANs.

EMC ProSphere is a newly launched tool with additional features specifically for the cloud computing environment. A future release of EMC ProSphere will include all the functionalities of EMC ControlCenter.

## EMC VPLEX

- Provides solution for block-level storage virtualization and data mobility – both within and across data centers
- Enables multiple hosts located across two locations to access a single copy of data
- Provides capability to mirror a virtual volume – both within and across locations
  - ▶ Allows hosts at different data centers to simultaneously access cache-coherent copies of the same virtual volume
- VPLEX family consists of three products
  - ▶ VPLEX Local
  - ▶ VPLEX Metro
  - ▶ VPLEX Geo

EMC VPLEX is the next-generation solution for block-level virtualization and data mobility both within and across datacenters. The VPLEX appliance resides between the servers and heterogeneous storage devices. It forms a pool of distributed block storage resources and enables creating virtual storage volumes from the pool. These virtual volumes are then allocated to the servers. The virtual-to-physical-storage mapping remains hidden to the servers.

VPLEX provides nondisruptive data mobility among physical storage devices to balance the application workload and to enable both local and remote data access. The mapping of virtual volumes to physical volumes can be changed dynamically by the administrator.

VPLEX uses a unique clustering architecture and distributed cache coherency that enable multiple hosts located across two locations to access a single copy of data. VPLEX also provides the capability to mirror data of a virtual volume both within and across locations. This enables hosts at different data centers to access cache-coherent copies of the same virtual volume. To avoid application downtime due to outage at a data center, the workload can be moved quickly to another data center. Applications continue accessing the same virtual volume and remain uninterrupted by the data mobility.

The VPLEX family consists of three products: VPLEX Local, VPLEX Metro, and VPLEX Geo.

EMC VPLEX Local delivers local federation, which provides simplified management and nondisruptive data mobility across heterogeneous arrays within a data center. EMC VPLEX Metro delivers distributed federation, which provides data access and mobility between two VPLEX clusters within synchronous distances that support round-trip latency up to 5 ms. EMC VPLEX Geo delivers data access and mobility between two VPLEX clusters within asynchronous distances (that support round-trip latency up to 50 ms).

## Module 5: Summary

Key points covered in this module:

- FC SAN components and connectivity options
- FC protocol stack and addressing
- Structure and organization of FC data
- Fabric services
- Fabric topologies
- Types of zoning
- Block-level storage virtualization and virtual SAN

This module covered FC SAN components – node port, cable, connector, interconnecting devices, and SAN management software; FC connectivity options – point-to-point, FC-AL, FC-SW; and fabric port types such as N\_Port, E\_Port, F\_Port, and G\_Port. It includes FC protocol stack and addressing, structure and organization of FC data, and fabric services and login types. This module also covered fabric topologies – core-edge and mesh; types of zoning – port, WWN, and mixed; block-level storage virtualization; and virtual SAN.

## Check Your Knowledge – 1

- Which cable type provides minimum signal attenuation over long distance?
  - A. Twisted-pair copper
  - B. Coaxial copper
  - C. Single-mode optical
  - D. Multimode optical
- What is an F\_Port in FC SAN?
  - A. Switch port that connects an E\_Port
  - B. Switch port that connects an N\_Port
  - C. Node port that connects an N\_Port
  - D. Node port that connects an E\_Port

## Check Your Knowledge – 2

- Which type of fabric login enables the exchange of upper layer protocol-related parameters between N\_Ports?
  - A. Fabric login
  - B. Port login
  - C. Process login
  - D. ULP login
- Which is a benefit of zoning in FC SAN?
  - A. Isolates fabric services
  - B. Restricts RSCN traffic
  - C. Enables online volume expansion
  - D. Provides non-disruptive data migration

## Check Your Knowledge – 3

- Which is a benefit of VSAN?
  - A. Eliminates need for fabric login process
  - B. Provides higher network bandwidth
  - C. Improves security by isolating traffic between VSANs
  - D. Enables VSANs to share fabric zoning service

## Exercise: FC SAN

- Current situation

- ▶ An organization's IT infrastructure consists of three storage arrays direct-attached to a heterogeneous mix of 45 servers
- ▶ Servers are dual-attached to the arrays to ensure high availability
- ▶ Each storage array has 32 front-end ports, which could support a maximum of 16 servers
- ▶ Each storage array has the disk capacity to support a maximum of 32 servers

### Current situation:

The IT infrastructure of an organization consists of three storage arrays direct-attached to a heterogeneous mix of 45 servers. All servers are dual-attached to the arrays for high availability. Because each storage array has 32 front-end ports, each could support a maximum of 16 servers. However, each existing storage array has the disk capacity to support a maximum of 32 servers.

## Exercise: FC SAN (contd..)

- Organization's challenges/requirements
  - ▶ Organization needs to deploy additional 45 servers to meet growth requirements
  - ▶ Existing storage arrays are poorly utilized and adding of new servers require to purchase new arrays
  - ▶ Organization wants to implement FC SAN to overcome the scalability and utilization challenges
  - ▶ Hop count must be minimized to meet performance requirement
- Task
  - ▶ Propose a FC switched fabric topology to address organization's challenges/requirements and justify your choice
  - ▶ If 72-port FC switches are available, determine the minimum number of switches required in the fabric

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 5: Fibre Channel Storage Area Network 44

### Organization's challenges/requirements:

The organization plans to purchase 45 more servers to meet its growth requirements. If it continues using direct-attached storage, the organization needs to purchase additional storage arrays to connect these new servers. The organization realizes that its existing storage arrays are poorly utilized; therefore, it plans to implement FC SAN to overcome the scalability and utilization challenges. The organization uses high-performance applications; therefore, it wants to minimize the hop count for the server's access to storage.

### Task:

Propose a FC switched fabric topology to address organization's challenges and requirements. Justify your choice of the fabric topology.

If 72-port switches are available for FC SAN implementation, determine the minimum number of switches required in the fabric.

# Module – 6

# IP SAN and FCoE



## Module 6: IP SAN and FCoE

Upon completion of this module, you should be able to:

- Describe IP SAN protocols, components, and topology
- Describe FCoE protocol, components, and topology

This module focuses on IP SAN protocols such as Internet SCSI (iSCSI) and Fibre Channel over IP (FCIP), infrastructure components, and topology. It also focuses on Fibre Channel over Ethernet (FCoE) protocol, infrastructure components, and topology.

# Module 6: IP SAN and FCoE

## Lesson 1: IP SAN

During this lesson the following topics are covered:

- Drivers for IP SAN
- IP SAN Protocols: iSCSI and FCIP
- Components, topologies, and protocol stack for iSCSI and FCIP

Two primary protocols that leverage IP as the transport mechanism are Internet SCSI (iSCSI) and Fibre Channel over IP (FCIP). This lesson covers the drivers for IP SAN and iSCSI components, topologies, protocol stack, and discovery methods. It also covers FCIP protocol stack and topology.

## Drivers for IP SAN

- IP SAN transports block-level data over IP network
- IP is being positioned as a storage networking option because:
  - ▶ Existing network infrastructure can be leveraged
  - ▶ Reduced cost compared to investing in new FC SAN hardware and software
  - ▶ Many long-distance disaster recovery solutions already leverage IP-based network
  - ▶ Many robust and mature security options are available for IP network

Traditional SAN enables the transfer of block I/O over Fibre Channel and provides high performance and scalability. These advantages of FC SAN come with the additional cost of buying FC components, such as FC HBA and switches. Organizations typically have an existing Internet Protocol (IP)-based infrastructure, which could be leveraged for storage networking. Advancements in technology have enabled IP to be used for transporting block I/O over the IP network. This technology of transporting block I/Os over an IP is referred to as IP SAN. IP is a mature technology, and using IP as a storage networking option provides several advantages. When block I/O is run over IP, the existing network infrastructure can be leveraged, which is more economical than investing in a new SAN infrastructure. In addition, many robust and mature security options are now available for IP networks. Many long-distance, disaster recovery (DR) solutions are already leveraging IP-based networks. With IP SAN, organizations can extend the geographical reach of their storage infrastructure.

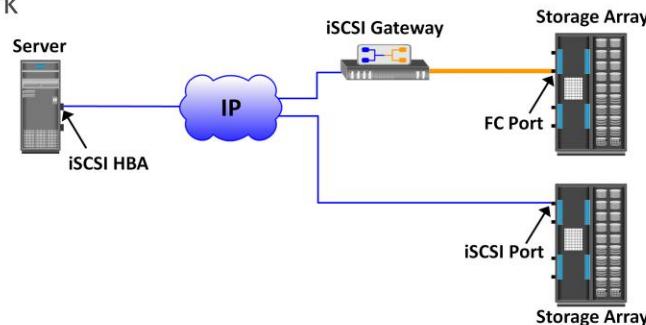
## IP SAN Protocol: iSCSI

- IP based protocol that is used to connect host and storage
- Encapsulates SCSI commands and data into an IP packet and transports them using TCP/IP

iSCSI is encapsulation of SCSI I/O over IP. iSCSI is an IP based protocol that establishes and manages connections between host and storage over IP. iSCSI encapsulates SCSI commands and data into an IP packet and transports them using TCP/IP. iSCSI is widely adopted for connecting servers to storage because it is relatively inexpensive and easy to implement, especially environments in which an FC SAN does not exist.

## Components of iSCSI

- iSCSI initiator
  - ▶ Example: iSCSI HBA
- iSCSI target
  - ▶ Storage array with iSCSI port
  - ▶ iSCSI gateway – enables communication with FC storage array
- IP network



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 6

An initiator (host), target (storage or iSCSI gateway), and an IP-based network are the key iSCSI components. If an iSCSI-capable storage array is deployed, then a host with the iSCSI initiator can directly communicate with the storage array over an IP network. However, in an implementation that uses an existing FC array for iSCSI communication, an iSCSI gateway is used. These devices perform the translation of IP packets to FC frames and vice versa, thereby bridging the connectivity between the IP and FC environments.

## iSCSI Host Connectivity Options

- Standard NIC with software iSCSI initiator
  - ▶ NIC provides network interface
  - ▶ Software initiator provides iSCSI functionality
  - ▶ Requires host CPU cycles for iSCSI and TCP/IP processing
- TCP Offload Engine (TOE) NIC with software iSCSI initiator
  - ▶ Moves TCP processing load off the host CPU onto the NIC card
  - ▶ Software initiator provides iSCSI functionality
  - ▶ Requires host CPU cycles for iSCSI processing
- iSCSI HBA
  - ▶ Offloads both iSCSI and TCP/IP processing from host CPU
  - ▶ Simplest option for boot from SAN

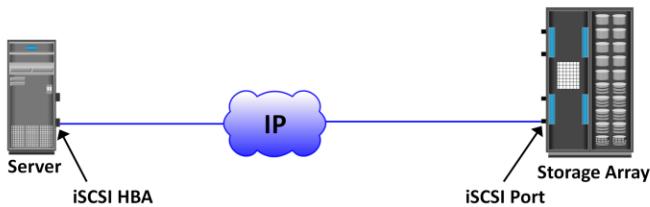
A standard NIC with software iSCSI initiator, a TCP offload engine (TOE) NIC with software iSCSI initiator, and an iSCSI HBA are the three iSCSI host connectivity options. The function of the iSCSI initiator is to route the SCSI commands over an IP network.

A standard NIC with a software iSCSI initiator is the simplest and least expensive connectivity option. It is easy to implement because most servers come with at least one, and in many cases two, embedded NICs. It requires only a software initiator for iSCSI functionality. Because NICs provide standard IP function, encapsulation of SCSI into IP packets and decapsulation are carried out by the host CPU. This places additional overhead on the host CPU. If a standard NIC is used in heavy I/O load situations, the host CPU might become a bottleneck. TOE NIC helps alleviate this burden. A TOE NIC offloads TCP management functions from the host and leaves only the iSCSI functionality to the host processor. The host passes the iSCSI information to the TOE card, and the TOE card sends the information to the destination using TCP/IP. Although this solution improves performance, the iSCSI functionality is still handled by a software initiator that requires host CPU cycles.

An iSCSI HBA is capable of providing performance benefits because it offloads the entire iSCSI and TCP/IP processing from the host processor. The use of an iSCSI HBA is also the simplest way to boot hosts from a SAN environment via iSCSI. If there is no iSCSI HBA, modifications must be made to the basic operating system to boot a host from the storage devices because the NIC needs to obtain an IP address before the operating system loads. The functionality of an iSCSI HBA is similar to the functionality of an FC HBA.

## iSCSI Topologies: Native iSCSI

- iSCSI initiators are either directly attached to storage array or connected through IP network
  - ▶ No FC component
- Storage array has iSCSI port
- Each iSCSI port is configured with an IP address



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

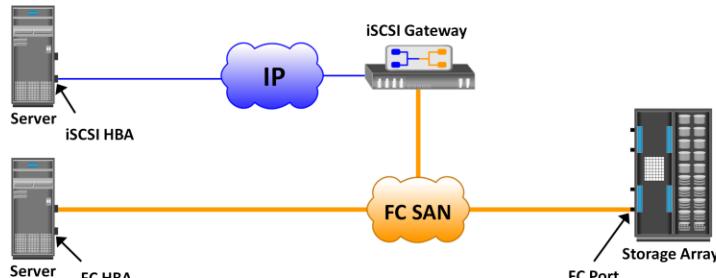
Module 6: IP SAN and FCoE 8

Two topologies of iSCSI implementations are native and bridged. Native topology does not have FC components. The initiators may be either directly attached to targets or connected through the IP network.

FC components are not required for iSCSI connectivity if an iSCSI-enabled array is deployed. In figure in the slide, the array has one or more iSCSI ports configured with an IP address and connected to a standard Ethernet switch. After an initiator is logged on to the network, it can access the available LUNs on the storage array. A single array port can service multiple hosts or initiators as long as the array port can handle the amount of storage traffic that the hosts generate.

## iSCSI Topologies: Bridged iSCSI

- iSCSI gateway is used to enable communication between iSCSI host and FC storage
- iSCSI gateway works as bridge between FC and IP network
  - ▶ Converts IP packets to FC frames and vice versa
- iSCSI initiator is configured with gateway's IP address as its target
- iSCSI gateway is configured as FC initiator to storage array



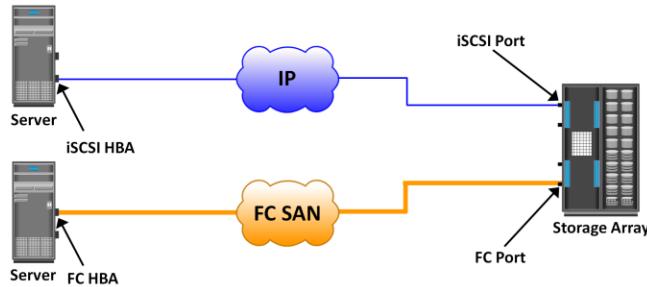
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 9

Bridged topology enables the coexistence of FC with IP by providing iSCSI-to-FC bridging functionality. Figure in the slide illustrates an iSCSI host connectivity to an FC storage array. In this case, the array does not have any iSCSI ports. Therefore, an external device, called a gateway or a multiprotocol router, must be used to facilitate the communication between the iSCSI host and FC storage. The gateway converts IP packets to FC frames and vice versa. The bridge devices contain both FC and Ethernet ports to facilitate the communication between the FC and IP environments. In bridged iSCSI implementation, the iSCSI initiator is configured with the gateway's IP address as its target destination. On the other side, the gateway is configured as an FC initiator to the storage array.

## Combining FC and Native iSCSI Connectivity

- Array provides both FC and iSCSI ports
  - ▶ Enable iSCSI and FC connectivity in the same environment
  - ▶ No bridge devices needed

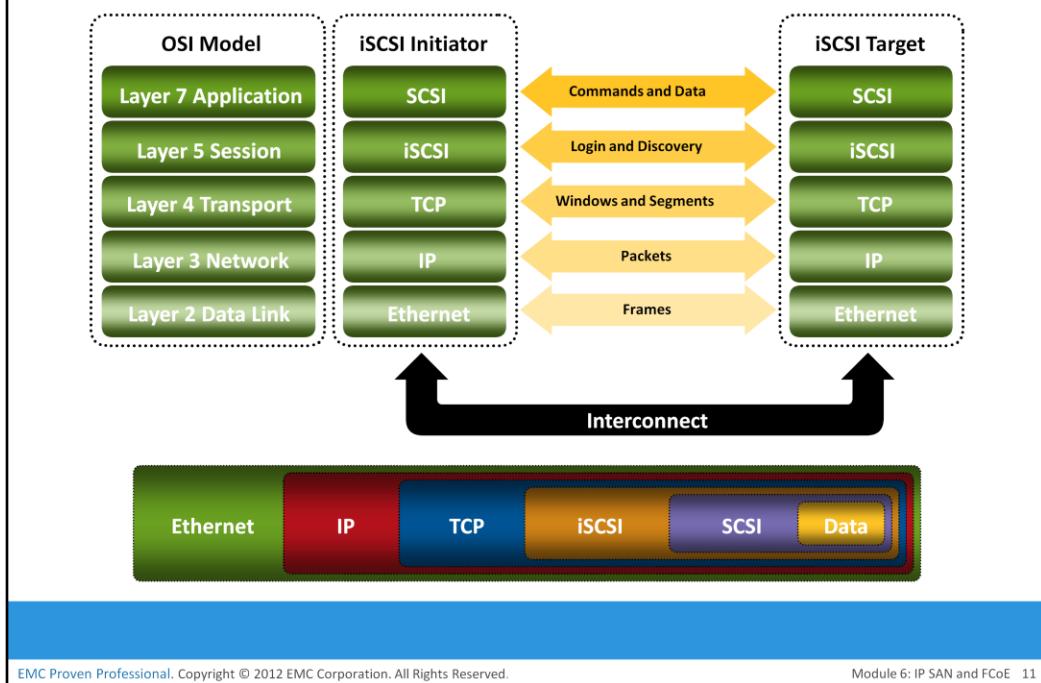


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 10

The most common topology is a combination of FC and native iSCSI. Typically, a storage array comes with both FC and iSCSI ports that enable iSCSI and FC connectivity in the same environment, as shown in the slide.

## iSCSI Protocol Stack



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 11

Figure in the slide displays a model of the iSCSI protocol layers and depicts the encapsulation order of the SCSI commands for their delivery through a physical carrier.

SCSI is the command protocol that works at the application layer of the Open System Interconnection (OSI) model. The initiators and targets use SCSI commands and responses to talk to each other. The SCSI command descriptor blocks, data, and status messages are encapsulated into TCP/IP and transmitted across the network between the initiators and targets.

iSCSI is the session-layer protocol that initiates a reliable session between devices that recognize SCSI commands and TCP/IP. The iSCSI session-layer interface is responsible for handling login, authentication, target discovery, and session management. TCP is used with iSCSI at the transport layer to provide reliable transmission.

TCP controls message flow, windowing, error recovery, and retransmission. It relies upon the network layer of the OSI model to provide global addressing and connectivity. The Layer 2 protocols at the data link layer of this model enable node-to-node communication through a physical network.

## iSCSI Discovery

- For iSCSI communication, initiator must discover location and name of target on a network
- iSCSI discovery takes place in two ways:
  - ▶ SendTargets discovery
    - ▶ Initiator is manually configured with the target's network portal
    - ▶ Initiator issues SendTargets command; target responds with required parameters
  - ▶ Internet Storage Name Service (iSNS)
    - ▶ Initiators and targets register themselves with iSNS server
    - ▶ Initiator can query iSNS server for a list of available targets

An initiator must discover the location of its targets on the network and the names of the targets available to it before it can establish a session. This discovery can take place in two ways: SendTargets discovery or internet Storage Name Service (iSNS).

In SendTargets discovery, the initiator is manually configured with the target's network portal to establish a discovery session. The initiator issues the SendTargets command, and the target network portal responds with the required parameters of the targets available to the host.

iSNS enables automatic discovery of iSCSI devices on an IP network. The initiators and targets can be configured to automatically register themselves with the iSNS server. Whenever an initiator wants to know the targets that it can access, it can query the iSNS server for a list of available targets.

## iSCSI Name

- iSCSI name is a unique iSCSI identifier that is used to identify initiators and targets within an iSCSI network
- Two common types of iSCSI names are:
  - ▶ iqn: iSCSI Qualified Name
    - ▶ iqn.2008-02.com.example:optional\_string
  - ▶ eui: Extended Unique Identifier
    - ▶ eui.0300732A32598D26

A unique worldwide iSCSI identifier, known as an iSCSI name, is used to identify the initiators and targets within an iSCSI network to facilitate communication. The unique identifier can be a combination of the names of the department, application, or manufacturer, serial number, asset number, or any tag that can be used to recognize and manage the devices. Following are two types of iSCSI names commonly used:

- **iSCSI Qualified Name (IQN):** An organization must own a registered domain name to generate iSCSI Qualified Names. This domain name does not need to be active or resolve to an address. It just needs to be reserved to prevent other organizations from using the same domain name to generate iSCSI names. A date is included in the name to avoid potential conflicts caused by the transfer of domain names. An example of an IQN is iqn.2008-02.com.example:*optional\_string*. The *optional\_string* provides a serial number, an asset number, or any other device identifiers. An iSCSI Qualified Name enables storage administrators to assign meaningful names to iSCSI devices, and therefore, manage those devices more easily.
- **Extended Unique Identifier (EUI):** An EUI is a globally unique identifier based on the IEEE EUI-64 naming standard. An EUI is composed of the eui prefix followed by a 16-character hexadecimal name, such as eui.0300732A32598D26.

In either format, the allowed special characters are dots, dashes, and blank spaces.

## IP SAN Protocol: FCIP

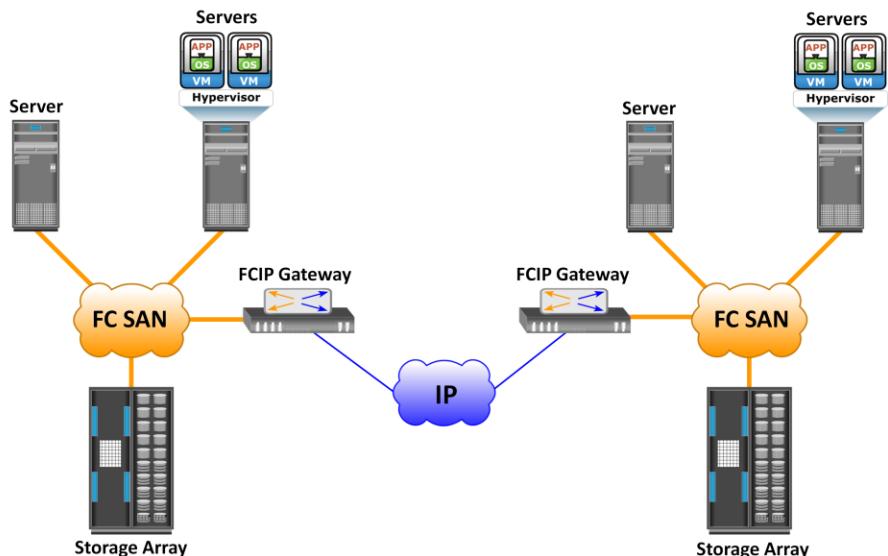
- IP-based protocol that is used to connect distributed FC SAN islands
- Creates virtual FC links over existing IP network that is used to transport FC data between different FC SANs
- Encapsulates FC frames onto IP packet
- Provides disaster recovery solution

FC SAN provides a high-performance infrastructure for localized data movement. Organizations are now looking for ways to transport data over a long distance between their disparate SANs at multiple geographic locations. One of the best ways to achieve this goal is to interconnect geographically dispersed SANs through reliable, high-speed links. This approach involves transporting the FC block data over the IP infrastructure. FCIP is a tunneling protocol that enables distributed FC SAN islands to be interconnected over the existing IP-based networks.

FCIP is a protocol in which FCIP entity such as FCIP gateway is used to tunnel FC fabrics through an IP network. In FCIP FC frames are encapsulated onto the IP payload. An FCIP implementation is capable to merge interconnected fabrics into a single fabric. Frequently, only a small subset of nodes at either end requires connectivity across fabrics. Thus, the majority of FCIP implementations today use switch-specific features such as IVR (Inter-VSAN Routing) or FCRS (Fibre Channel Routing Services) to create a tunnel. In this manner, traffic may be routed between specific nodes without actually merging the fabrics.

The FCIP standard has rapidly gained acceptance as a manageable, cost-effective way to blend the best of the two worlds: FC SAN and the proven, widely deployed IP infrastructure. As a result, organizations now have a better way to store, protect, and move their data by leveraging investments in their existing IP infrastructure. FCIP is extensively used in disaster recovery implementations in which data is duplicated to the storage located at a remote site.

## FCIP Topology

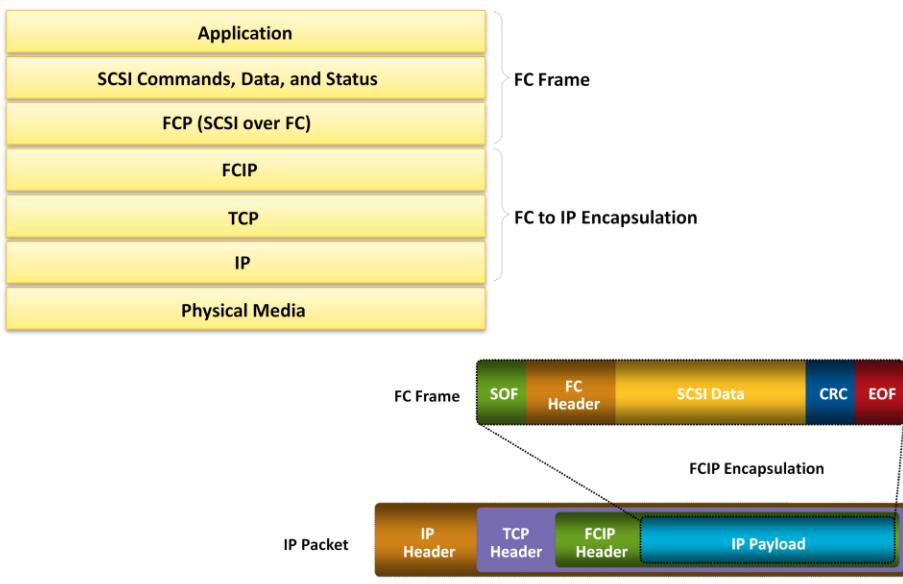


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 15

In an FCIP environment, an FCIP gateway is connected to each fabric via a standard FC connection. The FCIP gateway at one end of the IP network encapsulates the FC frames into IP packets. The gateway at the other end removes the IP wrapper and sends the FC data to the layer 2 fabric. The fabric treats these gateways as layer 2 fabric switches. An IP address is assigned to the port on the gateway, which is connected to an IP network. After the IP connectivity is established, the nodes in the two independent fabrics can communicate with other.

## FCIP Protocol Stack



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 16

The FCIP protocol stack is shown in the slide. Applications generate SCSI commands and data, which are processed by various layers of the protocol stack. The upper layer protocol SCSI includes the SCSI driver program that executes the read-and-write commands. Below the SCSI layer is the Fibre Channel Protocol (FCP) layer, which is simply a fibre channel frame whose payload is SCSI. The FCP layer rides on top of the Fibre Channel transport layer. This enables the FC frames to run natively within a SAN fabric environment. In addition, the FC frames can be encapsulated into the IP packet and sent to a remote SAN over the IP. The FCIP layer encapsulates the Fibre Channel frames onto the IP payload and passes them to the TCP layer. TCP and IP are used for transporting the encapsulated information across Ethernet, wireless, or other media that support the TCP/IP traffic.

Encapsulation of FC frame on to IP packet could cause the IP packet to be fragmented when the data link cannot support the maximum transmission unit (MTU) size of an IP packet. When an IP packet is fragmented, the required parts of the header must be copied by all fragments. When a TCP packet is segmented, normal TCP operations are responsible for receiving and re-sequencing the data prior to passing it on to the FC processing portion of the device.

## Module 6: IP SAN and FCoE

### Lesson 2: Fibre Channel over Ethernet (FCoE)

During this lesson the following topics are covered:

- Drivers for FCoE
- Components of FCoE network
- FCoE frame mapping
- Converged Enhanced Ethernet (CEE)

This lesson covers the drivers of FCoE, components of FCoE network, and FCoE frame mapping. It also covers converged enhanced ethernet (CEE).

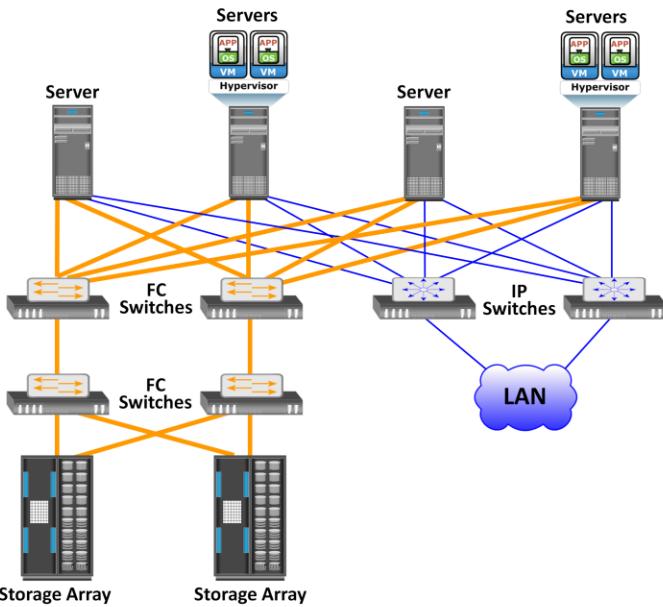
## Drivers for FCoE

- FCoE is a protocol that transports FC data over Ethernet network (Converged Enhanced Ethernet)
- FCoE is being positioned as a storage networking option because:
  - ▶ Enables consolidation of FC SAN traffic and Ethernet traffic onto a common Ethernet infrastructure
  - ▶ Reduces the number of adapters, switch ports, and cables
  - ▶ Reduces cost and eases data center management
  - ▶ Reduces power and cooling cost, and floor space

Data centers typically have multiple networks to handle various types of I/O traffic—for example, an Ethernet network for TCP/IP communication and an FC network for FC communication. TCP/IP is typically used for client-server communication, data backup, infrastructure management communication, and so on. FC is typically used for moving block-level data between storage and servers. To support multiple networks, servers in a data center are equipped with multiple redundant physical network interfaces—for example, multiple Ethernet and FC cards/adapters. In addition, to enable the communication, different types of networking switches and physical cabling infrastructure are implemented in data centers. The need for two different kinds of physical network infrastructure increases the overall cost and complexity of data center operation.

Fibre Channel over Ethernet (FCoE) protocol provides consolidation of LAN and SAN traffic over a single physical interface infrastructure. FCoE helps organizations address the challenges of having multiple discrete network infrastructures. FCoE uses the Converged Enhanced Ethernet (CEE) link (10 Gigabit Ethernet) to send FC frames over Ethernet.

## Data Center Infrastructure – Before Using FCoE

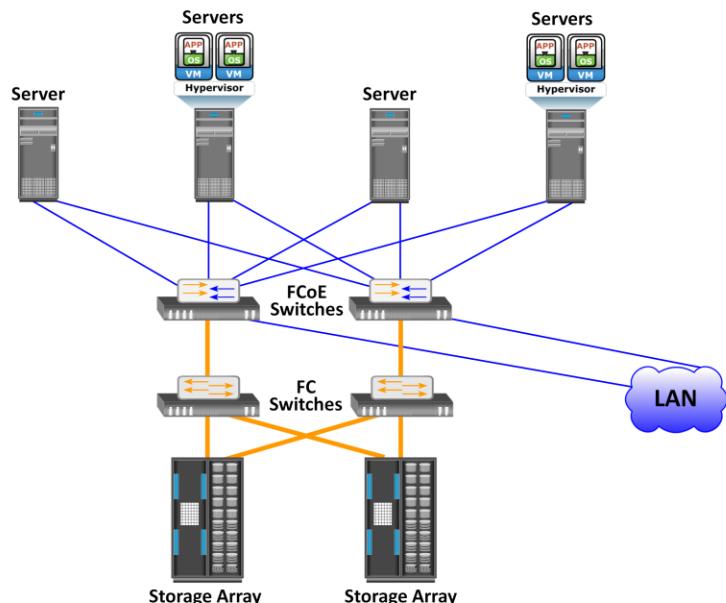


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 19

Figure in the slide represents the infrastructure before FCoE deployment. Here, the storage resources are accessed using HBAs, and the IP network resources are accessed using NICs by the servers. Typically, in a data center, a server is configured with 2 to 4 NIC cards and redundant HBA cards. If the data center has hundreds of servers, it would require a large number of adapters, cables, and switches. This leads to a complex environment, which is difficult to manage and scale. The cost of power, cooling, and floor space further adds to the challenge.

## Data Center Infrastructure – After Using FCoE



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 20

Figure in the slide shows the I/O consolidation with FCoE using FCoE switches and Converged Network Adapters (CNAs). A CNA replaces both HBAs and NICs in the server and consolidates both the IP and FC traffic. This reduces the requirement of multiple network adapters at the server to connect to different networks. Overall, this reduces the requirement of adapters, cables, and switches. This also considerably reduces the cost and management overhead.

## Components of an FCoE Network

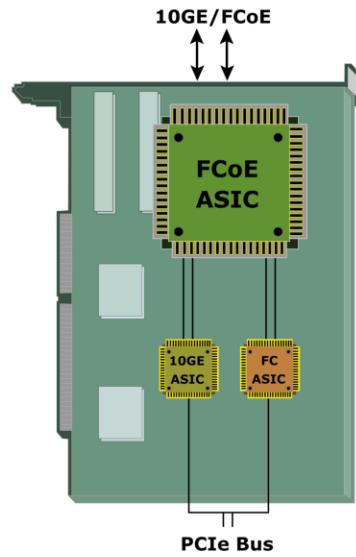
- Converged Network Adapter (CNA)
- Cable
- FCoE switch

The key FCoE components are:

- Converged Network Adapter (CNA)
- Cable
- FCoE switch

## Converged Network Adapter (CNA)

- Provides functionality of both – a standard NIC and an FC HBA
  - ▶ Eliminates the need to deploy separate adapters and cables for FC and Ethernet communications
- Contains separate modules for 10 Gigabit Ethernet, FC, and FCoE ASICs
  - ▶ FCoE ASIC encapsulates FC frames into Ethernet frames



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 22

A CNA provides the functionality of both a standard NIC and an FC HBA in a single adapter and consolidates both types of traffic. CNA eliminates the need to deploy separate adapters and cables for FC and Ethernet communications, thereby reducing the required number of server slots and switch ports. CNA offloads the FCoE protocol processing task from the server, thereby freeing the server CPU resources for application processing. A CNA contains separate modules for 10 Gigabit Ethernet, Fibre Channel, and FCoE Application Specific Integrated Circuits (ASICs). The FCoE ASIC encapsulate FC frames into Ethernet frames. One end of this ASIC is connected to 10GbE and FC ASICs for server connectivity, while the other end provides a 10GbE interface to connect to an FCoE switch.

## Cable

- Two options are available for FCoE cabling
  - ▶ Copper based Twinax cable
  - ▶ Standard fiber optical cable

Twinax Cable	Fiber Optical Cable
Suitable for shorter distances (up to 10 meters)	Can run over longer distances
Requires less power and are less expensive than fiber optical cable	Relatively more expensive than Twinax cables
Uses Small Form Factor Pluggable Plus (SFP+) connector	Uses Small Form Factor Pluggable Plus (SFP+) connector

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

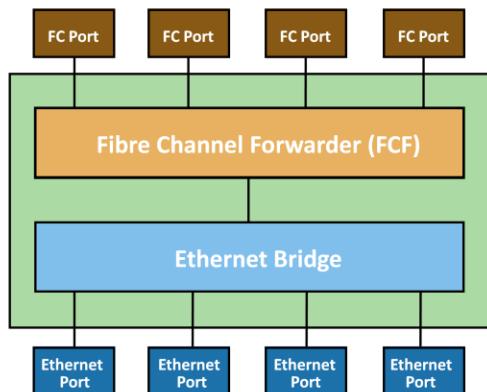
Module 6: IP SAN and FCoE 23

Currently two options are available for FCoE cabling: Copper based Twinax and standard fiber optical cables. A Twinax cable is composed of two pairs of copper cables covered with a shielded casing. The Twinax cable can transmit data at the speed of 10 Gbps over shorter distances up to 10 meters. Twinax cables require less power and are less expensive than fiber optic cables.

The Small Form Factor Pluggable Plus (SFP+) connector is the primary connector used for FCoE links and can be used with both optical and copper cables.

## FCoE Switch

- Provides both Ethernet and FC switch functionalities
- Consists of FCF, Ethernet bridge, and set of CEE ports and FC ports (optional)
  - ▶ FCF encapsulates and de-encapsulates FC frames
- Forwards frames based on Ethertype



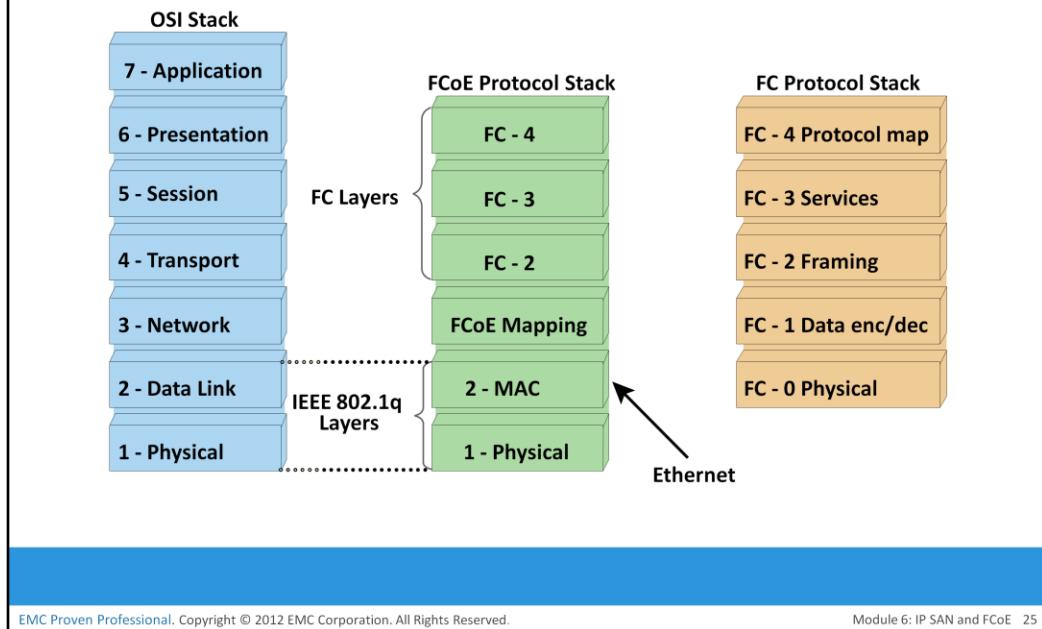
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 24

An FCoE switch has both Ethernet switch and Fibre Channel switch functionalities. The FCoE switch has a Fibre Channel Forwarder (FCF), Ethernet Bridge, and set of Ethernet ports and optional FC ports. The function of the FCF is to encapsulate the FC frames, received from the FC port, into the FCoE frames and also to de-encapsulate the FCoE frames, received from the Ethernet Bridge, to the FC frames.

Upon receiving the incoming traffic, the FCoE switch inspects the Ethertype (used to indicate which protocol is encapsulated in the payload of an Ethernet frame) of the incoming frames and uses that to determine the destination. If the Ethertype of the frame is FCoE, the switch recognizes that the frame contains an FC payload and forwards it to the FCF. From there, the FC is extracted from the FCoE frame and transmitted to FC SAN over the FC ports. If the Ethertype is not FCoE, the switch handles the traffic as usual Ethernet traffic and forwards it over the Ethernet ports.

## FCoE Frame Mapping



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 25

The encapsulation of the Fibre Channel frame occurs through the mapping of the FC frames onto Ethernet, as shown in the slide. Fibre Channel and traditional networks have stacks of layers where each layer in the stack represents a set of functionalities. The FC stack consists of five layers: FC-0 through FC-4. Ethernet is typically considered as a set of protocols that operates at the physical and data link layers in the seven-layer OSI stack. The FCoE protocol specification replaces the FC-0 and FC-1 layers of the FC stack with Ethernet. This provides the capability to carry the FC-2 to the FC-4 layer over the Ethernet layer.

A typical Fibre Channel data frame has a 2,112-byte payload, a 24-byte header, and an FCS. A standard Ethernet frame has a default payload capacity of 1,500 bytes. To maintain good performance, FCoE must use jumbo frames to prevent a Fibre Channel frame from being split into two Ethernet frames.

## Converged Enhanced Ethernet

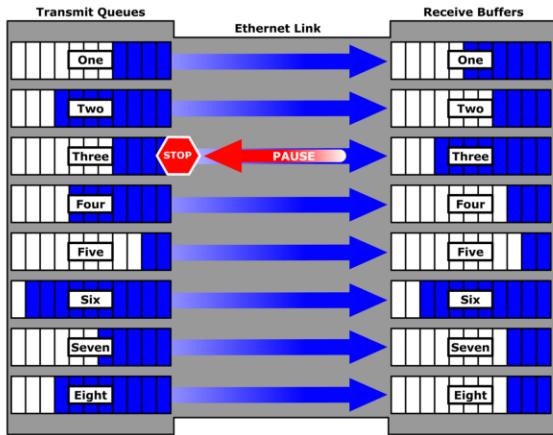
- Provides lossless Ethernet
- Lossless Ethernet requires following functionalities:
  - ▶ Priority-based flow control (PFC)
  - ▶ Enhanced transmission selection (ETS)
  - ▶ Congestion notification (CN)
  - ▶ Data center bridging exchange protocol(DCBX)

Conventional Ethernet is lossy in nature, which means that frames might be dropped or lost during transmission. Converged Enhanced Ethernet (CEE) or lossless Ethernet provides a new specification to the existing Ethernet standard that eliminates the lossy nature of Ethernet. This makes 10 Gb Ethernet a viable storage networking option, similar to FC. Lossless Ethernet requires certain functionalities. These functionalities are defined and maintained by the data center bridging (DCB) task group, which is a part of the IEEE 802.1 working group and they are:

- Priority-based flow control
- Enhanced transmission selection
- Congestion notification
- Data center bridging exchange protocol

## Priority-Based Flow Control (PFC)

- Creates eight virtual links on a single physical link
- Uses PAUSE capability of Ethernet for each virtual link
  - A virtual link can be paused and restarted independently
  - PAUSE mechanism is based on user priorities or classes of service



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 27

Traditional FC manages congestion through the use of a link-level, credit-based flow control that guarantees no loss of frames. Typical Ethernet, coupled with TCP/IP, uses a packet drop flow control mechanism. The packet drop flow control is not lossless. This challenge is eliminated by using an IEEE 802.3x Ethernet PAUSE control frame to create a lossless Ethernet. A receiver can send a PAUSE request to a sender when the receiver's buffer is filling up. Upon receiving a PAUSE frame, the sender stops transmitting frames, which guarantees no loss of frames. The downside of using the Ethernet PAUSE frame is that it operates on the entire link, which might be carrying multiple traffic flows.

PFC provides a link level flow control mechanism. PFC creates eight separate virtual links on a single physical link and allows any of these links to be paused and restarted independently. PFC enables the pause mechanism based on user priorities or classes of service. Enabling the pause based on priority allows creating lossless links for traffic, such as FCoE traffic. This PAUSE mechanism is typically implemented for FCoE while regular TCP/IP traffic continues to drop frames. Figure in the slide illustrates how a physical Ethernet link is divided into eight virtual links and allows a PAUSE for a single virtual link without affecting the traffic for the others.

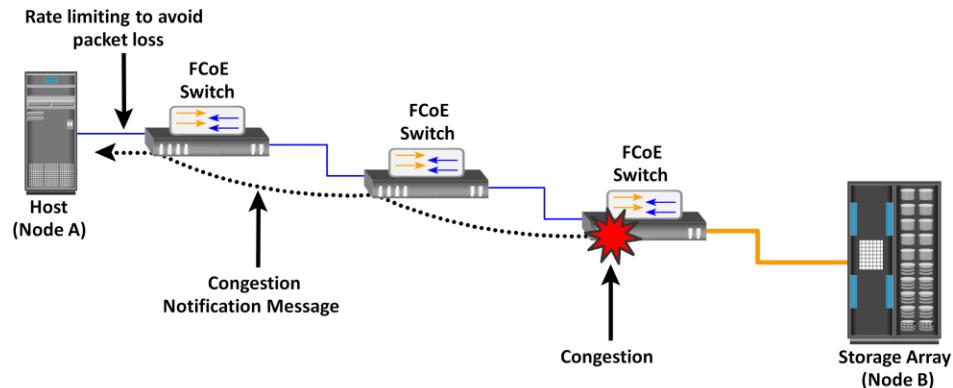
## Enhanced Transmission Selection (ETS)

- Allocates bandwidth to different traffic classes such as LAN, SAN, and Inter Process Communication (IPC)
- Provides available bandwidth to other classes of traffic when a particular class of traffic does not use its allocated bandwidth

Enhanced transmission selection provides a common management framework for the assignment of bandwidth to different traffic classes, such as LAN, SAN, and Inter Process Communication (IPC). When a particular class of traffic does not use its allocated bandwidth, ETS enables other traffic classes to use the available bandwidth.

## Congestion Notification (CN)

- Provides a mechanism for detecting congestion and notifying the source
  - ▶ Enables a switch to send a signal to other ports that need to stop or slow down their transmissions



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 6: IP SAN and FCoE 29

Congestion notification provides end-to-end congestion management for protocols, such as FCoE, that do not have built-in congestion control mechanisms. Link level congestion notification provides a mechanism for detecting congestion and notifying the source to move the traffic flow away from the congested links. Link level congestion notification enables a switch to send a signal to other ports that need to stop or slow down their transmissions. The process of congestion notification and its management is shown in the slide, which represents the communication between the nodes A (sender) and B (receiver). If congestion at the receiving end occurs, the algorithm running on the switch, generates a congestion notification (CN) message to the sending node (Node A). In response to the CN message, the sending end limits the rate of data transfer.

## Data Center Bridging Exchange Protocol (DCBX)

- Enables CEE devices to convey and configure their features with other CEE devices in the network
  - ▶ Allows a switch to distribute configuration values to attached adapters
- Ensures consistent configuration across network

DCBX protocol is a discovery and capability exchange protocol, which helps Converged Enhanced Ethernet devices to convey and configure their features with the other CEE devices in the network. DCBX is used to negotiate capabilities between the switches and the adapters, and it allows the switch to distribute the configuration values to all the attached adapters. This helps to ensure consistent configuration across the entire network.

## Module 6: Summary

Key points covered in this module:

- IP SAN protocols, their components, and topologies
- FCoE protocol, its components, and topology

This module covered IP SAN protocols such as iSCSI and FCIP, their components, and topologies. It also covered FCoE protocol, its components, and topology.

## Check Your Knowledge – 1

- Which iSCSI host connectivity option offloads both iSCSI and TCP/IP processing from the host CPU?
  - A. Standard NIC with iSCSI initiator software
  - B. TOE NIC
  - C. iSCSI HBA
  - D. CNA
- Which type of iSCSI name requires a registered domain name to generate unique iSCSI identifier?
  - A. eui
  - B. iqn
  - C. WWN
  - D. MAC

## Check Your Knowledge – 2

- Which protocol encapsulates FC frames onto IP packet?
  - A. FCoE
  - B. iSCSI
  - C. FCIP
  - D. CIFS
- Which is a feature of priority-based flow control?
  - A. A virtual link can be paused independently
  - B. All virtual links are paused together
  - C. Enables pausing virtual links based on their bandwidth
  - D. Enables pausing individual physical links based on their priority

## Check Your Knowledge – 3

- Which functionality enables allocation of bandwidth to different traffic classes in an FCoE environment?
  - A. Priority-based flow control
  - B. Enhanced transmission selection
  - C. Congestion notification
  - D. Data center bridging exchange

# Module – 7

# Network-Attached Storage (NAS)



## Module 7: Network-Attached Storage (NAS)

Upon completion of this module, you should be able to:

- Describe NAS, its benefits, and components
- Discuss NAS file-sharing protocols
- Describe different NAS implementations
- Describe file-level virtualization

This module focuses on benefits and components of network-attached storage (NAS). It also focuses on NAS file-sharing protocols, different NAS implementations, and file-level virtualization.

# Module 7: Network-Attached Storage (NAS)

## Lesson 1: NAS Components and Benefits

During this lesson the following topics are covered:

- File sharing technology evolution
- Benefits of NAS
- NAS components
- NAS file sharing protocols
- NAS I/O operations

This lesson covers a comparison of general purpose file server and NAS. It also describes key components of NAS, file sharing protocols (NFS and CIFS), and NAS I/O operations.

## File Sharing Environment

- File sharing enables users to share files with other users
- Creator or owner of a file determines the type of access to be given to other users
- File sharing environment ensures data integrity when multiple users access a shared file at the same time
- Examples of file sharing methods:
  - ▶ File Transfer Protocol (FTP)
  - ▶ Distributed File System (DFS)
  - ▶ Network File System (NFS) and Common Internet File System (CIFS)
  - ▶ Peer-to-Peer (P2P)

File sharing, as the name implies, enables users to share files with other users. In a file-sharing environment, a user who creates the file (the creator or owner of a file) determines the type of access (such as read, write, execute, append, delete) to be given to other users and controls changes to the file. When multiple users try to access a shared file at the same time, a locking scheme is required to maintain data integrity and, at the same time, make this sharing possible.

Some examples of file-sharing methods are; File Transfer Protocol (FTP), Distributed File System (DFS), client-server models that use file-sharing protocols such as NFS and CIFS, and the peer-to-peer (P2P) model.

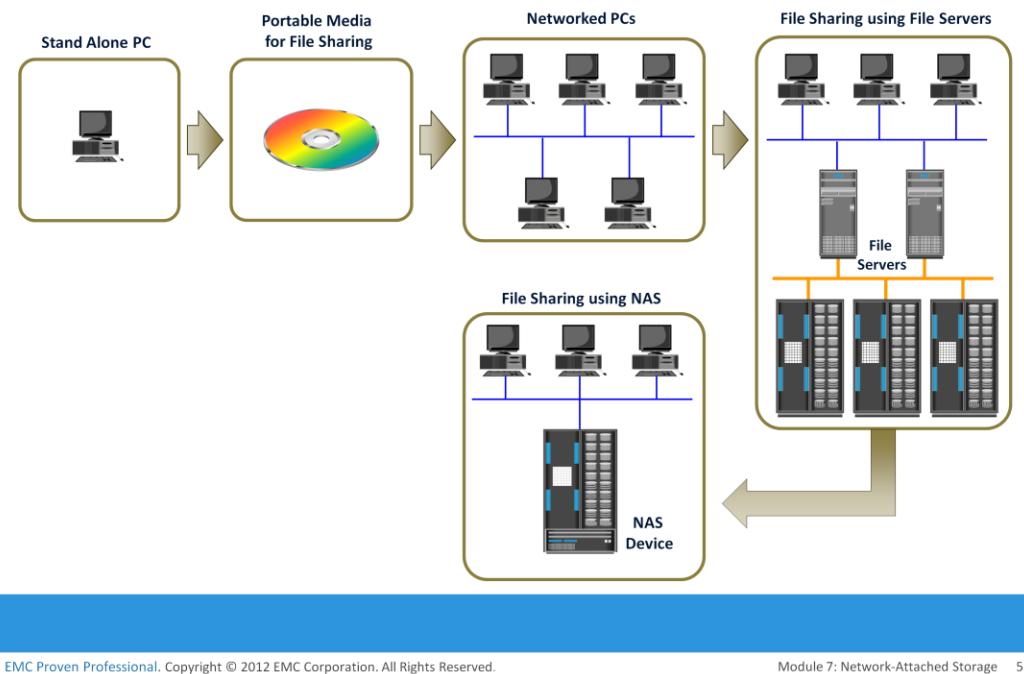
FTP is a client-server protocol that enables data transfer over a network. An FTP server and an FTP client communicate with each other using TCP as the transport protocol.

A distributed file system (DFS) is a file system that is distributed across several hosts. A DFS can provide hosts with direct access to the entire file system, while ensuring efficient management and data security.

The standard client-server file-sharing protocols, such as NFS and CIFS enable the owner of a file to set the required type of access, such as read-only or read-write, for a particular user or group of users. Using this protocol, the clients mount remote file systems that are available on dedicated file servers.

A peer-to-peer (P2P) file sharing model uses peer-to-peer network. P2P enables client machines to directly share files with each other over a network. Clients use a file sharing software that searches for other peer clients. This differs from client-server model that uses file servers to store files for sharing.

## File Sharing Technology Evolution



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 5

Traditional methods of file sharing involves copying of files to a portable media, such as floppy diskette, CD, DVD, or USB drives and delivering them to other users with whom it is being shared. However, this approach is not suitable in an enterprise environment in which a large number of users at different locations need access to common files.

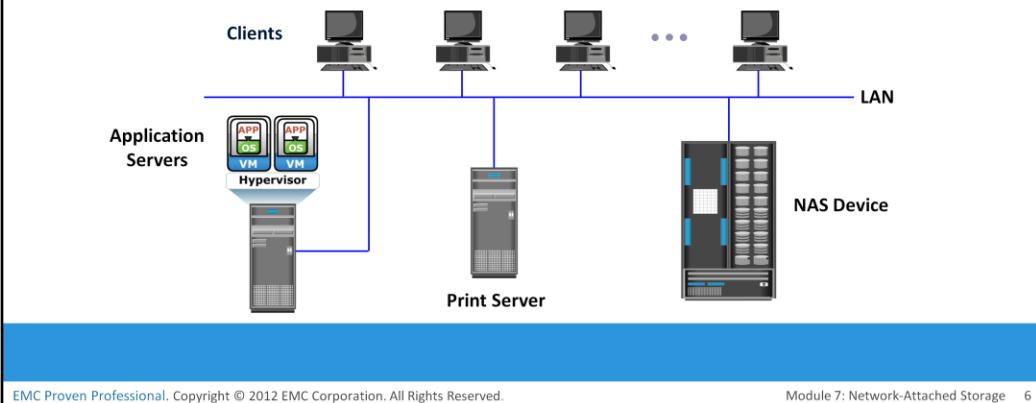
Network-based file sharing provides the flexibility to share files over long distances among a large number of users. File servers use client-server technology to enable file sharing over a network. To address the tremendous growth of file data in enterprise environments, organizations have been deploying large numbers of file servers. These servers are either connected to direct-attached storage (DAS) or storage area network (SAN)-attached storage. This has resulted in the proliferation of islands of over-utilized and under-utilized file servers and storage. In addition, such environments have poor scalability, higher management cost, and greater complexity. Network-attached storage (NAS) emerged as a solution to these challenges.

## What is NAS?

### NAS

It is an IP-based, dedicated, high-performance file sharing and storage device.

- Enables NAS clients to share files over IP network
- Uses specialized operating system that is optimized for file I/O
- Enables both UNIX and Windows users to share data



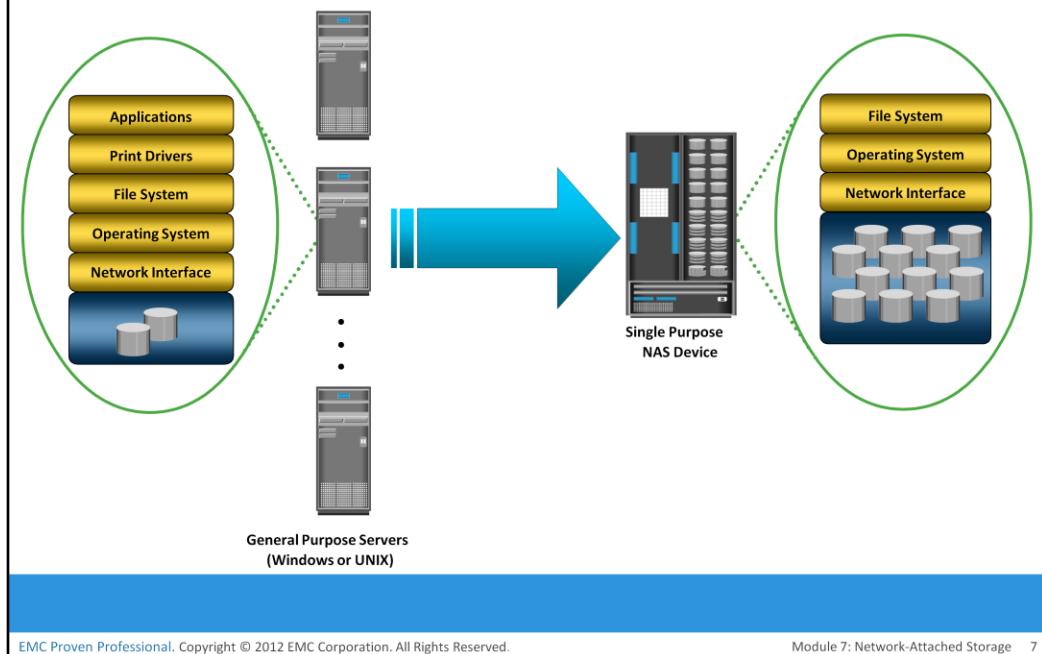
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 6

NAS is a dedicated, high-performance file sharing and storage device. NAS enables its clients to share files over an IP network. NAS provides the advantages of server consolidation by eliminating the need for multiple file servers. It also consolidates the storage used by the clients onto a single system, making it easier to manage the storage. NAS uses network and file-sharing protocols to provide access to the file data. These protocols include TCP/IP for data transfer, and Common Internet File System (CIFS) and Network File System (NFS) for network file service. NAS enables both UNIX and Microsoft Windows users to share the same data seamlessly.

A NAS device uses its own operating system and integrated hardware and software components to meet specific file-service needs. Its operating system is optimized for file I/O and, therefore, performs file I/O better than a general-purpose server. As a result, a NAS device can serve more clients than general-purpose servers and provide the benefit of server consolidation.

## General Purpose Servers Vs. NAS Devices



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 7

A NAS device is optimized for file-serving functions such as storing, retrieving, and accessing files for applications and clients. As shown in the slide, a general-purpose server can be used to host any application because it runs a general-purpose operating system. Unlike a general-purpose server, a NAS device is dedicated to file-serving. It has a specialized operating system dedicated to file serving by using industry standard protocols. Some NAS vendors support features, such as native clustering for high availability.

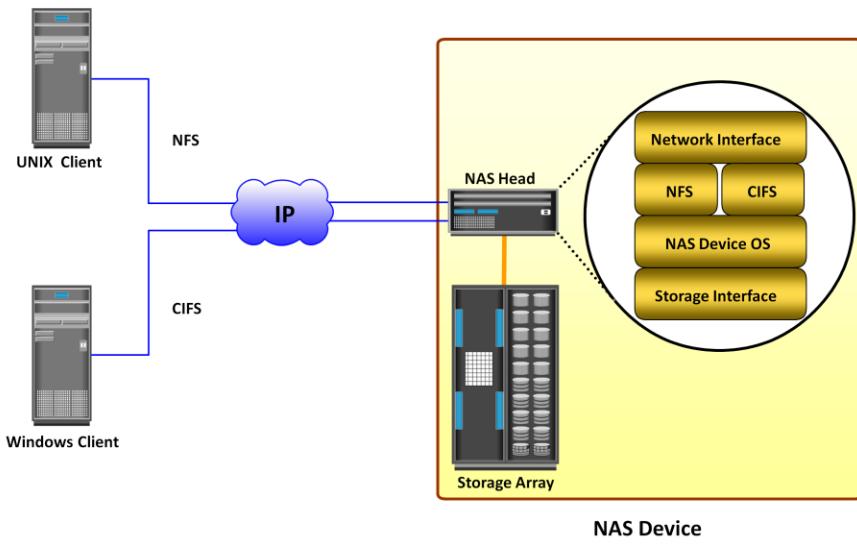
## Benefits of NAS

- Improved efficiency
- Improved flexibility
- Centralized storage
- Simplified management
- Scalability
- High availability – through native clustering and replication
- Security – authentication, authorization, and file locking in conjunction with industry-standard security
- Low cost
- Ease of deployment

NAS offers the following benefits:

- **Improved efficiency:** NAS delivers better performance compared to a general-purpose file server because NAS uses an operating system specialized for file serving.
- **Improved flexibility:** Compatible with clients on both UNIX and Windows platforms using industry-standard protocols. NAS is flexible and can serve requests from different types of clients from the same source.
- **Centralized storage:** Centralizes data storage to minimize data duplication on client workstations, and ensure greater data protection.
- **Simplified management:** Provides a centralized console that makes it possible to manage file systems efficiently.
- **Scalability:** Scales well with different utilization profiles and types of business applications because of the high-performance and low-latency design.
- **High availability:** Offers efficient replication and recovery options, enabling high data availability. NAS uses redundant components that provide maximum connectivity options. A NAS device supports clustering technology for failover.
- **Security:** Ensures security, user authentication, and file locking with industry-standard security schemas.
- **Low cost:** NAS uses commonly available, and inexpensive Ethernet components
- **Ease of deployment:** Configuration at the client is minimal, because the clients have required NAS connection software built in.

## Components of NAS



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 9

A NAS device has two key components: NAS head and storage. In some NAS implementations, the storage could be external to the NAS device and shared with other hosts. The NAS head includes the following components:

- CPU and memory
- One or more network interface cards (NICs), which provide connectivity to the client network. Examples of network protocols supported by NIC include Gigabit Ethernet, Fast Ethernet, ATM, and Fiber Distributed Data Interface (FDDI)
- An optimized operating system for managing the NAS functionality. It translates file-level requests into block-storage requests and further converts the data supplied at the block level to file data
- NFS, CIFS, and other protocols for file sharing
- Industry-standard storage protocols and ports to connect and manage physical disk resources

The NAS environment includes clients accessing a NAS device over an IP network using file-sharing protocols.

## NAS File Sharing Protocols

- Two common NAS file sharing protocols are:
  - ▶ Common Internet File System (CIFS)
  - ▶ Network File System (NFS)

Most NAS devices support multiple file-service protocols to handle file I/O requests to a remote file system. As discussed earlier, NFS and CIFS are the common protocols for file sharing. NAS devices enable users to share file data across different operating environments and provide a means for users to migrate transparently from one operating system to another.

## Common Internet File System

- Client-server application protocol
  - ▶ An open variation of the Server Message Block (SMB) protocol
- Enables clients to access files that are on a server over TCP/IP
- Stateful Protocol
  - ▶ Maintains connection information regarding every connected client
  - ▶ Can automatically restore connections and reopen files that were open prior to interruption

Common Internet File System (CIFS) is a client-server application protocol that enables client programs to make requests for files and services on remote computers over TCP/IP. It is a public, or open, variation of Server Message Block (SMB) protocol.

The CIFS protocol enables remote clients to gain access to files on a server. CIFS enables file sharing with other clients by using special locks. Filenames in CIFS are encoded using unicode characters. CIFS provides the following features to ensure data integrity:

- It uses file and record locking to prevent users from overwriting the work of another user on a file or a record.
- It supports fault tolerance and can automatically restore connections and reopen files that were open prior to an interruption. The fault tolerance features of CIFS depend on whether an application is written to take advantage of these features. Moreover, CIFS is a stateful protocol because the CIFS server maintains connection information regarding every connected client. If a network failure or CIFS server failure occurs, the client receives a disconnection notification. User disruption is minimized if the application has the embedded intelligence to restore the connection. However, if the embedded intelligence is missing, the user must take steps to reestablish the CIFS connection.

Users refer to remote file systems with an easy-to-use file-naming scheme:

\server\share or \\servername.domain.suffix\share.

## Network File System

- Client-server application protocol
- Enables clients to access files that are on a server
- Uses Remote Procedure Call (RPC) mechanism to provide access to remote file system
- Currently, three versions of NFS are in use:
  - ▶ NFS v2 is stateless and uses UDP as transport layer protocol
  - ▶ NFS v3 is stateless and uses UDP or optionally TCP as transport layer protocol
  - ▶ NFS v4 is stateful and uses TCP as transport layer protocol

Network File System (NFS) is a client-server protocol for file sharing that is commonly used on UNIX systems. NFS was originally based on the connectionless User Datagram Protocol (UDP). It uses a machine-independent model to represent user data. It also uses Remote Procedure Call (RPC) as a method of inter-process communication between two computers. The NFS protocol provides a set of RPCs to access a remote file system for the following operations:

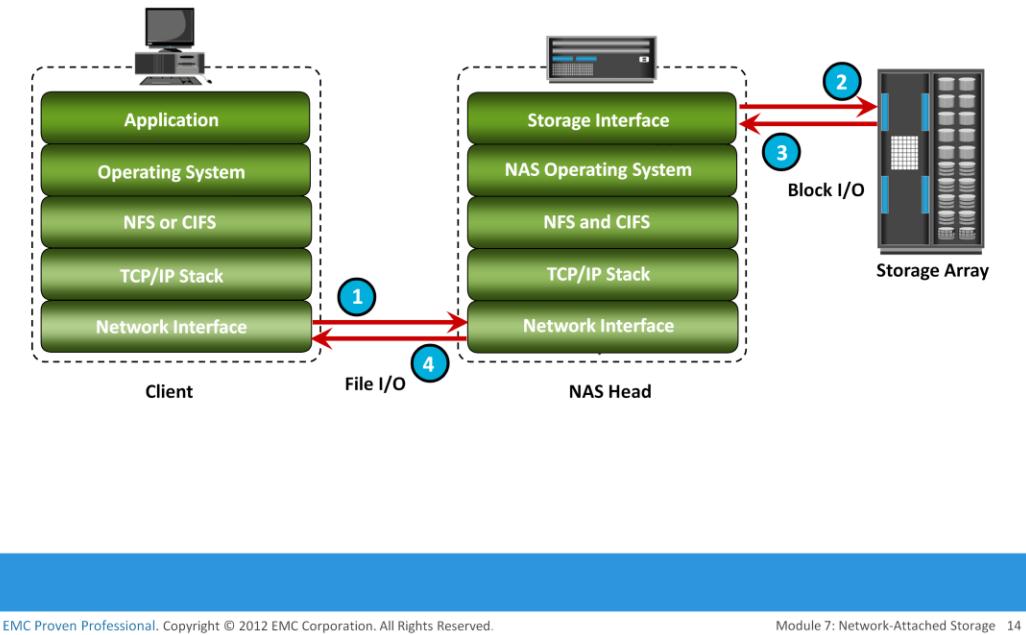
- Searching files and directories
- Opening, reading, writing to, and closing a file
- Changing file attributes
- Modifying file links and directories

NFS creates a connection between the client and the remote system to transfer data. NFS (NFSv3 and earlier) is a stateless protocol, which means that it does not maintain any kind of table to store information about open files and associated pointers. Therefore, each call provides a full set of arguments to access files on the server. These arguments include a file handle reference to the file, a particular position to read or write, and the versions of NFS.

Currently, three versions of NFS are in use:

- NFS version 2 (NFSv2): Uses UDP to provide a stateless network connection between a client and a server. Features, such as locking, are handled outside the protocol.
- NFS version 3 (NFSv3): The most commonly used version, which uses UDP or TCP, and is based on the stateless protocol design. It includes some new features, such as a 64-bit file size, asynchronous writes, and additional file attributes to reduce refetching.
- NFS version 4 (NFSv4): Uses TCP and is based on a stateful protocol design. It offers enhanced security. The latest NFS version 4.1 is the enhancement of NFSv4 and includes some new features, such as session model, parallel NFS (pNFS), and data retention.

## NAS I/O Operation



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 14

NAS provides file-level data access to its clients. File I/O is a high-level request that specifies the file to be accessed. For example, a client may request a file by specifying its name, location, or other attributes. The NAS operating system keeps track of the location of files on the disk volume and converts client file I/O into block-level I/O to retrieve data. The process of handling I/Os in a NAS environment is as follows:

1. The requestor (client) packages an I/O request into TCP/IP and forwards it through the network stack. The NAS device receives this request from the network.
2. The NAS device converts the I/O request into an appropriate physical storage request, which is a block-level I/O, and then performs the operation on the physical storage.
3. When the NAS device receives data from the storage, it processes and repackages the data into an appropriate file protocol response.
4. The NAS device packages this response into TCP/IP again and forwards it to the client through the network.

## Module 7: Network-Attached Storage (NAS)

### Lesson 2: NAS Implementation and File-level Virtualization

During this lesson the following topics are covered:

- NAS implementations
- NAS use cases
- File-level virtualization

This lesson describes three common NAS implementations: unified, gateway, and scale-out. It also covers server and storage consolidation use cases of NAS. Further it covers file-level virtualization and its benefits.

## NAS Implementation – Unified NAS

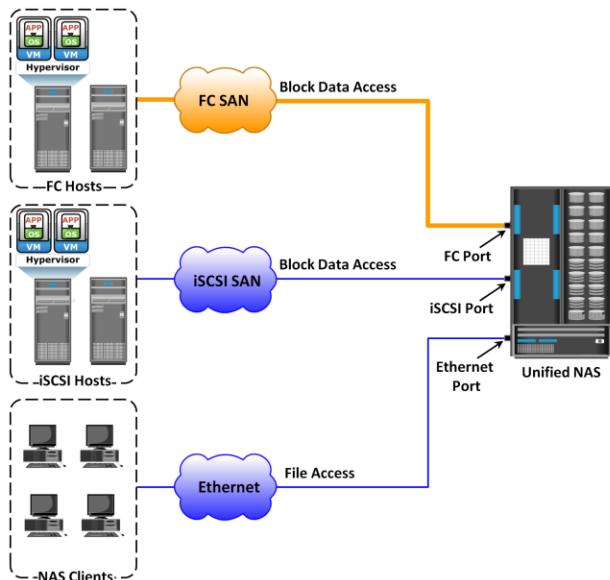
- Consolidates NAS-based (file-level) and SAN-based (block-level) access on a single storage platform
- Supports both CIFS and NFS protocols for file access and iSCSI and FC protocols for block level access
- Provides unified management for both NAS head and storage

The unified NAS consolidates NAS-based and SAN-based data access within a unified storage platform and provides a unified management interface for managing both the environments.

Unified NAS performs file serving and storing of file data, along with providing access to block-level data. It supports both CIFS and NFS protocols for file access and iSCSI and FC protocols for block level access. Due to consolidation of NAS-based and SAN-based access on a single storage platform, unified NAS reduces an organization's infrastructure and management costs.

A unified NAS contains one or more NAS heads and storage in a single system. NAS heads are connected to the storage controllers (SCs), which provide access to the storage. These storage controllers also provide connectivity to iSCSI and FC hosts. The storage may consist of different drive types, such as SAS, ATA, FC, and flash drives, to meet different workload requirements.

## Unified NAS Connectivity



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 17

Each NAS head in a unified NAS has front-end Ethernet ports, which connect to the IP network. The front-end ports provide connectivity to the clients and service the file I/O requests. Each NAS head has back-end ports, to provide connectivity to the storage controllers.

iSCSI and FC ports on a storage controller enable hosts to access the storage directly or through a storage network at the block level.

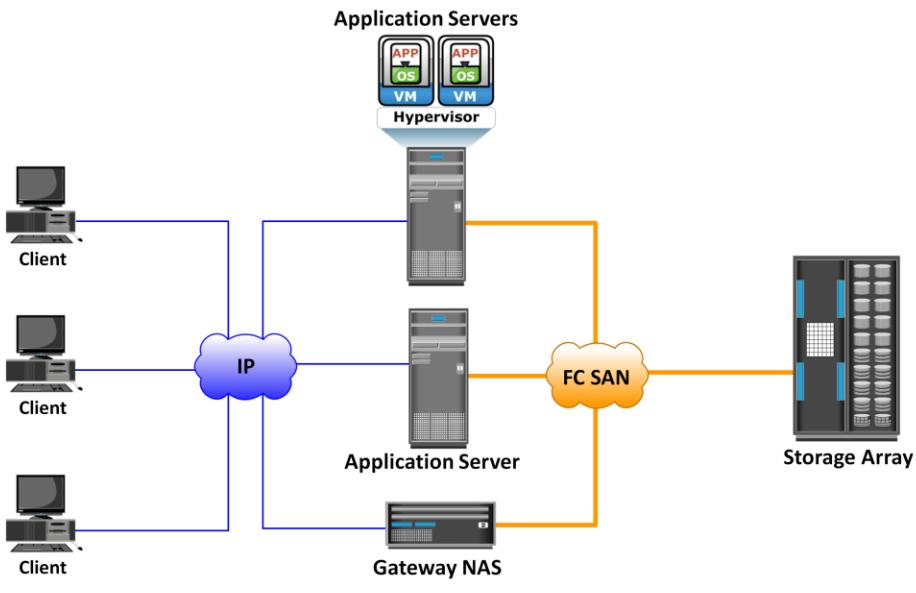
## NAS Implementation – Gateway NAS

- Uses external and independently-managed storage
  - ▶ NAS heads access SAN-attached or direct-attached storage arrays
- NAS heads share storage with other application servers that perform block I/O
- Requires separate management of NAS head and storage

A gateway NAS device consists of one or more NAS heads and uses external and independently managed storage. Similar to unified NAS, the storage is shared with other applications that use block-level I/O. Management functions in this type of solution are more complex than those in a unified NAS environment because there are separate administrative tasks for the NAS head and the storage. A gateway solution can use the FC infrastructure, such as switches and directors for accessing SAN-attached storage arrays or direct-attached storage arrays.

The gateway NAS is more scalable compared to unified NAS because NAS heads and storage arrays can be independently scaled up when required. For example, NAS heads can be added to scale up the NAS device performance. When the storage limit is reached, it can scale up, adding capacity on the SAN, independent of NAS heads. Similar to a unified NAS, a gateway NAS also enables high utilization of storage capacity by sharing it with the SAN environment.

## Gateway NAS Connectivity



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 19

In a gateway solution, the front-end connectivity is similar to that in a unified storage solution. Communication between the NAS gateway and the storage system in a gateway solution is achieved through a traditional FC SAN. To deploy a gateway NAS solution, factors, such as multiple paths for data, redundant fabrics, and load distribution, must be considered.

## NAS Implementation – Scale-out NAS

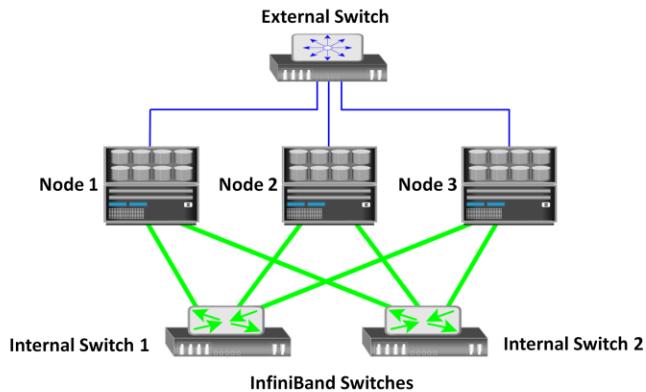
- Pools multiple nodes together in a cluster that works as a single NAS device
  - ▶ Pool is managed centrally
- Scales performance and/or capacity with addition of nodes to the pool non-disruptively
- Creates a single file system that runs on all nodes in the cluster
  - ▶ Clients, connected to any node, can access entire file system
  - ▶ File system grows dynamically as nodes are added
- Stripes data across all nodes in a pool along with mirror or parity protection

The scale-out NAS implementation pools multiple nodes together in a cluster. A node may consist of either the NAS head or storage or both. The cluster performs the NAS operation as a single entity.

A scale-out NAS provides the capability to scale its resources by simply adding nodes to a clustered NAS architecture. The cluster works as a single NAS device and is managed centrally. Nodes can be added to the cluster, when more performance or more capacity is needed, without causing any downtime. Scale-out NAS provides the flexibility to use many nodes of moderate performance and availability characteristics to produce a total system that has better aggregate performance and availability. It also provides ease of use, low cost, and theoretically unlimited scalability.

Scale-out NAS creates a single file system that runs on all nodes in the cluster. All information is shared among nodes, so the entire file system is accessible by clients connecting to any node in the cluster. Scale-out NAS stripes data across all nodes in a cluster along with mirror or parity protection. As data is sent from clients to the cluster, the data is divided and allocated to different nodes in parallel. When a client sends a request to read a file, the scale-out NAS retrieves the appropriate blocks from multiple nodes, recombines the blocks into a file, and presents the file to the client. As nodes are added, the file system grows dynamically and data is evenly distributed to every node. Each node added to the cluster increases the aggregate storage, memory, CPU, and network capacity. Hence, cluster performance also increases.

## Scale-out NAS Connectivity



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

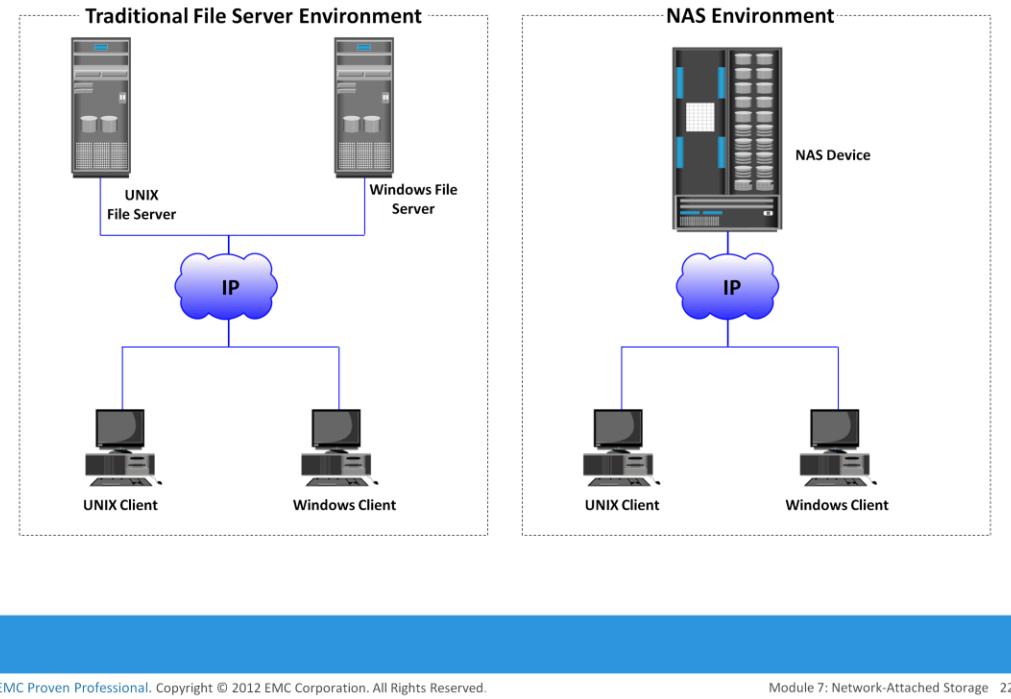
Module 7: Network-Attached Storage 21

Scale-out NAS clusters use separate internal and external networks for back-end and front-end connectivity, respectively. An internal network provides connections for intracluster communication, and an external network connection enables clients to access and share file data. Each node in the cluster connects to the internal network. The internal network offers high throughput and low latency and uses high-speed networking technology, such as InfiniBand or Gigabit Ethernet. To enable clients to access a node, the node must be connected to the external Ethernet network. Redundant internal or external networks may be used for high availability. This slide provides an example of scale-out NAS connectivity.

### Note:

InfiniBand is a networking technology that provides a low-latency, high-bandwidth communication link between hosts and peripherals. It provides serial connection and is often used for inter-server communications in high-performance computing environments. InfiniBand enables remote direct memory access (RDMA) that enables a device (host or peripheral) to access data directly from the memory of a remote device. InfiniBand also enables a single physical link to carry multiple channels of data simultaneously using a multiplexing technique.

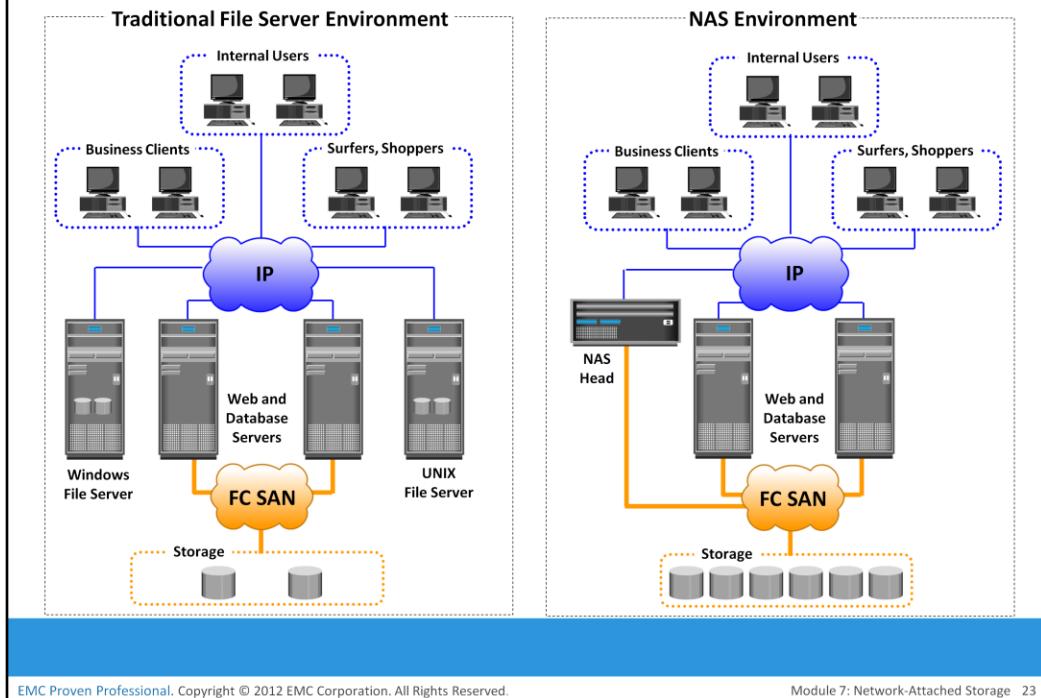
## NAS Use Case 1 – Server Consolidation with NAS



The slide provides a use case that illustrates how a NAS enables consolidation of file servers. Traditionally, network file system for UNIX and Microsoft Windows are housed on separate servers. This requires maintenance of both the environments.

By implementation of NAS, both Windows and UNIX file structures can be housed together in a single system, while still maintaining their integrity. Using NAS, the same file system can be accessed via different protocols, either NFS or CIFS, and still maintain the integrity of the data and security structures, as long as the applications used for both methodologies understand the data structures presented.

## NAS Use Case 2 – Storage Consolidation with NAS



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 23

The slide provides another use case that shows how storage resources in a traditional file server environment can be consolidated using NAS.

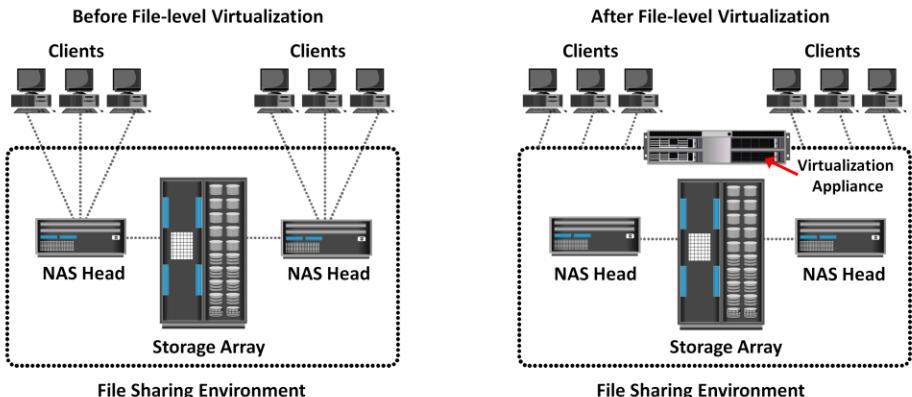
## File-level Virtualization

- Eliminates dependency between data accessed at the file-level and the location where the files are physically stored
- Enables users to use a logical path, rather than a physical path, to access files
- Uses global namespace that maps logical path of file resources to their physical path
- Provides non-disruptive file mobility across file servers or NAS devices

A network-based file sharing environment is composed of multiple file servers or NAS devices. It might be required to move the files from one device to another due to reasons such as cost or performance. File-level virtualization, implemented in NAS or the file server environment, provides a simple, nondisruptive file-mobility solution.

File-level virtualization eliminates the dependencies between the data accessed at the file level and the location where the files are physically stored. It creates a logical pool of storage, enabling users to use a logical path, rather than a physical path, to access files. A global namespace is used to map the logical path of a file to the physical path names. File-level virtualization enables the movement of files across NAS devices, even if the files are being accessed.

## Comparison: Before and After File-level Virtualization



- Dependency between client access and file location
- Underutilized storage resources
- Downtime is caused by data migrations

- Break dependencies between client access and file location
- Storage utilization is optimized
- Non-disruptive migrations

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 7: Network-Attached Storage 25

Before virtualization, each host knows exactly where its file resources are located. This environment leads to underutilized storage resources and capacity problems because files are bound to a specific NAS device or file server. It may be required to move the files from one server to another because of performance reasons or when the file server fills up. Moving files across the environment is not easy and may make files inaccessible during file movement. Moreover, hosts and applications need to be reconfigured to access the file at the new location. This makes it difficult for storage administrators to improve storage efficiency while maintaining the required service level.

File-level virtualization simplifies file mobility. It provides user or application independence from the location where the files are stored. File-level virtualization facilitates the movement of files across the online file servers or NAS devices. This means that while the files are being moved, clients can access their files nondisruptively. Clients can also read their files from the old location and write them back to the new location without realizing that the physical location has changed.

## Module 7: Network-Attached Storage (NAS)

### Concept in Practice:

- EMC Isilon
- EMC VNX Gateway

The Concept in Practice section covers EMC Isilon and VNX Gateway.

## EMC Isilon

- Scale-out NAS solution
- Includes ‘OneFS’ operating system that creates a single file system across Isilon cluster
- Provides ability to nondisruptively add nodes to Isilon cluster
- Includes ‘SmartPools’ that enables different node types to be mixed in a single cluster
- Monitors component health and transparently reallocates files
- Uses ‘Autobalance’ that rebalances data automatically, when a new node is added to the cluster
- Uses ‘FlexProtect’ that protects from up to four simultaneous failures of either nodes or individual drives

EMC Isilon is the scale-out NAS solution. Isilon has a specialized operating system called OneFS that enables the scale-out NAS architecture. OneFS combines the three layers of traditional storage architectures—file system, volume manager, and RAID—into one unified software layer, creating a single file system that spans across all nodes in an Isilon cluster. It also provides the ability to seamlessly add storage and other resources without system downtime.

OneFS enables different node types to be mixed in a single cluster through the addition of the SmartPools application software. SmartPools enables deploying a single file system to span multiple nodes that have different performance characteristics and capacities. Isilon offers different types of nodes, such as the X-Series, S-Series, NL-Series, and Accelerator.

OneFS constantly monitors the health of all files and disks within a cluster, and if components are at risk, the file system automatically flags the problem components for replacement and transparently reallocates those files to healthy components.

When a new storage node is added, the Autobalance feature of OneFS automatically moves data onto this new node via the Infiniband based internal network. This automatic rebalancing ensures that the new node does not become a hot spot for new data.

OneFS includes a core technology, called FlexProtect, to provide data protection. FlexProtect provides protection for up to four simultaneous failures of either nodes or individual drives per stripe. FlexProtect provides file-specific protection capabilities. Different protection levels can be assigned to individual files, directories, or to portions of a file system.

## EMC VNX Gateway

- Gateway NAS solution
- Provides multi-protocol network file system access, dynamic expansion of file systems, high availability, and high performance
- Comprises one or more NAS heads, called 'X-Blades' that run VNX operating environment
- Includes 'Control Station' that provides a single point for configuring X-Blades

The VNX Series Gateway contains one or more NAS heads, called X-Blades, that access external storage arrays, such as Symmetrix and block-based VNX via SAN. X-Blades run the VNX operating environment that is optimized for high-performance and multiprotocol network file system access. Each X-Blade consists of processors, redundant data paths, power supplies, Gigabit Ethernet, and 10-Gigabit Ethernet optical ports. All the X-Blades in a VNX gateway system are managed by Control Station, which provides a single point for configuring VNX Gateway. The VNX Gateway supports both pNFS and EMC patented Multi-Path File System (MPFS) protocols, which further improves the VNX Gateway performance.

VNX Series Gateway offers two models: VG2 and VG8. VG8 supports up to eight X-Blades, whereas VG2 supports up to two. X-Blades may be configured as either primary or standby. A primary X-Blade is the operating NAS head, whereas a standby X-Blade becomes operational if the primary X-Blade fails. The Control Station handles an X-Blade failover. The Control Station also provides other high-availability features, such as fault monitoring, fault reporting, call home, and remote diagnostics.

## Module 7: Summary

Key points covered in this module:

- NAS benefits
- NAS components
- NAS file sharing protocols
- NAS implementations
- File-level virtualization

This module covered the benefits and components of NAS. The key components of NAS are NAS head and storage. This module covered two common NAS file sharing protocols – CIFS and NFS – that enabled clients to access files located on a server. It also detailed on three common NAS implementations such as Unified, Gateway, and Scale-out that provides file sharing environment. Finally, it covered file-level virtualization that provides simple, nondisruptive file mobility solution.

## Check Your Knowledge – 1

- Which component of a NAS head translates file-level requests into block-storage requests?
  - A. Front-end ports
  - B. Optimized operating system
  - C. CIFS and NFS
  - D. Network Interface Card
- Which is a feature of scale-out NAS?
  - A. Uses general purpose operating system for file serving
  - B. Creates multiple file systems on each node in the cluster
  - C. Uses external and independently-managed nodes
  - D. Enables pooling of nodes that work as a single NAS device

## Check Your Knowledge – 2

- Which is a feature of gateway NAS?
  - A. Uses dedicated storage for each NAS head
  - B. NAS head and storage are managed independently
  - C. Creates a single file system that runs on all NAS heads
  - D. Provides connectivity to iSCSI and FC hosts
- Which NAS implementation consolidates file-based and block-based access on a single storage platform?
  - A. Scale-out
  - B. Gateway
  - C. Unified
  - D. Both gateway and scale-out

## Check Your Knowledge – 3

- Which is a benefit of file-level virtualization?
  - A. Enables users to use physical path, rather than logical path, to access files
  - B. Translates file-level request into block-storage request non disruptively
  - C. Consolidates NAS-based and SAN-based access on a single storage platform
  - D. Eliminates dependency between data accessed at file level and file location

# Module – 8

# Object-Based and Unified Storage



## Module 8: Object-based and Unified Storage

Upon completion of this module, you should be able to:

- Describe the object-based storage model
- List the key components of object-based storage
- Describe the storage and retrieval process in object-based storage
- Describe content-addressed storage
- List the key components of unified storage
- Describe the process of data access from unified storage

This module focuses on object-based storage device and unified storage device. It covers the object-based storage model. The module details on the function of the key components and process of storage and retrieval of object-based storage. This module also covers content-addressed storage, which is a special type of object-based storage and key features it supports.

This module also focus on the function of the key components of unified storage. Finally, this module covers the process of data access from unified storage.

# Module 8: Object-based and Unified Storage

## Lesson 1: Object-based Storage

During this lesson the following topics are covered:

- Comparison of hierarchical file system and flat address space
- Object-based storage model
- Key components of object-based storage
- Storage and retrieval process in object-based storage devices
- Content-addressed storage

This lesson covers a comparison of hierarchical file system with flat address space. The lesson describes the model and key components of object-based storage. This lesson also covers storage and retrieval process in the object-based storage. Finally, the lesson covers content-addressed storage.

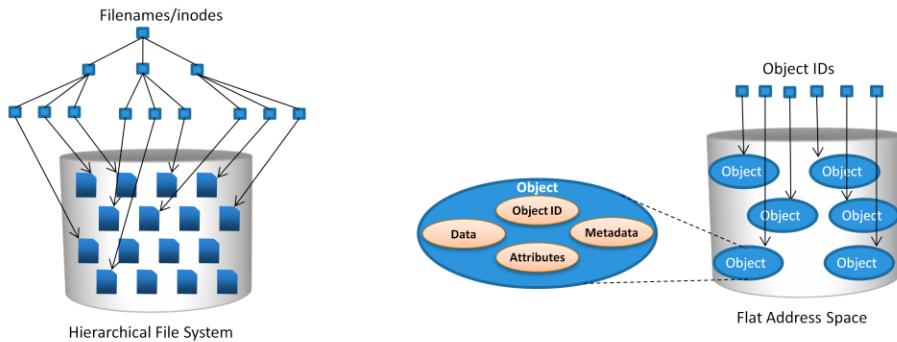
## Drivers for Object-based Storage

- More than 90% of the data being generated is unstructured
- Traditional solutions are inefficient to handle the growth
  - ▶ High overhead on NAS due to managing large number of permissions and nested directories
- These challenges demanded a smarter approach to manage unstructured data based on its content

*Object-based storage is a way to store file data in the form of objects on flat address space based on its content and attributes rather than the name and location*

Recent studies have shown that more than 90 percent of data generated is unstructured. This growth of unstructured data has posed new challenges to IT administrators and storage managers. With this growth, traditional NAS, which is a dominant solution for storing unstructured data, has become inefficient. Data growth adds high overhead to the network-attached storage (NAS) in terms of managing large number of permission and nested directories. In an enterprise environment, NAS also manages large amounts of metadata generated by hosts, storage systems, and individual applications. Typically this metadata is stored as part of the file and distributed throughout the environment. This adds to the complexity and latency in searching and retrieving files. These challenges demand a smarter approach to manage unstructured data based on its content rather than metadata about its name, location, and so on. *Object-based storage* is a way to store file data in the form of objects based on its content and other attributes rather than the name and location.

## Hierarchical File System Vs. Flat Address Space

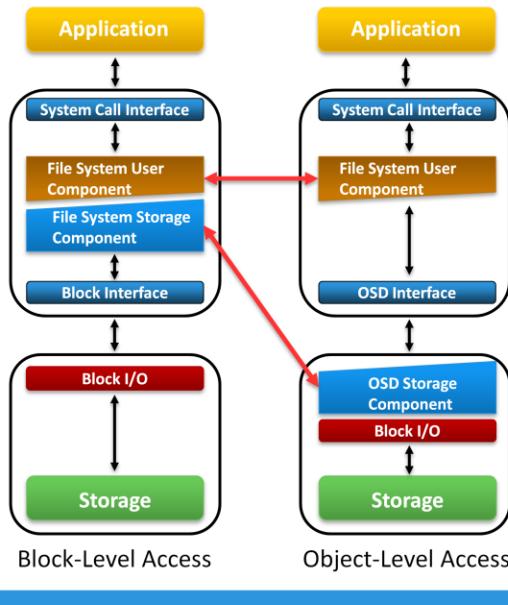


- Hierarchical file system organizes data in the form of files and directories
- Object-based storage devices store the data in the form of objects
  - ▶ It uses flat address space that enables storage of large number of objects
  - ▶ An object contains user data, related metadata, and other attributes
  - ▶ Each object has a unique object ID, generated using specialized algorithm

An OSD is a device that organizes and stores unstructured data, such as movies, office documents, and graphics, as objects. Object-based storage provides a scalable, self-managed, protected, and shared storage option. OSD stores data in the form of *objects*. OSD uses flat address space to store data. Therefore, there is no hierarchy of directories and files; as a result, a large number of objects can be stored in an OSD system .

An object might contain user data, related metadata (size, date, ownership, and so on), and other attributes of data (retention, access pattern, and so on). Each object stored in the system is identified by a unique ID called the *object ID*. The object ID is generated using specialized algorithms such as hash function on the data and guarantees that every object is uniquely identified.

## Traditional Vs. Object-based Storage Model



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

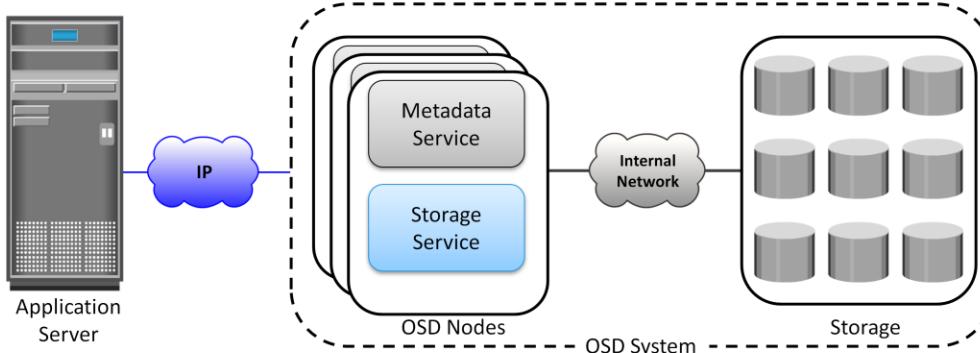
6

An I/O in the traditional block access method passes through various layers in the I/O path. The I/O generated by an application passes through the file system, the channel, or network and reaches the disk drive. When the file system receives the I/O from an application, the file system maps the incoming I/O to the disk blocks. The block interface is used for sending the I/O over the channel or network to the storage device. The I/O is then written to the block allocated on the disk drive.

The file system has two components: user component and storage component. The user component of the file system performs functions such as hierarchy management, naming, and user access control. The storage component maps the files to the physical location on the disk drive.

When an application accesses data stored in OSD, the request is sent to the file system user component. The file system user component communicates to the OSD interface, which in turn sends the request to the storage device. The storage device has the OSD storage component responsible for managing the access to the object on a storage device. After the object is stored, the OSD sends an acknowledgment to the application server. The OSD storage component manages all the required low-level storage and space management functions. It also manages security and access control functions for the objects.

## Key Components of Object-based Storage Device



- OSD system typically comprises three key components:
  - ▶ OSD nodes
  - ▶ Internal network
  - ▶ Storage

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

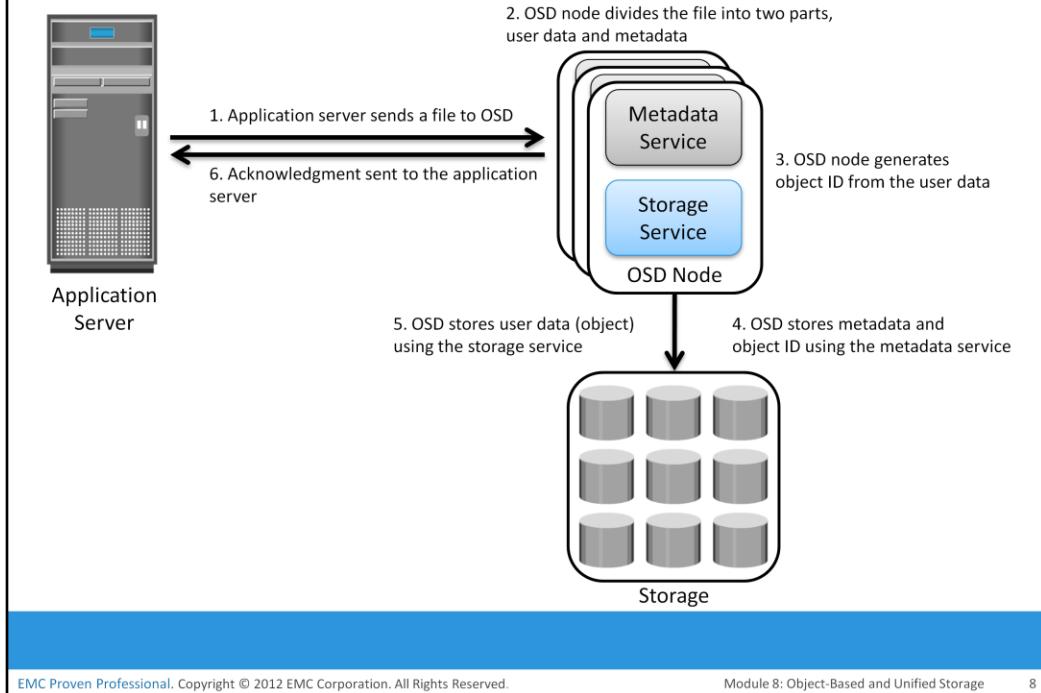
7

The OSD system is typically composed of three key components: nodes, private network, and storage.

The OSD system is composed of one or more *nodes*. A node is a server that runs the OSD operating environment and provides services to store, retrieve, and manage data in the system. The OSD node has two key services: metadata service and storage service. The metadata service is responsible for generating the object ID from the contents (may also include other attributes of data) of a file. It also maintains the mapping of the object IDs and the file system namespace. The storage service manages a set of disks on which the user data is stored. The OSD nodes connect to the storage via an internal network. The internal network provides node-to-node connectivity and node-to-storage connectivity. The application server accesses the node to store and retrieve data over an external network. In some implementations, such as CAS, the metadata service might reside on the application server or on a separate server.

OSD typically uses low-cost and high-density disk drives to store the objects. As more capacity is required, more disk drives can be added to the system.

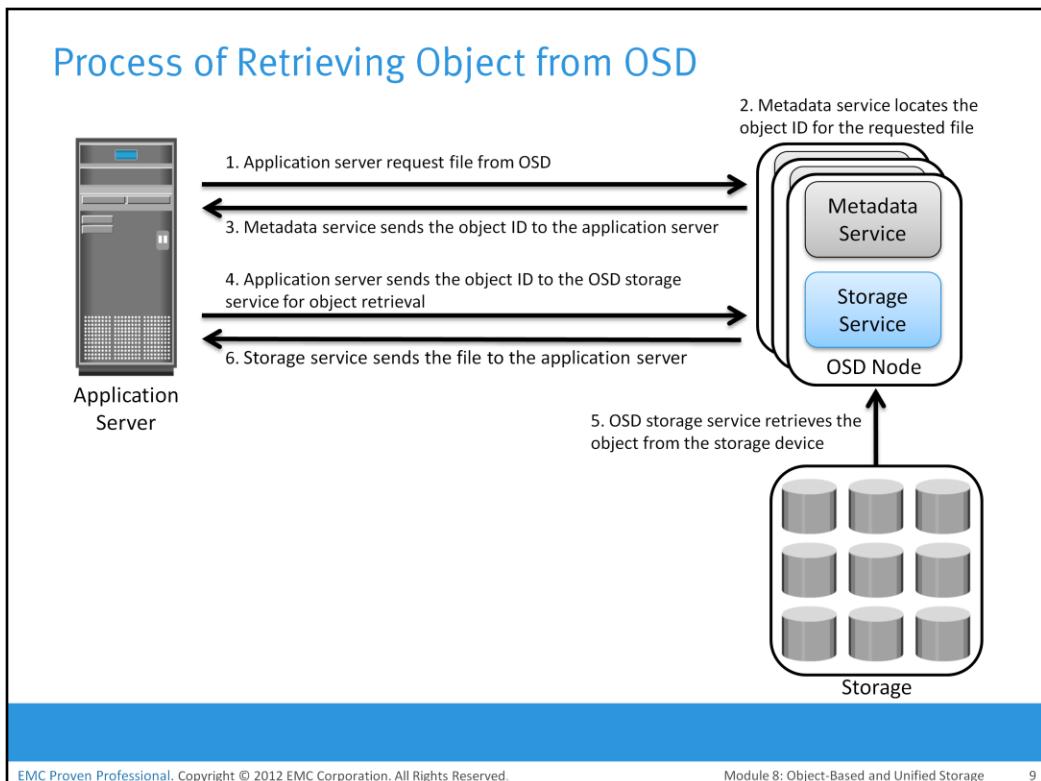
## Process of Storing Object in OSD



The process of storing objects in OSD is illustrated on the slide. The data storage process in an OSD system is as follows:

1. The application server presents the file to be stored to the OSD node.
2. The OSD node divides the file into two parts: user data and metadata.
3. The OSD node generates the object ID using a specialized algorithm. The algorithm is executed against the contents of the user data to derive an ID unique to this data.
4. For future access, the OSD node stores the metadata and object ID using the metadata service.
5. The OSD node stores the user data (objects) in the storage device using the storage service.
6. An acknowledgment is sent to the application server stating that the object is stored.

## Process of Retrieving Object from OSD



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

9

After an object is stored successfully, it is available for retrieval. A user accesses the data stored on OSD by the same filename. The application server retrieves the stored content using the object ID. This process is transparent to the user.

The process of retrieving objects in OSD is illustrated on the slide. The process of data retrieval from OSD is as follows:

1. The application server sends a read request to the OSD system.
2. The metadata service retrieves the object ID for the requested file.
3. The metadata service sends the object ID to the application server.
4. The application server sends the object ID to the OSD storage service for object retrieval.
5. The OSD storage service retrieves the object from the storage device.
6. The OSD storage service sends the file to the application server.

## Key Benefits of Object-based Storage

Benefits	Description
Security and reliability	<ul style="list-style-type: none"><li>Unique object ID generated by specialized algorithms ensures data integrity and content authenticity</li><li>Request authentication is performed at storage device</li></ul>
Platform independence	<ul style="list-style-type: none"><li>Because objects are abstract containers of data, it enables sharing of objects across heterogeneous platforms</li><li>This capability makes object-based storage suitable for cloud computing environment</li></ul>
Scalability	<ul style="list-style-type: none"><li>Both OSD nodes and storage can be independently scaled</li></ul>
Manageability	<ul style="list-style-type: none"><li>Have inherent intelligence to manage objects</li><li>Have self-healing capability</li><li>Policy based management capability enables OSD to handle routine jobs automatically</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

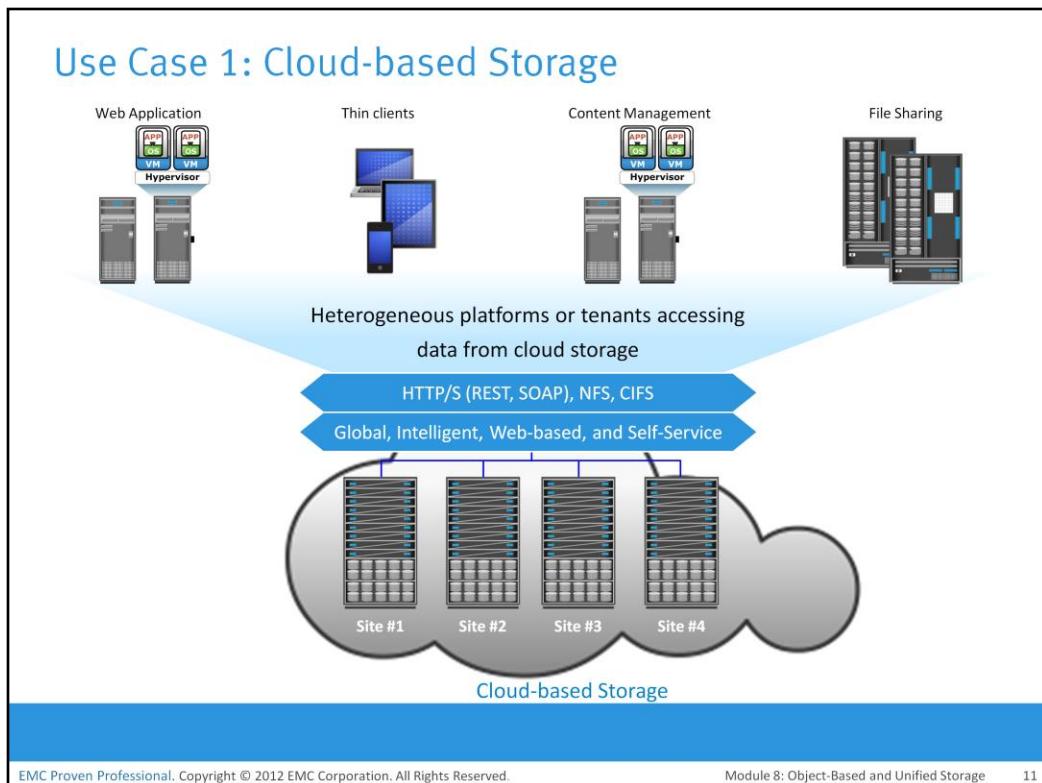
Module 8: Object-Based and Unified Storage

10

For unstructured data, object-based storage devices provide numerous benefits over traditional storage solutions. An ideal storage architecture should provide performance, scalability, security, and data sharing across multiple platforms. Traditional storage solutions, such as SAN, and NAS, do not offer all these benefits as a single solution. Object-based storage combines benefits of both the worlds. It provides platform and location independence, and at the same time, provides scalability, security and data-sharing capabilities. The key benefits of object-based storage are as follows:

- **Security and reliability:** Data integrity and content authenticity are the key features of object-based storage devices. OSD uses specialized algorithms to create objects that provide strong data encryption capability. In OSD, request authentication is performed at the storage device rather than with an external authentication mechanism.
- **Platform independence:** Objects are abstract containers of data, including metadata and attributes. This feature allows objects to be shared across heterogeneous platforms locally or remotely. This platform-independence capability makes object-based storage the best candidate for cloud computing environments.
- **Scalability:** Due to the use of flat address space, object-based storage can handle large amounts of data without impacting performance. Both storage and OSD nodes can be scaled independently in terms of performance and capacity.
- **Manageability:** Object-based storage has an inherent intelligence to manage and protect objects. It uses self-healing capability to protect and replicate objects. Policy-based management capability helps OSD to handle routine jobs automatically.

## Use Case 1: Cloud-based Storage



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

11

Cloud-based storage is a promising use case of OSD. OSD uses a web interface to access storage resources. OSD provides inherent security, scalability, and automated data management. It also enables data sharing across heterogeneous platforms or tenants while ensuring integrity of data. These capabilities make OSD a strong option for cloud-based storage. Cloud service providers can leverage OSD to offer storage-as-a-service. OSD supports web service access via *representational state transfer* (REST) and *simple object access protocol* (SOAP). REST and SOAP APIs can be easily integrated with business applications that access OSD over the web.

*Representational State Transfer or REST* is an architectural style developed for modern web applications. REST provides lightweight web services to access resources (for example, documents, blogs, and so on) on which a few basic operations can be performed, such as retrieving, modifying, creating, and deleting resources. REST-style web services are resource-oriented services. Resources can be uniquely located and identified by a Universal Resource Identifier (URI), and operations can be performed on those resources using an HTTP specification. For example, if a user accesses a blog using REST via a unique identifier, the request returns the representation of the blog in a particular format (XML or HTML). *Simple Object Access Protocol or SOAP* is a XML based protocol that enables communication between the web applications running on different OS and based on different programming languages. SOAP provides process to encode HTTP header and XML file to enable and pass information between different computers.

Cloud based storage is further discussed in module 13 Cloud Computing.

## Use Case 2: Content Address Storage (CAS)

- Storage designed to store fixed content
- Stores data as objects
- Each object is assigned a globally unique identifier, known as content address (CA)
  - ▶ CA is derived from the binary representation of the data
- CAS device can be accessed via the CAS API running on the application server

A data archival solution is a promising use case for OSD. Data integrity and protection is the primary requirement for any data archiving solution. Traditional archival solutions—CD and DVD-ROM—do not provide scalability and performance. OSD stores data in the form of objects, associates them with a unique object ID, and ensures high data integrity. Along with integrity, it provides scalability and data protection. These capabilities make OSD a viable option for long term data archiving for fixed content. Content addressed storage (CAS) is a special type of object-based storage device purposely built for storing fixed content.

CAS is an object-based storage device designed for secure online storage and retrieval of fixed content. CAS stores user data and its attributes as an object. The stored object is assigned a globally unique address, known as a *content address* (CA). This address is derived from the object's binary representation. CAS provides an optimized and centrally managed storage solution. Data access in CAS differs from other OSD devices. In CAS, the application server can access the CAS device only via the CAS API running on the application server. However, the way CAS stores data is similar to the other OSD systems.

## Key Features of CAS

- Content authenticity and integrity
- Location independence
- Single instance storage
- Retention enforcement
- Data protection
- Fast record retrieval
- Load balancing
- Scalability
- Self diagnosis and repair
- Audit trail and event notification

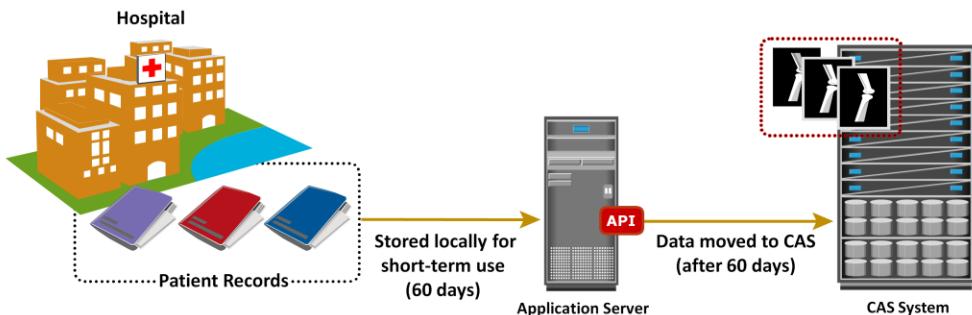
CAS provides all the features required for storing fixed content. The key features of CAS are as follows:

- *Content authenticity*: It assures the genuineness of stored content. This is achieved by generating a unique content address for each object and validating the content address for stored objects at regular intervals. Content authenticity is assured because the address assigned to each object is as unique as a fingerprint. Every time an object is read, CAS uses a hashing algorithm to recalculate the object's content address as a validation step and compares the result to its original content address. If the object fails validation, CAS rebuilds the object using a mirror or parity protection scheme.
- *Content integrity*: It provides assurance that the stored content has not been altered. CAS uses a hashing algorithm for content authenticity and integrity. If the fixed content is altered, CAS generates a new address for the altered content, rather than overwrite the original fixed content.
- *Location independence*: CAS uses a unique content address, rather than directory path names or URLs, to retrieve data. This makes the physical location of the stored data irrelevant to the application that requests the data.

Cont..

- *Single-instance storage (SIS)*: CAS uses a unique content address to guarantee the storage of only a single instance of an object. When a new object is written, the CAS system is polled to see whether an object is already available with the same content address. If the object is available in the system, it is not stored; instead, only a pointer to that object is created.
- *Retention enforcement*: Protecting and retaining objects is a core requirement of an archive storage system. After an object is stored in the CAS system and the retention policy is defined, CAS does not make the object available for deletion until the policy expires.
- *Data protection*: CAS ensures that the content stored on the CAS system is available even if a disk or a node fails. CAS provides both local and remote protection to the data objects stored on it. In the local protection option, data objects are either mirrored or parity protected. In mirror protection, two copies of the data object are stored on two different nodes in the same cluster. This decreases the total available capacity by 50 percent. In parity protection, the data object is split in multiple parts and parity is generated from them. Each part of the data and its parity are stored on a different node. This method consumes less capacity to protect the stored data, but takes slightly longer to regenerate the data if corruption of data occurs. In the remote replication option, data objects are copied to a secondary CAS at the remote location. In this case, the objects remain accessible from the secondary CAS if the primary CAS system fails.
- *Fast record retrieval*: CAS stores all objects on disks, which provides faster access to the objects compared to tapes and optical discs.
- *Load balancing*: CAS distributes objects across multiple nodes to provide maximum throughput and availability.
- *Scalability*: CAS allows the addition of more nodes to the cluster without any interruption to data access and with minimum administrative overhead.
- *Event notification*: CAS continuously monitors the state of the system and raises an alert for any event that requires the administrator's attention. The event notification is communicated to the administrator through SNMP, SMTP, or e-mail.
- *Self diagnosis and repair*: CAS automatically detects and repairs corrupted objects and alerts the administrator about the potential problem. CAS systems can be configured to alert remote support teams who can diagnose and repair the system remotely.
- *Audit trails*: CAS keeps track of management activities and any access or disposition of data. Audit trails are mandated by compliance requirements.

## Use Case 1: Healthcare Solution



- Each X-ray image size range from about 15MB to over 1GB
- Patient records are stored online for a period of 60 days
- Beyond 60 days patient records are archived to CAS

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

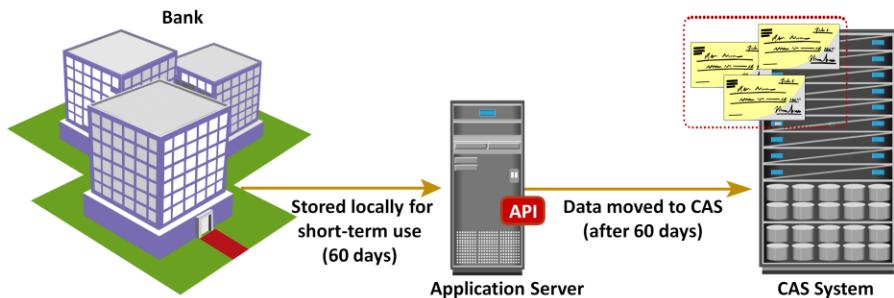
Module 8: Object-Based and Unified Storage

15

Large healthcare centers examine hundreds of patients every day and generate large volumes of medical records. Each record might be composed of one or more images that range in size from approximately 15 MB for a standard digital X-ray to more than 1 GB for oncology studies. The patient records are stored online for a specific period of time for immediate use by the attending physicians. Even if a patient's record is no longer needed, compliance requirements might stipulate that the records be kept in the original format for several years.

Medical image solution providers offer hospitals the capability to view medical records, such as X-ray images, with acceptable response times and resolution to enable rapid assessments of patients. Patients' records are retained on the primary storage for 60 days after which they are moved to the CAS system. CAS facilitates long-term storage and at the same time, provides immediate access to data, when needed.

## Use Case 2: Financial Solution



- Each check image size is about 25KB
- Check imaging service provider might process around 90 million check images per month
- Checks are stored online for a period of 60 days
- Beyond 60 days data is archived to CAS

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

16

In a typical banking scenario, images of checks, each approximately 25 KB in size, are created and sent to archive services over an IP network. A check imaging service provider might process approximately 90 million check images per month. Typically, check images are actively processed in transactional systems for about 5 days.

For the next 60 days, check images may be requested by banks or individual consumers for verification purposes and beyond 60 days, access requirements drop drastically. The check images are moved from the primary storage to the CAS system after 60 days, and can be held there for long term based on retention policy. Check imaging is one example of a financial service application that is best serviced with CAS. Customer transactions initiated by e-mail, contracts, and security transaction records might need to be kept online for 30 years; CAS is the preferred storage solution in such cases.

## Module 8: Object-based and Unified Storage

### Lesson 2: Unified Storage

During this lesson the following topics are covered:

- Key components of unified storage
- Data access from unified storage

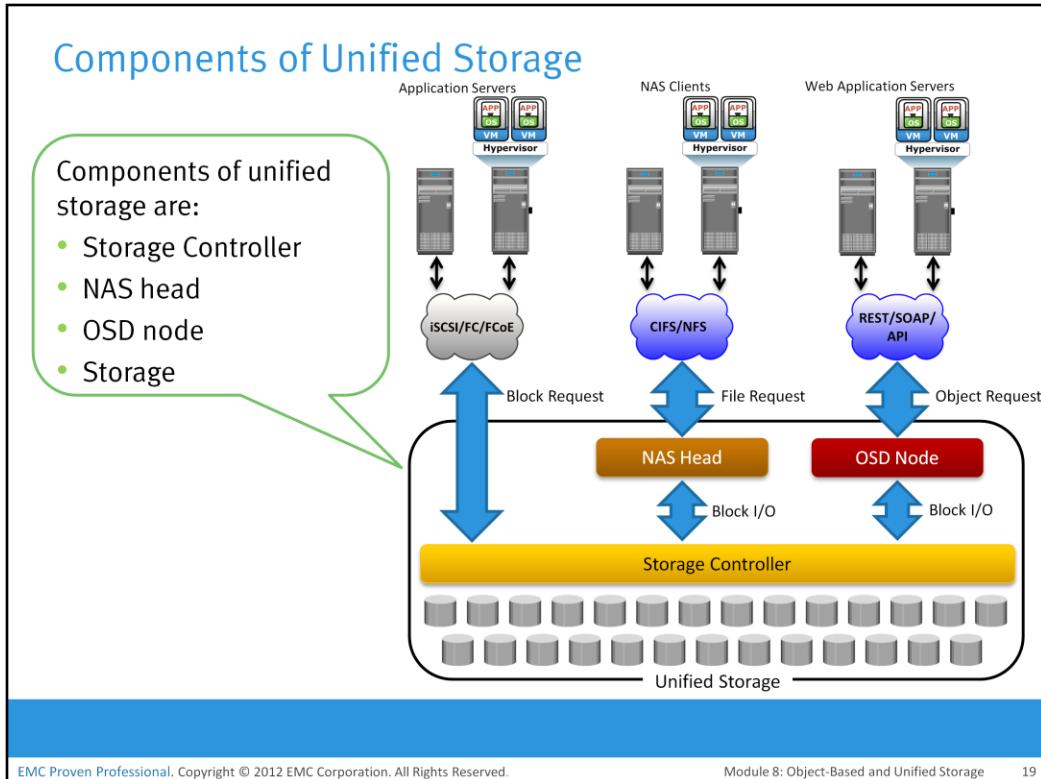
This lesson describes the function of key components of unified storage. This lesson also describes the process of data access from unified storage.

## Drivers for Unified Storage

- Deploying disparate storage solutions (SAN, NAS, and OSD) adds management cost, complexity, and environmental overhead
- Unified storage consolidates block, file, and object-based access within one unified platform
  - ▶ Supports multiple protocols for data access
  - ▶ Can be managed through single management interface

Due to varied application requirements, organizations have been deploying storage area networks (SANs), NAS, and object-based storage devices (OSDs) in their data centers. Deploying these disparate storage solutions adds management complexity, cost and environmental overheads. An ideal solution would be to have an integrated storage solution that supports block, file, and object access. Unified storage has emerged as a solution that consolidates block, file, and object-based access within one unified platform. It supports multiple protocols for data access and can be managed using a single management interface.

## Components of Unified Storage



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

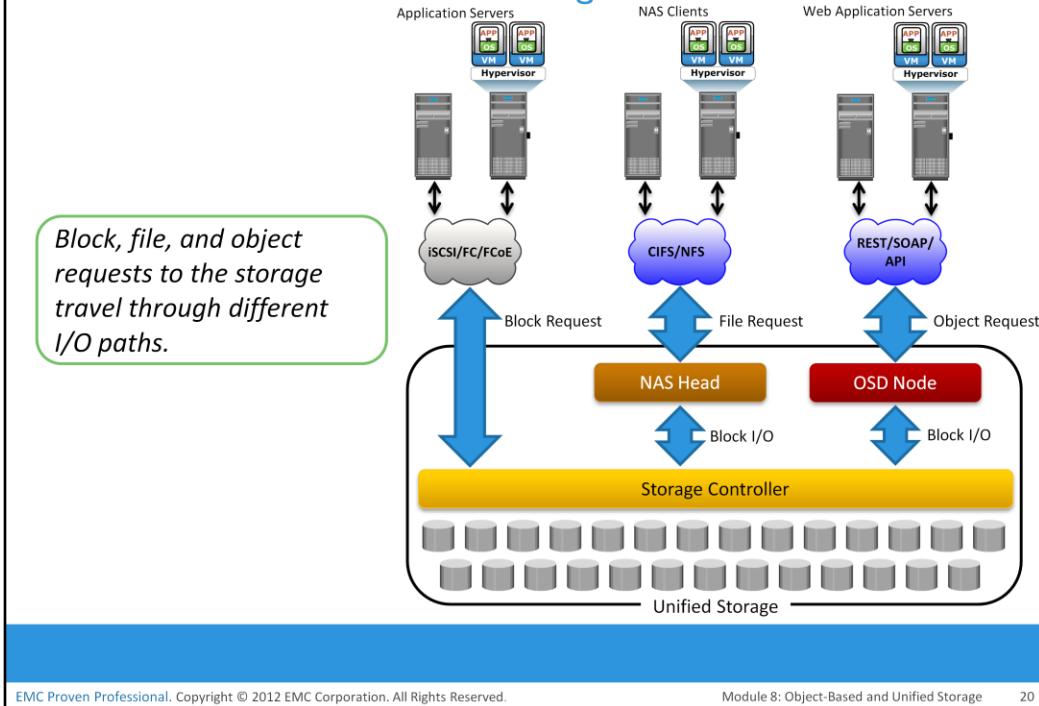
Module 8: Object-Based and Unified Storage

19

A unified storage system consists of following key components: storage controller, NAS head, OSD node, and storage.

- The *storage controller* provides block-level access to application servers through iSCSI, FC, or FCoE protocols. It contains iSCSI, FC, and FCoE front-end ports for direct block access. The storage controller is also responsible for managing the back-end storage pool in the storage system. The controller configures LUNs and presents them to application servers, NAS heads, and OSD nodes. The LUNs presented to the application server appear as local physical disks. A file system is configured on these LUNs and is made available to applications for storing data.
- A *NAS head* is a dedicated file server that provides file access to NAS clients. The NAS head is connected to the storage via the storage controller typically using a FC or FCoE connection. The system typically has two or more NAS heads for redundancy. The LUNs presented to the NAS head appear as physical disks. The NAS head configures the file systems on these disks, creates a NFS, CIFS, or mixed share, and exports the share to the NAS clients.
- The *OSD node* accesses the storage through the storage controller using a FC or FCoE connection. The LUNs assigned to the OSD node appear as physical disks. These disks are configured by the OSD nodes, enabling them to store the data from the web application servers.

## Data Access from Unified Storage



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

20

In a unified storage system, block, file, and object requests to the storage travel through different I/O paths.

- **Block I/O request:** The application servers are connected to an FC, iSCSI, or FCoE port on the storage controller. The server sends a block request over an FC, iSCSI, or FCoE connection. The storage controller processes the I/O and responds to the application server.
- **File I/O request:** The NAS clients (where the NAS share is mounted or mapped) send a file request to the NAS head using the NFS or CIFS protocol. The NAS head receives the request, converts it into a block request, and forwards it to the storage controller. Upon receiving the block data from the storage controller, the NAS head again converts the block request back to the file request and sends it to the clients.
- **Object I/O request:** The web application servers send an object request, typically using REST or SOAP protocols, to the OSD node. The OSD node receives the request, converts it into a block request, and sends it to the disk through the storage controller. The controller in turn processes the block request and responds back to the OSD node, which in turn provides the requested object to the web application server.

## Module 8: Object-based and Unified Storage

### Concept in Practice

- EMC Atmos
- EMC VNX
- EMC Centera

The Concept in Practice covers the product example of object-based and unified storage. It covers three products: EMC Atmos, EMC VNX, and EMC Centera.

## EMC Atmos

- Massively scalable objects-based storage
- Can be deployed in two ways: purpose-built hardware appliance or virtual machine (VM)
- Key features
  - ▶ Enable policy-based management
  - ▶ Provide protection with replication and parity
  - ▶ Provide services such as compression and deduplication
  - ▶ Support web service and legacy protocols
    - ▶ REST, SOAP, CIFS, NFS, and Installable File System
  - ▶ Enable automated system management
  - ▶ Supports multitenancy
  - ▶ Flexible administration

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

22

EMC Atmos supports object-based storage for unstructured data, such as pictures and videos. Atmos combines massive scalability with specialized intelligence to address the cost, distribution, and management challenges associated with vast amounts of unstructured data.

Atmos can be deployed in two ways: as a purpose-built hardware appliance or as software in VMware environments, where Atmos VE can leverage the existing servers and storage. The hardware appliance is comprised of servers (nodes) connected to standard disk enclosures. The rack includes a 24-port Gigabit Ethernet switch to provide internode communication. The Atmos software is installed on each node.

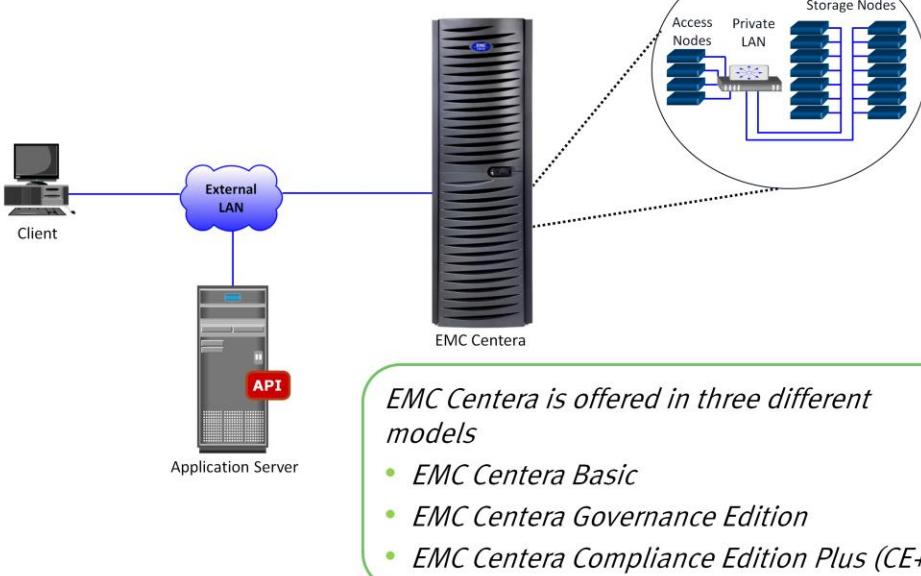
Atmos VE enables users to exploit the power of Atmos in a virtualized environment. It can be deployed on a virtual machine in VMware ESXi hosts and configured with the VMware certified back-end storage.

Cont..

Following are the key features offered by EMC Atmos:

- *Policy-based management*: EMC Atmos improves operational efficiency by automatically distributing content based on business policy. The administrator-defined policies dictate how, when, and where the information resides.
- *Protection*: Atmos offers two options to protect the objects, replication and Geo Parity:
  - *Replication* ensures that the content is available and accessible by creating redundant copies of an object at multiple designated locations.
  - *Geo Parity* ensures that the content is available and accessible by dividing objects into multiple segments plus parity segments and distributing them to one or more designated locations.
- *Data services*: EMC Atmos includes the data services, such as, compression, and deduplication. These features are native to Atmos and can be managed and accessed via a policy.
- *Web services and legacy protocols*: EMC Atmos provides flexible web services access (REST/SOAP) for web-scale applications and file access (CIFS/NFS/Installable File System/Centera API) for traditional applications.
- *Automated system management*: EMC Atmos provides auto-configuring, auto-managing, and auto-healing capabilities to reduce administration and downtime.
- *Multitenancy*: EMC Atmos enables multiple applications to be served from the same infrastructure. Each application is securely partitioned and cannot access the other application's data. Multitenancy is ideal for service providers or large enterprises that want to provide cloud computing services to multiple customers or departments allowing logical and secure separation within a single infrastructure.
- *Flexible administration*: EMC Atmos can be managed via a graphical user interface (GUI) or command line interface (CLI).

## EMC Centera



*EMC Centera is offered in three different models*

- *EMC Centera Basic*
- *EMC Centera Governance Edition*
- *EMC Centera Compliance Edition Plus (CE+)*

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

24

EMC Centera is a simple, affordable, and secure repository for information archiving. EMC Centera is designed and optimized specifically to deal with the storage and retrieval of fixed content by meeting performance, compliance, and regulatory requirements. Compared to traditional archive storage, EMC Centera provides faster record retrieval, Single instance storage (SIS), guaranteed content authenticity, self-healing, and support for numerous industry and regulatory standards.

EMC Centera is offered in three different models to meet different types of user requirements—EMC Centera Basic, EMC Centera Governance Edition, and EMC Centera Compliance Edition Plus (CE+):

- *EMC Centera Basic*: Provides all functionalities without the enforcement of retention periods.
- *EMC Centera Governance Edition*: Provides the retention capabilities required by organizations to manage digital records in addition to the features provided by EMC Centera Basic.
- *EMC Centera Compliance Edition Plus*: Provides extensive compliance capabilities. CE+ is designed to meet the requirements of the most stringent regulated business environments for electronic storage media, as established by regulations from the Securities and Exchange Commission (SEC), or other national and international regulatory groups.

Cont..

A client accesses the Centera over a LAN. The client can access Centera only through the server that runs the Centera API (application programming interface). The Centera API is responsible for performing functions that enable an application to store and retrieve the data.

Centera architecture is a *Redundant Array of Independent Nodes* (RAIN). It contains storage nodes and access nodes that are networked as a cluster by using a private LAN. The internal LAN reconfigures automatically when it detects configuration changes, such as the addition of storage or access nodes. The application server accesses the Centera via an external LAN.

The nodes are configured with low-cost, high-capacity SATA disk drives. These nodes run CentraStar, the operating environment for Centera, which provides the features and functionalities required in a Centera system.

When the nodes are installed, they are configured with a “role” that defines the functionality provided by the node. A node can be configured as a storage node, an access node, or a dual-role node.

*Storage nodes* store and protect data objects. They are sometimes referred to as *back-end nodes*.

*Access nodes* provide connectivity to application servers through an external LAN. They establish connectivity with the storage nodes in the cluster through a private LAN. The number of access nodes is determined by the amount of throughput required from the cluster. If a node is configured solely as an “access node,” its disk space cannot be used to store data objects. Storage and retrieval requests are sent to the access node via the external LAN.

*Dual-role nodes* provide both storage and access-node capabilities. This configuration is more common than a pure access-node configuration.

## EMC VNX

- Unified storage platform that consolidates block, file, and object accesses in one solution
  - ▶ Supports block access via storage processors
  - ▶ File access via X-Blade
  - ▶ Object access via EMC Atmos VE
- Components of VNX are:
  - ▶ Storage processor
  - ▶ X-Blade
  - ▶ Control station
  - ▶ Disk-array enclosures
  - ▶ Standby power supply



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 8: Object-Based and Unified Storage

26

EMC VNX is a unified storage platform that consolidates block, file, and object access in one solution. It implements a modular architecture that integrates hardware components for block, file, and object access. EMC VNX delivers file access (NAS) functionality via X-Blades (Data Movers) and block access functionality via storage processors. Optionally it offers object access to the storage using EMC Atmos Virtual Edition (Atmos VE). VNX storage systems include the following components:

- *Storage processors (SPs)* support block I/O access to storage with FC, iSCSI, and FCoE protocols.
- *X-Blades* access data from the back end and provide host access with NFS, CIFS, MPFS, pNFS, and FTP protocols. The X-Blades in each array are scalable and provide redundancy to ensure no single point of failure.
- *Control Stations* provide management functions to the X-Blades. The Control Station is also responsible for X-Blade failover. The Control Station may optionally be configured with a matching secondary Control Station to ensure management redundancy on the VNX array.
- *Disk-array enclosures (DAEs)* house the drives used in the array. Different sized DAEs are available that can each hold a maximum of 15, 25, or 60 drives. More DAEs can be added to meet growing storage demands.
- *Standby power supplies* provide enough power to each storage processor and first DAE to ensure that any data in flight is stored in the vault area if a power failure occurs. This ensures that no writes are lost.

## Module 8: Summary

Key points covered in this module:

- Object-based storage model
- Key components of object-based storage
- Process of storage and retrieval in object-based storage
- Content-addressed storage
- Key components of unified storage
- Process of data access from unified storage

This module covered the object-based storage model. Object-based storage has three key components such as OSD nodes, internal network, and storage. This module also covered the process of storage and retrieval in object-based storage and content-addressed storage.

Further, this module covered the key components of unified storage; storage processor, NAS head, OSD node, and storage. Finally, this module covered the process of data access from unified storage.

## Check Your Knowledge – 1

- What is an advantage of a flat address space over a hierarchical address space?
  - A. Highly scalable with minimal impact on performance
  - B. Provides access to data, based on retention policies
  - C. Provides access to block, file, and object with same interface
  - D. Consumes less bandwidth on network while accessing data
- What is a role of metadata service in an OSD node?
  - A. Responsible for storing data in the form of objects
  - B. Stores unique IDs generated for objects
  - C. Stores both objects and object IDs
  - D. Controls functioning of storage devices

## Check Your Knowledge – 2

- What is used to generate an object ID in a CAS system?
  - A. File metadata
  - B. Source and destination address
  - C. Binary representation of data
  - D. File system type and ownership
- What accurately describes block I/O access in a unified storage?
  - A. I/O traverse NAS head and storage controller to disk
  - B. I/O traverse OSD node and storage controller to disk
  - C. I/O traverse storage controller to disk
  - D. I/O is directly sent to the disk

## Check Your Knowledge – 3

- What accurately describes unified storage?
  - A. Provides block, file, and object-based access within one platform
  - B. Provides block and file storage access using objects
  - C. Supports block and file access using flat address space
  - D. Specialized storage device purposely built for archiving

# Module – 9

# Introduction to Business Continuity



## Module 9: Introduction to Business Continuity

Upon completion of this module, you should be able to:

- Define business continuity (BC) and information availability (IA)
- Explain the impact of information unavailability
- Describe BC planning process
- Explain business impact analysis (BIA)
- Explain BC technology solutions

This module focuses on the importance of business continuity, the factors that can affect information availability, and the consequences of information unavailability. This module also details on BC planning process and BC technology solutions, specifically on eliminating single points of failure.

# Module 9: Introduction to Business Continuity

## Lesson 1: Business Continuity Overview

During this lesson the following topics are covered:

- Business continuity
- Information availability metrics

This lesson covers the importance of business continuity to an organization, factors that can affect information availability and the consequences of information unavailability. Also this lesson focuses on information availability metrics namely mean time between failure (MTBF) and mean time to repair (MTTR).

## Why Business Continuity?

- Information is an organization's most important asset
- Continuous access to information ensures smooth functioning of business operations
- Cost of unavailability of information to an organization is greater than ever

In today's world, continuous access to information is a must for the smooth functioning of business operations. The cost of unavailability of information is greater than ever, and outages in key industries cost millions of dollars per hour. There are many threats to information availability, such as natural disasters, unplanned occurrences, and planned occurrences, that could result in the inaccessibility of information. Therefore it is critical for businesses to define appropriate strategies that can help them to overcome these crises. Business continuity is an important process to define and implement these strategies.

## What is Business Continuity?

### Business Continuity

It is a process that prepares for, responds to, and recovers from a system outage that can adversely affects business operations.

- An integrated and enterprise-wide process that includes set of activities to ensure “information availability”
- BC involves proactive measures and reactive countermeasures
- In a virtualized environment, BC solutions need to protect both physical and virtualized resources

*Business continuity (BC)* is an integrated and enterprise-wide process that includes all activities (internal and external to IT) that a business must perform to mitigate the impact of planned and unplanned downtime. BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. It involves proactive measures, such as business impact analysis, risk assessments, BC technology solutions deployment (backup and replication), and reactive measures, such as disaster recovery and restart, to be invoked in the event of a failure. The goal of a BC solution is to ensure the “information availability” required to conduct vital business operations.

In a virtualized environment, BC technology solutions need to protect both physical and virtualized resources. Virtualization considerably simplifies the implementation of BC strategy and solutions.

## Information Availability

### Information Availability

It is the ability of an IT infrastructure to function according to business expectations, during its specified time of operation.

- Information availability can be defined with the help of:
  - ▶ Accessibility
    - ▶ Information should be accessible to the right user when required
  - ▶ Reliability
    - ▶ Information should be reliable and correct in all aspects
  - ▶ Timeliness
    - ▶ Defines the time window during which information must be accessible

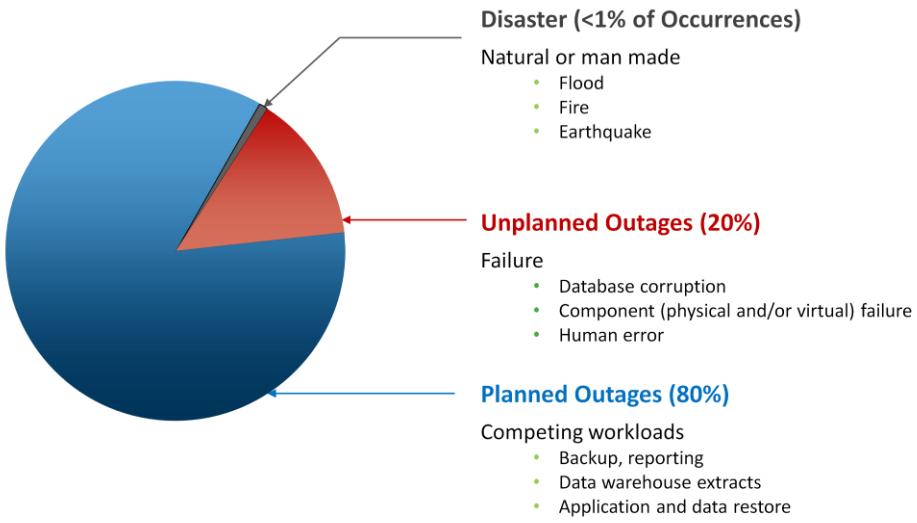
Information availability (IA) refers to the ability of an IT infrastructure to function according to business expectations during its specified time of operation. IA ensures that people (employees, customers, suppliers, and partners) can access information whenever they need it. IA can be defined in terms of the accessibility, reliability, and timeliness of the information.

**Accessibility:** Information should be accessible to the right user when required.

**Reliability:** Information should be reliable and correct in all aspects. It is “the same” as what was stored and there is no alteration or corruption to the information.

**Timeliness:** Defines the time window (a particular time of the day, week, month, and year as specified) during which information must be accessible. For example, if online access to an application is required between 8:00 am and 10:00 pm each day, any disruptions to data availability outside of this time slot are not considered to affect timeliness.

## Causes of Information Unavailability



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

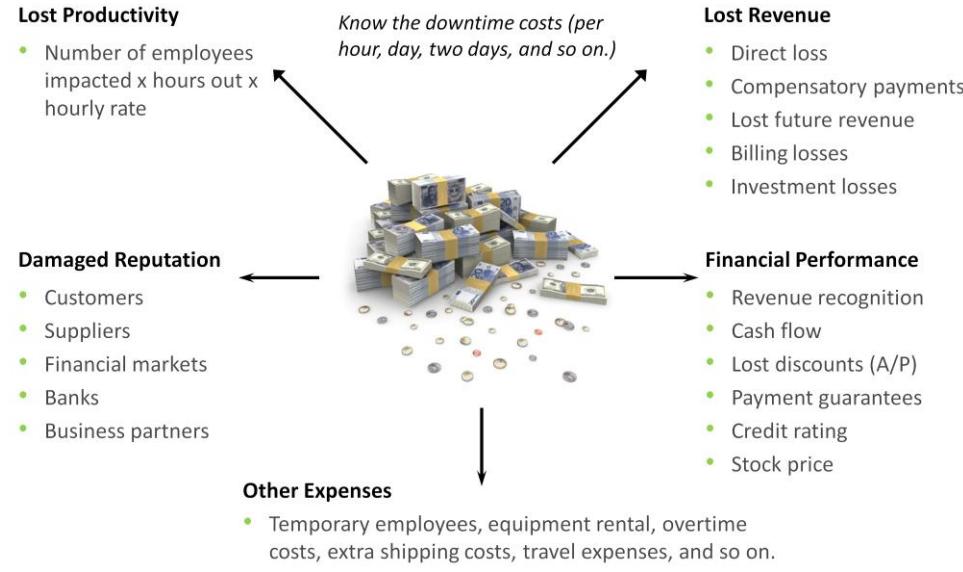
Module 9: Introduction to Business Continuity

7

Various planned and unplanned incidents result in information unavailability. *Planned outages* include installation/integration/maintenance of new hardware, software upgrades or patches, taking backups, application and data restores, facility operations (renovation and construction), and refresh/migration of the testing to the production environment. *Unplanned outages* include failure caused by human errors, database corruption, and failure of physical and virtual components.

Another type of incident that may cause data unavailability is natural or man-made disasters, such as flood, fire, earthquake, and so on. As illustrated in figure on the slide, the majority of outages are planned. Planned outages are expected and scheduled but still cause data to be unavailable. Statistically, the cause of information unavailability due to unforeseen disasters is less than 1 percent.

## Impact of Downtime



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity

8

Information unavailability or downtime results in loss of productivity, loss of revenue, poor financial performance, and damages to reputation. Loss of productivity include reduced output per unit of labor, equipment, and capital. Loss of revenue includes direct loss, compensatory payments, future revenue loss, billing loss, and investment loss. Poor financial performance affects revenue recognition, cash flow, discounts, payment guarantees, credit rating, and stock price. Damages to reputations may result in a loss of confidence or credibility with customers, suppliers, financial markets, banks, and business partners. Other possible consequences of downtime include the cost of additional equipment rental, overtime, and extra shipping.

The business impact of downtime is the sum of all losses sustained as a result of a given disruption. An important metric, *average cost of downtime per hour*, provides a key estimate in determining the appropriate BC solutions. It is calculated as follows:

Average cost of downtime per hour = average productivity loss per hour + average revenue loss per hour

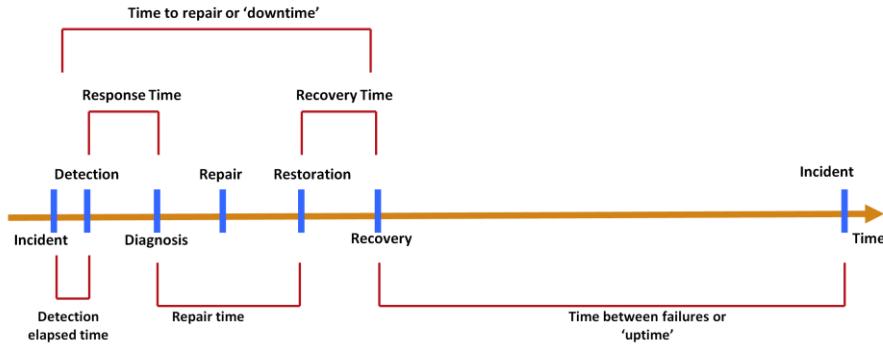
Where:

Productivity loss per hour = (total salaries and benefits of all employees per week) / (average number of working hours per week)

Average revenue loss per hour = (total revenue of an organization per week) / (average number of hours per week that an organization is open for business)

The average downtime cost per hour may also include estimates of projected revenue loss due to other consequences, such as damaged reputations, and the additional cost of repairing the system.

## Measuring Information Availability



- MTBF: Average time available for a system or component to perform its normal operations between failures

$$MTBF = \text{Total uptime}/\text{Number of failures}$$

- MTTR: Average time required to repair a failed component

$$MTTR = \text{Total downtime}/\text{Number of failures}$$

$$IA = \frac{MTBF}{(MTBF + MTTR)} \text{ or } IA = \frac{\text{uptime}}{(\text{uptime} + \text{downtime})}$$

Information availability relies on the availability of both physical and virtual components of a data center. Failure of these components might disrupt information availability. A failure is the termination of a component's ability to perform a required function. The component's ability can be restored by performing an external corrective actions, such as a manual reboot, a repair, or replacement of the failed component(s). Proactive risk analysis, performed as part of the BC planning process, considers the component failure rate and average repair time, which are measured by MTBF and MTTR:

**Mean Time Between Failure (MTBF):** It is the average time available for a system or component to perform its normal operations between failures. It is the measure of system or component reliability and is usually expressed in hours.

**Mean Time To Repair (MTTR):** It is the average time required to repair a failed component. MTTR includes the total time required to do the following activities: detect the fault, mobilize the maintenance team, diagnose the fault, obtain the spare parts, repair, test, and restore the data. MTTR is calculated as: Total downtime/Number of failures

IA can be expressed in terms of system uptime and downtime and measured as the amount or percentage of system uptime:

$$IA = \text{system uptime} / (\text{system uptime} + \text{system downtime})$$

Where *system uptime* is the period of time during which the system is in an accessible state; when it is not accessible, it is termed as *system downtime*.

In terms of MTBF and MTTR, IA could also be expressed as:  $IA = \frac{MTBF}{(MTBF + MTTR)}$

## Availability Measurement – Levels of ‘9s’ Availability

Uptime (%)	Downtime (%)	Downtime per Year	Downtime per Week
98	2	7.3 days	3hrs, 22 minutes
99	1	3.65 days	1 hr, 41 minutes
99.8	0.2	17 hrs, 31 minutes	20 minutes, 10 secs
99.9	0.1	8 hrs, 45 minutes	10 minutes, 5 secs
99.99	0.01	52.5 minutes	1 minute
99.999	0.001	5.25 minutes	6 secs
99.9999	0.0001	31.5 secs	0.6 secs

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity 10

Uptime per year is based on the exact timeliness requirements of the service. This calculation leads to the number of “9s” representation for availability metrics. Table on the slide lists the approximate amount of downtime allowed for a service to achieve certain levels of 9s availability.

For example, a service that is said to be “five 9s available” is available for 99.999 percent of the scheduled time in a year ( $24 \times 365$ ).

## Module 9: Introduction to Business Continuity

### Lesson 2: BC Planning and Technology Solutions

During this lesson the following topics are covered:

- BC terminologies
- BC planning
- Business impact analysis
- Single points of failure
- Multipathing software

This lesson covers various BC terminologies and BC planning. This lesson also focuses on eliminating single points of failure and multipathing software.

## BC Terminologies – 1

- Disaster recovery
  - ▶ Coordinated process of restoring systems, data, and infrastructure required to support business operations after a disaster occurs
  - ▶ Restoring previous copy of data and applying logs to that copy to bring it to a known point of consistency
  - ▶ Generally implies use of backup technology
- Disaster restart
  - ▶ Process of restarting business operations with mirrored consistent copies of data and applications
  - ▶ Generally implies use of replication technologies

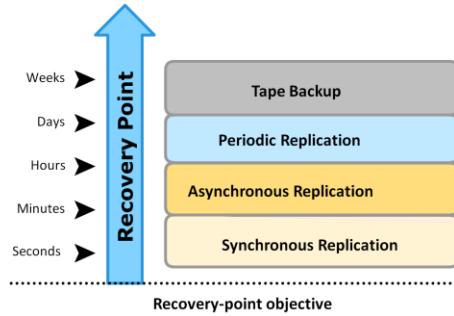
**Disaster recovery:** This is the coordinated process of restoring systems, data, and the infrastructure required to support ongoing business operations after a disaster occurs. It is the process of restoring a previous copy of the data and applying logs or other necessary processes to that copy to bring it to a known point of consistency. After all recovery efforts are completed, the data is validated to ensure that it is correct.

**Disaster restart:** This is the process of restarting business operations with mirrored consistent copies of data and applications.

## BC Terminologies – 2

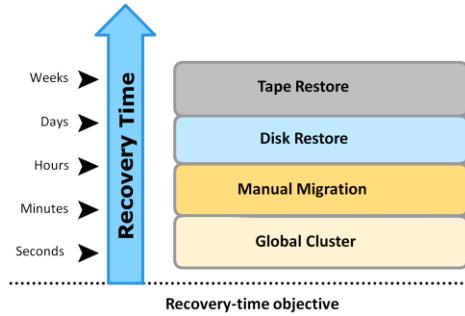
### Recovery-Point Objective (RPO)

- Point-in-time to which systems and data must be recovered after an outage
- Amount of data loss that a business can endure



### Recovery-Time Objective (RTO)

- Time within which systems and applications must be recovered after an outage
- Amount of downtime that a business can endure and survive



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity 13

**Recovery-Point Objective (RPO):** This is the point-in-time to which systems and data must be recovered after an outage. It defines the amount of data loss that a business can endure. Based on the RPO, organizations plan for the frequency with which a backup or replica must be made. An organization can plan for an appropriate BC technology solution on the basis of the RPO it sets. For example:

**RPO of 24 hours:** Backups are created at an offsite tape library every midnight. The corresponding recovery strategy is to restore data from the set of last backup tapes.

**RPO of 1 hour:** Shipping database logs to the remote site every hour. The corresponding recovery strategy is to recover the database to the point of the last log shipment.

**RPO in the order of minutes:** Mirroring data asynchronously to a remote site.

**RPO of zero:** Mirroring data synchronously to a remote site.

**Recovery-Time Objective (RTO):** The time within which systems and applications must be recovered after an outage. It defines the amount of downtime that a business can endure and survive. Some examples of RTOs and the recovery strategies to ensure data availability are listed below:

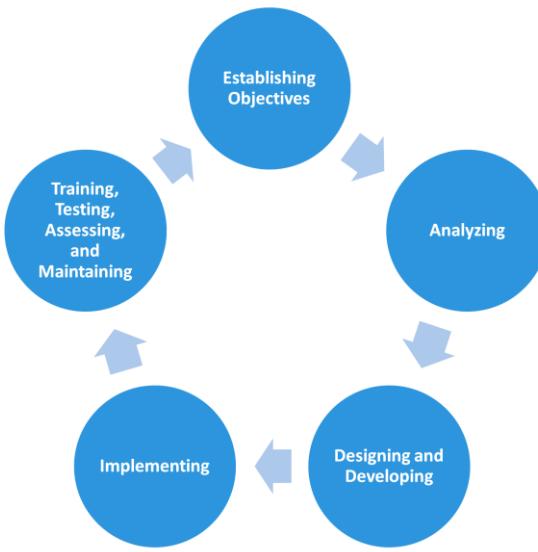
**RTO of 72 hours:** Restore from tapes available at a cold site.

**RTO of 12 hours:** Restore from tapes available at a hot site.

**RTO of few hours:** Use disk-based backup technology, which gives faster restore than a tape backup.

**RTO of a few seconds:** Cluster production servers with bidirectional mirroring, enabling the applications to run at both sites simultaneously.

## BC Planning Lifecycle



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity 14

BC planning must follow a disciplined approach like any other planning process. Organizations today dedicate specialized resources to develop and maintain BC plans. From the conceptualization to the realization of the BC plan, a lifecycle of activities can be defined for the BC process. The BC planning lifecycle includes five stages:

### 1. Establishing objectives

- Determine BC requirements.
- Estimate the scope and budget to achieve requirements.
- Select a BC team that includes subject matter experts from all areas of the business, whether internal or external.
- Create BC policies.

### 2. Analyzing

- Collect information on data profiles, business processes, infrastructure support, dependencies, and frequency of using business infrastructure.
- Conduct a business impact analysis.
- Identify critical business processes and assign recovery priorities.
- Perform risk analysis for critical functions and create mitigation strategies.
- Perform cost benefit analysis for available solutions based on the mitigation strategy.
- Evaluate options.

Cont..

### **3. Designing and developing**

- Define the team structure and assign individual roles and responsibilities. For example, different teams are formed for activities, such as emergency response, and infrastructure and application recovery.
- Design data protection strategies and develop infrastructure.
- Develop contingency solution and emergency response procedures.
- Detail recovery and restart procedures.

### **4. Implementing**

- Implement risk management and mitigation procedures that include backup, replication, and management of resources.
- Prepare the DR sites that can be utilized if a disaster affects the primary data center.
- Implement redundancy for every resource in a data center to avoid single points of failure.

### **5. Training, testing, assessing, and maintaining**

- Train the employees who are responsible for backup and replication of business-critical data on a regular basis or whenever there is a modification in the BC plan.
- Train employees on emergency response procedures when disasters are declared.
- Train the recovery team on recovery procedures based on contingency scenarios.
- Perform damage-assessment processes and review recovery plans.
- Test the BC plan regularly to evaluate its performance and identify its limitations.
- Assess the performance reports and identify limitations.
- Update the BC plans and recovery/restart procedures to reflect regular changes within the data center.

## Business Impact Analysis

- Identifies which business units and processes are essential to the survival of the business
- Estimates the cost of failure for each business process
- Calculates the maximum tolerable outage and defines RTO for each business process
- Businesses can prioritize and implement countermeasures to mitigate the likelihood of such disruptions

A *business impact analysis* (BIA) identifies which business units, operations, and processes are essential to the survival of the business. It evaluates the financial, operational, and service impacts of a disruption to essential business processes. Selected functional areas are evaluated to determine resilience of the infrastructure to support information availability. The BIA process leads to a report detailing the incidents and their impact over business functions. The impact may be specified in terms of money or in terms of time. Based on the potential impacts associated with downtime, businesses can prioritize and implement countermeasures to mitigate the likelihood of such disruptions. These are detailed in the BC plan. A BIA includes the following set of tasks:

- Determine the business areas.
- For each business area, identify the key business processes critical to its operation.
- Determine the attributes of the business process in terms of applications, databases, and hardware and software requirements.
- Estimate the costs of failure for each business process.
- Calculate the maximum tolerable outage and define RTO for each business process.
- Establish the minimum resources required for the operation of business processes.
- Determine recovery strategies and the cost for implementing them.
- Optimize the backup and business recovery strategy based on business priorities.
- Analyze the current state of BC readiness and optimize future BC planning.

## BC Technology Solutions

- Solutions that enable BC are:
  - ▶ Resolving single points of failure
  - ▶ Multipathing software
  - ▶ Backup and replication
    - ▶ Backup
    - ▶ Local replication
    - ▶ Remote replication

After analyzing the business impact of an outage, designing the appropriate solutions to recover from a failure is the next important activity. Following are the solutions and supporting technologies that enable business continuity and uninterrupted data availability:

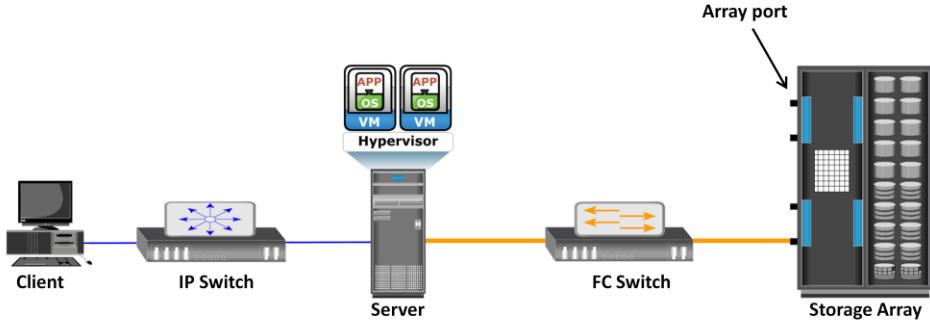
- Resolving single points of failure
- Multipathing software
- Backup and replication

*Note:* Backup and Replication will be discussed in forthcoming modules.

## Single Points of Failure

### Single Points of Failure

It refers to the failure of a component of a system that can terminate the availability of the entire system or IT service.

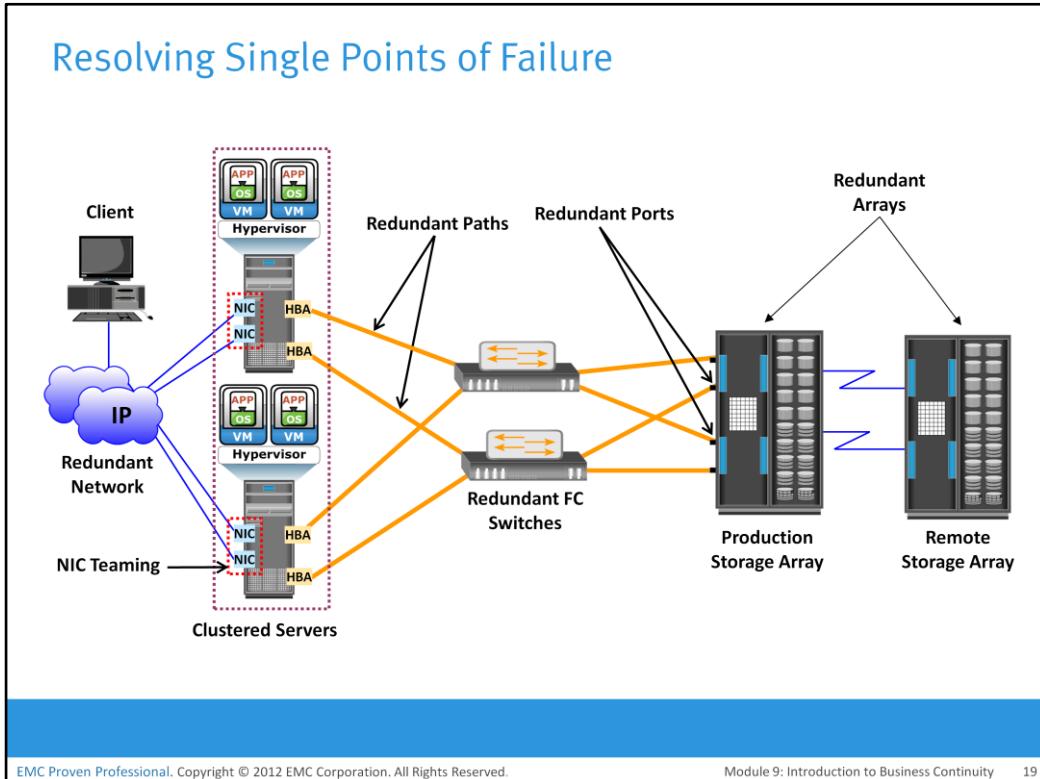


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity 18

A *single point of failure* refers to the failure of a component that can terminate the availability of the entire system or IT service. The figure depicts a system setup in which an application, running on a VM, provides an interface to the client and performs I/O operations. The client is connected to the server through an IP network, and the server is connected to the storage array through an FC connection. In this setup, each component must function as required to ensure data availability; the failure of a single physical or virtual component causes the unavailability of an application. This failure results in disruption of business operations. For example, failure of a hypervisor can affect all the running VMs and virtual network, which are hosted on it. In the figure on the slide, several single points of failure can be identified. A VM, a hypervisor, or an HBA/NIC on the server, the physical server itself, the IP network, the FC switch, the storage array port, or even the storage array could be a potential single point of failure.

## Resolving Single Points of Failure



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity 19

To mitigate single points of failure, systems are designed with redundancy, such that the system fails only if all the components in the redundancy group fail. This ensures that the failure of a single component does not affect data availability. Data centers follow stringent guidelines to implement fault tolerance for uninterrupted information availability. Careful analysis is performed to eliminate every single point of failure. The example shown in figure on the slide represents all enhancements in the infrastructure to mitigate single points of failure:

- Configuration of redundant HBAs at a server to mitigate single HBA failure.
- Configuration of NIC teaming at a server allows protection against single physical NIC failure. It allows grouping of two or more physical NICs and treating them as a single logical device. With NIC teaming, if one of the underlying physical NICs fails or its cable is unplugged, the traffic is redirected to another physical NIC in the team. Thus, NIC teaming eliminates the single point of failure associated with a single physical NIC.
- Configuration of redundant switches to account for a switch failure.
- Configuration of multiple storage array ports to mitigate a port failure.
- RAID and hot spare configuration to ensure continuous operation in the event of disk failure.

Cont..

- Implementation of a redundant storage array at a remote site to mitigate local site failure.
- Implementing server (or compute) clustering, a fault-tolerance mechanism whereby two or more servers in a cluster access the same set of data volumes. Clustered servers exchange *heartbeat* to inform each other about their health. If one of the servers or hypervisors fails, the other server or hypervisor can take up the workload.
- Implementing a VM Fault Tolerance mechanism ensures BC in the event of a server failure. This technique creates duplicate copies of each VM on another server so that when a VM failure is detected, the duplicate VM can be used for failover. The two VMs are kept in synchronization with each other in order to perform successful failover.

## Multipathing Software

- Recognizes and utilizes alternate I/O path to data
- Provides load balancing by distributing I/Os to all available, active paths:
  - ▶ Improves I/O performance and data path utilization
- Intelligently manages the paths to a device by sending I/O down the optimal path:
  - ▶ Based on the load balancing and failover policy setting for the device

Configuration of multiple paths increases the data availability through path failover. If servers are configured with one I/O path to the data, there will be no access to the data if that path fails. Redundant paths to the data eliminate the possibility of the path becoming a single point of failure. Multiple paths to data also improve I/O performance through load balancing among the paths and maximize server, storage, and data path utilization.

In practice, merely configuring multiple paths does not serve the purpose. Even with multiple paths, if one path fails, I/O does not reroute unless the system recognizes that it has an alternative path. Multipathing software provides the functionality to recognize and utilize alternative I/O paths to data. Multipathing software also manages the load balancing by distributing I/Os to all available, active paths.

Multipathing software intelligently manages the paths to a device by sending I/O down the optimal path based on the load balancing and failover policy setting for the device. It also takes into account path usage and availability before deciding the path through which to send the I/O. If a path to the device fails, it automatically reroutes the I/O to an alternative path.

In a virtual environment, multipathing is enabled either by using the hypervisor's built-in capability or by running a third-party software module, added to the hypervisor.

## Module 9: Introduction to Business Continuity

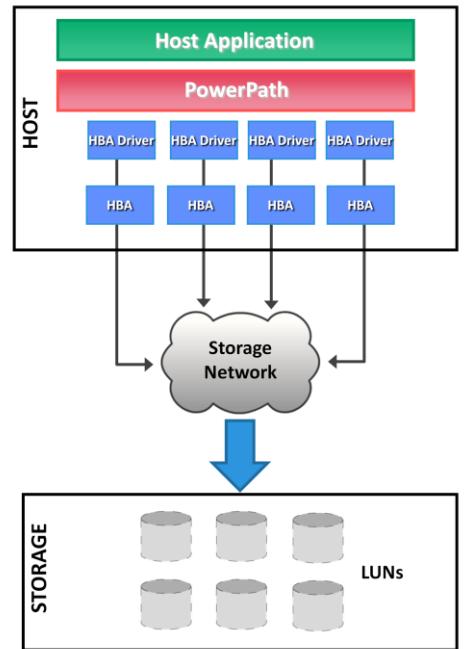
### Concept in Practice

- EMC PowerPath

The concept in practice section covers EMC PowerPath.

## EMC PowerPath

- Host-based multipathing software
- Provides path failover and load-balancing functionality
- Automatic detection and recovery from host-to-array path failures
- PowerPath/VE software allows optimizing virtual environments with PowerPath multipathing features



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 9: Introduction to Business Continuity 23

EMC PowerPath is host-based multipathing software. Every I/O from the host to the array must pass through the PowerPath software, which allows PowerPath to provide intelligent I/O path management. PowerPath provides path failover and dynamic load balancing. PowerPath/VE software allows optimizing virtual environments with PowerPath multipathing features.

## Module 9: Summary

- Importance of business continuity
- Impact of information unavailability
- Information availability metrics
- Business impact analysis
- Single points of failure
- Multipathing software

This module covered the importance of business continuity, impact of information unavailability, and information availability metrics. This module also focused on business continuity planning and business impact analysis. Further, this module detailed on single points of failure and multipathing software.

BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. Information unavailability or downtime results in loss of productivity, loss of revenue, poor financial performance, and damages to reputation.

Information availability metrics are MTBF and MTTR. MTBF defines average time available for a system or component to perform its normal operations between failures. MTTR defines the average time required to repair a failed component. A business impact analysis identifies which business units, operations, and processes are essential to the survival of the business.

A single point of failure refers to the failure of a component that can terminate the availability of the entire system or IT service. Multipathing software provides the functionality to recognize and utilize alternate I/O paths to data. Multipathing software also manages the load balancing by distributing I/Os to all available, active paths.

## Check Your Knowledge – 1

- Which statement is true if the recovery-point objective (RPO) of an application is 2 hours?
  - A. Time to resume application operations must be less than 2 hours
  - B. Time to resume application operations must equal to 2 hours
  - C. No more than 2 hours of production data can be lost
  - D. Mean time between application failure is 2 hours
- Which allows grouping of two or more physical NICs and treating them as a single logical device?
  - A. NIC streaming
  - B. NIC porting
  - C. NIC teaming
  - D. NIC zoning

## Check Your Knowledge – 2

- Which best describes recovery-time objective (RTO)?
  - A. Point-in-time to which data must be recovered after an outage
  - B. Time available for a system or component to perform its normal operations between failures
  - C. Time within which systems and applications must be recovered after an outage
  - D. Amount of data loss that a business can endure
- Which expression represents availability of a system in terms of MTBF and MTTR?
  - A. MTTR/(MTBF x MTTR)
  - B. MTBF/(MTBF x MTTR)
  - C. MTTR/(MTBF + MTTR)
  - D. MTBF/(MTTR + MTBF)

## Check Your Knowledge – 3

- A department requires access to the database application from Monday to Friday, 9 AM to 5 PM. Last Thursday at 1 PM the application crashed and it took six hours to fix the problem. What was the availability of the application during last week?
  - A. 85%
  - B. 90%
  - C. 95%
  - D. 100%

## Exercise 1: MTBF and MTTR

- A system has three components and requires all three to be operational for 24 hours from Monday to Friday. Failure of component 1 occurs as follows:
  - ▶ Monday = No failure
  - ▶ Tuesday = 5 am to 7 am
  - ▶ Wednesday = No failure
  - ▶ Thursday = 4 pm to 8 pm
  - ▶ Friday = 8 am to 11 am

Calculate the MTBF and MTTR of component 1.

## Exercise 2: Information Availability

- A system has three components and requires all three to be operational from 8 am to 5 pm business hours, Monday to Friday. Failure of component 2 occurs as follows:
  - ▶ Monday = 8 am to 11 am
  - ▶ Tuesday = No failure
  - ▶ Wednesday = 4 pm to 7 pm
  - ▶ Thursday = 5 pm to 8 pm
  - ▶ Friday = 1 pm to 2 pm

Calculate the availability of component 2.

This slide intentionally left blank.

# Module – 10

# Backup and Archive



## Module 10: Backup and Archive

Upon completion of this module, you should be able to:

- Describe backup granularities
- Explain backup and recovery operations
- Describe various backup targets
- Explain data deduplication
- Describe backup in virtualized environment
- Explain data archive

This module focuses on backup granularities and backup operations. This module also focuses on various backup targets and data deduplication. Further, this module details backup in virtualized environment. Additionally, this module focuses on data archive.

# Module 10: Backup and Archive

## Lesson 1: Backup Overview

During this lesson the following topics are covered:

- Backup granularity
- Backup method
- Backup architecture
- Backup and recovery operations

This lesson covers various backup granularities and backup method. This lesson also covers backup architecture and operations.

## What is Backup?

### Backup

It is an additional copy of production data that is created and retained for the sole purpose of recovering lost or corrupted data.

- Organization also takes backup to comply with regulatory requirements
- Backups are performed to serve three purposes:
  - ▶ Disaster recovery
  - ▶ Operational recovery
  - ▶ Archive

A *backup* is an additional copy of production data, created and retained for the sole purpose of recovering lost or corrupted data. With growing business and regulatory demands for data storage, retention, and availability, organizations are faced with the task of backing up an ever-increasing amount of data. This task becomes more challenging with the growth of information, stagnant IT budgets, and less time for taking backups. Moreover, organizations need a quick restore of backed up data to meet business service-level agreements (SLAs).

Backups are performed to serve three purposes: disaster recovery, operational recovery, and archival.

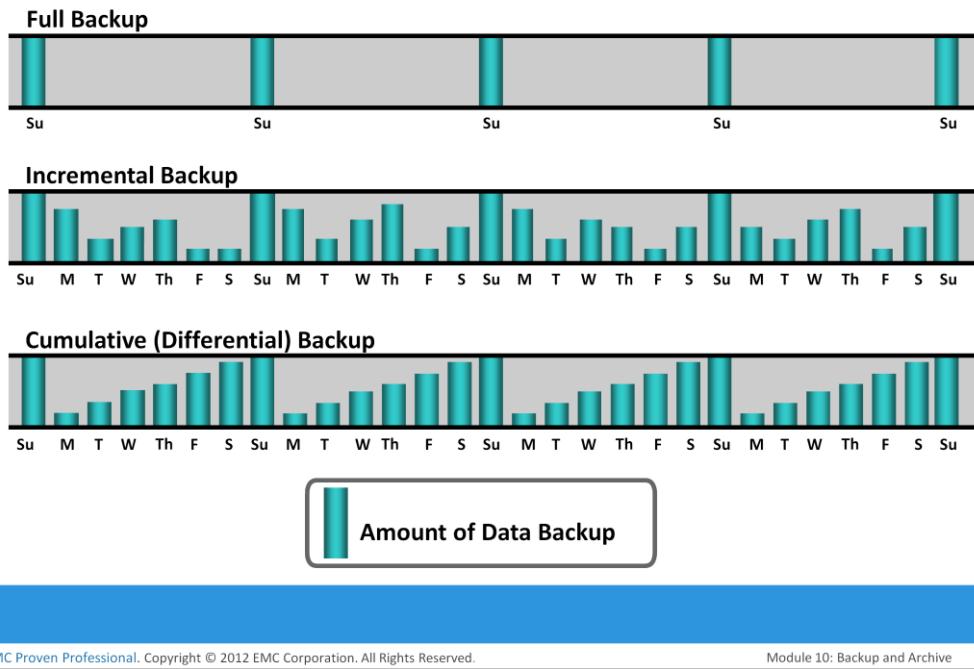
Backups can be performed to address disaster recovery needs. The backup copies are used for restoring data at an alternate site when the primary site is incapacitated due to a disaster. Based on Recovery-point objective (RPO) and Recovery-time objective (RTO) requirements, organizations use different data protection strategies for disaster recovery.

Data in the production environment changes with every business transaction and operation. Backups are used to restore data if data loss or logical corruption occurs during routine processing. The majority of restore requests in most organizations fall in this category. For example, it is common for a user to accidentally delete an important e-mail or for a file to become corrupted, which can be restored using backup data.

Backups are also performed to address archival requirements. Although content addressed storage (CAS) has emerged as the primary solution for archives (CAS is discussed in module 8), traditional backups are still used by small and medium enterprises for long-term preservation of transaction records, e-mail messages, and other business records required for regulatory compliance.

*Note:* Backup window is the period during which a source is available for performing a data backup.

## Backup Granularity



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

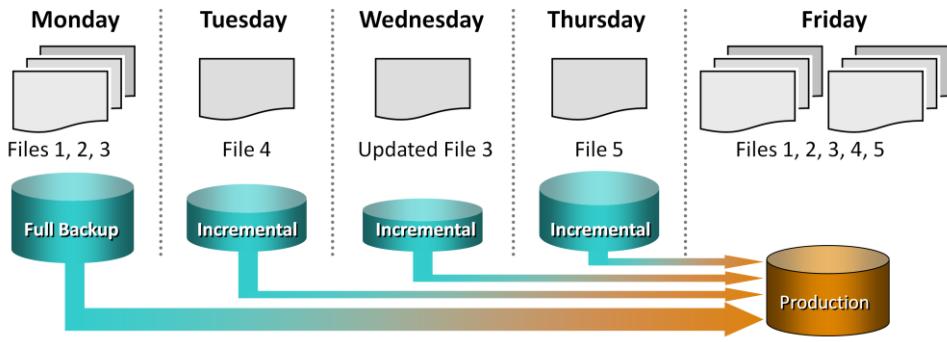
5

Backup granularity depends on business needs and the required RTO/RPO. Based on the granularity, backups can be categorized as full, incremental, and cumulative (or differential). Most organizations use a combination of these three backup types to meet their backup and recovery requirements. Figure on the slide depicts the different backup granularity levels.

*Full backup* is a backup of the complete data on the production volumes. A full backup copy is created by copying the data in the production volumes to a backup storage device. It provides a faster recovery but requires more storage space and also takes more time to back up. *Incremental backup* copies the data that has changed since the last full or incremental backup, whichever has occurred more recently. This is much faster than a full backup (because the volume of data backed up is restricted to the changed data only) but takes longer to restore. *Cumulative backup* copies the data that has changed since the last full backup. This method takes longer than an incremental backup but is faster to restore.

Another way to implement full backup is *synthetic* (or *constructed*) backup. This method is used when the production volume resources cannot be exclusively reserved for a backup process for extended periods to perform a full backup. It is usually created from the most recent full backup and all the incremental backups performed after that full backup. This backup is called *synthetic* because the backup is not created directly from production data. A synthetic full backup enables a full backup copy to be created offline without disrupting the I/O operation on the production volume. This also frees up network resources from the backup process, making them available for other production uses.

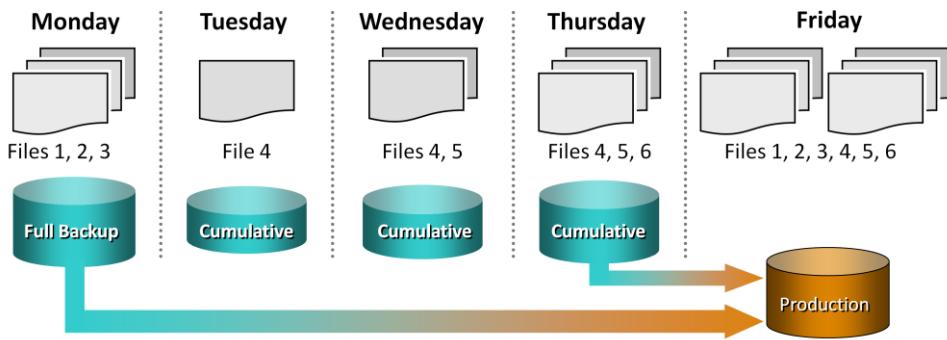
## Restoring from Incremental Backup



- Less number of files to be backed up, therefore, it takes less time to backup and requires less storage space
- Longer restore because last full and all subsequent incremental backups must be applied

The process of restoration from an incremental backup requires the last full backup and all the incremental backups available until the point of restoration. Consider an example, a full backup is performed on Monday evening. Each day after that, an incremental backup is performed. On Tuesday, a new file (File 4 as shown in the figure) is added, and no other files have changed. Consequently, only File 4 is copied during the incremental backup performed on Tuesday evening. On Wednesday, no new files are added, but File 3 has been modified. Therefore, only the modified File 3 is copied during the incremental backup on Wednesday evening. Similarly, the incremental backup on Thursday copies only File 5. On Friday morning, there is data corruption, which requires data restoration from the backup. The first step toward data restoration is restoring all data from the full backup of Monday evening. The next step is applying the incremental backups of Tuesday, Wednesday, and Thursday. In this manner, data can be successfully recovered to its previous state, as it existed on Thursday evening.

## Restoring from Cumulative Backup

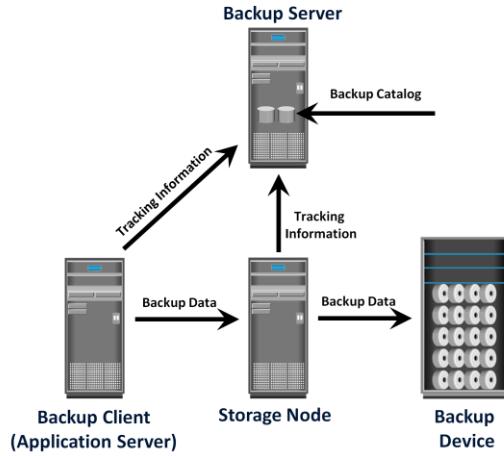


- More files to be backed up, therefore, it takes more time to backup and requires more storage space
- Faster restore because only the last full and the last cumulative backup must be applied

Consider an example, a full backup of the business data is taken on Monday evening. Each day after that, a cumulative backup is taken. On Tuesday, File 4 is added and no other data is modified since the previous full backup of Monday evening. Consequently, the cumulative backup on Tuesday evening copies only File 4. On Wednesday, File 5 is added. The cumulative backup taking place on Wednesday evening copies both File 4 and File 5 because these files have been added or modified since the last full backup. Similarly, on Thursday, File 6 is added. Therefore, the cumulative backup on Thursday evening copies all three files: File 4, File 5, and File 6. On Friday morning, data corruption occurs that requires data restoration using backup copies. The first step in restoring data is to restore all the data from the full backup of Monday evening. The next step is to apply only the latest cumulative backup, which is taken on Thursday evening. In this way, the production data can be recovered faster because its needs only two copies of data—the last full backup and the latest cumulative backup.

## Backup Architecture

- Backup client
  - ▶ Gathers the data that is to be backed up and send it to storage node
- Backup server
  - ▶ Manages backup operations and maintains backup catalog
- Storage node
  - ▶ Responsible for writing data to backup device
  - ▶ Manages the backup device



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

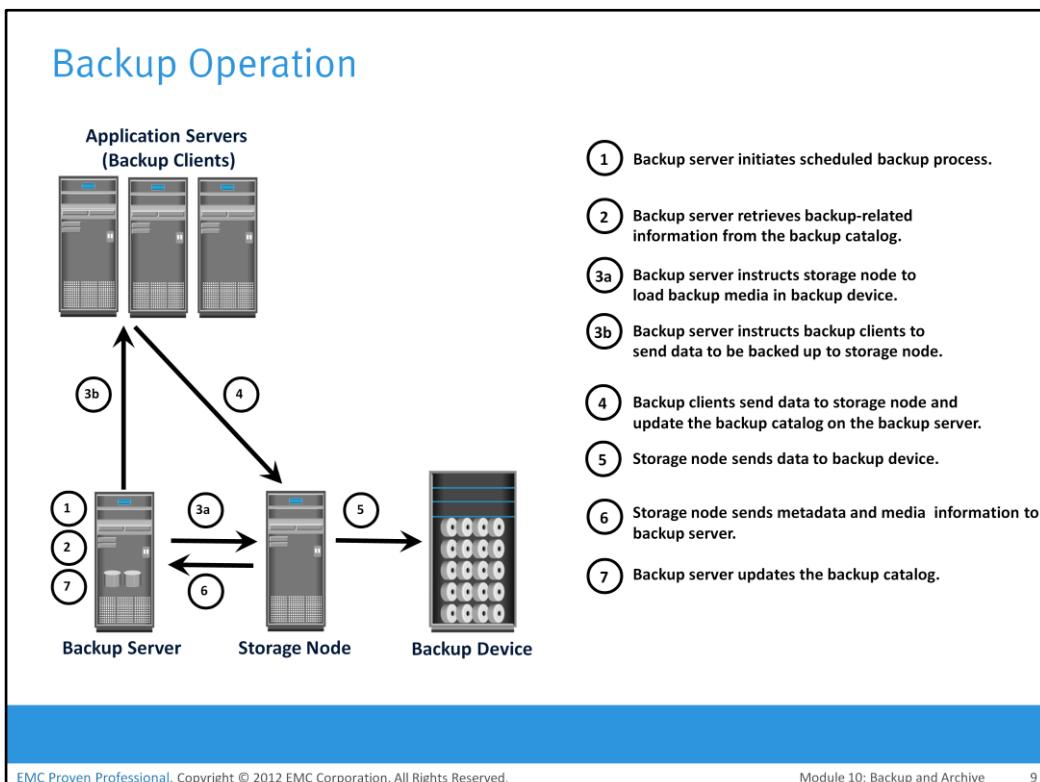
Module 10: Backup and Archive

8

A backup system commonly uses the client-server architecture with a backup server and multiple backup clients. Figure on the slide illustrates the backup architecture. The backup server manages the backup operations and maintains the backup catalog, which contains information about the backup configuration and backup metadata. Backup configuration contains information about when to run backups, which client data to be backed up, and so on, and the backup metadata contains information about the backed up data. The role of a backup client is to gather the data that is to be backed up and send it to the storage node. It also sends the tracking information to the backup server.

The storage node is responsible for writing the data, to the backup device. In a backup environment, a *storage node* is a host that controls backup devices. The storage node also sends tracking information to the backup server. In many cases, the storage node is integrated with the backup server, and both are hosted on the same physical platform. A backup device is attached directly or through a network to the storage node's host platform. Some backup architecture refers the storage node as the *media server* because it manages the storage device.

## Backup Operation



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

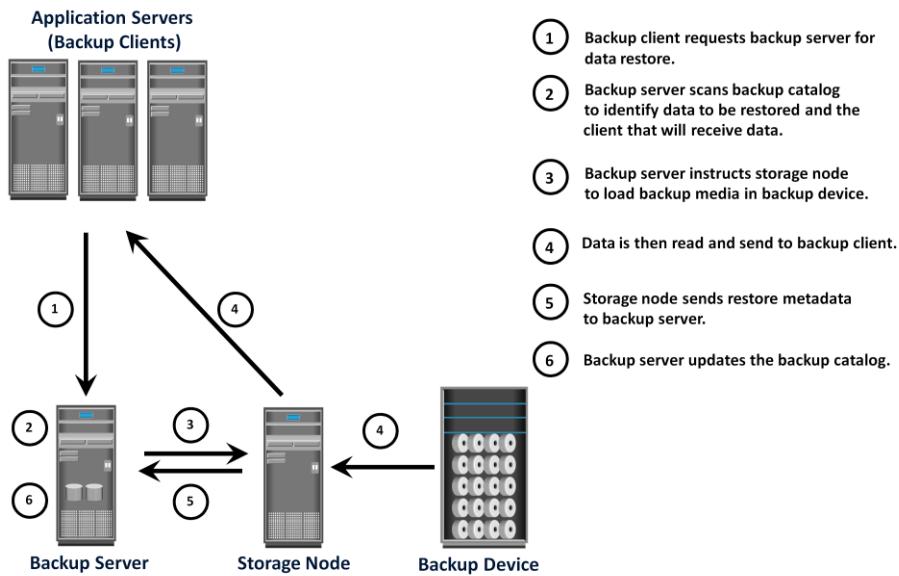
Module 10: Backup and Archive

9

When a backup operation is initiated, significant network communication takes place between the different components of a backup infrastructure. The backup operation is typically initiated by a server, but it can also be initiated by a client. The backup server initiates the backup process for different clients based on the backup schedule configured for them. For example, the backup for a group of clients may be scheduled to start at 3:00 a.m. every day.

The backup server coordinates the backup process with all the components in a backup environment. The backup server maintains the information about backup clients to be backed up and storage nodes to be used in a backup operation. The backup server retrieves the backup-related information from the backup catalog and, based on this information, instructs the storage node to load the appropriate backup media into the backup devices. Simultaneously, it instructs the backup clients to gather the data to be backed up and send it over the network to the assigned storage node. After the backup data is sent to the storage node, the client sends some backup metadata (the number of files, name of the files, storage node details, and so on) to the backup server. The storage node receives the client data, organizes it, and sends it to the backup device. The storage node then sends additional backup metadata (location of the data on the backup device, time of backup, and so on) to the backup server. The backup server updates the backup catalog with this information.

## Recovery Operation



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

10

After the data is backed up, it can be restored when required. A restore process must be manually initiated from the client. Some backup software has a separate application for restore operations. These restore applications are usually accessible only to the administrators or backup operators. Figure on the slide depicts a restore operation.

Upon receiving a restore request, an administrator opens the restore application to view the list of clients that have been backed up. While selecting the client for which a restore request has been made, the administrator also needs to identify the client that will receive the restored data. Data can be restored on the same client for whom the restore request has been made or on any other client. The administrator then selects the data to be restored and the specified point in time to which the data has to be restored based on the RPO. Because all this information comes from the backup catalog, the restore application needs to communicate with the backup server.

The backup server instructs the appropriate storage node to mount the specific backup media onto the backup device. Data is then read and sent to the client that has been identified to receive the restored data.

Some restorations are successfully accomplished by recovering only the requested production data. For example, the recovery process of a spreadsheet is completed when the specific file is restored. In database restorations, additional data, such as log files, must be restored along with the production data. This ensures consistency for the restored data. In these cases, the RTO is extended due to the additional steps in the restore operation.

## Backup Methods

- Two methods of backup, based on the state of the application when the backup is performed
  - ▶ Hot or Online
    - ▶ Application is up and running, with users accessing their data during backup
    - ▶ Open file agent can be used to backup open files
  - ▶ Cold or Offline
    - ▶ Requires application to be shutdown during the backup process
- Bare-metal recovery
  - ▶ OS, hardware, and application configurations are appropriately backed up for a full system recovery
  - ▶ Server configuration backup (SCB) can also recover a server onto dissimilar hardware

Hot backup and cold backup are the two methods deployed for backup. They are based on the state of the application when the backup is performed. In a *hot backup*, the application is up-and-running, with users accessing their data during the backup process. This method of backup is also referred to as online backup. A *cold backup* requires the application to be shutdown during the backup process. Hence, this method is also referred to as offline backup.

The hot backup of online production data is challenging because data is actively being used and changed. If a file is open, it is normally not backed up during the backup process. In such situations, an *open file agent* is required to back up the open file. These agents interact directly with the operating system or application and enable the creation of consistent copies of open files. The disadvantage associated with a hot backup is that the agents usually affect the overall application performance. Consistent backups of databases can also be done by using a cold backup. This requires the database to remain inactive during the backup. Of course, the disadvantage of a cold backup is that the database is inaccessible to users during the backup process. All the files must be backed up in the same state for consistent backup of a database that comprises many files.

In a disaster recovery environment, *bare-metal recovery* (BMR) refers to a backup in which OS, hardware, and application configurations are appropriately backed up for a full system recovery. BMR builds the base system, which includes partitioning, the file system layout, the operating system, the applications, and all the relevant configurations. BMR recovers the base system first before starting the recovery of data files. Some BMR technologies—for example server configuration backup (SCB)—can recover a server even onto dissimilar hardware.

## Server Configuration Backup

- Creates and backs up server configuration profiles, based on user-defined schedules
  - ▶ Profiles are used to configure the recovery server in case of production server failure
  - ▶ Profiles include OS configurations, network configurations, security configurations, registry settings, application configurations
- Two types of profiles used
  - ▶ Base profile
    - ▶ Contains the key elements of the OS required to recover the server
  - ▶ Extended profile
    - ▶ Typically larger than base profile and contains all necessary information to rebuild application environment

Most organizations spend a considerable amount of time and money protecting their application data but give less attention to protecting their server configurations. During disaster recovery, server configurations must be re-created before the application and data are accessible to the user. The process of system recovery involves reinstalling the operating system, applications, and server settings and then recovering the data. During a normal data backup operation, server configurations required for the system restore are not backed up. *Server configuration backup (SCB)* creates and backs up server configuration profiles based on user-defined schedules. The backed up profiles are used to configure the recovery server in case of production-server failure. SCB has the capability to recover a server onto dissimilar hardware.

In a server configuration backup, the process of taking a snapshot of the application server's configuration (both system and application configurations) is known as *profiling*. The profile data includes operating system configurations, network configurations, security configurations, registry settings, application configurations, and so on. Thus, profiling allows recovering the configuration of the failed system to a new server regardless of the underlying hardware.

There are two types of profiles generated in the server configuration backup environment: base profile and extended profile. The base profile contains the key elements of the operating system required to recover the server. The extended profile is typically larger than the base profile and contains all the necessary information to rebuild the application environment.

## Key Backup/Restore Considerations

- Customer business needs determine:
  - ▶ What are the restore requirements – RPO & RTO?
  - ▶ Which data needs to be backed up?
  - ▶ How frequently should data be backed up?
  - ▶ How long will it take to backup?
  - ▶ How many copies to create?
  - ▶ How long to retain backup copies?
  - ▶ Location, size, and number of files?

The amount of data loss and downtime that a business can endure in terms of RPO and RTO are the primary considerations in selecting and implementing a specific backup strategy. The RPO determines backup frequency. For example, if an application requires an RPO of 1 day, it would need the data to be backed up at least once every day. Another consideration is the retention period, which defines the duration for which a business needs to retain the backup copies.

The backup media type or backup target is another consideration that is driven by RTO and impacts the data recovery time. Organizations must also consider the granularity of backups. The development of a backup strategy must include a decision about the most appropriate time for performing a backup to minimize any disruption to production operations. The size, number of files and data compression should also be considered because they might affect the backup process. Backing up large-size files (for example, ten 1 MB files) takes less time, compared to backing up an equal amount of data composed of small-size files (for example, ten thousand 1 KB files). Data compression and data deduplication (discussed later in the module) are widely used in the backup environment because these technologies save space on the media.

Location is an important consideration for the data to be backed up. Many organizations have dozens of heterogeneous platforms locally and remotely supporting their business. The backup process must address these sources for transactional and content integrity.

## Module 10: Backup and Archive

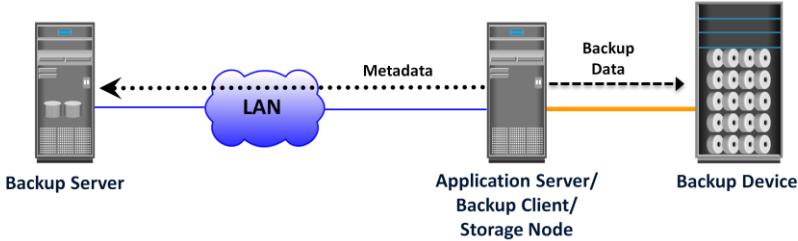
### Lesson 2: Backup Topologies and Backup in NAS Environment

During this lesson the following topics are covered:

- Common backup topologies
- Backup in NAS environment

This lesson covers various backup topologies such as Direct-attached, LAN-based, SAN-based and mixed backup. This lesson also covers backup in NAS environment.

## Direct-Attached Backup



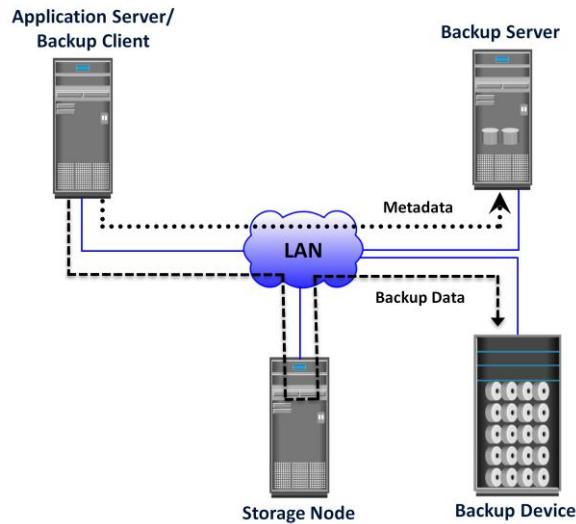
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

15

In a *direct-attached backup*, the storage node is configured on a backup client, and the backup device is attached directly to the client. Only the metadata is sent to the backup server through the LAN. This configuration frees the LAN from backup traffic. As the environment grows, there will be a need for centralized management and sharing of backup devices to optimize costs. An appropriate solution is required to share the backup devices among multiple servers. Network-based topologies (LAN-based and SAN-based) provide the solution to optimize the utilization of backup devices.

## LAN-based Backup



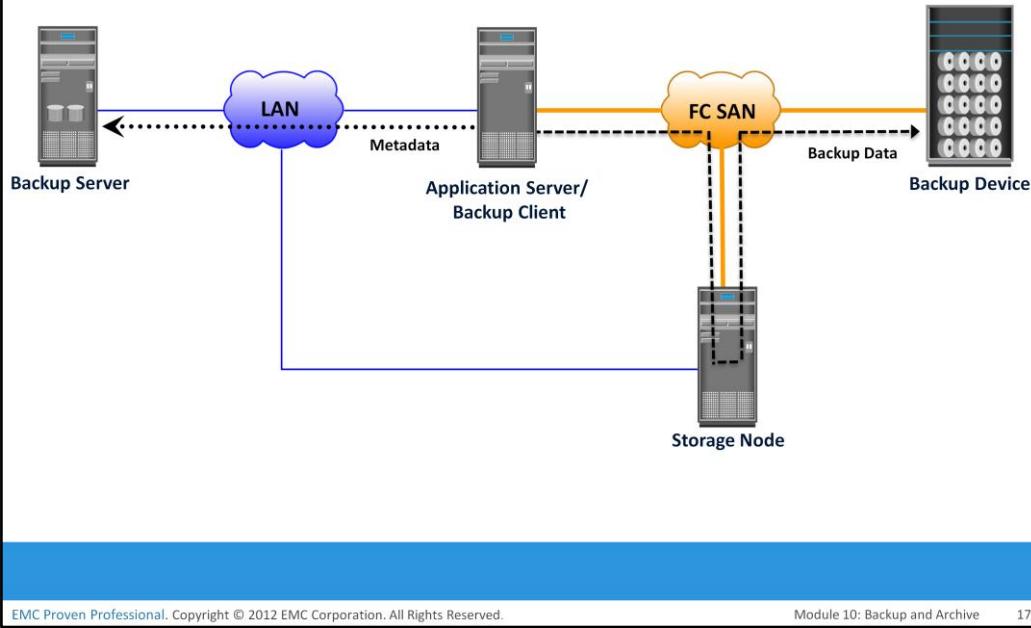
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

16

In a *LAN-based backup*, the clients, backup server, storage node, and backup device are connected to the LAN. The data to be backed up is transferred from the backup client (source) to the backup device (destination) over the LAN, which might affect network performance. This impact can be minimized by adopting a number of measures, such as configuring separate networks for backup and installing dedicated storage nodes for some application servers.

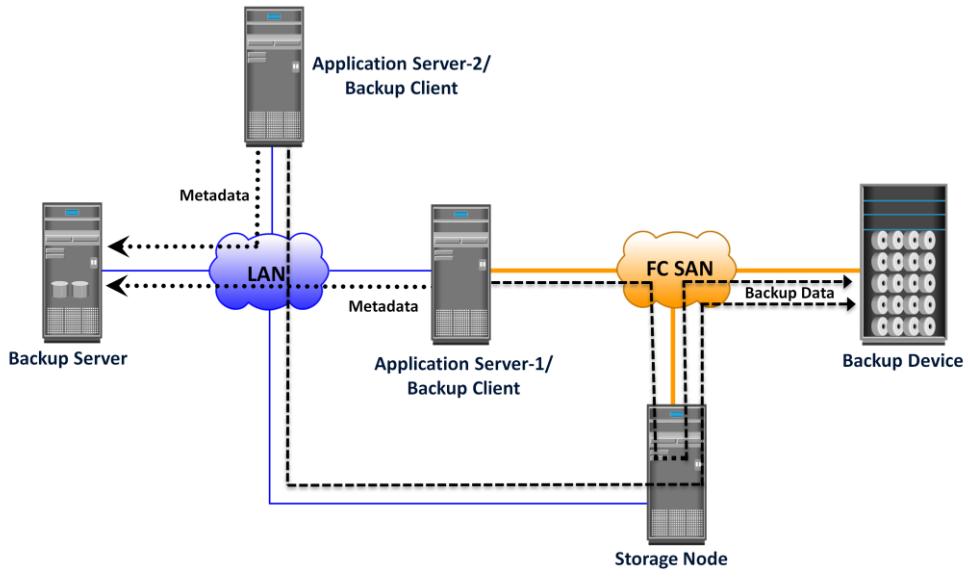
## SAN-based Backup



A *SAN-based backup* is also known as a *LAN-free backup*. The SAN-based backup topology is the most appropriate solution when a backup device needs to be shared among clients. In this case, the backup device and clients are attached to the SAN. In the figure shown on the slide, a client sends the data to be backed up to the backup device over the SAN. Therefore, the backup data traffic is restricted to the SAN, and only the backup metadata is transported over the LAN. The volume of metadata is insignificant when compared to the production data; the LAN performance is not degraded in this configuration.

The emergence of low-cost disks as a backup medium has enabled disk arrays to be attached to the SAN and used as backup devices. A tape backup of these data backups on the disks can be created and shipped offsite for disaster recovery and long-term retention.

## Mixed Backup Topology



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

18

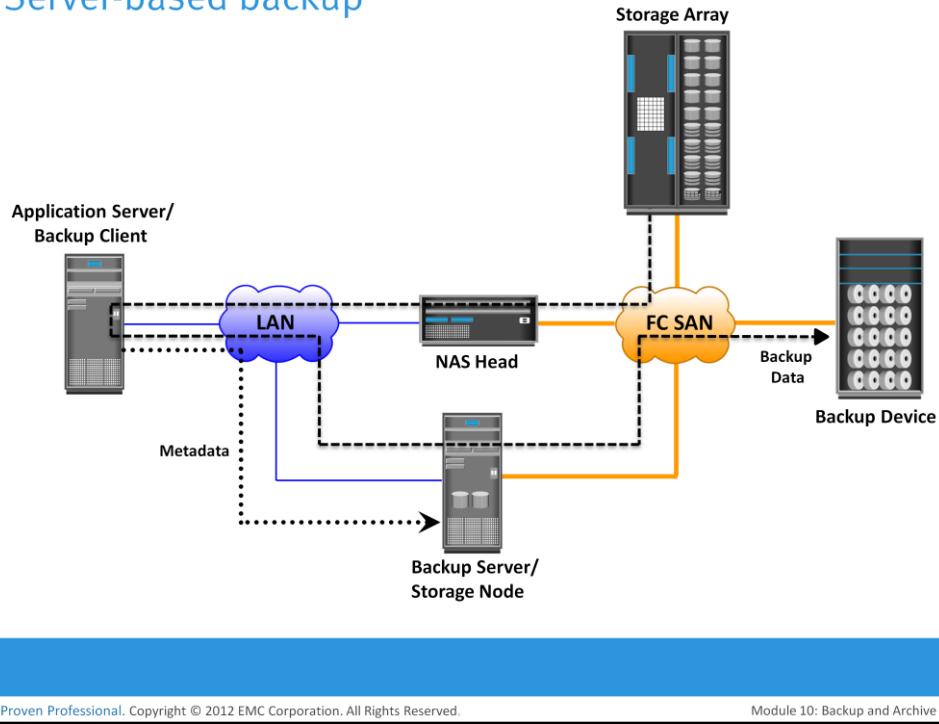
The *mixed topology* uses both the LAN-based and SAN-based topologies. This topology might be implemented for several reasons, including cost, server location, reduction in administrative overhead, and performance considerations.

## Backup in NAS Environment

- Common backup implementations in a NAS environment are:
  - ▶ Server-based backup
  - ▶ Serverless backup
  - ▶ NDMP 2-way backup
  - ▶ NDMP 3-way backup

The use of a NAS head imposes a new set of considerations on the backup and recovery strategy in NAS environments. NAS heads use a proprietary operating system and file system structure that supports multiple file-sharing protocols. In the NAS environment, backups can be implemented in different ways: server-based, serverless, or using Network Data Management Protocol (NDMP). Common implementations are NDMP 2-way and NDMP 3-way.

## Server-based backup



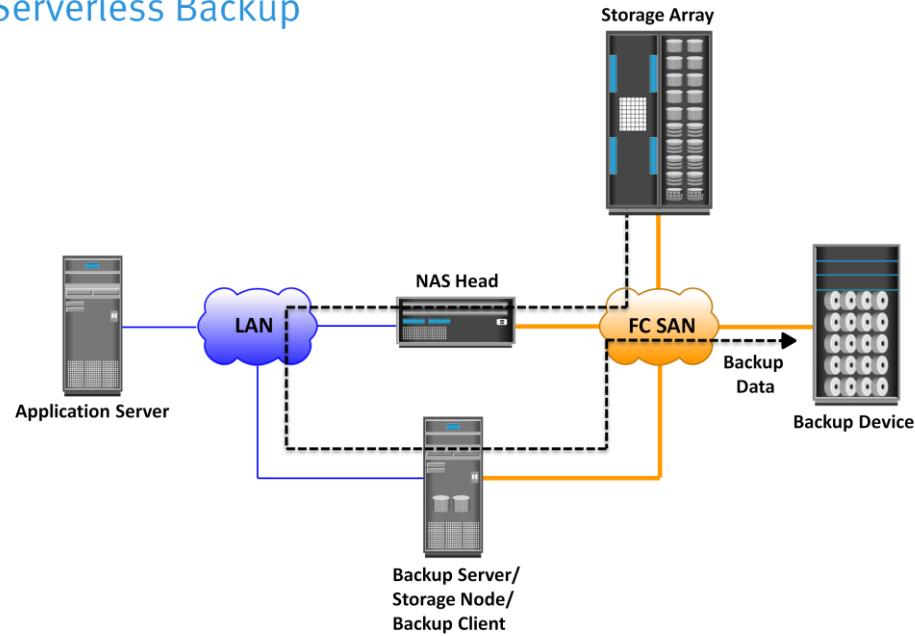
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

20

In an *application server-based backup*, the NAS head retrieves data from a storage array over the network and transfers it to the backup client running on the application server. The backup client sends this data to the storage node, which in turn writes the data to the backup device. This results in overloading the network with the backup data and using application server resources to move the backup data.

## Serverless Backup



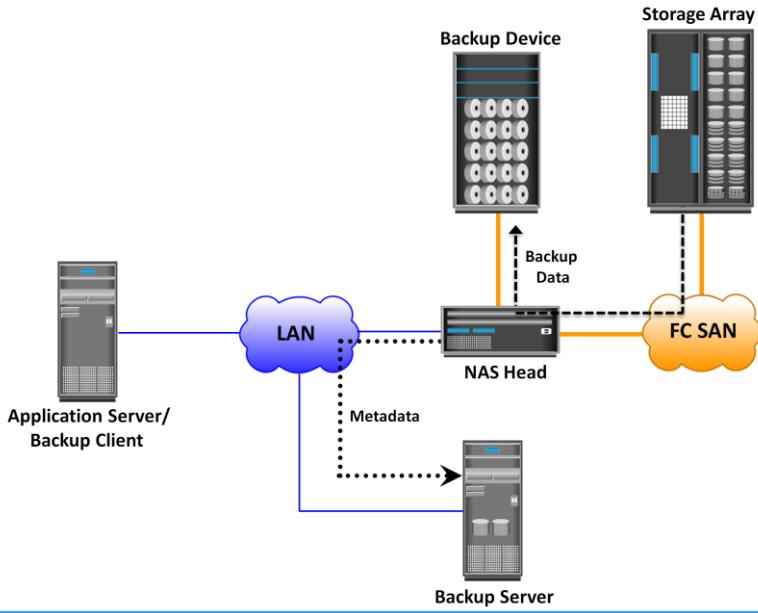
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

21

In a *serverless backup*, the network share is mounted directly on the storage node. This avoids overloading the network during the backup process and eliminates the need to use resources on the application server. In this scenario, the storage node, which is also a backup client, reads the data from the NAS head and writes it to the backup device without involving the application server. Compared to the previous solution, this eliminates one network hop.

## NDMP 2-way Backup



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

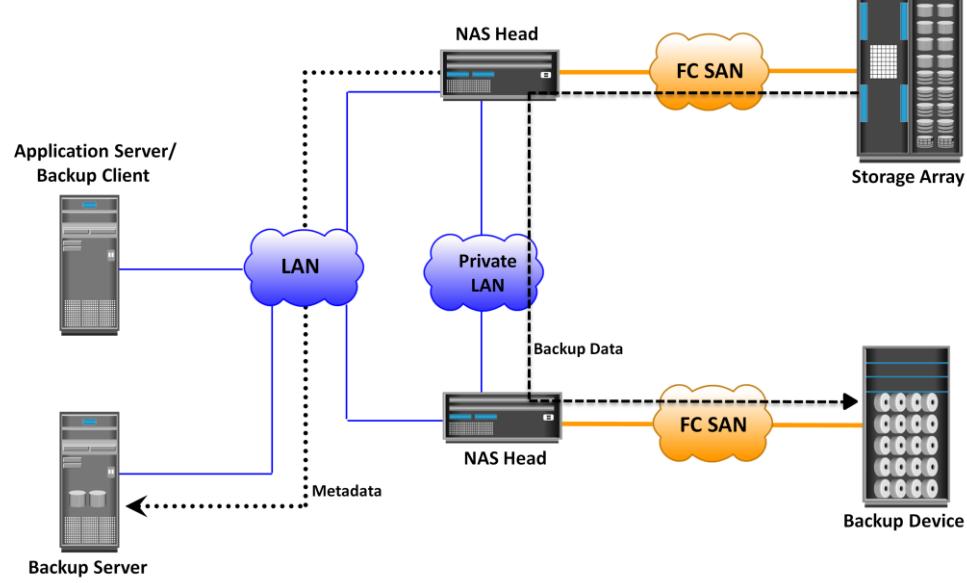
Module 10: Backup and Archive

22

**NDMP** is an industry-standard TCP/IP-based protocol specifically designed for a backup in a NAS environment. It communicates with several elements in the backup environment (NAS head, backup devices, backup server, and so on) for data transfer and enables vendors to use a common protocol for the backup architecture. Data can be backed up using NDMP regardless of the operating system or platform. Due to its flexibility, it is no longer necessary to transport data through the application server, which reduces the load on the application server and improves the backup speed. NDMP optimizes backup and restore by leveraging the high-speed connection between the backup devices and the NAS head. In NDMP, backup data is sent directly from the NAS head to the backup device, whereas metadata is sent to the backup server.

Figure on the slide illustrates backup in the NAS environment using NDMP 2-way. In this model, network traffic is minimized by isolating data movement from the NAS head to the locally attached backup device. Only metadata is transported on the network. The backup device is dedicated to the NAS device, and hence, this method does not support centralized management of all backup devices.

## NDMP 3-way Backup



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

23

In the *NDMP 3-way* method, to avoid the backup data traveling on the production LAN, a separate private backup network must be established between all NAS heads and the NAS head connected to the backup device. Metadata and NDMP control data are still transferred across the public network. Figure on the slide depicts NDMP 3-way backup. An NDMP 3-way is useful when backup devices need to be shared among NAS heads. It enables the NAS head to control the backup device and share it with other NAS heads by receiving the backup data through the NDMP.

## Module 10: Backup and Archive

### Lesson 3: Backup Targets

During this lesson the following topics are covered:

- Backup to Tape
- Backup to Disk
- Backup to Virtual Tape

This lesson covers various backup targets such as physical tape, disk and virtual tape.

## Backup to Tape

- Traditionally low cost solution
- Tape drives are used to read/write data from/to a tape
- Sequential/linear access
- Multiple streaming to improve media performance
  - ▶ Writes data from multiple streams on a single tape
- Limitation of tape
  - ▶ Backup and recovery operations are slow due to sequential access
  - ▶ Wear and tear of tape
  - ▶ Shipping/handling challenges
  - ▶ Controlled environment is required for tape storage
  - ▶ Causes “shoe shining effect” or “backhitching”

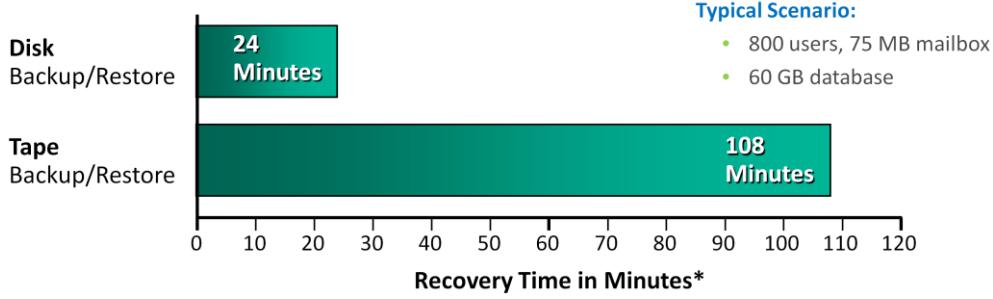
Tapes, a low-cost solution, are used extensively for backup. Tape drives are used to read/write data from/to a tape cartridge (or cassette). Tape drives are referred to as sequential, or linear, access devices because the data is written or read sequentially. A tape cartridge is composed of magnetic tapes in a plastic enclosure. Tape Mounting is the process of inserting a tape cartridge into a tape drive. The tape drive has motorized controls to move the magnetic tape around, enabling the head to read or write data.

Tape drive *streaming* or *multiple streaming* writes data from multiple streams on a single tape to keep the drive busy. Multiple streaming improves media performance, but it has an associated disadvantage. The backup data is interleaved because data from multiple streams is written on it. Consequently, the data recovery time is increased because all the extra data from the other streams must be read and discarded while recovering a single stream.

Data access in a tape is sequential, which can slow backup and recovery operations. Tapes are primarily used for long-term offsite storage because of their low cost. Tapes must be stored in locations with a controlled environment to ensure preservation of the media and to prevent data corruption. Tapes are highly susceptible to wear and tear and usually have shorter shelf life. Physical transportation of the tapes to offsite locations also adds to management overhead and increases the possibility of loss of tapes during offsite shipment. Many times, even the buffering and speed adjustment features of a tape drive fail to prevent the gaps, causing the “*shoe shining effect*” or “*backhitching*.” *Shoe shining* is the repeated back and forth motion a tape drive makes when there is an interruption in the backup data stream. This repeated back-and-forth motion not only causes a degradation of service, but also excessive wear and tear to tapes.

## Backup to Disk

- Enhanced overall backup and recovery performance
  - ▶ Random access
- More reliable
- Can be accessed by multiple hosts simultaneously



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

26

Availability of low cost disks have now replaced tapes as the primary device for storing backup data because of their performance advantages. Backup-to-disk systems offer ease of implementation, reduced TCO, and improved quality of service. Apart from performance benefits in terms of data transfer rates, disks also offer faster recovery when compared to tapes.

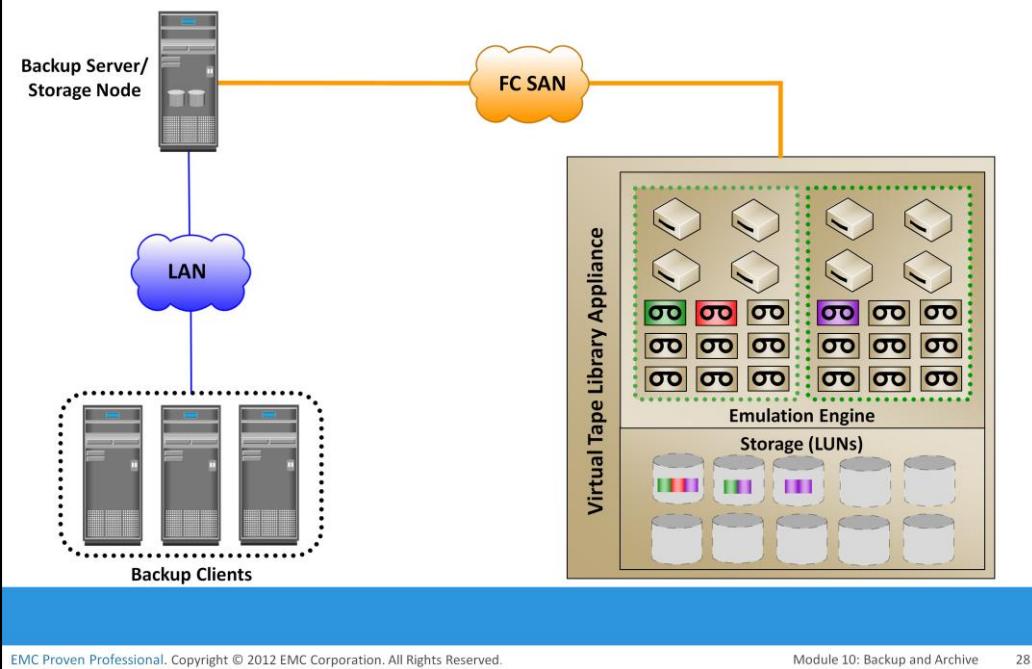
Backing up to disk storage systems offers clear advantages due to their inherent random access and RAID-protection capabilities. In most backup environments, backup to disk is used as a staging area where the data is copied temporarily before transferring or staging it to tapes. This enhances backup performance. Some backup products allow for backup images to remain on the disk for a period of time even after they have been staged. This enables a much faster restore. Figure on the slide illustrates a recovery scenario comparing tape versus disk in a Microsoft Exchange environment that supports 800 users with a 75 MB mailbox size and a 60 GB database. As shown in the figure, a restore from the disk took 24 minutes compared to the restore from a tape, which took 108 minutes for the same environment.

## Backup to Virtual Tape

- Disks are emulated and presented as tapes to backup software
- Does not require any additional modules or changes in the legacy backup software
- Provides better single stream performance and reliability over physical tape
- Online and random disk access
  - ▶ Provides faster backup and recovery

*Virtual tapes* are disk drives emulated and presented as tapes to the backup software. The key benefit of using a virtual tape is that it does not require any additional modules, configuration, or changes in the legacy backup software. This preserves the investment made in the backup software. Compared to physical tapes, virtual tapes offer better single stream performance, better reliability, and random disk access characteristics. Backup and restore operations are benefited from the disk's random access characteristics because they are online and provide faster backup and recovery. A virtual tape drive does not require the usual maintenance tasks associated with a physical tape drive, such as periodic cleaning and drive calibration. Compared to backup-to-disk devices, a virtual tape library offers easy installation and administration because it is preconfigured by the manufacturer.

## Virtual Tape Library



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

28

A *virtual tape library* (VTL) has the same components as that of a physical tape library, except that the majority of the components are presented as virtual resources. For the backup software, there is no difference between a physical tape library and a virtual tape library. Figure on the slide shows a virtual tape library. Virtual tape libraries use disks as backup media. Emulation software has a database with a list of virtual tapes, and each virtual tape is assigned space on a LUN. A virtual tape can span multiple LUNs if required. File system awareness is not required while backing up because the virtual tape solution typically uses raw devices.

Similar to a physical tape library, a robot mount is virtually performed when a backup process starts in a virtual tape library. However, unlike a physical tape library, where this process involves some mechanical delays, in a virtual tape library it is almost instantaneous. Even the *load to ready* time is much less than a physical tape library. After the virtual tape is mounted and the virtual tape drive is positioned, the virtual tape is ready to be used, and backup data can be written to it. In most cases, data is written to the virtual tape immediately. Unlike a physical tape library, the virtual tape library is not constrained by the sequential access and shoe shining effect. When the operation is complete, the backup software issues a rewind command. This rewind is also instantaneous. The virtual tape is then unmounted, and the virtual robotic arm is instructed to move it back to a virtual slot.

The steps to restore data are similar to those in a physical tape library, but the restore operation is nearly instantaneous. Even though virtual tapes are based on disks, which provide random access, they still emulate the tape behavior.

## Backup Target Comparison

	Tape	Disk	Virtual Tape
Offsite Replication Capabilities	No	Yes	Yes
Reliability	No inherent protection methods	RAID, spare	RAID, spare
Performance	Low	High	High
Use	Backup only	Multiple (backup and production)	Backup only

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

29

The table on the slide provides the comparison among various backup targets.

## Module 10: Backup and Archive

### Lesson 4: Data Deduplication

During this lesson the following topics are covered:

- Deduplication overview
- Deduplication methods
- Deduplication implementations
- Key benefits of deduplication

This lesson covers different deduplication method. This lesson also covers deduplication implementation such as source-based deduplication and target-based deduplication. Further this lesson details various key benefits of data deduplication.

## What is Data Deduplication?

### Data Deduplication

It is a process of identifying and eliminating redundant data.

- Deduplication methods
  - ▶ File level
  - ▶ Subfile level
- Deduplication implementations
  - ▶ Source-based
  - ▶ Target-based

Traditional backup solutions do not provide any inherent capability to prevent duplicate data from being backed up. With the growth of information and 24x7 application availability requirements, backup windows are shrinking. Traditional backup processes back up a lot of duplicate data. Backing up duplicate data significantly increases the backup window size requirements and results in unnecessary consumption of resources, such as storage space and network bandwidth.

*Data deduplication* is the process of identifying and eliminating redundant data. When duplicate data is detected during backup, the data is discarded and only the pointer is created to refer the copy of the data that is already backed up. Data deduplication helps to reduce the storage requirement for backup, shorten the backup window, and remove the network burden. It also helps to store more backups on the disk and retain the data on the disk for a longer time.

There are two methods of data deduplication, file level and subfile level. Determining the uniqueness by implementing either method offers benefits; however, results can vary. The differences exist in the amount of data reduction each method produces and the time each approach takes to determine the unique content.

Deduplication can occur close to where the data is created, which is often referred to as “Source-based Deduplication”. It can also occur close to where the data is stored, which is commonly called “Target-based Deduplication”.

## Data Deduplication Methods

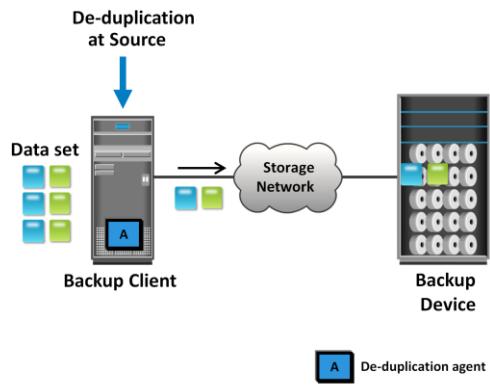
- File-level deduplication (single-instance storage)
  - ▶ Detects and removes redundant copies of identical files
  - ▶ After a file is stored, all other references to the same file refer to the original copy
- Subfile deduplication
  - ▶ Detects redundant data within and across files
  - ▶ Two methods
    - ▶ Fixed-length block
    - ▶ Variable-length segment

*File-level deduplication* (also called *single-instance storage*) detects and removes redundant copies of identical files. It enables storing only one copy of the file; the subsequent copies are replaced with a pointer that points to the original file. File-level deduplication is simple and fast but does not address the problem of duplicate content inside the files. For example, two 10-MB PowerPoint presentations with a difference in just the title page are not considered as duplicate files, and each file will be stored separately.

*Subfile deduplication* breaks the file into smaller chunks and then uses specialized algorithm to detect redundant data within and across the file. As a result, subfile deduplication eliminates duplicate data across files. There are two forms of subfile deduplication: fixed-length block and variable-length segment. The *fixed-length block deduplication* divides the files into fixed-length blocks and uses a hash algorithm to find the duplicate data. Although simple in design, fixed-length blocks might miss many opportunities to discover redundant data because the block boundary of similar data might be different. Consider the addition of a person's name to a document's title page. This shifts the whole document, and all the blocks appear to have changed, causing the failure of the deduplication method to detect equivalencies. In variable-length segment deduplication, if there is a change in the segment, the boundary for only that segment is adjusted, leaving the remaining segments unchanged. This method vastly improves the ability to find duplicate data segments compared to fixed-block.

## Data Deduplication Implementation – Source-based

- Data is deduplicated at the source (backup client)
- Backup client sends only new, unique segments across the network
- Reduced storage capacity and network bandwidth requirements
- Increased overhead on the backup client



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

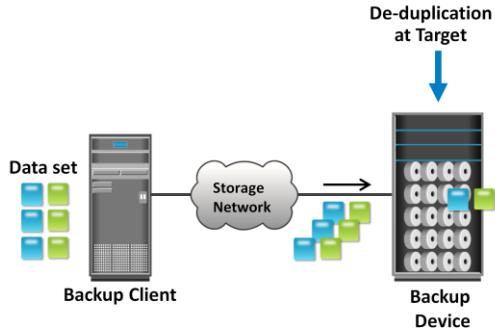
Module 10: Backup and Archive

33

*Source-based data deduplication* eliminates redundant data at the source before it transmits to the backup device. Source-based data deduplication can dramatically reduce the amount of backup data sent over the network during backup processes. It provides the benefits of a shorter backup window and requires less network bandwidth. There is also a substantial reduction in the capacity required to store the backup images. Source-based deduplication increases the overhead on the backup client, which impacts the performance of the backup and application running on the client. Source-based deduplication might also require a change of backup software if it is not supported by backup software.

## Data Deduplication Implementation – Target-based

- Data is deduplicated at the target
  - ▶ Inline
  - ▶ Post-process
- Offloads the backup client from deduplication process
- All the backup data traverse the network



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

34

*Target-based data deduplication* is an alternative to source-based data deduplication. Target-based data deduplication occurs at the backup device, which offloads the backup client from the deduplication process. Figure on the slide illustrates target-based data deduplication. In this case, the backup client sends the data to the backup device and the data is deduplicated at the backup device, either immediately (Inline) or at a scheduled time (Post-process).

*Inline deduplication* performs deduplication on the backup data before it is stored on the backup device. Hence, this method reduces the storage capacity needed for the backup. Inline deduplication introduces overhead in the form of the time required to identify and remove duplication in the data. So, this method is best suited for an environment with a large backup window.

*Post-process deduplication* enables the backup data to be stored on the backup device first and then deduplicated later. This method is suitable for situations with tighter backup windows. However, post-process deduplication requires more storage capacity to store the backup images before they are deduplicated.

Because deduplication occurs at the target, all the backup data need to be transferred over the network, which increases network bandwidth requirements. Target-based data deduplication does not require any changes in the existing backup software.

## Data Deduplication – Key Benefits

- Reduces infrastructure costs
  - ▶ By eliminating redundant data, less storage is required to hold the backup images
- Enables longer retention periods
  - ▶ Reduces the amount of redundant content in the daily backup, and hence, users can extend their retention policies
- Reduces backup window
  - ▶ Less data to be backed up, which reduces backup window
- Reduces backup bandwidth requirement
  - ▶ Source based de-duplication eliminates redundant data before data is sent over the network

**Reduces infrastructure costs:** By eliminating redundant data from the backup, far less infrastructure is required to hold the backup images. Data de-duplication directly results in reduced storage capacities to hold backup images. Smaller capacity requirements means lower acquisition costs as well as reduced power and cooling costs.

**Enables longer retention periods:** As data de-duplication reduces the amount of content in the daily backup, users can extend their retention policies. This can have a significant benefit to users who currently require longer retention.

**Reduces backup window:** Data de-duplication eliminates redundant content of backup data, which makes less data to be backed up and reduces backup window.

**Reduces backup bandwidth requirement:** By utilizing data de-duplication at the client (source-based), redundant data is removed before the data is transferred over the network. This considerably reduces the network bandwidth required for backup.

## Use Case: Remote Office/Branch Office Backup

- Protecting data at an organization's branch and remote offices, across multiple locations, is critical for business
- Backing up data from remote offices to a centralized data center was restricted due to
  - ▶ Time and cost involved in sending huge volumes of data over the network
- Disk-based backup solution, along with source-based deduplication, eliminates the challenges in centrally backing up remote-office data
  - ▶ Reduces the network bandwidth requirement
  - ▶ Reduces the backup window

Today, businesses have their remote or branch offices spread over multiple locations. Typically, these remote offices have their local IT infrastructure. This infrastructure includes file, print, Web, or email servers, workstations, and desktops, and might also house some applications and databases. Too often, business-critical data at remote offices are inadequately protected, exposing the business to the risk of lost data and productivity. As a result, protecting the data of an organization's branch and remote offices across multiple locations is critical for business. Traditionally, remote-office data backup was done manually using tapes, which were transported to offsite locations for DR support. Some of the challenges with this approach were lack of skilled onsite technical resources to manage backups and risk of sending tapes to offsite locations, which could result in loss or theft of sensitive data. Backing up data from remote offices to a centralized data center was restricted due to the time and cost involved in sending huge volumes of data over the WAN. Therefore, organizations needed an effective solution to address the data backup and recovery challenges of remote and branch offices.

Disk-based backup solutions along with source-based deduplication eliminate the challenges associated with centrally backing up remote-office data. Deduplication considerably reduces the required network bandwidth and enables remote-office data backup using the existing network. Organizations can now centrally manage and automate remote-office backups while reducing the required backup window.

## Module 10: Backup and Archive

### Lesson 5: Backup in Virtualized Environment

During this lesson the following topics are covered:

- Traditional backup approach
- Image-based backup

This lesson covers traditional backup approach and image-based backup in virtualized environment.

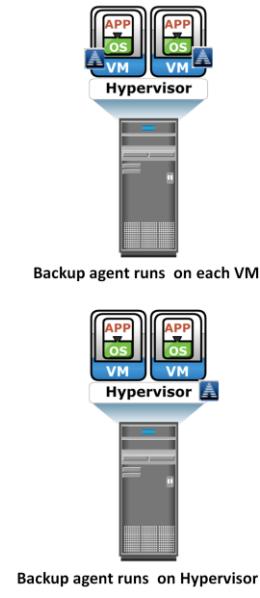
## Backup in Virtualized Environment Overview

- Backup options
  - ▶ Traditional backup approach
  - ▶ Image-based backup approach
- Backup optimization
  - ▶ Deduplication

In a virtualized environment, it is imperative to back up the virtual machine data (OS, application data, and configuration) to prevent its loss or corruption due to human or technical errors. There are two approaches for performing a backup in a virtualized environment: the traditional backup approach and the image-based backup approach. Owing to the increased capacity requirements in a virtualized environment, backup optimization methods are necessary. The use of deduplication techniques significantly reduces the amount of data to be backed up in a virtualized environment. The effectiveness of deduplication is identified when VMs with similar configurations are deployed in a data center. The deduplication types and methods used in a virtualized environment are the same as in the physical environment.

## Traditional Backup Approaches

- Backup agent on VM
  - ▶ Requires installing a backup agent on each VM running on a hypervisor
  - ▶ Can only backup virtual disk data
  - ▶ Does not capture VM files such as VM swap file, configuration file
  - ▶ Challenge in VM restore
- Backup agent on Hypervisor
  - ▶ Requires installing backup agent only on hypervisor
  - ▶ Backs up all the VM files



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

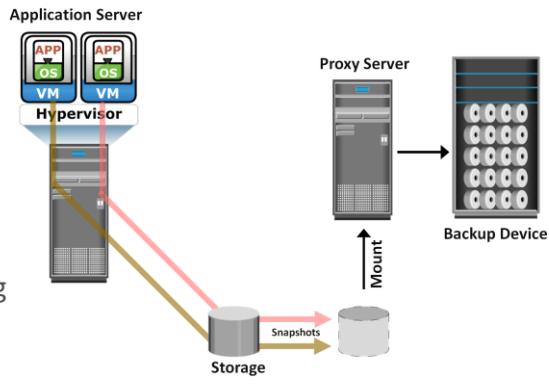
Module 10: Backup and Archive 39

In the *traditional backup approach*, a backup agent is installed either on the virtual machine (VM) or on the hypervisor. If the backup agent is installed on a VM, the VM appears as a physical server to the agent. The backup agent installed on the VM backs up the VM data to the backup device. The agent does not capture VM files, such as the virtual BIOS file, VM swap file, logs, and configuration files. Therefore, for a VM restore, a user needs to manually re-create the VM and then restore data on to it.

If the backup agent is installed on the hypervisor, the VMs appear as a set of files to the agent. So, VM files can be backed up by performing a file system backup from a hypervisor. This approach is relatively simple because it requires having the agent just on the hypervisor instead of having on all the VMs. The traditional backup method can cause high CPU utilization on the server being backed up. In the traditional approach, the backup should be performed when the server resources are idle or during a low activity period on the network. Also consider allocating enough resources to manage the backup on each server when a large number of VMs are in the environment.

## Image-based Backup

- Creates a copy of the guest OS, its data, VM state, and configurations
  - The backup is saved as a single file – “image”
  - Mounts image on a proxy server
  - Offloads backup processing from the hypervisor
- Enables quick restoration of VM



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

40

*Image-based backup* operates at the hypervisor level and essentially takes a snapshot of the VM. It creates a copy of the guest OS and all the data associated with it (snapshot of VM disk files), including the VM state and application configurations. The backup is saved as a single file called an “image” and this image is mounted on the proxy server (acts as a backup client). The backup software then backs up these image files normally. This effectively offloads the backup processing from the hypervisor and transfers the load on the proxy server, thereby reducing the impact to VMs running on the hypervisor. Image-based backup enables quick restoration of a VM.

## Module 10: Backup and Archive

### Lesson 6: Data Archive

During this lesson the following topics are covered:

- Fixed content
- Data archive
- Archive solution architecture

This lesson covers fixed content and challenges in storing fixed content. This lesson also focuses on data archive solution architecture.

## Fixed Content

- Fixed content is growing at more than 90% annually
  - ▶ Significant amount of newly created information falls into this category
  - ▶ New regulations require retention and data protection

Examples of Fixed Content		
<p><b>Electronic Documents</b></p> <ul style="list-style-type: none"><li>• Contracts and claims</li><li>• Email attachments</li><li>• Financial spread sheets</li><li>• CAD/CAM designs</li><li>• Presentations</li></ul>	<p><b>Digital Records</b></p> <ul style="list-style-type: none"><li>• Documents<ul style="list-style-type: none"><li>• Checks, securities trades</li><li>• Historical preservation</li></ul></li><li>• Photographs<ul style="list-style-type: none"><li>• Personal/professional</li></ul></li><li>• Surveys<ul style="list-style-type: none"><li>• Seismic, astronomic, geographic</li></ul></li></ul>	<p><b>Rich Media</b></p> <ul style="list-style-type: none"><li>• Medical<ul style="list-style-type: none"><li>• X-rays, MRIs, CT Scan</li></ul></li><li>• Video<ul style="list-style-type: none"><li>• News/media, movies</li><li>• Security surveillance</li></ul></li><li>• Audio<ul style="list-style-type: none"><li>• Voicemail</li><li>• Radio</li></ul></li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

42

In the life cycle of information, data is actively created, accessed, and changed. As data ages, it is less likely to be changed and eventually becomes “fixed” but continues to be accessed by applications and users. This data is called *fixed content*. All organizations may require to retain their data for an extended period of time due to government regulations and legal/contractual obligations. Organizations also make use of this fixed content to generate new revenue strategies and improve service levels.

Currently, fixed content data is the fastest growing sector of the data storage market. Assets such as X-rays, MRIs, CAD/CAM designs, surveillance video, MP3s and financial documents are just a few examples of an important class of data that is growing at over 90% annually.

## Data Archive

- A repository where fixed content is stored
- Enables organizations retaining their data for an extended period of time in order to
  - ▶ Meet regulatory compliance
  - ▶ Plan new revenue strategies
- Archive can be implemented as
  - ▶ Online
  - ▶ Nearline
  - ▶ Offline

Data archive is a repository where fixed content is stored. It enables organizations retaining their data for an extended period of time in order to meet regulatory compliance and generate new revenue strategies. An archive can be implemented as an online, nearline, or offline solution:

**Online archive:** A storage device directly connected to a host that makes the data immediately accessible.

**Nearline archive:** A storage device connected to a host, but the device where the data is stored must be mounted or loaded to access the data.

**Offline archive:** A storage device that is not ready to use. Manual intervention is required to connect, mount, or load the storage device before data can be accessed.

## Challenges of Traditional Archiving Solutions

- Both tape and optical are susceptible to wear and tear
  - ▶ Involve operational, management, and maintenance overhead
- Have no intelligence to identify duplicate data
  - ▶ Same content could be archived many times
- Inadequate for long-term preservation (years-decades)
- Unable to provide online and fast access to fixed content

Typically, long-term preservation is required (years-decades) for fixed content and it is also important to have simultaneous, fast online access. The increase in fixed content is driven by regulatory requirements. Because of this explosive growth and changes to user requirements, traditional storage options are inadequate.

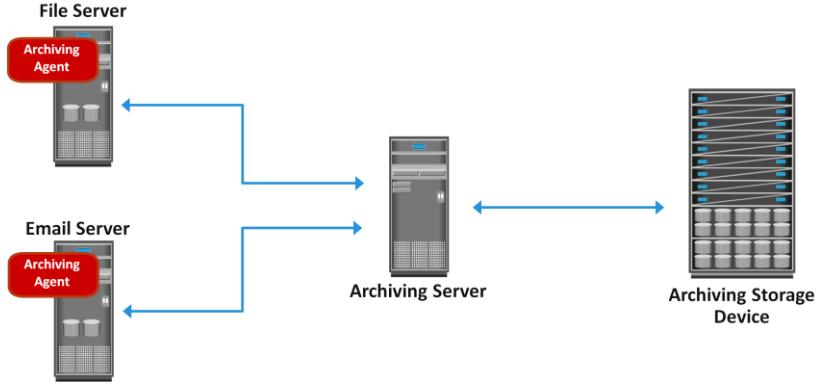
Optical media are typically *write once read many* (WORM) devices that protect the original file from being overwritten. Some tape devices also provide this functionality by implementing file-locking capabilities. Although these devices are inexpensive, they involve operational, management, and maintenance overhead. The traditional solutions using optical discs and tapes is not optimized to recognize the content, so that the same content could be stored several times. Additional costs are involved in offsite storage of media and media management. Tapes and optical media are also susceptible to wear and tear. Frequent changes in these device technologies lead to the overhead of converting the media into new formats to enable access and retrieval. Government agencies and industry regulators are establishing new laws and regulations to enforce the protection of archives from unauthorized destruction and modification. These regulations and standards have established new requirements for preserving the integrity of information in the archives. These requirements have exposed the shortcomings of the traditional tape and optical media archive solutions.

## Content Addressed Storage – An Archival Solution

- Disk-based storage that has emerged as an alternative to traditional archiving solutions
- Provides online accessibility to archive data
- Enables organization to meet the required SLAs
- Provides features that are required for storing archive data
  - ▶ Content authenticity and content integrity
  - ▶ Location independence
  - ▶ Single-instance storage
  - ▶ Retention enforcement
  - ▶ Data protection

Content addressed storage (CAS) a disk based storage that has emerged as an alternative to tape and optical solutions. CAS meets the demand to improve data accessibility and to protect, dispose off, and ensure service level agreements (SLAs) for archive data. CAS is detailed in module 8.

## Archiving Solution Architecture



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 10: Backup and Archive

46

Archiving solution architecture consists of three key components; archiving agent, archiving server, and archiving storage device.

An *archiving agent* is software installed on the application server. The agent is responsible for scanning the data that can be archived based on the policy defined on the archiving server. After the data is identified for archiving, the agent sends the data to the archiving server. Then the original data on the application server is replaced with a stub file. The stub file contains the address of the archived data. The size of this file is small and significantly saves space on primary storage. This stub file is used to retrieve the file from the archive storage device.

An *archiving server* is software installed on a host that enables administrators to configure the policies for archiving data. Policies can be defined based on file size, file type, or creation/modification/access time. The archiving server receives the data to be archived from the agent and sends it to the archive storage device.

An *archive storage device* stores fixed content.

## Use Case: Email Archiving

- Moves the emails from primary to archive storage, based on policy
- Saves space on primary storage
- Enables to retain emails in the archive for longer period to meet regulatory requirements
- Gives end users virtually unlimited mailbox space
- File archiving is another use case that benefits from an archival solution

E-mail is an example of an application that benefits most by an archival solution. Typically, a system administrator configures small mailboxes that store a limited number of e-mails. This is because large mailboxes with a large number of e-mails can make management difficult, increase primary storage cost, and degrade system performance. When an e-mail server is configured with a large number of mailboxes, the system administrator typically configures a quota on each mailbox to limit its size. Configuring fixed quota on mailboxes impacts end users. A fixed quota for a mailbox forces users to delete e-mails as they approach the quota size. End users often need to access e-mails that are weeks, months, or even years old.

E-mail archiving provides an excellent solution that overcomes the preceding challenges. Archiving solutions move e-mails that have been identified as candidates for archive from primary storage to the archive storage device based on a policy—for example, “e-mails that are 90 days old should be archived.” After the e-mail is archived, it is retained for years based on the retention policy. This considerably saves space on primary storage and enables organizations to meet regulatory requirements. Implementation of an archiving solution gives end users virtually unlimited mailbox space.

A file sharing environment is another environment that benefits from an archival solution. Typically, users store a large number of files in the shared location. Most of these files are old and rarely accessed. Administrators configure quotas on the file share that forces the users to delete these files. This impact users because they may require access to files that may be months or even years old. In some cases the user may request an increase in the size of the file share. This in turn increases the cost of primary storage. A file archiving solution archives the files based on the policy such as age of files, size of files, and so on. This considerably reduces the primary storage requirement and also enables users to retain the files in the archive for longer periods.

## Module 10: Backup and Archive

### Concepts in Practice

- EMC NetWorker
- EMC Avamar
- EMC Data Domain

The concept in practice section covers various EMC backup and archive products.

## EMC NetWorker

- Centralizes, automates, and accelerates data backup and recovery operations across the enterprise
- Key features
  - ▶ Supports heterogeneous platforms such as Windows, UNIX, Linux, and also supports virtual environments
  - ▶ Supports different backup targets – tapes, disks, and virtual tapes
  - ▶ Supports Multiplexing (or multi-streaming) of data
  - ▶ Provides both source-based and target-based deduplication capabilities by integrating with EMC Avamar and EMC Data Domain respectively
  - ▶ Cloud-backup option enables backing up data to cloud

The EMC NetWorker backup and recovery software centralizes, automates, and accelerates data backup and recovery operations across the enterprise. The features of EMC NetWorker are listed on the slide.

## EMC Avamar

- Disk-based backup and recovery solution that provides source-based data deduplication
- Three major components include Avamar server, Avamar backup clients, and Avamar administrator
- Avamar server includes
  - ▶ Software only, Avamar Data Store, Avamar Virtual Edition

EMC Avamar is a disk-based backup and recovery solution that provides inherent source-based data deduplication. With its unique global data deduplication feature, Avamar differs from traditional backup and recovery solutions, by identifying and storing only unique sub-file data objects. Redundant data is identified at the source, the amount of data that travels across the network is drastically reduced, and the backup storage requirement is also considerably reduced. The three major components of an Avamar system includes Avamar server, Avamar backup clients, and Avamar administrator. Avamar server provides the essential processes and services required for client access and remote system administration. The Avamar client software runs on each computer or network server that is being backed up. Avamar administrator is a user management console application that is used to remotely administer an Avamar system. The three Avamar server editions include software only, Avamar Data Store, and Avamar Virtual Edition. The features of EMC Avamar are as follows:

- **Fault tolerance:** Uses RAID, RAIN, checkpoints, and replication to provide data integrity and protection.
- **Standard IP network leveraging:** Optimizes the use of network for backup; dedicated backup networks are not required. Daily full backups are possible using the existing networks and infrastructure.
- **Scalable server architecture:** Additional storage nodes can be added non-disruptively to accommodate increased backup storage requirements.
- **Centralized management:** Enables remote management of Avamar servers from a centralized location and through the use of the Avamar Enterprise Manager and Avamar Administrator interfaces.

## EMC Data Domain

- Target-based deduplication solution
- Provides technological advantages
  - ▶ Data invulnerability architecture
  - ▶ Data Domain Stream-Informed Segment Layout (SISL) scaling architecture
  - ▶ Support native replication technology
  - ▶ Global compression
- EMC Data Domain Archiver
  - ▶ Solution for long term retention of backup and archive data
  - ▶ Designed with internal tiering approach
  - ▶ Supports deduplication technology

The EMC Data Domain deduplication storage system is a target-based data deduplication solution. Using high-speed, inline deduplication technology, the Data Domain system provides a storage footprint that is significantly smaller on an average, than that of the original data set. Data Domain systems can scale from smaller remote office appliances to large data-center systems. These systems are available as integrated appliances or as gateways that use external storage.

Data Domain deduplication storage systems provide the following unique advantages:

**Data invulnerability architecture:** Provides unprecedented levels of data integrity, data verification, and self-healing capabilities, such as RAID 6 protection. Continuous fault detection, healing, and write verification ensure that the backup is accurately stored, available, and recoverable.

**Data Domain SISL (Stream-Informed Segment Layout) scaling architecture:** Enables scaling of CPUs to add a direct benefit to system throughput scalability.

**Support native replication technology:** Enables automatic, secure transfer of compressed data over the wide area network (WAN) with minimum bandwidth requirement.

**Global compression:** Highly efficient deduplication and compression technology, which radically changes storage economics.

EMC Data Domain Archiver is a solution for long term retention of backup and archive data. It is designed with internal tiering approach to enable cost effective, long term retention of data on disk by implementing deduplication technology.

## Module 10: Summary

Key points covered in this module:

- Backup granularity
- Backup and recovery operations
- Backup topologies
- Backup targets
- Data deduplication
- Backup in virtualized environment
- Data archive

This module covered various backup granularities and backup operations. This module also discussed various backup topologies and backup targets. Further this module covered data deduplication and backup in virtualized environment. Additionally, this module also covered data archive in detail.

A *backup* is an additional copy of the production data, created and retained for the sole purpose of recovering lost or corrupted data. Based on the granularity, backups can be categorized as full, cumulative, and incremental.

The three basic topologies used in a backup environment are direct-attached backup, LAN-based backup, and SAN-based backup.

A wide range of technology solutions are currently available for backup targets. Tape and disk libraries are the two most commonly used backup targets. Virtual tape library (VTL) is one of the options that uses disks as backup medium. VTL emulates tapes and provides enhanced backup and recovery capabilities.

Data deduplication is the process of identifying and eliminating redundant data. When duplicate data is detected during backup, the data is discarded and only the pointer is created to refer the copy of the data which is already backed up.

In a virtualized environment, it is imperative to back up the virtual machine data (OS, application data, and configuration) to prevent its loss or corruption due to human or technical errors.

Data archive is a repository where fixed content is stored. It enables organizations to retain their data for an extended period of time, in order to meet regulatory compliance and generate new revenue strategies.

## Check Your Knowledge – 1

- Which is true about incremental backup?
  - A. Restore requires only last full and last incremental backup
  - B. Restore requires only last incremental backup
  - C. Copies the data that has changed since last full or incremental backup
  - D. Copies the data that has changed since last full backup
- What is an advantage of image-based backup over traditional backup approach in a virtualized environment?
  - A. Offloads backup processing from the hypervisor
  - B. Faster because it copies only virtual machine disk data
  - C. Faster because it copies only virtual machine configuration data
  - D. Space required is a fraction of total backup data

## Check Your Knowledge – 2

- Which accurately describes the role of a backup server?
  - A. Gathers the data that is to be backed up and send it to storage node
  - B. Responsible for writing the data, which client sends, to backup device
  - C. Manages the backup operation and maintains backup catalog
  - D. Controls the robotic arm in the tape library
- In backup to tape environment, what does ‘shoe shining’ mean?
  - A. Writing data from multiple streams on a single tape
  - B. Process of emulating disk drives and presenting as tapes to backup software
  - C. Repeated back and forth motion that a tape drive makes when there is an interruption in the backup data stream
  - D. Process of deleting redundant content in the backup data

## Check Your Knowledge – 3

- What is an advantage of source-based deduplication?
  - A. Improves the performance of backup client
  - B. Improves the performance of backup server
  - C. Reduces backup window to zero
  - D. Reduces the network bandwidth requirement for backup

## Exercise: Backup/Recovery

- Current situation
  - ▶ Full backup is performed on every Sunday and incremental on remaining days
  - ▶ Database have to be shut down during backup
  - ▶ Multiple redundant copies of backup data
  - ▶ Network bandwidth constraint
- Business requirement
  - ▶ Eliminate the need to shutdown the database for backup
  - ▶ Need faster backup and restore
  - ▶ Eliminate redundant copies of backup data
- Task
  - ▶ Suggest a solution and justify

### **Business profile:**

An organization uses tape as their primary backup storage media for their applications.

### **Current situation:**

- Full backup is performed on every Sunday and incremental on remaining days
- Database have to be shut down during backup
- Multiple redundant copies of backup data
- Network bandwidth constraint

### **Requirements:**

- Eliminate the need to shutdown the database for backup
- Need faster backup and restore
- Eliminate redundant copies of backup data

### **Task:**

Propose a solution to address the organization's concern and justify your solution.

# Module – 11

# Local Replication



## Module 11: Local Replication

Upon completion of this module, you should be able to:

- Describe various uses of local replica
- Describe how consistency is ensured in file system and database replication
- Describe host-based, array-based, and network-based local replication technologies
- Explain restore and restart considerations
- Describe local replication in virtualized environment

This module focuses on various uses of local replica along with file system and database consistency in replication. This module also focuses on various host-based, storage-based, and network-based local replication technologies. Further, this module focuses on restore and restart considerations. Additionally, this module details local replication in a virtualized environment.

# Module 11: Local Replication

## Lesson 1: Local Replication Overview

During this lesson the following topics are covered:

- Uses of local replica
- File system and database consistency

This lesson covers various uses of local replica and also covers how consistency is ensured in file system and database replication.

## What is Replication?

### Replication

It is a process of creating an exact copy (replica) of data.

- Replication can be classified as
  - ▶ Local replication
    - ▶ Replicating data within the same array or data center
  - ▶ Remote replication
    - ▶ Replicating data at remote site



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

4

In today's business environment, it is imperative for an organization to protect mission-critical data and minimize the risk of business disruption. If a local outage or disaster occurs, fast data restore and restart is essential to ensure business continuity (BC). Replication is one of the ways to ensure BC. It is the process to create an exact copy (replica) of data. These replica copies are used for restore and restart operations if data loss occurs. These replicas can also be assigned to other hosts to perform various business operations, such as backup, reporting, and testing.

Replication can be classified into two major categories: local and remote. Local replication refers to replicating data within the same array or the same data center. Remote replication refers to replicating data at a remote site. Remote replication is discussed in the next module.

## Uses of Local Replica

- Alternate source for backup
- Fast recovery
- Decision support activities
- Testing platform
- Data Migration

One or more local replicas of the source data may be created for various purposes, including the following:

**Alternative source for backup:** Under normal backup operations, data is read from the production volumes (LUNs) and written to the backup device. This places an additional burden on the production infrastructure because production LUNs are simultaneously involved in production operations and servicing data for backup operations. The local replica contains an exact point-in-time (PIT) copy of the source data, and therefore can be used as a source to perform backup operations. This alleviates the backup I/O workload on the production volumes. Another benefit of using local replicas for backup is that it reduces the *backup window* to zero.

**Fast recovery:** If data loss or data corruption occurs on the source, a local replica might be used to recover the lost or corrupted data. If a complete failure of the source occurs, some replication solutions enable replica to be used to restore data on to a different set of source devices or production can be restarted on the replica. In either case, this method provides faster recovery and minimal RTO compared to traditional recovery from tape backups.

Cont..

**Decision-support activities, such as reporting or data warehousing:** Running the reports using the data on the replicas greatly reduces the I/O burden placed on the production device. Local replicas are also used for data-warehousing applications. The data-warehouse application may be populated by the data on the replica and thus avoid the impact on the production environment.

**Testing platform:** Local replicas are also used for testing new applications or upgrades. For example, an organization may use the replica to test the production application upgrade; if the test is successful, the upgrade may be implemented on the production environment.

**Data migration:** Another use for a local replica is data migration. Data migrations are performed for various reasons, such as migrating from a smaller capacity LUN to one of a larger capacity for newer versions of the application.

## Replica Characteristics

- Recoverability/Restartability
  - ▶ Replica should be able to restore data on the source device
  - ▶ Restart business operation from replica
- Consistency
  - ▶ Replica must be consistent with the source
- Choice of replica tie back into RPO
  - ▶ Point-in-Time (PIT)
    - ▶ Non-zero RPO
  - ▶ Continuous
    - ▶ Near-zero RPO

A replica should have following characteristics:

**Recoverability:** Enables restoration of data from the replicas to the source if data loss or corruption occurs.

**Restartability:** Enables restarting business operations using the replicas.

**Consistency:** Replica must be consistent with the source so that it is usable for both recovery and restart operations. Ensuring consistency is the primary requirement for all the replication technologies.

Replicas can either be point-in-time (PIT) or continuous:

**Point-in-Time:** The data on the replica is an identical image of the production at some specific timestamp. For example, a replica of a file system is created at 4:00 PM on Monday. This replica would then be referred to as the Monday 4:00 PM PIT copy. The RPO will map to the time when the PIT was created to the time when any kind of failure on the production occurred. If there is a failure on the production at 8:00 PM and there is a 4:00 PM PIT available, the RPO would be 4 hours ( $8 - 4 = 4$ ). To minimize RPO, take periodic PITs.

**Continuous replica:** The data on the replica is in-sync with the production data at all times. The objective with any continuous replication is to reduce the RPO to zero or near-zero.

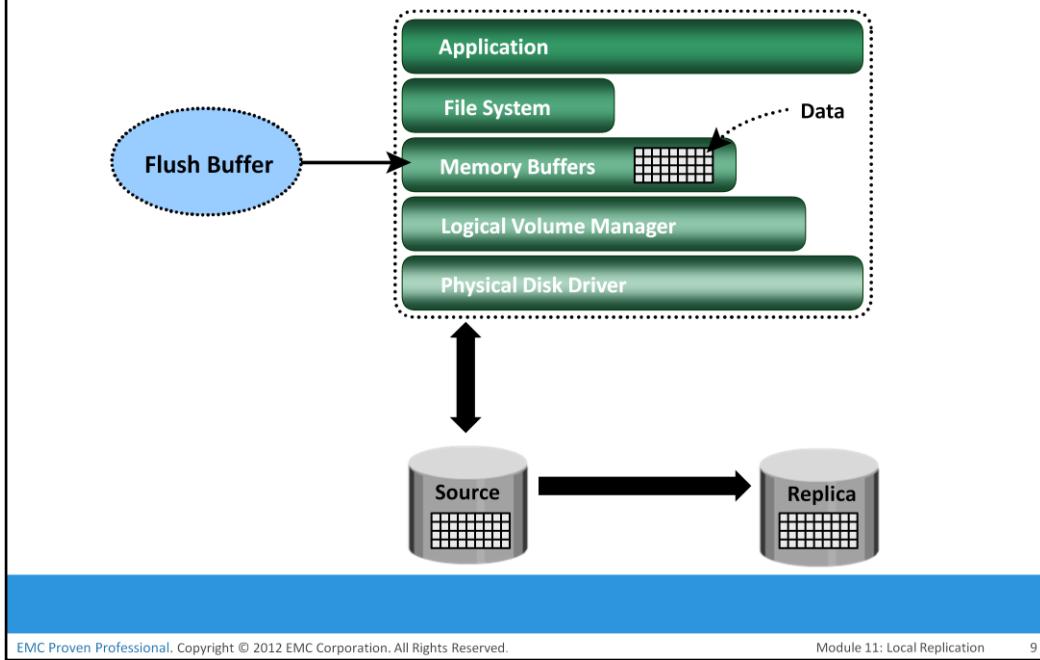
## Understanding Consistency

- Consistency ensures the usability of replica
- Consistency can be achieved in various ways for file system and database

	Offline	Online
File System	Unmount file system	Flushing host buffers
Database	Shutdown database	a) Using dependent write I/O principle b) Holding I/Os to source before creating replica

Consistency is a primary requirement to ensure the usability of replica device. In case of file systems, consistency can be achieved either by taking FS offline i.e. by un-mounting FS or by keeping FS online by flushing host buffers before creating replica. Similarly in case of databases, consistency can be achieved either by taking database offline for creating consistent replica or by keeping online. Consistent replica of online database can be created by using dependent write I/O principle or by holding I/Os before creating replica.

## File System Consistency: Flushing Host Buffer



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

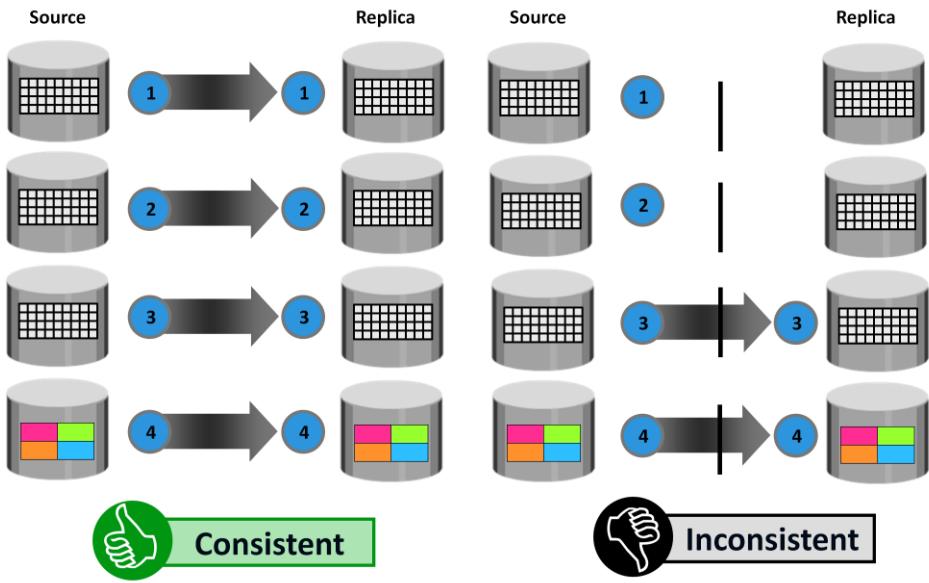
Module 11: Local Replication

9

File systems buffer the data in the host memory to improve the application response time. The buffered data is periodically written to the disk. In UNIX operating systems, *sync daemon* is the process that flushes the buffers to the disk at set intervals. In some cases, the replica is created between the set intervals, which might result in the creation of an inconsistent replica. Therefore, host memory buffers must be flushed to ensure data consistency on the replica, prior to its creation. If the host memory buffers are not flushed, the data on the replica will not contain the information that was buffered in the host. If the file system is unmounted before creating the replica, the buffers will be automatically flushed and the data will be consistent on the replica.

If a mounted file system is replicated, some level of recovery, such as *fsck* or *log replay*, is required on the replicated file system. When the file system replication and check process are completed, the replica file system can be mounted for operational use.

## Database Consistency: Dependent Write I/O Principle



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

10

If the database is online, it is available for I/O operations, and transactions to the database update the data continuously. When a database is replicated while it is online, changes made to the database at this time must be applied to the replica to make it consistent. A consistent replica of an online database is created by using the dependent write I/O principle or by holding I/Os momentarily to the source before creating the replica.

A *dependent write I/O* principle is inherent in many applications and database management systems (DBMS) to ensure consistency. According to this principle, a write I/O is not issued by an application until a prior related write I/O has completed. For example, a data write is dependent on the successful completion of the prior log write.

For a transaction to be deemed complete, databases require a series of writes to have occurred in a particular order. These writes will be recorded on the various devices/file systems.

When the replica is created, all the writes to the source devices must be captured on the replica devices to ensure data consistency. Figure on the slide illustrates the process of replication from the source to the replica. The I/O transactions 1 to 4 must be carried out for the data to be consistent on the replica. It is possible that I/O transactions 3 and 4 were copied to the replica devices, but I/O transactions 1 and 2 were not copied. Figure on the slide also shows this situation. In this case, the data on the replica is inconsistent with the data on the source. If a restart were to be performed on the replica devices, I/O 4, which is available on the replica, might indicate that a particular transaction is complete, but all the data associated with the transaction will be unavailable on the replica, making the replica inconsistent.

Another way to ensure consistency is to make sure that the write I/O to all source devices is held for the duration of creating the replica. This creates a consistent image on the replica. However, databases and applications might time out if the I/O is held for too long.

## Module 11: Local Replication

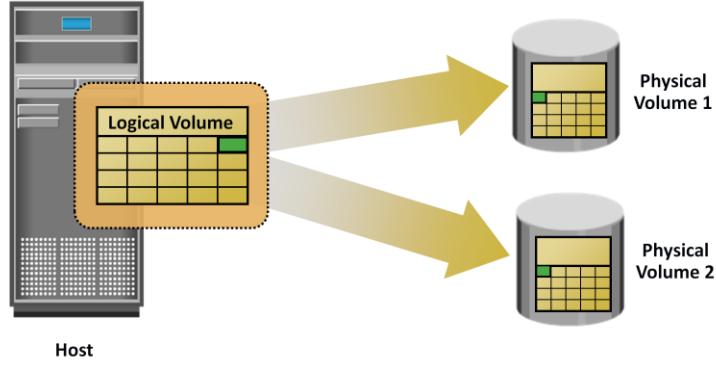
### Lesson 2: Local Replication Technologies

During this lesson the following topics are covered:

- Local replication technologies
- Restore and restart considerations

This lesson covers various host-based, network-based and storage array-based local replication technologies and also covers restore and restart considerations.

## Host-based Replication: LVM-based Mirroring



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

12

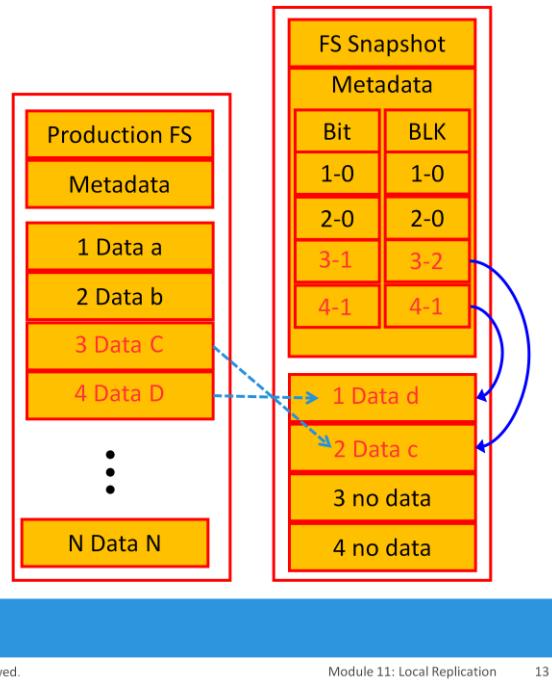
In *LVM-based replication*, the logical volume manager is responsible for creating and controlling the host-level logical volumes. An LVM has three components: physical volumes (physical disk), volume groups, and logical volumes. A *volume group* is created by grouping one or more physical volumes. *Logical volumes* are created within a given volume group. A volume group can have multiple logical volumes. In LVM-based replication, each *logical block* in a logical volume is mapped to two physical blocks on two different physical volumes, as shown in figure on the slide. An application write to a logical volume is written to the two physical volumes by the LVM device driver. This is also known as *LVM mirroring*. Mirrors can be split, and the data contained therein can be independently accessed.

**Advantages:** The LVM-based replication technology is not dependent on a vendor-specific storage system. Typically, LVM is part of the operating system, and no additional license is required to deploy LVM mirroring.

**Limitations:** Every write generated by an application translates into two writes on the disk, and thus, an additional burden is placed on the host CPU. This can degrade application performance. Presenting an LVM-based local replica to another host is usually not possible because the replica will still be part of the volume group, which is usually accessed by one host at any given time. If the devices are already protected by some level of RAID on the array, then the additional protection that the LVM mirroring provides is unnecessary. This solution does not scale to provide replicas of federated databases and applications. Both the replica and source are stored within the same volume group. Therefore, the replica might become unavailable if there is an error in the volume group. If the server fails, both the source and replica are unavailable until the server is brought back online.

## Host-based Replication: File System Snapshot

- Pointer-based replication
- Uses Copy on First Write (CoFW) principle
- Uses bitmap and block map
- Requires a fraction of the space used by the production FS



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

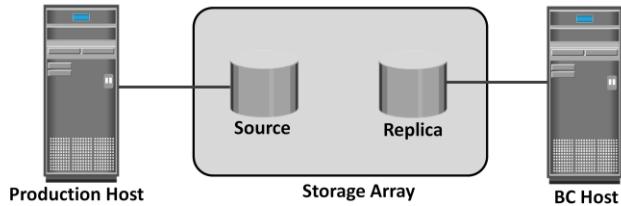
Module 11: Local Replication

13

File system (FS) snapshot is a pointer-based replica that requires a fraction of the space used by the production FS. It uses the Copy on First Write (CoFW) principle to create snapshots. When a snapshot is created, a bitmap and blockmap are created in the metadata of the Snap FS. The bitmap is used to keep track of blocks that are changed on the production FS after the snap creation. The blockmap is used to indicate the exact address from which the data is to be read when the data is accessed from the Snap FS. Immediately after the creation of the FS Snapshot, all reads from the snapshot are actually served by reading the production FS. In a CoFW mechanism, if a write I/O is issued to the production FS for the first time after the creation of a snapshot, the I/O is held and the original data of production FS corresponding to that location is moved to the Snap FS. Then, the write is allowed to the production FS. The bitmap and blockmap are updated accordingly. Subsequent writes to the same location will not initiate the CoFW activity. To read from the Snap FS, the bitmap is consulted. If the bit is 0, then the read is directed to the production FS. If the bit is 1, then the block address is obtained from the blockmap and the data is read from that address on the snap FS. Read requests from the production FS work as normal.

## Storage Array-based Local Replication

- Replication performed by the array operating environment
- Source and replica are on the same array
- Types of array-based replication
  - ▶ Full-volume mirroring
  - ▶ Pointer-based full-volume replication
  - ▶ Pointer-based virtual replication



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

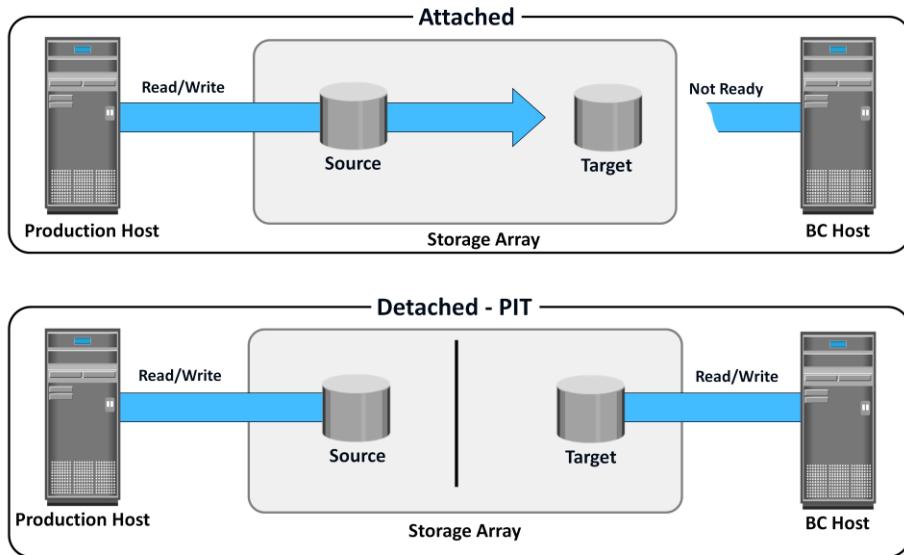
Module 11: Local Replication

14

In *storage array-based local replication*, the array-operating environment performs the local replication process. The host resources, such as the CPU and memory are not used in the replication process. Consequently, the host is not burdened by the replication operations. The replica can be accessed by an alternate host for other business operations.

In this replication, the required number of replica devices should be selected on the same array and then data should be replicated between the source-replica pairs. Figure on the slide shows a storage array-based local replication, where the source and target (replica) are in the same array and accessed by different hosts. Storage array-based local replication is commonly implemented in three ways full-volume mirroring, pointer-based full-volume replication, and pointer-based virtual replication.

## Full-Volume Mirroring



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

15

In *full-volume mirroring*, the target is attached to the source and established as a mirror of the source. The data on the source is copied to the target. New updates to the source are also updated on the target. After all the data is copied and both the source and the target contain identical data, the target can be considered as a mirror of the source. While the target is attached to the source it remains unavailable to any other host. However, the production host continues to access the source.

After the synchronization is complete, the target can be detached from the source and made available for other business operations. Both the source and the target can be accessed for read and write operations by the production and business continuity hosts respectively. After detaching from the source, the target becomes a point-in-time (PIT) copy of the source. The PIT of a replica is determined by the time when the target is detached from the source. For example, if the time of detachment is 4:00 pm., the PIT for the target is 4:00 pm. After detachment, changes made to both the source and replica can be tracked at some predefined granularity. This enables incremental resynchronization (source to target) or incremental restore (target to source). The granularity of the data change can range from 512 byte blocks to 64 KB blocks or higher.

## Pointer-based Full-Volume Replication

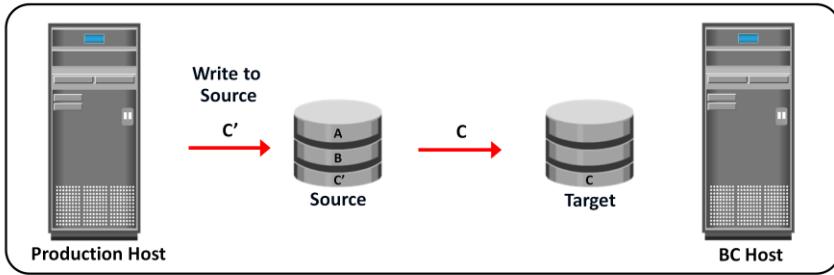
- Provides full copy of source data on the target
- Target device is immediately accessible by the BC host after the replication session is activated
- PIT is determined by time of session activation
- Target device is at least as large as the source device
- Two modes
  - ▶ Full copy mode
    - ▶ After session starts, all the data from source is copied to the target in the background
  - ▶ Copy on First Access (deferred)

Another method of array-based local replication is *pointer-based full-volume replication*. Similar to full-volume mirroring, this technology can provide full copies of the source data on the targets. Unlike full-volume mirroring, the target is immediately accessible by the BC host after the replication session is activated. Therefore, data synchronization and detachment of the target is not required to access it. Here, the time of replication session activation defines the PIT copy of the source.

Pointer-based, full-volume replication can be activated in either Copy on First Access (CoFA) mode or Full Copy mode. In either case, at the time of activation, a protection bitmap is created for all data on the source devices. The protection bitmap keeps track of the changes at the source device. The pointers on the target are initialized to map the corresponding data blocks on the source. The data is then copied from the source to the target based on the mode of activation.

In a Full Copy mode, all data from the source is copied to the target in the background. Data is copied regardless of access. If access to a block that has not yet been copied to the target is required, this block is preferentially copied to the target. In a complete cycle of the Full Copy mode, all data from the source is copied to the target. If the replication session is terminated now, the target contains all the original data from the source at the point-in-time of activation. This makes the target a viable copy for restore or other business continuity operations.

## Copy on First Access: Write to the Source



- When a write is issued to the source for the first time after replication session activation:
  - Original data at that address is copied to the target
  - Then the new data is updated on the source
  - This ensures that original data at the point-in-time of activation is preserved on the target

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

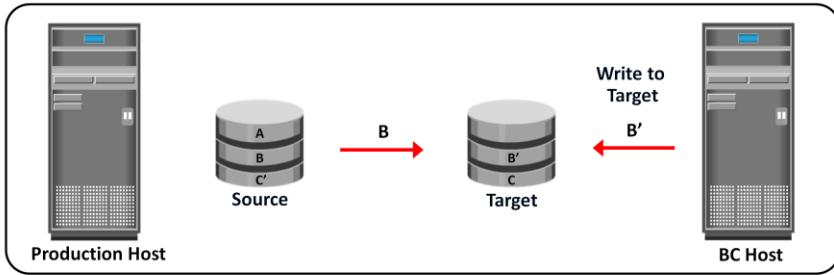
17

In CoFA, after the replication session is initiated, the data is copied from the source to the target only when the following condition occurs:

- A write I/O is issued to a specific address on the source for the first time.
- A read or write I/O is issued to a specific address on the target for the first time.

When a write is issued to the source for the first time after replication session activation, the original data at that address is copied to the target. After this operation, the new data is updated on the source. This ensures that the original data at the point-in-time of activation is preserved on the target.

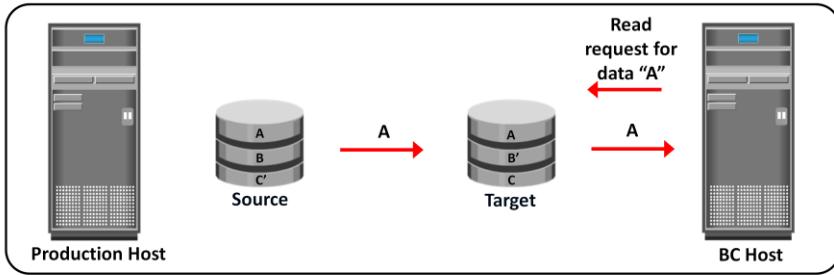
## Copy on First Access: Write to the Target



- When a write is issued to the target for the first time after replication session activation:
  - The original data is copied from the source to the target
  - Then the new data is updated on the target

When a write is issued to the target for the first time after the replication session activation, the original data is copied from the source to the target. After this, the new data is updated on the target (see figure on the slide).

## Copy on First Access: Read from Target



- When a read is issued to the target for the first time after replication session activation:
  - The original data is copied from the source to the target and is made available to the BC host

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

19

When a read is issued to the target for the first time after replication session activation, the original data is copied from the source to the target and is made available to the BC host.

In all cases, the protection bit for the data block on the source is reset to indicate that the original data has been copied over to the target. The pointer to the source data can now be discarded. Subsequent writes to the same data block on the source, and the reads or writes to the same data blocks on the target, do not trigger a copy operation, therefore this method is termed “Copy on First Access.”

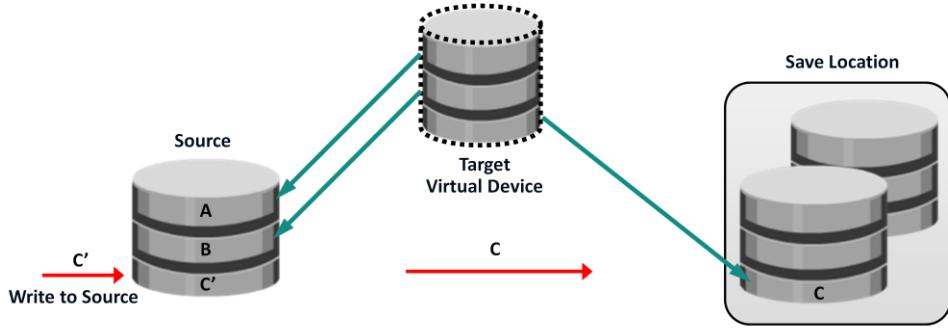
If the replication session is terminated, then the target device has only the data that was accessed until the termination, not the entire contents of the source at the point-in-time. In this case, the data on the target cannot be used for restore because it is not a full replica of the source.

## Pointer-based Virtual Replication

- Targets do not hold data, but hold pointers to where the data is located
  - ▶ At the start of the session the target device holds pointers to data on source device
  - ▶ Target requires a small fraction of the size of the source volumes
- Target devices are accessible immediately when the session is started
- Uses CoFW principle
- This method is recommended, if the changes to the source are typically less than 30%

In *pointer-based virtual replication*, at the time of the replication session activation, the target contains pointers to the location of the data on the source. The target does not contain data at any time. Therefore, the target is known as a *virtual replica*. Similar to pointer-based full-volume replication, the target is immediately accessible after the replication session activation. This replication method uses CoFW technology and typically recommended when the changes to the source are less than 30%.

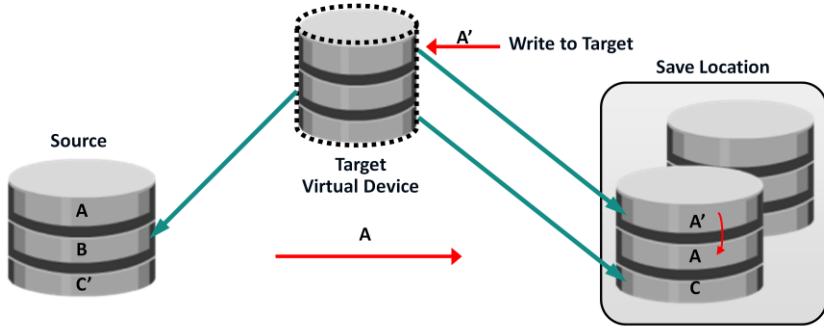
## Pointer-based Virtual Replication (CoFW): Write to Source



- When a write is issued to the source for the first time after replication session activation:
  - ▶ Original data at that address is copied to save location
  - ▶ The pointer in the target is updated to point to this data in the save location
  - ▶ Finally, the new write is updated on the source

When a write is issued to the source for the first time after the replication session activation, the original data at that address is copied to a predefined area in the array. This area is generally known as the *save location*. The pointer in the target is updated to point to this data in the save location. After this, the new write is updated on the source.

## Pointer-based Virtual Replication (CoFW): Write to Target



- When a write is issued to the target for the first time after replication session activation:
  - Original data from the source device is copied to the save location
  - The pointer is updated to the data in save location
  - Another copy of the original data is created in the save location before the new write is updated on the save location

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

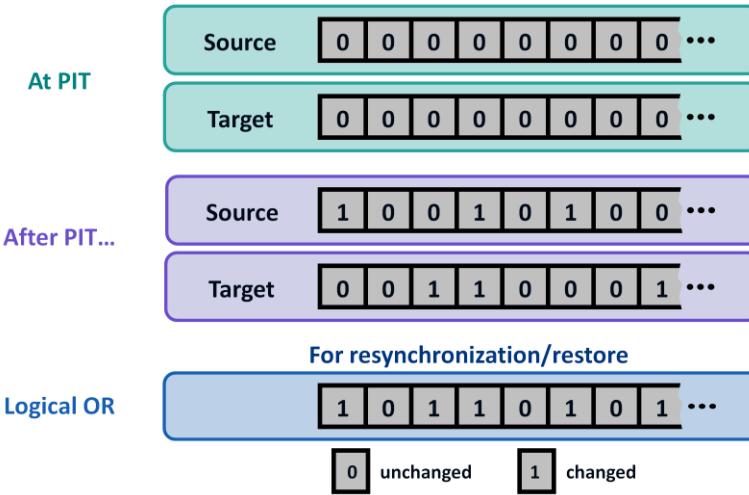
22

When a write is issued to the target for the first time after replication session activation, the data is copied from the source to the save location, and the pointer is updated to the data in the save location. Another copy of the original data is created in the save location before the new write is updated on the save location. Subsequent writes to the same data block on the source or target do not trigger a copy operation.

When reads are issued to the target, unchanged data blocks since the session activation are read from the source, whereas data blocks that have changed are read from the save location.

Data on the target is a combined view of unchanged data on the source and data on the save location. Unavailability of the source device invalidates the data on the target. The target contains only pointers to the data, and therefore, the physical capacity required for the target is a fraction of the source device. The capacity required for the save location depends on the amount of the expected data change.

## Tracking Changes to Source and Target



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

23

Updates can occur on the source device after the creation of point-in-time local replicas. If the primary purpose of local replication is to have a viable point-in-time copy for data recovery or restore operations, then the replica devices should not be modified. Changes can occur on the replica device if it is used for other business operations. To enable incremental resynchronization or restore operations, changes to both the source and replica devices after the point-in-time should be tracked. This is typically done using bitmaps, where each bit represents a block of data. The data block sizes can range from 512 bytes to 64 KB or greater. For example, if the block size is 32 KB, then a 1 GB device would require 32,768 bits (1GB divided by 32KB). The size of the bitmap would be 4 KB. If the data in any 32 KB block is changed, the corresponding bit in the bitmap is flagged. If the block size is reduced for tracking purposes, then the bitmap size increases correspondingly.

The bits in the source and target bitmaps are all set to 0 (zero) when the replica is created. Any changes to the source or replica are then flagged by setting the appropriate bits to 1 in the bitmap. When resynchronization or restore is required, a *logical OR* operation between the source bitmap and the target bitmap is performed. The bitmap resulting from this operation references all blocks that have been modified in either the source or replica. This enables an optimized resynchronization or a restore operation, because it eliminates the need to copy all the blocks between the source and the replica. The direction of data movement depends on whether a resynchronization or a restore operation is performed. If resynchronization is required, changes to the replica are overwritten with the corresponding blocks from the source. If a restore is required, changes to the source are overwritten with the corresponding blocks from the replica. In either case, changes to both the source and the target cannot be simultaneously preserved.

## Restore and Restart Considerations

- Source has a failure
  - ▶ Logical corruption or physical failure of source devices
- Solution
  - ▶ Restore data from target to source
    - ▶ Restore would typically be done incrementally
    - ▶ Applications can be restarted even before synchronization is complete

-----OR-----

- ▶ Start production on target
  - ▶ Create a “Gold” copy of target device before restarting on target
  - ▶ Resolve issues with source while continuing operations on target
  - ▶ After resolving the issue, restore latest data on target to source

Local replicas are used to restore data to production devices. Alternatively, applications can be restarted using the consistent point-in-time replicas.

Replicas are used to restore data to the production devices if logical corruption of data on production devices occurs—that is, the devices are available but the data on them is invalid. Examples of logical corruption include accidental deletion of data (tables or entries in a database), incorrect data entry, and incorrect data updates. Restore operations from a replica are incremental and provide a small RTO. In some instances, the applications can be resumed on the production devices prior to the completion of the data copy. Prior to the restore operation, access to production and replica devices should be stopped.

Production devices might also become unavailable due to physical failures, such as production server or physical drive failure. In this case, applications can be restarted using the data on the latest replica. As a protection against further failures, a “Gold Copy” (another copy of replica device) of the replica device should be created to preserve a copy of data in the event of failure or corruption of the replica devices. After the issue has been resolved, the data from the replica devices can be restored back to the production devices.

Full-volume replicas (both full-volume mirrors and pointer-based in Full Copy mode) can be restored to the original source devices or to a new set of source devices. Restores to the original source devices can be incremental, but restores to a new set of devices are full-volume copy operations.

In pointer-based virtual and pointer-based full-volume replication in CoFA mode, access to data on the replica is dependent on the health and accessibility of the source volumes. If the source volume is inaccessible for any reason, these replicas cannot be used for a restore or a restart operation.

## Comparison of Local Replication Technologies

Factor	Full-Volume Mirroring	Pointer-based Full-Volume Replication	Pointer-based Virtual Replication
Performance impact on source due to replica	No impact	Full copy mode – no impact CoFA mode – some impact	High impact
Size of target	At least the same as the source	At least the same as the source	Small fraction of the source
Availability of source for restoration	Not required	Full copy mode – not required CoFA mode – required	Required
Accessibility to target	Only after synchronization and detachment from the source	Immediately accessible	Immediately accessible

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

25

This table summarizes the comparison between full-volume mirroring, pointer-based full-volume, and pointer-based virtual replication technologies.

*Note:*

Most array based replication technologies allow the source devices to maintain replication relationships with multiple targets.

- More frequent replicas will reduce the RPO.
- Each PIT could be used for a different BC activity and also as restore points.

## Network-based Local Replication: Continuous Data Protection

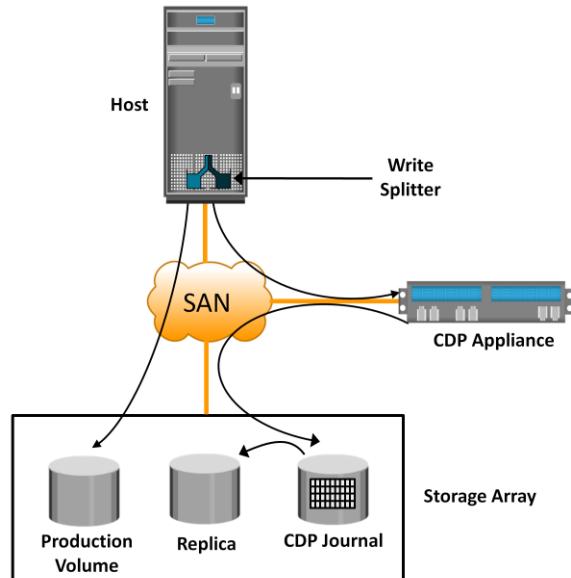
- Replication occurs at the network layer between the hosts and storage arrays
  - ▶ Ideal for highly heterogeneous environment
- Typically provides the ability to restore data to any previous point-in-time
  - ▶ RPOs are random and do not need to be defined in advance
- Data changes are continuously captured and stored in a separate location from the production data
- CDP is implemented by using
  - ▶ Journal volume
  - ▶ CDP appliance
  - ▶ Write splitter

In network-based replication, the replication occurs at the network layer between the hosts and storage arrays. Network-based replication combines the benefits of array-based and host-based replications. By offloading replication from servers and arrays, network-based replication can work across a large number of server platforms and storage arrays, making it ideal for highly heterogeneous environments. *Continuous data protection* (CDP) is a technology used for network-based local and remote replications. CDP for remote replication is detailed in module 12.

In a data center environment, mission-critical applications often require instant and unlimited data recovery points. Traditional data protection technologies offer limited recovery points. If data loss occurs, the system can be rolled back only to the last available recovery point. Mirroring offers continuous replication; however, if logical corruption occurs to the production data, the error might propagate to the mirror, which makes the replica unusable. In normal operation, CDP provides the ability to restore data to any previous PIT. It enables this capability by tracking all the changes to the production devices and maintaining consistent point-in-time images.

In CDP, data changes are continuously captured and stored in a separate location from the primary storage. Moreover, RPOs are random and do not need to be defined in advance. With CDP, recovery from data corruption poses no problem because it allows going back to a PIT image prior to the data corruption incident. CDP uses a *journal volume* to store all data changes on the primary storage. The journal volume contains all the data that has changed from the time the replication session started. The amount of space that is configured for the journal determines how far back the recovery points can go. CDP also uses *CDP appliance* and *write splitter*. CDP implementation may also be host-based in which CDP software is installed on a separate host machine. CDP appliance is an intelligent hardware platform that runs the CDP software and manages local and remote data replications. Write splitters intercept writes to the production volume from the host and split each write into two copies. Write splitting can be performed at the host, fabric, or storage array.

## CDP Local Replication Operation



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

27

Figure on the slide describes CDP local replication. In this method, before the start of replication, the replica is synchronized with the source and then the replication process starts. After the replication starts, all the writes to the source are split into two copies. One of the copies is sent to the CDP appliance and the other to the production volume. When the CDP appliance receives a copy of a write, it is written to the journal volume along with its timestamp. As a next step, data from the journal volume is sent to the replica at predefined intervals.

While recovering data to the source, the CDP appliance restores the data from the replica and applies journal entries up to the point in time chosen for recovery.

## Module 11: Local Replication

### Lesson 3: Local Replication in Virtualized Environment

During this lesson the following topics are covered:

- Mirroring of a virtual volume
- Replication of virtual machines

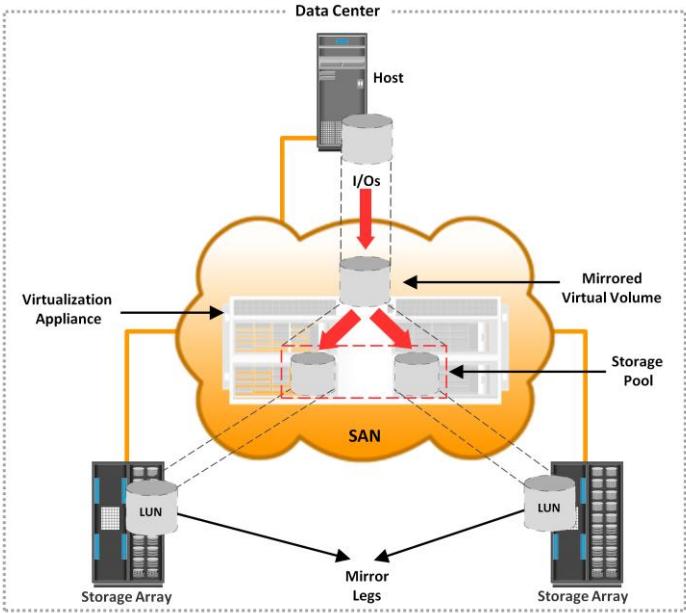
This lesson covers mirroring of a virtual volume and also replication of virtual machines.

## Local Replication in Virtualized Environment

- Local replication (mirroring) of a virtual volume assigned to a host
  - ▶ Mirroring is performed by a virtualization appliance
- Replication of virtual machines
  - ▶ VM snapshot
  - ▶ VM clone

The discussion so far has focused on local replication in a physical infrastructure environment. In a virtualized environment, along with replicating storage volumes, replication of virtual machine (VM) is performed by the hypervisor. For hypervisor-based local replication, two options are available - VM Snapshot, and VM Clone. However, in a virtualized environment, the data in the virtual volumes can also be replicated with the help of a virtualization layer in a SAN.

## Local Replication of Virtual Volume



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

30

In a virtualized environment, the virtualization appliance at the SAN abstracts the LUNs and creates virtual volumes. These virtual volumes are presented to the host. The virtualization appliance has the ability to mirror the data of a virtual volume between the LUNs. Figure on the slide provides an illustration of a virtual volume that is mirrored between arrays within a data center. Each I/O to the virtual volume is mirrored to the underlying LUNs on the arrays. If one of the arrays incurs an outage, the virtualization appliance will be able to continue processing I/O on the surviving mirror leg. Upon restoration of the failed storage array, the data from the surviving LUN is resynchronized to the recovered leg.

## VM Snapshot

- Captures the state and data of a running VM at a specific PIT
- Uses a separate delta file to record all the changes to the virtual disk since the snapshot session is activated
- Restores all settings configured in a guest OS to the PIT

VM Snapshot captures the state and data of a running virtual machine at a specific point in time. The VM state includes VM files, such as BIOS, network configuration, and its power state (powered-on, powered-off, or suspended). The VM data includes all the files that make up the VM, including virtual disks and memory. A VM Snapshot uses a separate delta file to record all the changes to the virtual disk since the snapshot session is activated. Snapshots are useful when a VM needs to be reverted to the previous state in the event of logical corruptions. Reverting a VM to a previous state causes all settings configured in the guest OS to be reverted to that PIT when that snapshot was created. There are some challenges associated with the VM Snapshot technology. It does not support data replication if a virtual machine accesses the data by using raw disks. Also, using the hypervisor to perform snapshots increases the load on the compute and impacts the compute performance.

## VM Clone

- An identical copy of an existing VM
  - ▶ Clones are created for different use such as testing
  - ▶ Changes made to a clone VM do not affect the parent VM and vice versa
- Clone VM is assigned a separate network identity
  - ▶ Clone has its own separate MAC address
- Useful when multiple identical VMs need to deploy

VM Clone is another method that creates an identical copy of a virtual machine. When the cloning operation is complete, the clone becomes a separate VM from its parent VM. The clone has its own MAC address, and changes made to a clone do not affect the parent VM. Similarly, changes made to the parent VM do not appear in the clone. VM Clone is a useful method when there is a need to deploy many identical VMs. Installing guest OS and applications on multiple VMs is a time-consuming task; VM Clone helps to simplify this process.

## Module 11: Local Replication

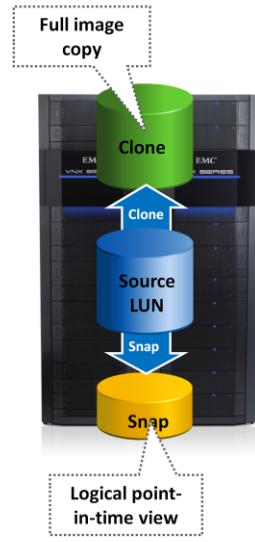
### Concept in Practice

- EMC SnapView
- EMC TimeFinder
- EMC RecoverPoint

The concepts in practice section covers various EMC local replication products such as EMC SnapView, EMC TimeFinder, and EMC RecoverPoint.

## EMC SnapView

- SnapView Snapshot
  - ▶ Logical view of the production volume
  - ▶ Uses CoFW principle
- SnapView Clone
  - ▶ Full volume copies that require same disk space as the source
  - ▶ Becomes a PIT copy once the clone is fractured from the source



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 11: Local Replication

34

SnapView is an EMC VNX array-based local replication software that creates pointer-based virtual copy and full-volume mirror of the source using SnapView snapshot and SnapView clone respectively.

### SnapView Snapshot

A SnapView snapshot is not a full copy of the production volume; it is a logical view of the production volume based on the time at which the snapshot was created. Snapshots are created in seconds, and can be retired when no longer needed. A snapshot “roll back” feature provides instant restore to the source volume. The key terminologies of SnapView snapshot are as follows:

*SnapView session:* The SnapView snapshot mechanism is activated when a session starts and deactivated when a session stops. A snapshot appears “offline” until there is an active session. Multiple snapshots can be included in a session.

*Reserved LUN pool:* This is a private area, also called a save area, used to contain CoFW data. The “Reserved” part of the name refers to the fact that the LUNs are reserved and therefore cannot be assigned to a host.

### SnapView Clone

SnapView Clones are full-volume copies that require the same disk space as the source. These PIT copies can be used for other business operations, such as backup and testing. SnapView Clone enables incremental resynchronization between the source and replica. Clone fracture is the process of breaking off a clone from its source. After the clone is fractured, it becomes a PIT copy and available for other business operations.

## EMC TimeFinder

- TimeFinder/Snap
  - ▶ Creates space-saving, logical PIT (snapshots)
  - ▶ Allows creating multiple snapshots from a single source
- TimeFinder/Clone
  - ▶ Creates PIT copy of the source volume
  - ▶ Uses pointer-based full-volume replication technology
  - ▶ Allows creating multiple clones from a single source device

The TimeFinder family of products consists of two base solutions and four add-on solutions. The base solutions are TimeFinder/Clone and TimeFinder/Snap. The add-on solutions are TimeFinder/Clone Emulation, TimeFinder/Consistency Groups, TimeFinder/Exchange Integration Module, and TimeFinder/SQL Integration Module. TimeFinder is available for both open systems and mainframes. The base solutions support the different storage array-based local replication technologies discussed in this module. The add-on solutions are customizations of the replicas for specific application or database environments.

### TimeFinder/Snap

TimeFinder/Snap creates space-saving, logical PIT images called snapshots. The snapshots are not full copies, but contain pointers to the source data. The target device used by TimeFinder/Snap is called a virtual device (VDEV). It keeps pointers to the source device or SAVE devices. The SAVE devices keep the point-in-time data that has changed on the source after the start of the replication session. TimeFinder/Snap allows creating multiple snapshots from a single source device.

### TimeFinder/Clone

TimeFinder/Clone creates a PIT copy of the source volume that can be used for backups, decision support, or any other process that requires parallel access to production data. TimeFinder/Clone uses pointer-based full-volume replication technology. TimeFinder/Clone allows creating multiple clones from a single production device, and all the clones are available immediately for read and write access.

## EMC RecoverPoint

- Provides continuous data protection and recovery to any PIT
- Uses splitting technology at server, fabric, or array to mirror a write to a RecoverPoint appliance
- Provides automatic RecoverPoint appliance failover
- Family of product includes
  - ▶ RecoverPoint/CL
  - ▶ RecoverPoint/EX
  - ▶ RecoverPoint/SE

RecoverPoint is a high-performance, cost-effective, single product that provides local and remote data protection for both physical and virtual environments. It provides faster recovery and unlimited recovery points. RecoverPoint provides continuous data protection and performs replication between the LUNs. RecoverPoint uses lightweight splitting technology either at the application server, fabric, or arrays to mirror a write to a RecoverPoint appliance. The RecoverPoint family of products include RecoverPoint/CL, RecoverPoint/EX, and RecoverPoint/SE.

*RecoverPoint/CL* is a replication product for a heterogeneous server and storage environment. It supports both EMC and non-EMC storage arrays. This product supports host-based, fabric-based, and array-based write splitters. *RecoverPoint/EX* supports replication between EMC storage arrays and allows only array-based write splitting.

*RecoverPoint/SE* is a version of RecoverPoint that is targeted for VNX series arrays and enables only Windows-based host and array-based write splitting.

## Module 11: Summary

Key points covered in this module:

- Uses of local replicas
- Consistency in file system and database replication
- Host-based, storage array-based, and network-based replication
- Restore and restart considerations
- Local replication of a virtual volume
- VM snapshot and VM clone

This module covered various uses of local replica along with file system and database consistency in replication. This module also covered various host-based, storage-based, and network-based local replication technologies. Further, this module focused restore and restart considerations for local replica. Finally, this module detailed local replication in virtualized environment.

Local replicas are used for several purposes such as alternate source for backup, fast recovery, decision support, testing platform, and data migration.

Replica must be consistent with the source so that it is usable for both recovery and restart operations. Ensuring consistency is the primary requirement for all the replication technologies.

Host-based, storage array-based, and network-based replications are the major technologies used for local replication. File system replication and LVM-based replication are examples of host-based local replication. Storage array-based replication can be implemented with distinct solutions, namely, full-volume mirroring, pointer-based full-volume replication, and pointer-based virtual replication. Continuous Data Protection (CDP) is an example of network-based replication.

In a virtualized environment, virtual machine (VM) replication is a critical requirement for successful BC process. Typically, local replication of VMs (VM snapshot and VM clone) is performed by the hypervisor at the compute level. However, it can also be performed at the storage level using array-based local replication, similar to the physical environment.

## Check Your Knowledge – 1

- What is an advantage of pointer-based virtual replication?
  - A. Source device need not be healthy for restore
  - B. Save location is not required to activate session
  - C. Less storage space is required for creating replica
  - D. No performance impact on source due to replication
- Which is true about pointer-based full-volume replication?
  - A. Size of the target is a small fraction of the source
  - B. Size of the target is at least the same as the source
  - C. Target can be accessed only after synchronization and detachment from source
  - D. Target contains only pointers to save location at all time

## Check Your Knowledge – 2

- In continuous data protection technology, which factor determines how far back the recovery points can go?
  - A. Amount of space that is configured for the replica
  - B. Type of write splitter used in replication
  - C. Amount of space that is configured for the journal
  - D. Rate of changes happening to the replica
- When is the data copied from source to target in CoFA mode of replication?
  - A. A read occurs for the first time from a location on the source
  - B. A read or write occurs for the first time to a location on the source
  - C. A read or write occurs for the first time to a location on the target
  - D. All writes issued to a location on the target

## Check Your Knowledge – 3

- What occurs to the guest OS configuration when a VM is reverted from its snapshot?
  - A. Guest OS configurations are reverted to the PIT of snapshot creation
  - B. Current guest OS configurations are preserved
  - C. Guest OS configurations are duplicated to a new VM
  - D. Guest OS settings are lost and need manual configuration

## Exercise: Local Replication

- Scenario
  - ▶ Organization's mission critical data is stored on RAID 1 volumes
  - ▶ Database application uses 1TB storage
  - ▶ Average data that changes in 24 hours is 60 GB
- Requirements
  - ▶ Need solution to address logical corruption of database
  - ▶ Maximum RPO of 1 hour
  - ▶ Solution should support restore request for up to 8 hours old data
  - ▶ Minimize the amount of storage used for data protection
- Task
  - ▶ Suggest an appropriate local replication solution to meet RPO requirement with minimum amount of storage
  - ▶ Estimate the physical storage required by this solution

### **Scenario:**

A manufacturing organization stores data of their mission critical applications on a high end storage array with RAID 1 configuration. The database application has 1 TB of storage and needs 24x7 availability. Average data that changes in 24 hours is 60 GB.

### **Requirements:**

Need solution to address logical corruption of database

Maximum RPO of 1 hour

Solution should support restore request for up to 8 hours old data

Minimize the amount of storage used for data protection

### **Task:**

Suggest an appropriate local replication solution to meet RPO requirement with minimum amount of storage. Estimate the physical storage required by this solution.

This slide intentionally left blank.

# Module – 12

# Remote Replication



## Module 12: Remote Replication

Upon completion of this module, you should be able to:

- Explain synchronous and asynchronous replication mode
- Describe host-based, array-based, and network-based remote replication technologies
- Describe three-site remote replication
- Explain data migration solution
- Describe remote replication and migration in virtualized environment

This module focuses on synchronous and asynchronous remote replication mode. This module also focuses on host-based, array-based, and network-based remote replication technologies. Further, this module details three-site remote replication and data migration solution. Additionally, this module details remote replication and migration in virtualized environment.

# Module 12: Remote Replication

## Lesson 1: Remote Replication Overview

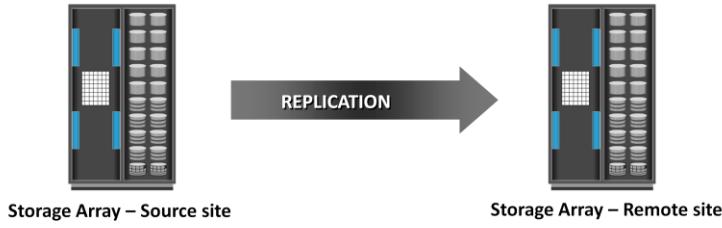
During this lesson the following topics are covered:

- Synchronous and asynchronous remote replication
- Bandwidth requirement for synchronous and asynchronous remote replication

This lesson covers synchronous and asynchronous remote replication mode. This lesson also covers on bandwidth requirement for both synchronous and asynchronous remote replication.

## What is Remote Replication?

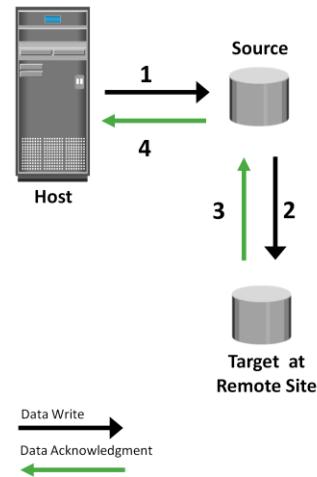
- Process of creating replicas at remote sites
  - ▶ Addresses risk associated with regionally driven outages
- Modes of remote replication
  - ▶ Synchronous
  - ▶ Asynchronous



*Remote replication* is the process to create replicas of information assets at remote sites (locations). Remote replication helps organizations mitigate the risks associated with regionally driven outages resulting from natural or human-made disasters. During disasters, the workload can be moved to a remote site to ensure continuous business operation. Similar to local replicas, remote replicas can also be used for other business operations. Two modes of remote replications are synchronous and asynchronous replication.

## Synchronous Replication – 1

- A write is committed to both source and remote replica before it is acknowledged to the host
- Ensures source and replica have identical data at all times
  - ▶ Maintains write ordering
- Provides near-zero RPO



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

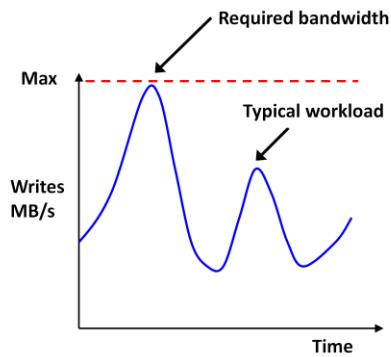
Module 12: Remote Replication

5

In *synchronous remote replication*, writes must be committed to the source and remote replica (or target), prior to acknowledging “write complete” to the host. Additional writes on the source cannot occur until each preceding write has been completed and acknowledged. This ensures that data is identical on the source and replica at all times. Further, writes are transmitted to the remote site exactly in the order in which they are received at the source. Therefore, write ordering is maintained. If a source-site failure occurs, synchronous remote replication provides zero or near-zero recovery-point objective (RPO).

## Synchronous Replication – 2

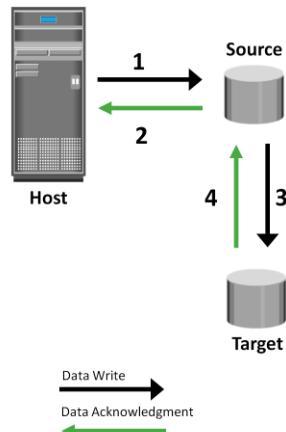
- Response time depends on bandwidth and distance
- Requires bandwidth more than the maximum write workload
- Typically deployed for distance less than 200 km (125 miles) between two sites



Application response time is increased with synchronous remote replication because writes must be committed on both the source and target before sending the “write complete” acknowledgment to the host. The degree of impact on response time depends primarily on the distance between sites, bandwidth and quality of service (QOS) of, the network connectivity infrastructure. If the bandwidth provided for synchronous remote replication is less than the maximum write workload, there will be times during the day when the response time might be excessively elongated, causing applications to time out. The distances over which synchronous replication can be deployed depend on the application’s capability to tolerate extensions in response time. Typically, it is deployed for distances less than 200 KM (125 miles) between the two sites.

## Asynchronous Replication – 1

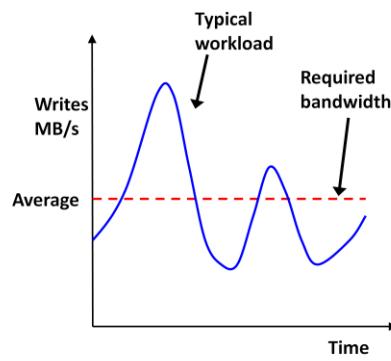
- A write is committed to the source and immediately acknowledged to the host
- Data is buffered at the source and transmitted to the remote site later
- Finite RPO
  - ▶ Replica will be behind the source by a finite amount



In *asynchronous remote replication*, a write is committed to the source and immediately acknowledged to the host. In this mode, data is buffered at the source and transmitted to the remote site later. Data at the remote site will be behind the source by at least the size of the buffer. Hence, asynchronous remote replication provides a finite (nonzero) RPO disaster recovery solution.

## Asynchronous Replication – 2

- RPO depends on size of buffer and available network bandwidth
- Requires bandwidth equal to or greater than average write workload
- Sufficient buffer capacity should be provisioned
- Can be deployed over long distances



Asynchronous replication eliminates the impact to the application's response time because the writes are acknowledged immediately to the source host. This enables deployment of asynchronous replication over distances ranging from several hundred to several thousand kilometers between the primary and remote sites. In this case, the required bandwidth can be provisioned equal to or greater than the average write workload. Data can be buffered during times when the bandwidth is not enough and moved later to the remote site.

Therefore, sufficient buffer capacity should be provisioned. RPO depends on the size of the buffer, the available network bandwidth, and the write workload to the source.

Asynchronous replication implementation can take advantage of *locality of reference* (repeated writes to the same location). If the same location is written multiple times in the buffer prior to transmission to the remote site, only the final version of the data is transmitted. This feature conserves link bandwidth.

**Note:** In both synchronous and asynchronous modes of replication, only writes to the source are replicated; reads are still served from the source.

## Module 12: Remote Replication

### Lesson 2: Remote Replication Technologies

During this lesson the following topics are covered:

- Host-based, storage array-based, and network-based remote replication technologies
- Three-site remote replication
- Data migration solution
- Remote replication in virtualized environment

This lesson focuses on host-based, array-based, and network-based remote replication technologies. Further this lesson focuses on three-site remote replication and data migration solution. Additionally this lesson also details remote replication and migration in virtualized environment.

## Host-based Remote Replication

- Replication is performed by host-based software
- LVM-based replication
  - ▶ All writes to the source volume group are replicated to the target volume group by the LVM
  - ▶ Can be synchronous or asynchronous
- Log shipping
  - ▶ Commonly used in a database environment
  - ▶ All relevant components of source and target databases are synchronized prior to the start of replication
  - ▶ Transactions to source database are captured in logs and periodically transferred to remote host

Host-based remote replication uses the host resources to perform and manage the replication operation. There are two basic approaches to host-based remote replication: Logical volume manager (LVM) based replication and database replication via log shipping.

*LVM-based remote replication* is performed and managed at the volume group level. Writes to the source volumes are transmitted to the remote host by the LVM. The LVM on the remote host receives the writes and commits them to the remote volume group. Prior to the start of replication, identical volume groups, logical volumes, and file systems are created at the source and target sites. Initial synchronization of data between the source and replica is performed. One method to perform initial synchronization is to backup the source data and restore the data to the remote replica. Alternatively, it can be performed by replicating over the IP network. Until the completion of the initial synchronization, production work on the source volumes is typically halted. After the initial synchronization, production work can be started on the source volumes and replication of data can be performed over an existing standard IP network. LVM-based remote replication supports both synchronous and asynchronous modes of replication. If a failure occurs at the source site, applications can be restarted on the remote host, using the data on the remote replicas.

Cont...

LVM-based remote replication is independent of the storage arrays and therefore supports replication between heterogeneous storage arrays. Most operating systems are shipped with LVMs, so additional licenses and specialized hardware are not typically required. The replication process adds overhead on the host CPUs. CPU resources on the source host are shared between replication tasks and applications. This might cause performance degradation to the applications running on the host.

Database replication via log shipping is a host-based replication technology supported by most databases. Transactions to the source database are captured in logs, which are periodically transmitted by the source host to the remote host. The remote host receives the logs and applies them to the remote database. Prior to starting production work and replication of log files, all relevant components of the source database are replicated to the remote site. This is done while the source database is shut down.

After this step, production work is started on the source database. The remote database is started in a standby mode. Typically, in standby mode, the database is not available for transactions. All DBMSs switch log files at preconfigured time intervals or when a log file is full. The current log file is closed at the time of log switching, and a new log file is opened. When a log switch occurs, the closed log file is transmitted by the source host to the remote host. The remote host receives the log and updates the standby database. This process ensures that the standby database is consistent up to the last committed log. RPO at the remote site is finite and depends on the size of the log and the frequency of log switching. Available network bandwidth, latency, rate of updates to the source database, and the frequency of log switching should be considered when determining the optimal size of the log file.

Similar to LVM-based remote replication, the existing standard IP network can be used for replicating log files. Host-based log shipping requires low network bandwidth because it transmits only the log files at regular intervals.

## Storage Array-based Remote Replication – 1

- Replication is performed by array-operating environment
- Three replication methods: synchronous, asynchronous, and disk buffered
- Synchronous
  - ▶ Writes are committed to both source and replica before it is acknowledged to host
- Asynchronous
  - ▶ Writes are committed to source and immediately acknowledged to host
  - ▶ Data is buffered at source and transmitted to remote site later

In *storage array-based remote replication*, the array-operating environment and resources perform and manage data replication. This relieves the burden on the host CPUs, which can be better used for applications running on the host. A source and its replica device reside on different storage arrays. Data can be transmitted from the source storage array to the target storage array over a shared or a dedicated network. Replication between arrays may be performed in synchronous, asynchronous, or disk-buffered modes.

In array-based synchronous remote replication, writes must be committed to the source and the target prior to acknowledging “write complete” to the production host. Additional writes on that source cannot occur until each preceding write has been completed and acknowledged. To optimize the replication process and to minimize the impact on application response time, the write is placed on cache of the two arrays. The storage arrays destage these writes to the appropriate disks later. If the network links fail, replication is suspended; however, production work can continue uninterrupted on the source storage array. The array operating environment keeps track of the writes that are not transmitted to the remote storage array. When the network links are restored, the accumulated data is transmitted to the remote storage array. During the time of network link outage, if there is a failure at the source site, some data will be lost, and the RPO at the target will not be zero.

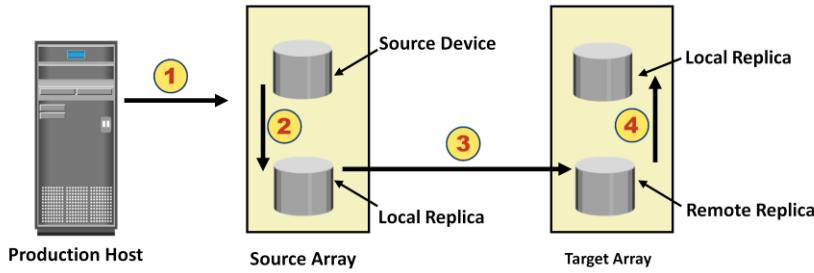
Cont...

In *asynchronous remote replication mode*, a write is committed to the source and immediately acknowledged to the host. Data is buffered at the source and transmitted to the remote site later. The source and the target devices do not contain identical data at all times. The data on the target device is behind that of the source, so the RPO in this case is not zero. Similar to synchronous replication, asynchronous replication writes are placed in cache on the two arrays and are later destaged to the appropriate disks. Some implementations of asynchronous remote replication maintain write ordering. A timestamp and sequence number are attached to each write when it is received by the source. Writes are then transmitted to the remote array, where they are committed to the remote replica in the exact order in which they were buffered at the source. This implicitly guarantees consistency of data on the remote replicas.

In asynchronous remote replication, the writes are buffered for a predefined period of time. At the end of this duration, the buffer is closed, and a new buffer is opened for subsequent writes. All writes in the closed buffer are transmitted together and committed to the remote replica. Asynchronous remote replication provides network bandwidth cost-savings because the required bandwidth is lower than the peak write workload. During times when the write workload exceeds the average bandwidth, sufficient buffer space must be configured on the source storage array to hold these writes.

## Storage Array-based Remote Replication – 2

- Disk-buffered



- ① Production host writes data to source device.
- ② A consistent PIT local replica of the source device is created.
- ③ Data from local replica is transmitted to the remote replica at target.
- ④ Optionally a PIT local replica of the remote replica on the target is created.

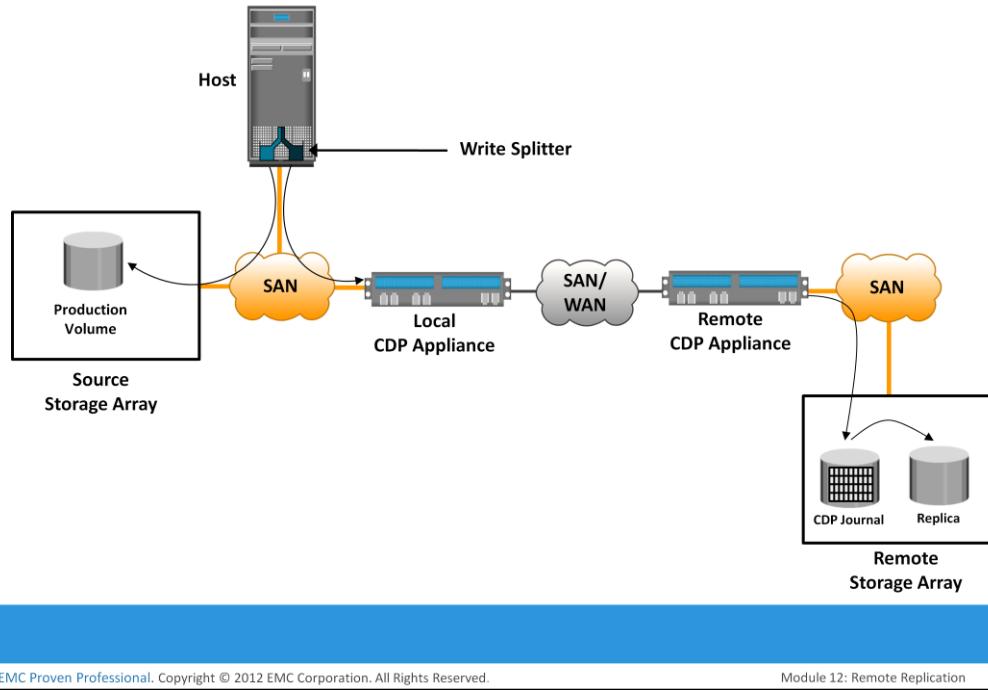
*Disk-buffered replication* is a combination of local and remote replication technologies. A consistent PIT local replica of the source device is first created. This is then replicated to a remote replica on the target array. Figure on the slide shows the sequence of operations in a disk-buffered remote replication. At the beginning of the cycle, the network links between the two arrays are suspended, and there is no transmission of data. While production application runs on the source device, a consistent PIT local replica of the source device is created. The network links are enabled, and data on the local replica in the source array transmits to its remote replica in the target array. After synchronization of this pair, the network link is suspended, and the next local replica of the source is created. Optionally, a local PIT replica of the remote device on the target array can be created. The frequency of this cycle of operations depends on the available link bandwidth and the data change rate on the source device. Because disk-buffered technology uses local replication, changes made to the source and its replica are possible to track. Therefore, all the resynchronization operations between the source and target can be done incrementally. When compared to synchronous and asynchronous replications, disk-buffered remote replication requires less bandwidth.

## Network-based Replication – Continuous Data Protection

- Provides any-point-in-time recovery capability during its normal operation
- Components of CDP
  - ▶ CDP appliance
  - ▶ Write splitter
  - ▶ Journal volume
- CDP appliances are present at both source and remote sites
- Supports both synchronous and asynchronous replication modes

In network-based remote replication, the replication occurs at the network layer between the host and storage array. Continuous data protection technology, discussed in the previous module, also provides solutions for network-based remote replication. In normal operation CDP remote replication, provides any-point-in-time recovery capability, which enables the target LUNs to be rolled back to any previous point in time. Similar to CDP local replication, CDP remote replication typically uses *journal volume*, *CDP appliance* or CDP software installed on a separate host (*host based CDP*), and *write splitter* to perform replication between sites. The CDP appliance is maintained at both source and remote sites.

## CDP Remote Replication Operation



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

16

Figure on the slide describes CDP remote replication. In this method, the replica is synchronized with the source, and then the replication process starts. After the replication starts, all the writes from the host to the source are split into two copies. One of the copies is sent to the local CDP appliance at the source site, and the other copy is sent to the production volume. After receiving the write, the appliance at the source site sends it to the appliance at the remote site. Then, the write is applied to the journal volume at the remote site. For an asynchronous operation, writes at the source CDP appliance are accumulated, and redundant blocks are eliminated. Then, the writes are sequenced and stored with their corresponding timestamp. The data is then compressed, and a checksum is generated. It is then scheduled for delivery across the IP or FC network to the remote CDP appliance. After the data is received, the remote appliance verifies the checksum to ensure the integrity of the data. The data is then uncompressed and written to the remote journal volume. As a next step, data from the journal volume is sent to the replica at predefined intervals.

In the asynchronous mode, the local CDP appliance instantly acknowledges a write as soon as it is received. In the synchronous replication mode, the host application waits for an acknowledgment from the CDP appliance at the remote site before initiating the next write. The synchronous replication mode impacts the application's performance under heavy write loads.

## Three-site Replication

- Data from source site is replicated to two remote sites
  - ▶ Replication is synchronous to one of the remote sites and asynchronous or disk buffered to the other remote site
- Mitigates the risk in two site replication
  - ▶ No DR protection after source or remote site failure
- Implemented in two ways:
  - ▶ Cascade/multihop
  - ▶ Triangle/multitarget

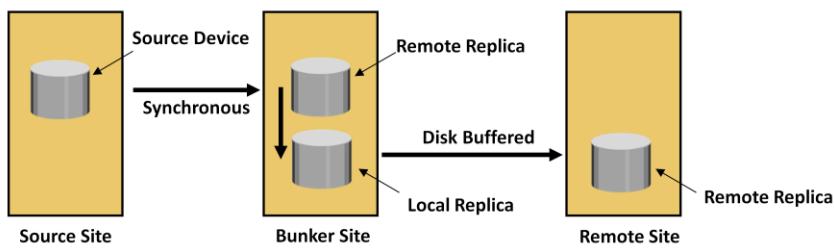
In synchronous replication, the source and target sites are usually within a short distance. Therefore, if a regional disaster occurs, both the source and the target sites might become unavailable. This can lead to extended RPO and RTO because the last known good copy of data would need to come from another source, such as an offsite tape library.

A regional disaster will not affect the target site in asynchronous replication because the sites are typically several hundred or several thousand kilometers apart. If the source site fails, production can be shifted to the target site, but there is no further remote protection of data until the failure is resolved.

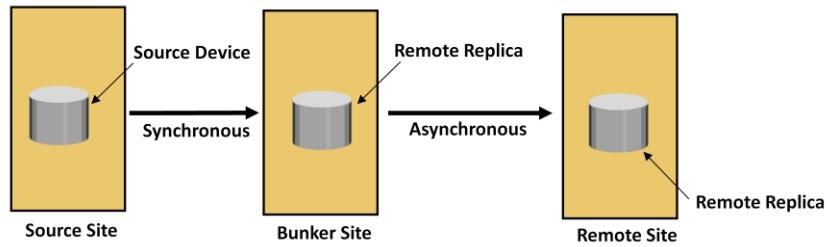
*Three-site replication* mitigates the risks identified in two-site replication. In a three-site replication, data from the source site is replicated to two remote sites. Replication can be synchronous to one of the two sites, providing a near zero-RPO solution, and it can be asynchronous or disk buffered to the other remote site, providing a finite RPO. Three-site remote replication can be implemented as a cascade/multihop or a triangle/multitarget solution.

## Three-site Replication: Cascade/Multihop

- Synchronous + Disk Buffered



- Synchronous + Asynchronous



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

18

In the *cascade/multihop* three-site replication, data flows from the source to the intermediate storage array, known as a *bunker*, in the first hop, and then, from a bunker to a storage array at a remote site in the second hop. Replication can be performed in two ways, synchronous+ disk-buffered or synchronous+asynchronous.

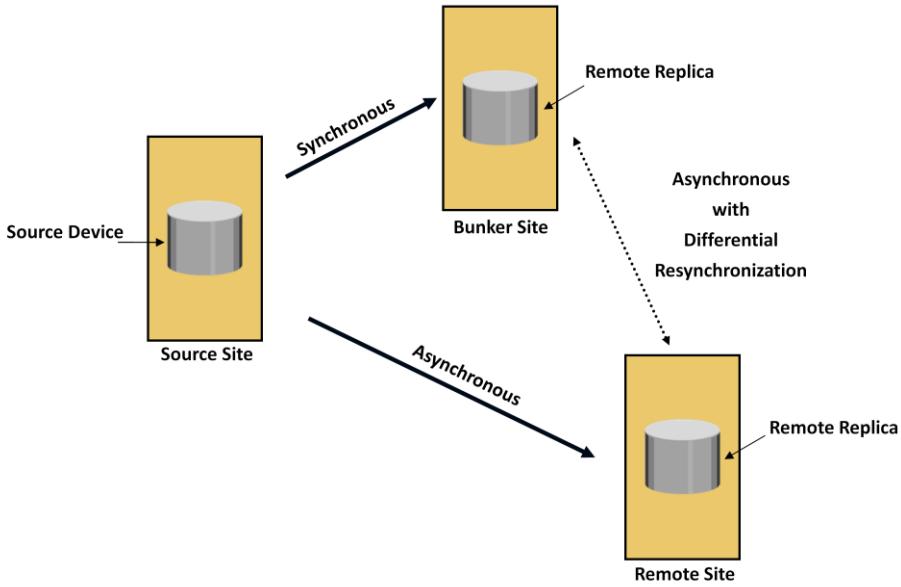
**Synchronous + Disk Buffered:** This method employs a combination of local and remote replication technologies. Synchronous replication occurs between the source and the bunker; a consistent PIT local replica is created at the bunker. Data is transmitted from the local replica at the bunker to the remote replica at the remote site. Optionally, a local replica can be created at the remote site after data is received from the bunker. In this method, a minimum of 4 storage volumes are required (including the source) to replicate one storage device. RPO at the remote site is usually in the order of hours for this implementation.

Cont..

**Synchronous + Asynchronous:** Synchronous replication occurs between the source and the bunker. Asynchronous replication occurs between the bunker and the remote site. The replica in the bunker acts as the source for asynchronous replication to create a remote replica at the remote site. RPO at the remote site is usually in the order of minutes. In this method, a minimum of 3 storage volumes are required (including the source). If a disaster occurs at the source, production operations are failed over to the bunker site with zero or near-zero data loss. If there is a disaster at the bunker site or if there is a network link failure between the source and bunker sites, the source site will continue to operate as normal but without any remote replication. This situation is very similar to remote site failure in a two-site replication solution. The updates to the remote site cannot occur due to the failure in the bunker site. Therefore, the data at the remote site keeps falling behind, but the advantage here is that if the source fails during this time, operations can be resumed at the remote site. RPO at the remote site depends on the time difference between the bunker site failure and source site failure.

A *regional disaster* in cascade/multihop replication is similar to a source site failure in two-site asynchronous replication. Operations are failover to the remote site with an RPO in the order of minutes. There is no remote protection until the regional disaster is resolved. Local replication technologies could be used at the remote site during this time. If a disaster occurs at the remote site, or if the network links between the bunker and the remote site fail, the source site continues to work as normal with disaster recovery protection provided at the bunker site.

## Three-site Replication: Triangle/Multitarget



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

20

In *three-site triangle/multitarget replication*, data at the source storage array is concurrently replicated to two different arrays at two different sites, as shown in the figure on the slide. The source-to-bunker site (target 1) replication is synchronous with a near-zero RPO. The source-to-remote site (target 2) replication is asynchronous with an RPO in the order of minutes. The distance between the source and the remote sites could be thousands of miles. This implementation does not depend on the bunker site for updating data on the remote site because data is asynchronously copied to the remote site directly from the source. The triangle/multitarget configuration provides consistent RPO unlike cascade/multihop solutions in which the failure of the bunker site results in the remote site falling behind and the RPO increasing. The key benefit of three-site triangle/multitarget replication is the ability to failover to either of the two remote sites in the case of source-site failure, with disaster recovery (asynchronous) protection between the bunker and remote sites. Disaster recovery protection is always available if any one-site failure occurs.

During normal operations, all three sites are available and the production workload is at the source site. At any given instant, the data at the bunker and the source is identical. The data at the remote site is behind the data at the source and the bunker. The replication network links between the bunker and remote sites will be in place but not in use. The difference in the data between the bunker and remote sites is tracked, so that if a source site disaster occurs, operations can be resumed at the bunker or the remote sites with incremental resynchronization between these two sites.

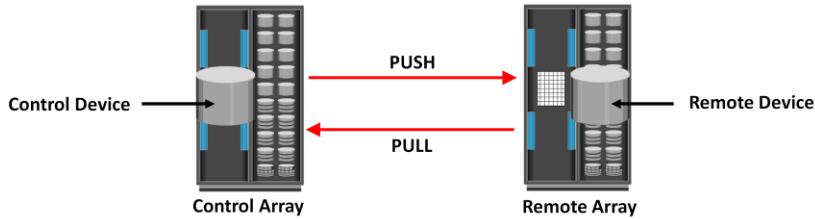
Cont..

A *regional disaster* in three-site triangle/multitarget replication is very similar to a source site failure in two-site asynchronous replication. If failure occurs, operations failover to the remote site with an RPO in the order of minutes. There is no remote protection until the regional disaster is resolved.

A failure of the bunker or the remote site is not really considered as a disaster because the operation can continue uninterrupted at the source site while remote DR protection is still available. A network link failure to either the source-to-bunker or the source-to-remote site does not impact production at the source site while remote DR protection is still available with the site that can be reached.

## Data Migration Solution

- Specialized replication technique that enables creating remote point-in-time copies
  - ▶ Used for data mobility, migration, and disaster recovery
- Moves data between heterogeneous storage arrays
  - ▶ Array performing replication operations is called control array
    - ▶ Push: Data is pushed from control array to remote array
    - ▶ Pull: Data is pulled to the control array from remote array



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

22

A *data migration and mobility solution* is a specialized replication technique that enables creating remote point-in-time copies. These copies can be used for data mobility, migration, content distribution, and disaster recovery. This solution moves data between heterogeneous storage arrays. This technology is application- and server-operating-system independent because the replication operations are performed by one of the storage arrays.

The array performing the replication operations is called the *control array*. Data can be moved from/to devices in the control array to/from a remote array. The terms control or remote do not indicate the direction of data flow; they indicate only the array that is performing the replication operation. Data migration solutions perform *push* and *pull operations* for data movement. These terms are defined from the perspective of the control array. In the *push* operation, data is moved from the control array to the remote array. In the *pull* operation, data is moved from the remote array to the control array.

Cont..

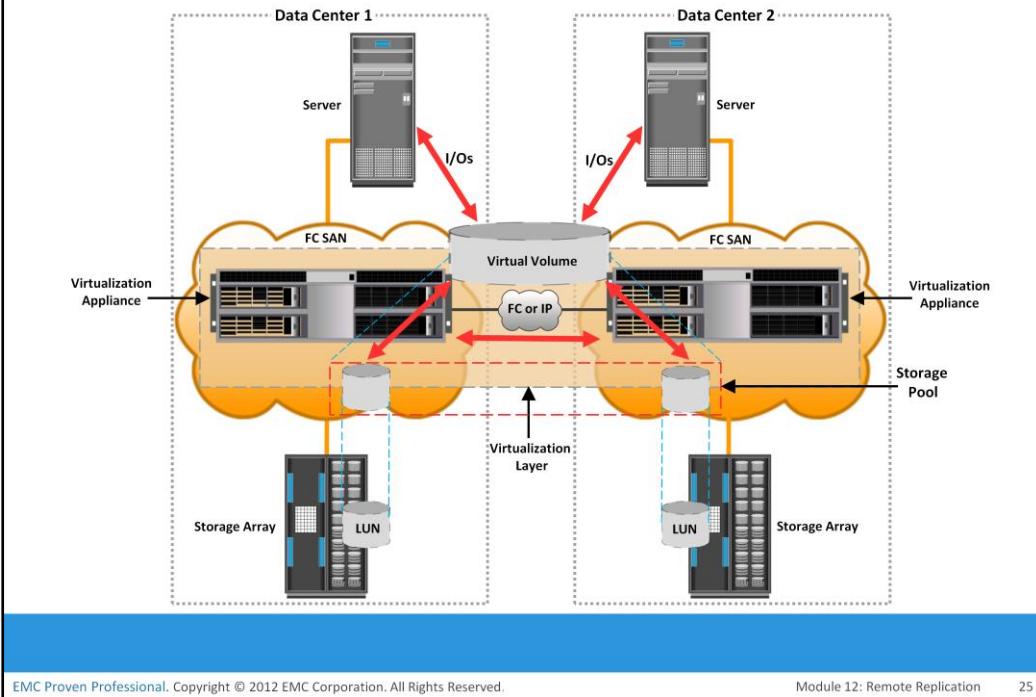
When a push or pull operation is initiated, the control array creates a protection bitmap to track the replication process. Each bit in the protection bitmap represents a data chunk on the control device. When the replication operation is initiated, all the bits are set to one, indicating that all the contents of the source device need to be copied to the target device. As the replication process copies data, the bits are changed to zero, indicating that a particular chunk has been copied. During the push and pull operations, host access to the remote device is not allowed because the control array has no control over the remote array and cannot track any change on the remote device. Data integrity cannot be guaranteed if changes are made to the remote device during the push and pull operations. The push/pull operations can be either *hot* or *cold*. These terms apply to the control devices only. In a cold operation, the control device is inaccessible to the host during replication. Cold operations guarantee data consistency because both the control and the remote devices are offline. In a hot operation, the control device is online for host operations. During hot push/pull operations, changes can be made to the control device because the control array can keep track of all changes, and thus ensure data integrity.

## Remote Replication/Migration in Virtualized Environment

- Remote mirroring of virtual volume
  - ▶ Virtual volumes assigned to hosts are mirrored to two different sites
- VM migration
  - ▶ Moving VMs from one location to another without powering off VMs
  - ▶ Commonly used techniques for VM migration are:
    - ▶ Hypervisor-to-hypervisor
    - ▶ Array-to-array

In a virtualized environment, the data residing in a virtual volume is replicated (mirrored) to two storage arrays located at two different sites. Virtual machine migration is another technique used to ensure business continuity in case of hypervisor failure or scheduled maintenance. VM migration is the process of moving VMs from one location to another without powering off the virtual machines. VM migration also helps in load balancing when multiple virtual machines running on the same hypervisor contend for resources. Two commonly used techniques for VM migration are hypervisor-to-hypervisor and array-to-array migration.

## Remote Mirroring of Virtual Volume



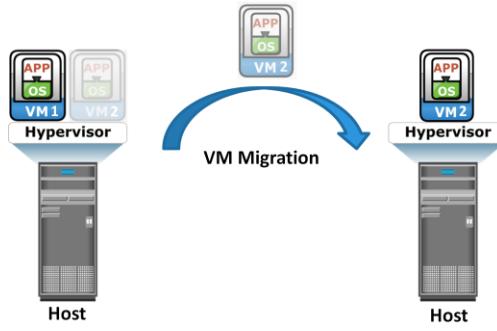
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

25

Virtualization enables mirroring of a virtual volume data between data centers. It provides the capability to connect the virtualization layers (appliance) at multiple data centers. The connected virtualization layers are managed centrally and work as a single virtualization layer that is stretched across data centers. The virtual volumes are created from the federated storage resources across data centers. The virtualization appliance has the ability to mirror the data of a virtual volume between the LUNs located in two different storage arrays at different locations. Figure on the slide provides an illustration of a virtual volume that is mirrored between arrays across data centers. Each I/O from a host to the virtual volume is mirrored to the underlying LUNs on the arrays. If an outage occurs at one of the data centers, the virtualization appliance will be able to continue processing I/O on the surviving mirror leg. Data on the virtual volume can be mirrored synchronously or asynchronously.

## VM Migration: Hypervisor-to-Hypervisor

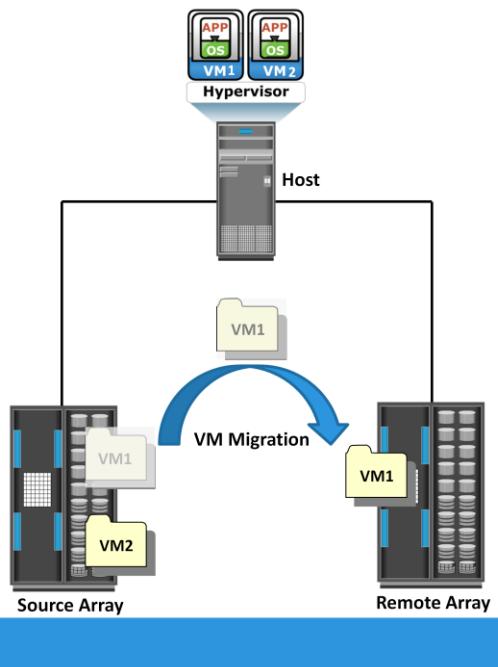


- Active state of a VM is moved from one hypervisor to another
  - ▶ Copies the contents of virtual machine memory from the source hypervisor to the target
- This technique requires source and target hypervisor access to the same storage

In hypervisor-to-hypervisor VM migration, the entire active state of a VM is moved from one hypervisor to another. This method involves copying the contents of virtual machine memory from the source hypervisor to the target and then transferring the control of the VM's disk files to the target hypervisor. Because the virtual disks of the VMs are not migrated, this technique requires both source and target hypervisor access to the same storage.

## VM Migration: Array-to-Array

- VM files are moved from source array to remote array
- Can move VMs across dissimilar storage arrays
- Balances storage utilization by redistributing VMs to different storage arrays



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

27

In array-to-array VM migration, VM files are moved from the source array to the remote array. This approach enables the administrator to move VMs across dissimilar storage arrays. Array-to-array migration starts by copying the metadata about the VM from the source array to the target. The metadata essentially consists of configuration, swap, and log files. After the metadata is copied, the VM disk file is replicated to the new location. During replication, there might be a chance that the source is updated; therefore, it is necessary to track the changes on the source to maintain data integrity. After the replication is complete, the blocks that have changed since the replication started are replicated to the new location. Array-to-array VM migration improves performance and balances the storage capacity by redistributing virtual disks to different storage devices.

## Module 12: Remote Replication

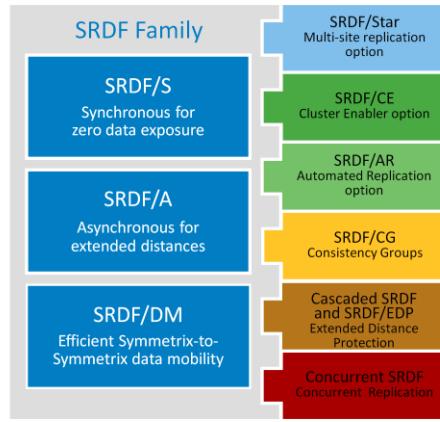
### Concept in Practice

- EMC Symmetrix Remote Data Facility (SRDF)
- EMC MirrorView
- EMC RecoverPoint

The concepts in practice section covers various EMC remote replication products such as EMC SRDF, EMC MirrorView, and EMC RecoverPoint.

## EMC SRDF

- Offers a family of solutions to implement array-based remote replication
- Minimizes performance impact on applications and hosts



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 12: Remote Replication

29

SRDF offers a family of technology solutions to implement storage array-based remote replication. The SRDF family of software includes the following:

**SRDF/Synchronous (SRDF/S):** A remote replication solution that creates a synchronous replica at one or more Symmetrix targets located within campus, metropolitan, or regional distances. SRDF/S provides a no-data-loss solution (near zero RPO) if a local disaster occurs.

**SRDF/Asynchronous (SRDF/A):** A remote replication solution that enables the source to asynchronously replicate data. It incorporates delta set technology, which enables write ordering by employing a buffering mechanism. SRDF/A provides minimal data loss if a regional disaster occurs.

**SRDF/DM:** A data migration solution that enables data migration from the source to target volume over extended distances.

**SRDF/Automated Replication (SRDF/AR):** A remote replication solution that uses both SRDF and TimeFinder/Mirror to implement disk-buffered replication technology. It is offered as SRDF/AR Single-hop for two-site replication and SRDF/AR Multi-hop for three-site cascade replication. SRDF/AR provides long distance solution with RPO in the order of hours.

**SRDF/Star:** Three-site multi-target remote replication solution that consists of primary (production), secondary (bunker), and tertiary (remote) sites. The replication between the primary and secondary sites is synchronous, whereas the replication between the primary and tertiary sites is asynchronous. In the event of primary site outage, EMC's SRDF/Star solution allows to quickly move operations and re-establish remote replication between the remaining two sites.

## EMC MirrorView

- Replicates data from a primary volume to a secondary volume that reside on different VNX storage systems
- Uses a bitmap to track host writes while the link to the secondary array is down
  - ▶ When secondary is available, sends only changed data
- MirrorView family consists of:
  - ▶ MirrorView/Synchronous (MirrorView/S)
  - ▶ MirrorView/Asynchronous (MirrorView/A)

The MirrorView software enables EMC VNX storage array-based remote replication. It replicates the contents of a primary volume to a secondary volume that resides on a different VNX storage system. The MirrorView family consists of MirrorView/Synchronous (MirrorView/S) and MirrorView/Asynchronous (MirrorView/A) solutions.

MirrorView/S is a synchronous product that mirrors data between local and remote storage systems. MirrorView/A is an asynchronous product that offers extended distance replication based on periodic incremental update model. It periodically updates the remote copy of the data with all the changes that occurred on the primary copy since the last update.

## EMC RecoverPoint

- RecoverPoint Continuous Remote Replication (CRR) provides both synchronous and asynchronous remote replication
- Dynamically switches between synchronous and asynchronous replication
  - ▶ Based on the policy for performance and latency
- Capable to recover data remotely to any PIT

EMC RecoverPoint Continuous Remote Replication (CRR) provides bi-directional synchronous and asynchronous replication. In normal operations, RecoverPoint CRR enables users to recover data remotely to any point in time. RecoverPoint dynamically switches between synchronous and asynchronous replication based on the policy for performance and latency.

## Module 12: Summary

Key points covered in this module:

- Synchronous and asynchronous replication mode
- Host-based, array-based, and network-based remote replication
- Three-site remote replication
- Data migration solution
- Remote replication and migration in virtualized environment

This module covered various host-based, network-based, and array-based remote replication technologies. This module also covered three-site remote replication and data migration solution. Finally, this module covered remote replication in a virtualized environment.

Remote replication is the process of creating replicas of information assets at remote sites (locations). Two basic modes of remote replications are synchronous and asynchronous replication. Remote replication of data can be handled by the hosts or storage arrays. Other options include specialized network-based appliances to replicate data over the LAN or SAN.

In host-based remote replication, there are two basic approaches for performing replication: Logical volume manager (LVM) based replication and database replication via log shipping. In storage array-based remote replication, replication between arrays can be performed in synchronous, asynchronous, or disk-buffered modes. In network-based remote replication, continuous data protection technology is used for performing remote replication.

Three-site replication is used to mitigate the risks identified in two-site replication. In a three-site replication, data from the source site is replicated to two remote sites.

Data Migration and mobility solution is a specialized replication technique that enables creating remote point-in-time copies. These copies can be used for data mobility, migration, content distribution, and disaster recovery.

In a virtualized environment, the data residing in virtual volume is replicated (mirrored) to two storage arrays located in two different sites. Virtual machine migration is another technique used to ensure business continuity in case of hypervisor failure or scheduled maintenance. VM migration is the process of moving VMs from one location to another without powering off the virtual machines.

## Check Your Knowledge – 1

- Which factor influences the write response time in a synchronous remote replication solution?
  - A. Number of LUNs on source array
  - B. Number of LUNs on remote array
  - C. Distance between source and remote site
  - D. RAID level configured on the remote array
- What accurately describes asynchronous remote replication?
  - A. Provides near zero RPO in the event of disaster
  - B. Not suitable for distance beyond 200 km
  - C. Writes are committed to both source and replica before acknowledging host
  - D. Writes are committed to source and immediately acknowledged to host

## Check Your Knowledge – 2

- Which determines the RPO in an asynchronous remote replication?
  - A. Application response time
  - B. Average read workload
  - C. Size of the buffer
  - D. Number of LUNs on remote array
- What is the minimum number of storage volumes required in the cascade/multihop (synchronous+disk buffered) three-site replication?
  - A. 2
  - B. 3
  - C. 4
  - D. 5

## Check Your Knowledge – 3

- Which is true about hypervisor-to-hypervisor VM migration?
  - A. Moves VM files from the source array to the remote array
  - B. Enables to move VMs across dissimilar storage arrays
  - C. Requires both source and target hypervisor access to the same storage
  - D. Requires VMs to go offline during the migration

This slide intentionally left blank.

# Module – 13

# Cloud Computing



## Module 13: Cloud Computing

Upon completion of this module, you should be able to:

- Explain the characteristics of cloud computing
- Describe cloud services and deployment models
- Describe cloud computing infrastructure
- Discuss the challenges of cloud computing
- Discuss cloud adoption considerations

This module focuses on the essential characteristics of cloud computing, cloud services and deployment models, and cloud computing infrastructure. It also focuses on the challenges of cloud computing and cloud adoption considerations.

# Module 13: Cloud Computing

## Lesson 1: Cloud Computing Overview

During this lesson the following topics are covered:

- Definition of cloud computing
- Essential characteristics of cloud computing
- Benefits of cloud computing
- Cloud enabling technologies

This lesson covers definition and essential characteristics of cloud computing. It also covers benefits of cloud computing and cloud enabling technologies.

## Drivers for Cloud Computing

- Business requirements
  - ▶ Transformation of IT processes to achieve more with less
  - ▶ Better agility and higher availability at reduced expenditure
  - ▶ Reduced time-to-market
  - ▶ Accelerated pace of innovation
- IT challenges to meet business requirements are:
  - ▶ Serving customers worldwide round the clock, refreshing technology quickly, faster provisioning of IT resources – all at reduced cost
- These challenges are addressed with the emergence of cloud computing

In today's competitive environment, organizations are under increasing pressure to improve efficiency and transform their IT processes to achieve more with less. Businesses need reduced time-to-market, better agility, higher availability, and reduced expenditures to meet the changing business requirements and accelerated pace of innovation. These business requirements are posing several challenges to IT teams. Some of the key challenges are serving customers worldwide round the clock, refreshing technology quickly, and faster provisioning of IT resources—all at reduced costs.

These long-standing challenges are addressed with the emergence of a new computing style, called cloud computing, which enables organizations and individuals to obtain and provision IT resources as a service.

# What is Cloud Computing?

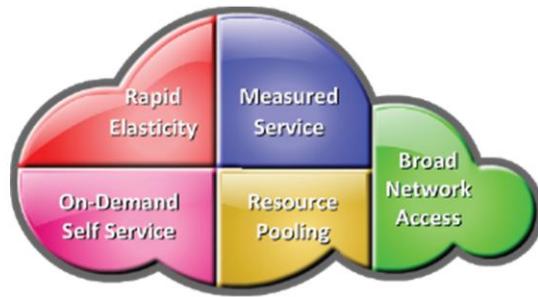
## Cloud Computing

A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., servers, storage, networks, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

– NIST

- Essential Cloud characteristics

- ▶ On-demand self-service
- ▶ Broad network access
- ▶ Resource pooling
- ▶ Rapid elasticity
- ▶ Measured service



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 5

With cloud computing users can browse and select relevant cloud services, such as compute, software, information storage, or a combination of these resources, via a portal. Cloud computing automates delivery of selected cloud services to the users. It helps organizations and individuals deploy IT resources at reduced total cost of ownership with faster provisioning and compliance adherence. A widely adopted definition of cloud computing comes from the U.S. National Institute of Standards and Technology (NIST Special Publication 800-145) as provided in the slide.

A computing infrastructure used for cloud services must meet certain capabilities or characteristics. According to NIST, the cloud infrastructure should have five essential characteristics:

- On-Demand Self-Service
- Broad Network Access
- Resource Pooling
- Rapid Elasticity
- Measured Service

## On-demand Self-service



- Enables consumers to unilaterally provision computing capabilities (examples: server time and storage capacity) as needed automatically
- Consumers view service catalogue via a Web-based user interface and use it to request for a service

A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

A cloud service provider publishes a service catalogue, which contains information about all cloud services available to consumers. The service catalogue includes information about service attributes, prices, and request processes. Consumers view the service catalogue via a web-based user interface and use it to request for a service. Consumers can either leverage the “ready-to-use” services or change a few service parameters to customize the services.

## Broad Network Access



- Computing capabilities are available over the network
- Computing capabilities are accessed from a broad range of client platforms such as:
  - ▶ Desktop computer
  - ▶ Laptop
  - ▶ Tablet
  - ▶ Mobile device

Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (example, mobile phones, tablets, laptops, and workstations)

## Resource Pooling



- Provider's computing resources are pooled to serve multiple consumers using a multitenant model
- Resources are assigned from the pool according to consumer demand
- Consumers have no control or knowledge over the exact location of the provided resources

The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (example, country, state, or data center). Examples of resources include storage, processing, memory, and network bandwidth.

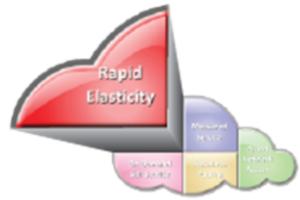
### Note:

Multitenancy refers to an architecture in which multiple independent consumers (tenants) are serviced using a single set of resources. This lowers the cost of services for consumers.

Virtualization enables resource pooling and multitenancy in the cloud. For example, multiple virtual machines from different consumers can run simultaneously on the same physical server that runs the hypervisor.

## Rapid Elasticity

- Computing capabilities can be elastically provisioned and released
- Computing capabilities are scaled rapidly, commensurate with consumer's demand
  - ▶ Provides a sense of unlimited scalability



Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Consumers can leverage rapid elasticity of the cloud when they have a fluctuation in their IT resource requirements. For example, an organization might require doubling the number of web and application servers for a specific duration to accomplish a specific task. For the remaining period, they might want to release idle server resources to cut down the expenses. The cloud enables consumers to grow and shrink the demand for resources dynamically.

## Measured Service



- Cloud computing provides a metering system that continuously monitors resource consumption and generates reports
  - ▶ Helps to control and optimize resource use
  - ▶ Helps to generate billing and chargeback reports

Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (example, storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

## Benefits of Cloud Computing

Benefits	Description
Reduced IT cost	<ul style="list-style-type: none"><li>Reduces the up-front capital expenditure (CAPEX)</li></ul>
Business agility	<ul style="list-style-type: none"><li>Provides the ability to deploy new resources quickly</li><li>Enables businesses to reduce time-to-market</li></ul>
Flexible scaling	<ul style="list-style-type: none"><li>Enables consumers to scale up, scale down, scale out, or scale in the demand for computing resources easily</li><li>Consumers can unilaterally and automatically scale computing resources</li></ul>
High availability	<ul style="list-style-type: none"><li>Ensures resource availability at varying levels, depending on consumer's policy and priority</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 11

Cloud computing offers the following key benefits:

- **Reduced IT cost:** Cloud services can be purchased based on pay-per-use or subscription pricing. This reduces or eliminates consumer's IT capital expenditure (CAPEX).
- **Business agility:** Cloud computing provides the capability to allocate and scale computing capacity quickly. Cloud can reduce the time required to provision and deploy new applications and services from months to minutes. This enables businesses to respond more quickly to market changes and reduce time-to-market.
- **Flexible scaling:** Cloud computing enables consumers to scale up, scale down, scale out, or scale in the demand for computing resources easily. Consumers can unilaterally and automatically scale computing resources without any interaction with cloud service providers. The flexible service provisioning capability of cloud often provides a sense of unlimited scalability to the cloud service consumers.
- **High availability:** Cloud computing has the ability to ensure resource availability at varying levels depending on the consumer's policy and priority. Redundant infrastructure components (servers, network paths, and storage equipments, along with clustered software) enable fault tolerance for cloud deployments. These techniques can encompass multiple datacenters located in different geographic regions, which prevents data unavailability due to regional failures.

## Cloud Enabling Technologies

Technologies	Description
Grid computing	<ul style="list-style-type: none"><li>• Form of distributed computing</li><li>• Enables resources of numerous computers in a network to work on a single task at the same time</li></ul>
Utility computing	<ul style="list-style-type: none"><li>• Service provisioning model that offers computing resources as a metered service</li></ul>
Virtualization	<ul style="list-style-type: none"><li>• Abstracts physical characteristics of IT resources from resource users</li><li>• Enables resource pooling and creating virtual resources from pooled resources</li></ul>
Service-oriented architecture (SOA)	<ul style="list-style-type: none"><li>• Provides a set of services that can communicate with each other</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 12

Grid computing, utility computing, virtualization, and service-oriented architecture are enabling technologies of cloud computing.

Grid computing is a form of distributed computing that enables the resources of numerous heterogeneous computers in a network to work together on a single task at the same time. Grid computing enables parallel computing and is best for large workloads.

Utility computing is a service-provisioning model in which a service provider makes computing resources available to customers, as required, and charges them based on usage. This is analogous to other utility services, such as electricity, where charges are based on the consumption.

Virtualization is a technique that abstracts the physical characteristics of IT resources from resource users. It enables the resources to be viewed and managed as a pool and lets users create virtual resources from the pool. Virtualization provides better flexibility for provisioning of IT resources compared to provisioning in a non-virtualized environment. It helps optimizing resource utilization and delivering resources more efficiently.

Service-oriented architecture (SOA) provides a set of services that can communicate with each other. These services work together to perform some activity or simply passes data among services.

# Module 13: Cloud Computing

## Lesson 2: Cloud Service and Deployment Models

During this lesson the following topics are covered:

- Cloud service models
- Cloud deployment models

This lesson covers three primary cloud service models – Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). This lesson also covers cloud deployment models – Public, Private, Community, and Hybrid.

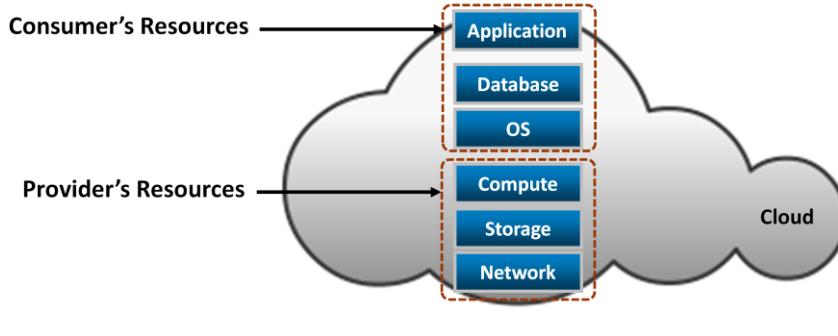
## Cloud Service Models

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS)

According to NIST, cloud service offerings are classified primarily into three models: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS).

## Infrastructure-as-a-Service

- Consumers deploy their software, including OS and application on provider's infrastructure
  - ▶ Computing resources such as processing power, memory, storage, and networking components are offered as service
  - ▶ Example: Amazon Elastic Compute Cloud
- Consumers have control over the OSs and deployed applications



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 15

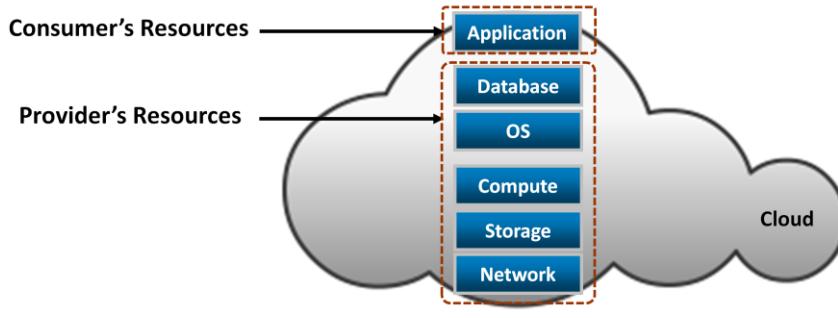
The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems and deployed applications; and possibly limited control of select networking components (example, host firewalls).

IaaS is the base layer of the cloud services stack. It serves as the foundation for both the SaaS and PaaS.

Amazon Elastic Compute Cloud (Amazon EC2) is an example of IaaS that provides scalable compute capacity, on-demand, in the cloud. It enables consumers to leverage Amazon's massive computing infrastructure with no up-front capital investment.

## Platform-as-a-Service

- Consumers deploy consumer-created or acquired applications onto provider's computing platform
  - ▶ Computing platform is offered as a service
  - ▶ Example: Google App Engine and Microsoft Windows Azure Platform
- Consumer has control over deployed applications



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

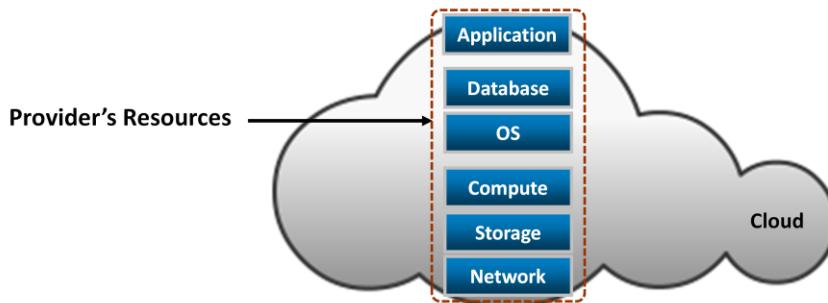
Module 13: Cloud Computing 16

The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

PaaS is also used as an application development environment, offered as a service by the cloud service provider. The consumer may use these platforms to code their applications and then deploy the applications on the cloud. Because the workload to the deployed applications varies, the scalability of computing resources is usually guaranteed by the computing platform, transparently. Google App Engine and Microsoft Windows Azure Platform are examples of PaaS.

## Software-as-a-Service

- Consumers use provider's applications running on the cloud infrastructure
  - ▶ Applications are offered as a service
  - ▶ Examples: EMC Mozy and Salesforce.com
- Service providers exclusively manage computing infrastructure and software to support services



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 17

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (example, web-based e-mail), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

In a SaaS model, applications, such as Customer Relationship Management (CRM), e-mail, and Instant Messaging (IM), are offered as a service by the cloud service providers. The cloud service providers exclusively manage the required computing infrastructure and software to support these services. The consumers may be allowed to change a few application configuration settings to customize the applications.

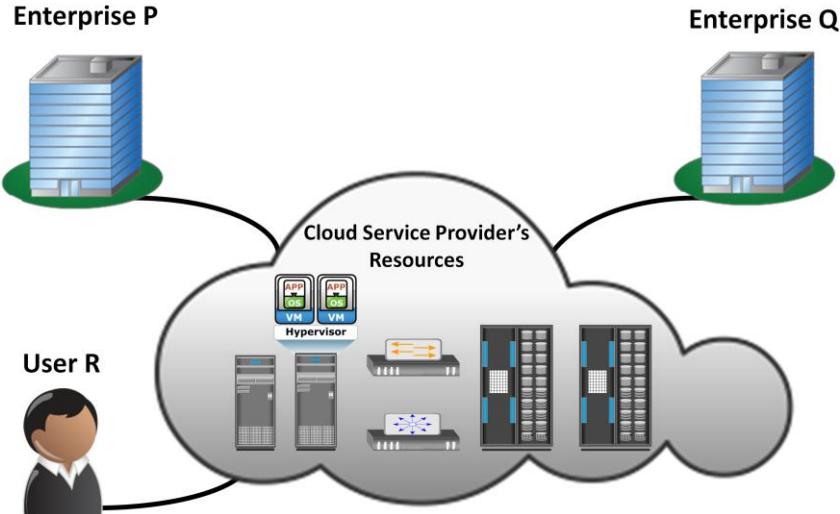
EMC Mozy is an example of Software-as-a-Service. Consumers can leverage the Mozy console to perform automatic, secured, online backup and recovery of their data with ease. Salesforce.com is a provider of SaaS-based CRM applications, such as Sales Cloud and Service Cloud.

## Cloud Deployment Models

- Public
- Private
- Community
- Hybrid

According to NIST, cloud computing is classified into four deployment models—public, private, community, and hybrid—which provide the basis for how cloud infrastructures are constructed and consumed.

## Public Cloud



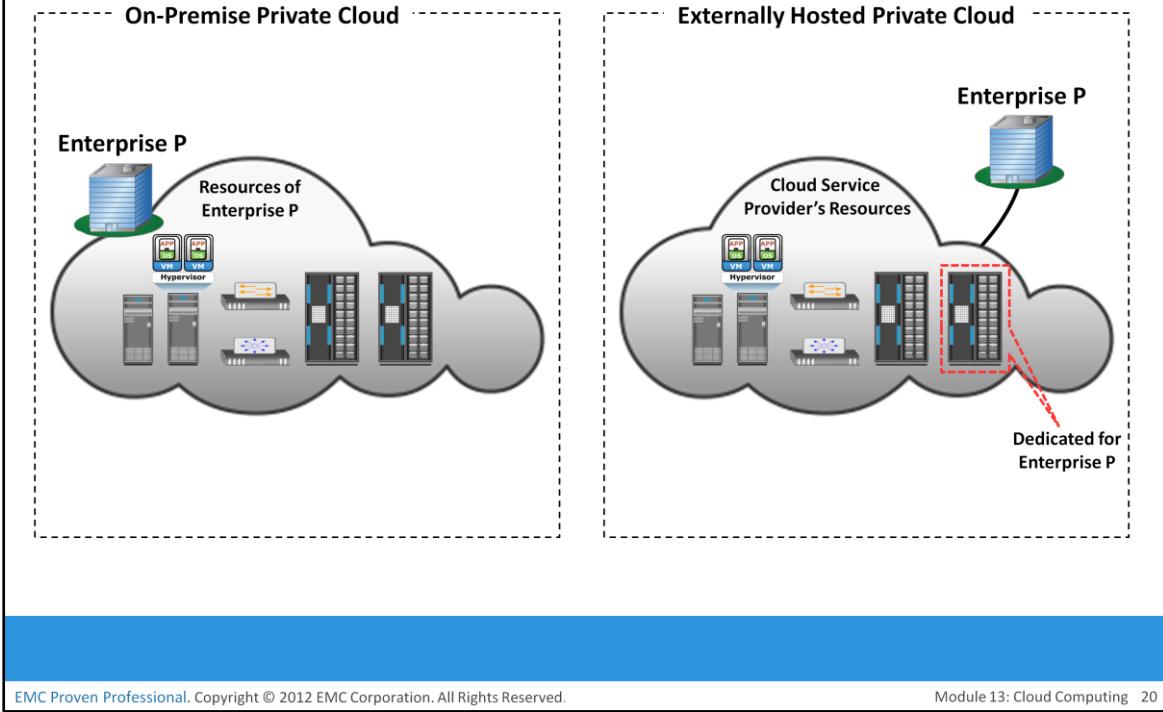
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 19

In a public cloud model, the cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

Consumers use the cloud services offered by the providers via the Internet and pay metered usage charges or subscription fees. An advantage of the public cloud is its low capital cost with enormous scalability. However, for consumers, these benefits come with certain risks: no control over the resources in the cloud, the security of confidential data, network performance, and interoperability issues. Popular public cloud service providers are Amazon, Google, and Salesforce.com. Figure in the slide shows a public cloud that provides cloud services to organizations and individuals.

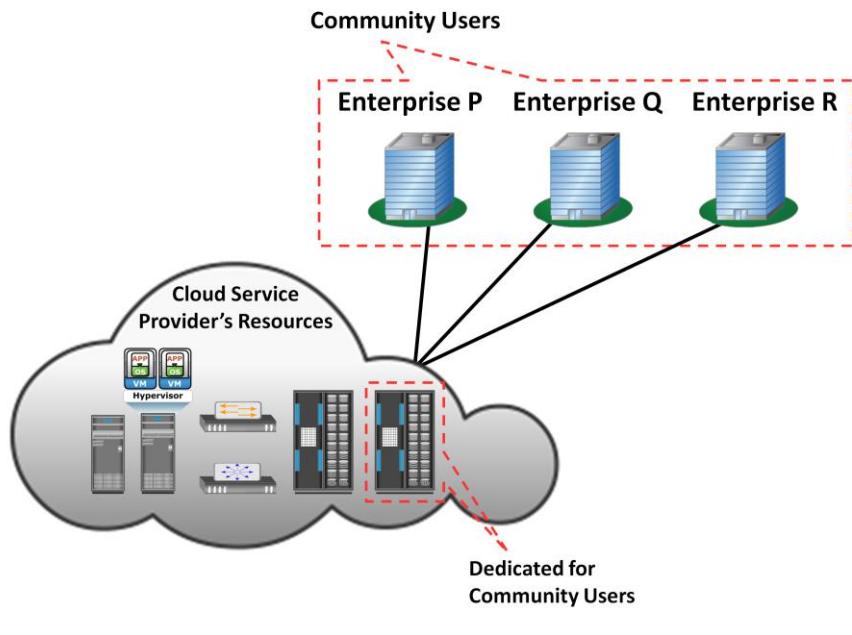
# Private Cloud



In a private cloud model, the cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (example, business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises. Following are two variations to the private cloud model:

- **On-premise private cloud:** The on-premise private cloud, also known as internal cloud, is hosted by an organization within its own data centers. This model enables organizations to standardize their cloud service management processes and security, although this model has limitations in terms of size and resource scalability. Organizations would also need to incur the capital and operational costs for the physical resources. This is best suited for organizations that require complete control over their applications, infrastructure configurations, and security mechanisms.
- **Externally hosted private cloud:** This type of private cloud is hosted external to an organization and is managed by a third-party organization. The third-party organization facilitates an exclusive cloud environment for a specific organization with full guarantee of privacy and confidentiality.

## Community Cloud



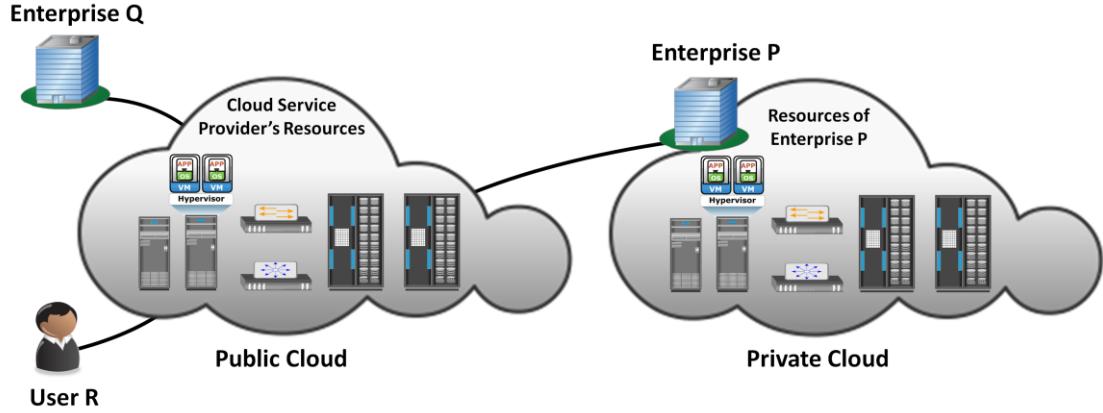
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 21

In a community cloud model, the cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (example, mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

In a community cloud, the costs spread over to fewer consumers than a public cloud. Hence, this option is more expensive but might offer a higher level of privacy, security, and compliance. The community cloud also offers organizations access to a vast pool of resources compared to the private cloud. An example in which a community cloud could be useful is government agencies. If various agencies within the government operate under similar guidelines, they could all share the same infrastructure and lower their individual agency's investment.

## Hybrid Cloud



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 22

In a hybrid cloud model, the cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (example, cloud bursting for load balancing between clouds).

The hybrid model allows an organization to deploy less critical applications and data to the public cloud, leveraging the scalability and cost-effectiveness of the public cloud. The organization's mission-critical applications and data remain on the private cloud that provides greater security. Figure in the slide shows an example of a hybrid cloud.

# Module 13: Cloud Computing

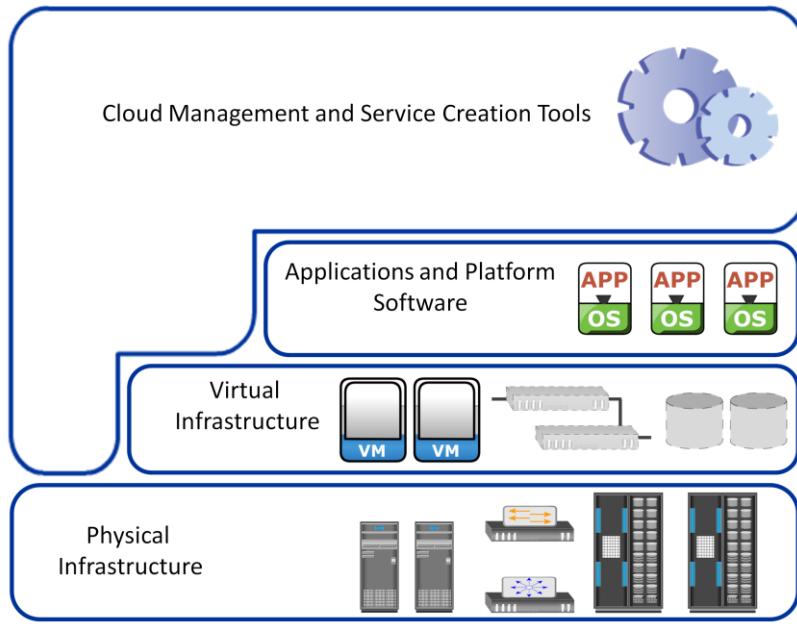
## Lesson 3: Cloud Infrastructure, Challenges, and Considerations

During this lesson the following topics are covered:

- Cloud infrastructure
- Challenges of cloud computing
- Cloud adoption considerations

This lesson covers the cloud computing infrastructure, challenges of cloud computing, and cloud adoption considerations.

## Cloud Infrastructure Framework



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 24

A cloud computing infrastructure is the collection of hardware and software that enables the five essential characteristics of cloud computing. Cloud computing infrastructure usually consists of the following layers:

- Physical infrastructure
- Virtual infrastructure
- Applications and platform software
- Cloud management and service creation tools

The resources of these layers are aggregated and coordinated to provide cloud services to the consumers.

## Physical Infrastructure

- Physical infrastructure includes physical IT resources
  - ▶ Physical servers
  - ▶ Storage systems
  - ▶ Networks
- Physical servers are connected to each other, to the storage systems, and to clients via networks
- Physical resources may be located in a single data center or distributed across multiple data centers

The physical infrastructure consists of physical computing resources, which include physical servers, storage systems, and networks. Physical servers are connected to each other, to the storage systems, and to the clients via networks, such as IP, FC SAN, IP SAN, or FCoE networks.

Cloud service providers may use physical computing resources from one or more data centers to provide services. If the computing resources are distributed across multiple data centers, connectivity must be established among them. The connectivity enables the data centers in different locations to work as a single large data center. This enables migration of business applications and data across data centers and provisioning cloud services using the resources from multiple data centers.

## Virtual Infrastructure

- Virtual infrastructure consists of:
  - ▶ Resource pools
    - ▶ CPU, memory, network bandwidth, storage pools
  - ▶ Identity pools
    - ▶ VLAN ID and VSAN ID pools
  - ▶ Virtual IT resources
    - ▶ Virtual Machines (VMs), virtual storage volumes, virtual networks (VLAN and VSAN)
- Virtual IT resources obtain capacity and identity from resource and identity pools respectively

Cloud service providers employ virtualization technologies to build a virtual infrastructure layer on the top of the physical infrastructure. Virtualization enables fulfilling some of the cloud characteristics, such as resource pooling and rapid elasticity. It also helps reduce the cost of providing the cloud services. Some cloud service providers may not have completely virtualized their physical infrastructure yet, but they are adopting virtualization for better efficiency and optimization.

Virtualization abstracts physical computing resources and provides a consolidated view of the resource capacity. The consolidated resources are managed as a single entity called a resource pool. For example, a resource pool might group CPUs of physical servers within a cluster. The capacity of the resource pool is the sum of the power of all CPUs (for example, 10,000 megahertz) available in the cluster. In addition to the CPU pool, the virtual infrastructure includes other types of resource pools, such as memory pool, network pool, and storage pool. Apart from resource pools, the virtual infrastructure also includes identity pools, such as VLAN ID pools and VSAN ID pools. The number of each type of pool and the pool capacity depend on the cloud service provider's requirement to create different cloud services.

Virtual infrastructure also includes virtual computing resources, such as virtual machines, virtual storage volumes, and virtual networks. These resources obtain capacities, such as CPU power, memory, network bandwidth, and storage space from the resource pools. The capacity is allocated to the virtual computing resources easily and flexibly based on the service requirement. Virtual networks are created using network identifiers, such as VLAN IDs and VSAN IDs from the respective identity pools. Virtual computing resources are used for creating cloud infrastructure services.

## Applications and Platform Software

- Suite of software that may include:
  - ▶ Business applications
  - ▶ Platform software such as OS and database
    - ▶ Provide environments for applications to run
- Applications and platform software are hosted on VMs
  - ▶ To create software-as-a-service (SaaS) and platform-as-a-service (PaaS)

This layer includes a suite of business applications and platform software such as the OS and database. Platform software provides the environment on which business applications to run. Applications and platform software are hosted on virtual machines to create SaaS and PaaS. For SaaS, both the application and platform software are provided by cloud service providers. In the case of PaaS, only the platform software is provided by cloud service providers; consumers export their applications to the cloud.

## Cloud Management and Service Creation Tools

- Include three types of software:
  - ▶ Physical and virtual infrastructure management software
  - ▶ Unified management software
  - ▶ User-access management software
- These software interact among themselves to automate provisioning of cloud services

The cloud management and service creation tools layer includes three types of software:

- Physical and virtual infrastructure management software
- Unified management software
- User-access management software

This classification is based on the different functions performed by these software. These software interact with each other to automate provisioning of cloud services.

The physical and virtual infrastructure management software is offered by the vendors of various infrastructure resources and third-party organizations. For example, a storage array has its own management software. Similarly, network and physical servers are managed independently using network and compute management software respectively. These software provide interfaces to construct a virtual infrastructure from the underlying physical infrastructure.

Unified management software interacts with all standalone physical and virtual infrastructure management software. It collects information on the existing physical and virtual infrastructure configurations, connectivity, and utilization. Unified management software compiles this information and provides a consolidated view of infrastructure resources scattered across one or more data centers. It allows an administrator to monitor performance, capacity, and availability of physical and virtual resources centrally. Unified management software also provides a single management interface to configure physical and virtual infrastructure and integrate the compute (both CPU and memory), network, and storage pools. The integration allows a group of compute pools to use the storage and network pools for storing and transferring data respectively.

Cont..

The unified management software passes configuration commands to respective physical and virtual infrastructure management software, which executes the instructions. This eliminates the administration of compute, storage, and network resources separately using native management software.

The key function of the unified management software is to automate the creation of cloud services. It enables administrators to define service attributes such as CPU power, memory, network bandwidth, storage capacity, name and description of applications and platform software, resource location, and backup policy. When the unified management software receives consumer requests for cloud services, it creates the service based on predefined service attributes.

The user-access management software provides a web-based user interface to consumers. Consumers can use the interface to browse the service catalogue and request cloud services. The user-access management software authenticates users before forwarding their request to the unified management software. It also monitors allocation or usage of resources associated to the cloud service instances. Based on the allocation or usage of resources, it generates a chargeback report. The chargeback report is visible to consumers and provides transparency between consumers and providers.

## Cloud-optimized Storage

- Provides rapid elasticity, global access, and storage capacity on-demand
- Leverages object-based storage technology
- Enables self-service and fully metered access to storage resources
- Key characteristics of cloud-optimized storage solution are:
  - ▶ Massively scalable
  - ▶ Unified namespace
  - ▶ Metadata and policy-based information management
  - ▶ Secure multitenancy
  - ▶ Multiple access mechanisms (through REST and SOAP web service APIs and file-based access)

Content-rich applications combined with the growth of unstructured data are challenging to manage with traditional approach of storing data at scale. This combination of massive growth, new information types, and the need to serve multiple locations and users around the world, has led to requirements for information storage and management at a global scale. Cloud-optimized storage is a solution to meet these requirements. It delivers scalable and flexible architecture that provides rapid elasticity, global access and storage capacity on-demand. It also addresses the constraints of rigid, mount-point based interaction between storage and consumer by presenting a singular access point to the entire storage infrastructure. It leverages a built-in multi-tenancy model and enables self-service, fully metered access to storage resources thereby delivers storage-as-a-service on a shared infrastructure. Cloud-optimized storage typically leverages object-based storage technology that uses customizable, value-driven metadata to drive storage placement, protection and lifecycle policies. Key characteristics of cloud-optimized storage solution are:

- Massively scalable infrastructure that supports large number of objects across a globally distributed infrastructure
- Unified namespace that eliminates capacity, location, and other file system limitations
- Metadata and policy-based information management capabilities that optimizes data protection, availability and cost, based on service levels
- Secure multitenancy that enables multiple applications to be securely served from the same infrastructure. Each application is securely partitioned and data is neither co-mingled nor accessible by other tenants
- Provide access through REST and SOAP web service APIs and file-based access using variety of client devices

## Cloud Challenges – Consumer's Perspective

- Security and regulation
  - ▶ Consumers are indecisive to transfer control of sensitive data
  - ▶ Regulation may prevent organizations to use cloud services
- Network latency
  - ▶ Real time applications may suffer due to network latency and limited bandwidth
- Supportability
  - ▶ Service provider might not support proprietary environments
  - ▶ Incompatible hypervisors could impact VM migration
- Vendor lock-in
  - ▶ Restricts consumers from changing their cloud service providers
  - ▶ Lack of standardization across cloud-based platforms

Business critical data requires protection and continuous monitoring of its access. If the data moves to a cloud model other than an on-premise private cloud, consumers could lose absolute control of their sensitive data. Although most of the cloud service providers offer enhanced data security, consumers might not be willing to transfer control of their business-critical data to the cloud.

Cloud service providers might use multiple data centers located in different countries to provide cloud services. They might replicate or move data across these data centers to ensure high availability and load distribution. Consumers may or may not know in which country their data is stored. Some cloud service providers allow consumers to select the location for storing their data. Data privacy concerns and regulatory compliance requirements, such as the EU Data Protection Directive and the U.S. Safe Harbor program, create challenges for the consumers in adopting cloud computing.

Cloud services can be accessed from anywhere via a network. However, network latency increases when the cloud infrastructure is not close to the access point. A high network latency can either increase the application response time or cause the application to timeout. This can be addressed by implementing stringent Service Level Agreements (SLAs) with the cloud service providers.

Cont..

Another challenge is cloud platform services may not support consumer's desired applications. For example, service provider might not be able to support highly specialized or proprietary environments, such as compatible OSs and preferred programming languages, required to develop and run the consumer's application. Also a mismatch between hypervisors could impact migration of virtual machines into or between clouds.

Another challenge is vendor lock-in: the difficulty for consumers to change their cloud service provider. A lack of interoperability between the APIs of different cloud service providers could also create complexity and high migration costs when moving from one service provider to another.

## Cloud Challenges – Provider's Perspective

- Service warranty and service cost
  - ▶ Resources must be kept ready to meet unpredictable demand
  - ▶ Hefty penalty, if SLAs are not fulfilled
- Complexity in deploying vendor software in the cloud
  - ▶ Many vendors do not provide cloud-ready software licenses
  - ▶ Higher cost of cloud-ready software licenses
- No standard cloud access interface
  - ▶ Cloud consumers want open APIs
  - ▶ Need agreement among cloud providers for standardization

Cloud service providers usually publish a service-level agreement (SLA) so that their consumers know about the availability of service, quality of service, downtime compensation, and legal and regulatory clauses. Alternatively, customer-specific SLAs may be signed between a cloud service provider and a consumer. SLAs typically mention a penalty amount if cloud service providers fail to provide the services levels. Therefore, cloud service providers must ensure that they have adequate resources to provide the required levels of services. Because the cloud resources are distributed and services demand fluctuates, it is a challenge for cloud service providers to provision physical resources for peak demand of all consumers and estimate the actual cost of providing the services.

Many software vendors do not have a cloud-ready software licensing model. Some of the software vendors offer standardized cloud licenses at a higher price compared to traditional licensing models. The cloud software licensing complexity has been causing challenges in deploying vendor software in the cloud. This is also a challenge to the consumer.

Cloud service providers usually offer proprietary APIs to access their cloud. However, consumers might want open APIs or standard APIs to become the tenant of multiple clouds. This is a challenge for cloud service providers because this requires agreement among cloud service providers.

## Cloud Adoption Considerations

- CIOs/IT Managers seeking to move to the cloud face several questions:
  - ▶ Which deployment model fits organization's requirements?
    - ▶ Private, public, hybrid
  - ▶ Which are the applications suitable for cloud?
  - ▶ How do I choose the cloud service provider?
  - ▶ Is the cloud infrastructure capable of providing the required Quality of Service (QoS)?
    - ▶ Performance, availability, and security
  - ▶ What is the financial benefit in adopting cloud?

Organizations that decide to adopt cloud computing always face this question “How does the cloud fit the organization’s environment?” Most organizations are not ready to abandon their existing IT investments to move all of their business processes to the cloud at once. Instead, they need to consider various factors before moving their business processes to the cloud. Even individuals seeking to use cloud services need to understand some cloud adoption considerations.

## What Deployment Model Fits for You?



### Public cloud

- Convenience outweighs risk
- Low cost or free
- Ex: Picasa, Google apps

### Hybrid cloud

- Tier 1 apps: private cloud
- Tier 2-4 apps (backup, archive, testing): public cloud

### Public cloud

- Convenience outweighs risk

### Private cloud

- Tier 2-4: private cloud
- Tier 1: may continue to run in a traditional data center environment

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 35

Risk versus convenience is a key consideration for deciding on a cloud adoption strategy. This consideration also forms the basis for choosing the right cloud deployment model. A public cloud is usually preferred by individuals and start-up businesses. For them, the cost reduction offered by the public cloud outweighs the security or availability risks in the cloud. Small- and medium-sized businesses (SMBs) have a moderate customer base, and any anomaly in customer data and service levels might impact their business. Therefore, they may not be willing to deploy their tier 1 applications, such as Online Transaction Processing (OLTP), in the public cloud. A hybrid cloud model fits in this case. The tier 1 applications should run on the private cloud, whereas less critical applications such as backup, archive, and testing can be deployed in the public cloud. Enterprises typically have a strong customer base worldwide. They usually enforce strict security policies to safeguard critical customer data. Because they are financially capable, they might prefer building their own private clouds.

## Choosing Applications for Public Cloud

- Some key questions to ask before migrating a consumer application to the public cloud:
  - ▶ Is the application compatible to cloud platform software? Is it a legacy application?
  - ▶ Is the application proprietary and mission-critical? Does the application provide competitive advantage?
  - ▶ Is the application workload network traffic intensive? Will application performance be impacted by network latency and limited network bandwidth?
  - ▶ Does the application communicate with other data center resources or applications?

Not all applications are good candidates for a public cloud. This may be due to the incompatibility between the cloud platform software and the consumer applications, or maybe the organization plans to move a legacy application to the cloud. Proprietary and mission-critical applications are core and essential to the business. They are usually designed, developed, and maintained in-house. These applications often provide competitive advantages. Due to high security risk, organizations are unlikely to move these applications to the public cloud. These applications are good candidate for an on-premise private cloud. Nonproprietary and nonmission critical applications are suitable for deployment in the public cloud. If an application workload is network traffic-intensive, its performance might not be optimal if deployed in the public cloud. Also if the application communicates with other data center resources or applications, it might experience performance issues.

## Financial Advantage

- Require analysis of financial benefits in adopting cloud
- Consider CAPEX and OPEX to deploy and maintain own infrastructure versus cloud-adoption cost

Cost of Owning Infrastructure		Cloud Adoption Cost
CAPEX	OPEX	OPEX
<ul style="list-style-type: none"><li>• Servers</li><li>• Storage</li><li>• Operating system (OS)</li><li>• Application</li><li>• Network equipments</li><li>• Real estate</li></ul>	<ul style="list-style-type: none"><li>• Power and cooling</li><li>• Personnel</li><li>• Bandwidth</li><li>• Maintenance</li><li>• Support</li><li>• Backup</li></ul>	<ul style="list-style-type: none"><li>• Migration</li><li>• Compliance and security</li><li>• Subscription fee</li></ul>

A careful analysis of financial benefits provides a clear picture about the cost-savings in adopting the cloud. The analysis should compare both the Total Cost of Ownership (TCO) and the Return on Investment (ROI) in the cloud and noncloud environment and identify the potential cost benefit. While calculating TCO and ROI, organizations and individuals should consider the expenditure to deploy and maintain their own infrastructure versus cloud-adoption costs. While calculating the expenditures for owning infrastructure resources, organizations should include both the capital expenditure (CAPEX) and operation expenditure (OPEX). The CAPEX includes the cost of servers, storage, OS, application, network equipment, real estate, and so on. The OPEX includes the cost incurred for power and cooling, personnel, maintenance, backup, and so on. These expenditures should be compared with the operation cost incurred in adopting cloud computing. The cloud adoption cost includes the cost of migrating to the cloud, cost to ensure compliance and security, and usage or subscription fees. Moving applications to the cloud reduces CAPEX, except when the cloud is built on-premise.

## Selecting a Public Cloud Service Provider

- Some key questions to ask before selecting a provider:
  - ▶ How long and how well has the provider been delivering the services?
  - ▶ How well does the provider meet the organization's current and future requirements?
  - ▶ How easy is it to add or remove services?
  - ▶ How easy is it to move to another provider, when required?
  - ▶ What happens when the provider upgrades their software? Is it forced on everyone? Can you upgrade on your own schedule?
  - ▶ Does the provider offer the required security services?
  - ▶ Does the provider meet your legal and privacy requirements?
  - ▶ Does the provider have good customer service support?

The selection of the provider is important for a public cloud. Consumers need to find out how long and how well the provider has been delivering the services. They also need to determine how easy it is to add or terminate cloud services with the service provider. The consumer should know how easy it is to move to another provider, when required. They must assess how the provider fulfills the security, legal, and privacy requirements. They should also check whether the provider offers good customer service support.

## QoS Considerations

- Consumers should check whether the QoS attributes meet their requirements
- SLA is a contract between the cloud service provider and consumers that defines QoS attributes
  - ▶ Attributes examples: throughput, uptime, and so on

Cloud service providers typically mention quality of service (QoS) attributes such as throughput and uptime, along with cloud services. The QoS attributes are generally part of an SLA, which is the service contract between the provider and the consumers. The SLA serves as the foundation for the expected level of service between the consumer and the provider. Before adopting the cloud services, consumers should check whether the QoS attributes meet their requirements.

## Module 13: Cloud Computing

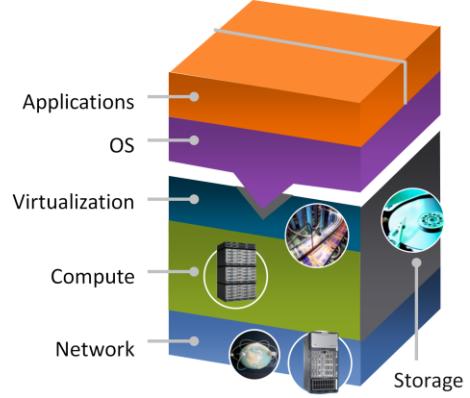
### Concept in Practice:

- Vblock

The Concept in Practice section covers Vblock.

## Vblock

- Integrated cloud infrastructure package
  - ▶ Includes compute, storage, network, and virtualization products
  - ▶ Delivered by EMC, VMware, and Cisco
- Enables building virtualized data center and cloud infrastructure
- Pre-architected, preconfigured, pretested, and ready to be deployed
  - ▶ Saves cost and deployment time



**Vblock Infrastructure Package**

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 13: Cloud Computing 41

Vblock is completely integrated cloud infrastructure offering that includes compute, storage, network, and virtualization products. These products are provided by EMC, VMware, and Cisco, who have formed a coalition to deliver Vblocks.

Vblocks enables organizations to build virtualized data centers and cloud infrastructures. Vblocks are pre-architected, preconfigured, pretested and have defined performance and availability attributes. Rather than customers buying and assembling individual cloud infrastructure components, Vblock provides a validated cloud infrastructure solution and is factory-ready for deployment and production. This saves significant cost and deployment time.

EMC Unified Infrastructure Manager (UIM) is the unified management solution for Vblocks. UIM provides a single point of management for Vblocks and manages multiple Vblocks. With UIM, cloud infrastructure services can be provisioned automatically and based on provisioning best practices.

## Module 13: Summary

Key points covered in this module:

- Characteristics of cloud computing
- Cloud services and deployment models
- Cloud computing infrastructure
- Challenges of cloud computing
- Cloud adoption considerations

This module covered the essential characteristics of cloud computing such as on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service.

It also covered various cloud service models such as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS).

This module also covered cloud deployment models such as public, private, community, and hybrid.

It also covered cloud computing infrastructure that includes physical infrastructure, virtual infrastructure, applications and platform software, and cloud management and service creation tools.

Finally, it covered the challenges of cloud computing and cloud adoption considerations.

## Check Your Knowledge – 1

- Which capability is provided by cloud computing?
  - A. Unilateral provisioning
  - B. Human interaction with service provider
  - C. Increased time-to-market
  - D. Vendor lock-in
  
- According to NIST, which is an essential cloud characteristic?
  - A. Open API
  - B. Policy-based monitoring
  - C. Measured service
  - D. Software-as-a-service

## Check Your Knowledge – 2

- What is offered in infrastructure-as-a-service?
  - A. Database Management System
  - B. Operating system
  - C. Application
  - D. Storage
- Which is a component of virtual infrastructure in cloud?
  - A. Management software
  - B. Storage array
  - C. Network identity pool
  - D. Service catalogue

## Check Your Knowledge – 3

- Which is a characteristic of cloud-optimized storage?
  - A. Server-centric
  - B. Location dependent
  - C. Secure multitenancy
  - D. Single access mechanism

This slide intentionally left blank.

# Module – 14

# Securing the Storage Infrastructure



## Module 14: Securing the Storage Infrastructure

Upon completion of this module, you should be able to:

- Describe information security framework
- Explain various storage security domains
- Discuss security implementations in SAN, NAS, and IP SAN
- Explain security in virtualized and cloud environments

This module focuses on information security framework and various storage security domains. This module also focuses on security implementation in SAN, NAS, and IP SAN. Further, this module focuses on security in virtualized and cloud environment.

# Module 14: Securing the Storage Infrastructure

## Lesson 1: Information Security Framework

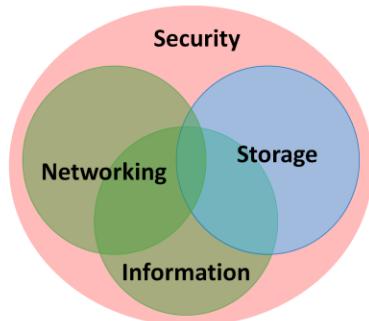
During this lesson the following topics are covered:

- Building information security framework
- Risk triad
- Security elements
- Security controls

This lesson covers building information security framework and risk triad. This lesson also covers security elements such as assets, threats and vulnerabilities. Additionally this lesson also focuses on security controls.

## Storage Security

- Process of applying information security principles and practices within the domain of storage networking technologies
- Storage security focuses on securing access to information by implementing safeguards or controls
- Storage security begins with building ‘information security framework’



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 4

Valuable information, including intellectual property, personal identities, and financial transactions, is routinely processed and stored in storage arrays, which are accessed through the network. As a result, storage is now more exposed to various security threats that can potentially damage business-critical data and disrupt critical services. Securing storage infrastructure has become an integral component of the storage management process in traditional and virtualized data centers. It is an intensive and necessary task, essential to managing and protecting vital information.

Storage security is the process of applying information security principles and practices within the domain of storage networking technologies. Storage security implements various kinds of safeguards or controls, in order to lessen the risk of an exploitation or a vulnerability in the storage network which could otherwise cause a significant impact to organization’s business. From this perspective, security is an ongoing process, not static and requires continuing revalidation and modification. Storage security begins with building a framework.

## Information Security Framework

- A systematic way of defining security requirements
- Framework should incorporate:
  - ▶ Anticipated security attacks
    - ▶ Actions that compromise the security of information
  - ▶ Security measures
    - ▶ Control designed to protect from these security attacks
- Security framework is built to achieve four security goals:
  - ▶ Confidentiality
  - ▶ Integrity
  - ▶ Availability
  - ▶ Accountability
- Securing infrastructure begins with understanding the risk

The basic information security framework is built to achieve four security goals, confidentiality, integrity, and availability (CIA) along with accountability. This framework incorporates all security standards, procedures and controls, required to mitigate threats in the storage infrastructure environment.

**Confidentiality :** Provides the required secrecy of information and ensures that only authorized users have access to data. This requires authentication of users who need to access information. Data in transit (data transmitted over cables) and data at rest (data residing on a primary storage, backup media, or in the archives) can be encrypted to maintain its confidentiality. In addition to restricting unauthorized users from accessing information, confidentiality also requires to implement traffic flow protection measures as part of the security protocol. These protection measures generally include hiding source and destination addresses, frequency of data being sent, and amount of data sent.

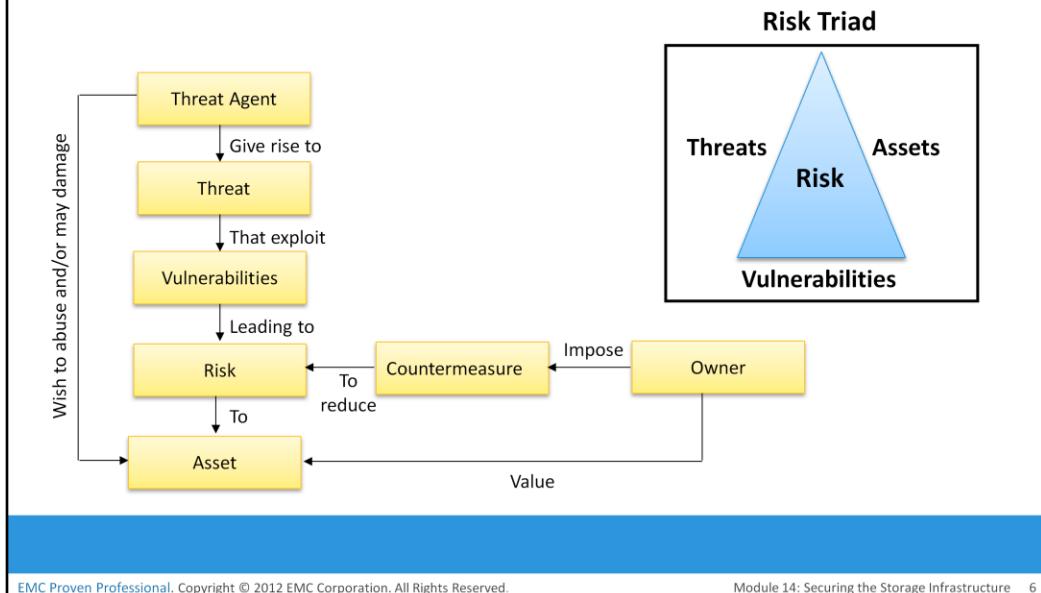
**Integrity:** Ensures that the information is unaltered. Ensuring integrity requires detection and protection against unauthorized alteration or deletion of information. Ensuring integrity stipulate measures such as error detection and correction for both data and systems.

**Availability:** This ensures that authorized users have reliable and timely access to systems, data and applications residing on these systems. Availability requires protection against unauthorized deletion of data and denial of service. Availability also implies that sufficient resources are available to provide a service.

**Accountability:** Refers to accounting for all the events and operations that take place in the data center infrastructure. The accountability service maintains a log of events that can be audited or traced later for the purpose of security.

## Risk Triad

- Defines risk in terms of threats, assets, and vulnerabilities



Risk triad defines risk in terms of threats, assets, and vulnerabilities. Risk arises when a threat agent (an attacker) uses an existing vulnerability to compromise the security services of an asset, for example, if a sensitive document is transmitted without any protection over an insecure channel, an attacker might get unauthorized access to the document and may violate its confidentiality and integrity. This may, in turn, result in business loss for the organization. In this scenario potential business loss is the risk, which arises because an attacker uses the vulnerability of the unprotected communication to access the document and tamper with it.

To manage risks, organizations primarily focus on vulnerabilities because they cannot eliminate threat agents that appear in various forms and sources to its assets. Organizations can enforce countermeasures to reduce the possibility of occurrence of attacks and the severity of their impact.

Risk assessment is the first step to determine the extent of potential threats and risks in an IT infrastructure. The process assesses risk and helps to identify appropriate controls to mitigate or eliminate risks. Based on value of assets, risk assessment helps to prioritize the investment and provisioning of security measures. To determine the probability of an adverse event occurring, threats to an IT system must be analyzed with the potential vulnerabilities and the existing security controls.

The severity of an adverse event is estimated by the impact that it may have on critical business activities. Based on this analysis, a relative value of criticality and sensitivity can be assigned to IT assets and resources. For example, a particular IT system component may be assigned a high-criticality value if an attack on this particular component can cause a complete termination of mission-critical services.

## Assets

- “Information” – the most important asset for any organization
  - ▶ Other assets include hardware, software, and network infrastructure
- Protecting assets is the primary concern
- Security considerations
  - ▶ Must provide easy access to assets for authorized users
  - ▶ Cost of securing the assets should be a fraction of the value of the assets
  - ▶ Make it difficult for potential attackers to access and compromise the assets
    - ▶ Should cost heavily to a potential attacker in terms of money, effort, and time

Information is one of the most important *assets* for any organization. Other assets include hardware, software, and other infrastructure components required to access the information. To protect these assets, organizations must develop a set of parameters to ensure the availability of the resources to authorized users and trusted networks. These parameters apply to storage resources, network infrastructure, and organizational policies.

Security methods have two objectives. The first objective is to ensure that the network is easily accessible to authorized users. It should also be reliable and stable under disparate environmental conditions and volumes of usage. The second objective is to make it difficult for potential attackers to access and compromise the system.

The security methods should provide adequate protection against unauthorized access, viruses, worms, trojans, and other malicious software programs. Security measures should also include options to encrypt critical data and disable unused services to minimize the number of potential security gaps. The security method must ensure that updates to the operating system and other software are installed regularly. At the same time, it must provide adequate redundancy in the form of replication and mirroring of the production data to prevent catastrophic data loss if there is an unexpected data compromise. For the security system to function smoothly, all users are informed about the policies governing the use of the network.

The effectiveness of a storage security methodology can be measured by two key criteria. One, the cost of implementing the system should be a fraction of the value of the protected data. Two, it should cost heavily to a potential attacker, in terms of money, effort, and time.

## Threats

- Potential attacks that can be carried out on an IT infrastructure
- Attacks can be classified as passive or active
  - ▶ Passive attacks
    - ▶ Attempt to gain unauthorized access into the system
    - ▶ Attempt to threaten the confidentiality of information
  - ▶ Active attacks
    - ▶ Attempt data modification, Denial of Service (DoS), and repudiation attacks
    - ▶ Attempt to threaten data integrity, availability, and accountability

Threats are the potential attacks that can be carried out on an IT infrastructure. These attacks can be classified as active or passive. Passive attacks are attempts to gain unauthorized access into the system. They pose threats to confidentiality of information. Active attacks include data modification, denial of service (DoS), and repudiation attacks. They pose threats to data integrity, availability, and accountability.

In a data modification attack, the unauthorized user attempts to modify information for malicious purposes. A modification attack can target the data at rest or the data in transit. These attacks pose a threat to data integrity.

Denial of service (DoS) attacks prevent legitimate users from accessing resources and services. These attacks generally do not involve access to or modification of information. Instead, they pose a threat to data availability. The intentional flooding of a network or website to prevent legitimate access to authorized users is one example of a DoS attack.

Repudiation is an attack against the accountability of information. It attempts to provide false information by either impersonating someone's identity or denying that an event or a transaction has taken place. For example, a repudiation attack may involve performing an action and eliminating any evidence that could prove the identity of the user (attacker) who performed that action. Repudiation attacks include circumventing the logging of security events or tampering with the security log to conceal the identity of the attacker.

## Vulnerabilities

- Paths that provide access to information are vulnerable to potential attacks
- Requires implementation of “defense in depth”
- Factors to consider when assessing the extent to which an environment is vulnerable:
  - ▶ Attack surface
  - ▶ Attack vectors
  - ▶ Work factor
- Managing vulnerabilities
  - ▶ Minimize the attack surface and maximize the work factor
  - ▶ Install controls (or countermeasures)

The paths that provide access to information are vulnerable to potential attacks. Each of the paths may contain various access points, which provide different levels of access to the storage resources. It is important to implement adequate security controls at all the access points on an access path. Implementing security controls at each access point of every access path is known as *defense in depth*. Defense in depth recommends using multiple security measures to reduce the risk of security threats if one component of the protection is compromised. It is also known as a “layered approach to security”. Because there are multiple measures for security at different levels and defense in depth gives additional time to detect and respond to an attack. This can reduce the scope or impact of a security breach. *Attack surface*, *attack vector*, and *work factor* are the three factors to consider when assessing the extent to which an environment is vulnerable to security threats. *Attack surface* refers to the various entry points that an attacker can use to launch an attack. Each component of a storage network is a source of potential vulnerability. An attacker can use all the external interfaces supported by that component, such as the hardware and the management interfaces, to execute various attacks. These interfaces form the attack surface for the attacker. Even unused network services, if enabled, can become a part of the attack surface.

Cont..

An *attack vector* is a step or a series of steps necessary to complete an attack. For example, an attacker might exploit a bug in the management interface to execute a snoop attack whereby the attacker can modify the configuration of the storage device to allow the traffic to be accessed from one more host. This redirected traffic can be used to snoop the data in transit. *Work factor* refers to the amount of time and effort required to exploit an attack vector. For example, if attackers attempt to retrieve sensitive information, they consider the time and effort that would be required for executing an attack on a database. This may include determining privileged accounts, determining the database schema, and writing SQL queries. Instead, based on the work factor, they may consider a less effort-intensive way to exploit the storage array by attaching to it directly and reading from the raw disk blocks.

## Security Controls

- Reduces the impact of vulnerabilities
- Any control measure should involve all the three aspects of infrastructure
  - ▶ People, process, and technology
- Controls can be technical or non-technical
  - ▶ Technical: antivirus, firewalls, and intrusion detection system
  - ▶ Non-technical: administrative policies and physical controls
- Controls are categorized as:
  - ▶ Preventive
  - ▶ Corrective
  - ▶ Detective

Having assessed the vulnerability of the environment, organizations can deploy specific control measures. Any control measures should involve all the three aspects of infrastructure: people, process and technology, and their relationship. To secure people, first step is to establish and assure their identity. Based on their identity, selective controls can be implemented for their access to data and resources. The effectiveness of any security measure is primarily governed by the process and policies. The processes should be based on a thorough understanding of risks in the environment and recognize the relative sensitivity of different types of data, the needs of various stakeholders to access the data. Without an effective process, the deployment of technology is neither cost-effective nor aligned to organizations' priorities. And finally, the technologies or controls that are deployed should ensure compliance with the processes, policies, and people for its effectiveness. These security technologies are directed at reducing vulnerability by minimizing attack surfaces and maximizing the work factors. These controls can be technical or nontechnical. Technical controls are usually implemented through computer systems, whereas nontechnical controls are implemented through administrative and physical controls. Administrative controls include security and personnel policies or standard procedures to direct the safe execution of various operations. Physical controls include setting up physical barriers, such as security guards, fences, or locks.

Cont..

Based on the roles they play, controls are categorized as preventive, detective, and corrective. The preventive control attempts to prevent an attack; the detective control detects whether an attack is in progress; and after an attack is discovered, the corrective controls are implemented. *Preventive controls* avert the vulnerabilities from being exploited and prevent an attack or reduce its impact. *Corrective controls* reduce the effect of an attack, whereas *detective controls* discover attacks and trigger preventive or corrective controls. For example, an Intrusion Detection/Intrusion Prevention System (IDS/IPS) is a detective control that determines whether an attack is underway and then attempts to stop it by terminating a network connection or invoking a firewall rule to block traffic.

# Module 14: Securing the Storage Infrastructure

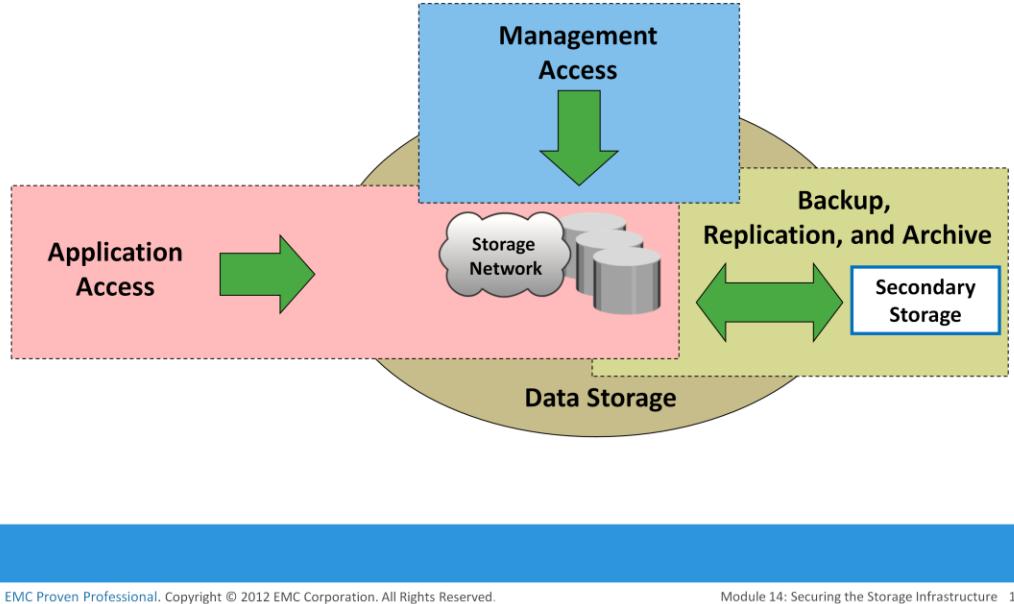
## Lesson 2: Storage Security Domains

During this lesson the following topics are covered:

- Storage security domains
- Security threats in each domain
- Controls applied to reduce the risk in each domain

This lesson covers various storage security domains such as application access, management access, and back, replication, and archive. This lesson also covers security threats and control in each domain.

## Storage Security Domains



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 14

Storage devices connected to a network raises the risk level and more exposed to security threats via networks. However, with increasing use of networking in storage environments, storage devices are becoming highly exposed to security threats from a variety of sources. Specific controls must be implemented to secure a storage networking environment. This requires a closer look at storage networking security and a clear understanding of the access paths leading to storage resources. If a particular path is unauthorized and needs to be prohibited by technical controls, ensure that these controls are not compromised. If each component within the storage network is considered a potential access point, the attack surface of all these access points must be analyzed to identify the associated vulnerabilities. To identify the threats that apply to a storage network, access paths to data storage can be categorized into three security domains: *application access*, *management access*, and *backup, replication, and archive*. Figure on the slide depicts the three security domains of a storage system environment. The first security domain involves application access to the stored data through the storage network. The second security domain includes management access to storage and interconnect devices and to the data residing on those devices. This domain is primarily accessed by storage administrators who configure and manage the environment. The third domain consists of backup, replication, and archive access. Along with the access points in this domain, the backup media also needs to be secured. To secure the storage networking environment, identify the existing threats within each of the security domains and classify the threats based on the type of security services—availability, confidentiality, integrity, and accountability. The next step is to select and implement various controls as countermeasures to the threats.

## Securing the Application Access Domain

- Protect data and access to the data

Common Threats	Available Controls	Examples
<ul style="list-style-type: none"><li>• Spoofing user or host identity</li><li>• Elevation of privileges</li><li>• Tampering with data in-flight and at rest</li><li>• Network snooping</li><li>• Denial of service</li><li>• Media theft</li></ul>	<ul style="list-style-type: none"><li>• Strong user and host authentication and authorization</li><li>• Access control to storage objects</li><li>• Data encryption</li><li>• Storage network encryption</li></ul>	<ul style="list-style-type: none"><li>• Multi-factor authentication</li><li>• RBAC, DH-CHAP</li><li>• Zoning, LUN masking</li><li>• Storage encryption</li><li>• IP-Sec, FC security protocol</li><li>• Antivirus</li><li>• Controlling physical access to data center</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 15

Access control services regulate user and host access to data. These services mitigate the threats of spoofing user identity and elevating their privileges. Both these threats affect data integrity and confidentiality. Access control mechanisms used in application access domain are user and host authentication (technical control) and authorization (administrative control). These mechanisms may lie outside the boundaries of the storage network and require various systems to interconnect with other enterprise identity management and authentication systems. NAS devices support the creation of *access control lists* that regulates user access to specific files. The Enterprise Content Management application enforces access to data by using Information Rights Management (IRM) that specifies which users have what rights to a document.

Restricting access at the host level starts with authenticating a node when it tries to connect to a network. Different storage networking technologies, such as iSCSI, FC, and IP-based storage, use various authentication mechanisms, such as Challenge-Handshake Authentication Protocol (CHAP), Fibre Channel Security Protocol (FC-SP), and IPsec, respectively, to authenticate host access. After a host has been authenticated, the next step is to specify security controls for the storage resources, such as ports, volumes, or storage pools, that the host is authorized to access. *Zoning* is a control mechanism on the switches that segments the network into specific paths to be used for data traffic; *LUN masking* determines which hosts can access which storage devices.

Cont..

It is also important to ensure that administrative controls are implemented. Regular auditing is required to ensure proper functioning of administrative controls. This is enabled by logging significant events on all participating devices. Event logs should also be protected from unauthorized access because they may fail to achieve their goals if the logged content is exposed to unauthorized modifications by an attacker.

Security controls for protecting the storage infrastructure address the threats of unauthorized tampering of data in transit that leads to a loss of data integrity, denial of service that compromises availability, and network snooping that may result in loss of confidentiality.

The security controls for protecting the network fall into two general categories: network infrastructure integrity and storage network encryption. Controls for ensuring the infrastructure integrity include a fabric switch function that ensures fabric integrity. This is achieved by preventing a host from being added to the SAN fabric without proper authorization. Storage network encryption methods include the use of IPSec for protecting IP-based storage networks, and FC-SP for protecting FC networks. In secure storage environments, root or administrator privileges for a specific device are not granted to every user. Instead, *role-based access control* (RBAC) is deployed to assign necessary privileges to users, enabling them to perform their roles. A role may represent a job function, for example, an administrator. Privileges are associated with the roles and users acquire these privileges based upon their roles. It is also advisable to consider administrative controls, such as “separation of duties,” when defining data center procedures. Management networks for storage systems should be logically separate from other enterprise networks. Finally, physical access to the device console and the cabling of FC switches must be controlled to ensure protection of the storage infrastructure.

The most important aspect of securing data is protecting data held inside the storage arrays. Threats at this level include tampering with data, which violates data integrity, and media theft, which compromises data availability and confidentiality. To protect against these threats, encrypt the data held on the storage media or encrypt the data prior to being transferred to the disk. Data should be encrypted as close to its origin as possible. If it is not possible to perform encryption on the host device, an encryption appliance can be used for encrypting data at the point of entry into the storage network. It is also critical to decide upon a method for ensuring that data deleted at the end of its lifecycle has been completely erased from the disks and cannot be reconstructed for malicious purposes. On NAS devices, adding antivirus checks and file extension controls can further enhance data integrity. In the case of CAS, use of MD5 or SHA-256 cryptographic algorithms guarantees data integrity by detecting any change in content bit patterns. In addition, the data erasure service ensures that the data has been completely overwritten by bit sequence before the disk is discarded.

## Securing the Management Access Domain

- Involves protecting administrative access and management infrastructure
- Common threats
  - ▶ Spoofing administrator's identity
  - ▶ Elevating administrative privileges
  - ▶ Network snooping and DoS
- Available controls
  - ▶ Authentication, authorization, and management access control
  - ▶ Private management network
  - ▶ Disable unnecessary network services
  - ▶ Encryption of management traffic

Management access, whether monitoring, provisioning, or managing storage resources, is associated with every device within the storage network. Implementing appropriate controls for securing storage management applications is important because the damage that can be caused by using these applications can be far more extensive.

Controlling administrative access to storage aims to safeguard against the threats of an attacker spoofing an administrator's identity or elevating privileges to gain administrative access. To protect against these threats, administrative access regulation and various auditing techniques are used to enforce accountability of users and processes. Access control should be enforced for each storage component. In some storage environments, it may be necessary to integrate storage devices with third-party authentication directories, such as Lightweight Directory Access Protocol (LDAP) or Active Directory. Security best practices stipulate that no single user should have ultimate control over all aspects of the system. It is better to assign various administrative functions by using RBAC. Auditing logged events is a critical control measure to track the activities of an administrator. However, access to administrative log files and their content must be protected. In addition, having a Security Information Management (SIM) solution supports effective analysis of the event log files.

Mechanisms to protect the management network infrastructure include encrypting management traffic, enforcing management access controls, and applying IP network security best practices. Restricting network activity and access to a limited set of hosts minimizes the threat of an unauthorized device attaching to the network and gaining access to the management interfaces. Access controls need to be enforced at the storage-array level to specify which host has management access to which array. A separate private management network is highly recommended for the management traffic. If possible, management traffic should not be mixed with either production data traffic or other LAN traffic used in the enterprise. Unused network services must be disabled on every device within the storage network. This decreases the attack surface for that device by minimizing the number of interfaces through which the device can be accessed.

## Securing Backup, Replication, and Archive Domain

- Involves protecting backup, replication, and archive infrastructure
- Common threats
  - ▶ Spoofing DR site identity
  - ▶ Tampering with data in-flight and at rest
  - ▶ Network snooping
- Available controls
  - ▶ Access control – primary to secondary storage
  - ▶ Backup encryption
  - ▶ Replication network encryption

Backup, replication, and archive is the third domain that needs to be secured against an attack. A backup involves copying the data from a storage array to backup media, such as tapes or disks. Securing backup is complex and is based on the backup software that accesses the storage arrays. It also depends on the configuration of the storage environments at the primary and secondary sites, especially with remote backup solutions performed directly on a remote tape device or using array-based remote replication.

Organizations must ensure that the disaster recovery (DR) site maintains the same level of security for the backed up data. Protecting the backup, replication, and archive infrastructure requires addressing several threats, including spoofing the legitimate identity of a DR site, tampering with data, network snooping, DoS attacks, and media theft. Such threats represent potential violations of integrity, confidentiality, and availability. In a remote backup solution where the storage components are separated by a network, the threats at the transmission layer need to be countered. Otherwise, an attacker can spoof the identity of the backup server and request the host to send its data. The unauthorized host claiming to be the backup server may lead to a remote backup being performed to an unauthorized and unknown site. In addition, attackers can use the DR network connection to tamper with data, snoop the network, and create a DoS attack against the storage devices.

The physical threat of a backup tape being lost, stolen, or misplaced, especially if the tapes contain highly confidential information, is another type of threat. Backup-to-tape applications are vulnerable to severe security implications if they do not encrypt data while backing it up.

## Module 14: Securing the Storage Infrastructure

### Lesson 3: Security Implementations in Storage Networking

During this lesson the following topics are covered:

- SAN security implementations
- NAS security implementations
- IP SAN security implementations

This lesson covers various security implementation in SAN, NAS and IP SAN environment.

## Security Implementation in SAN

- Common SAN security mechanisms are:
  - ▶ LUN masking and zoning
  - ▶ Securing FC switch ports
  - ▶ Switch-wide and fabric-wide access control
  - ▶ Logical partitioning of a fabric: VSAN

Traditional FC SANs enjoy an inherent security advantage over IP-based networks. An FC SAN is configured as an isolated private environment with fewer nodes than an IP network. Consequently, FC SANs impose fewer security threats. However, this scenario has changed with converged network, storage consolidation, driving rapid growth and necessitating designs for large, complex SANs that span multiple sites across the enterprise. Today, no single comprehensive security solution is available for FC SANs. Many FC SAN security mechanisms have evolved from their counterpart in IP networking, thereby bringing in matured security solutions.

*Fibre Channel Security Protocol* (FC-SP) standards (T11 standards), published in 2006, align security mechanisms and algorithms between IP and FC interconnects. These standards describe protocols used to implement security measures in a FC fabric, among fabric elements and N\_Ports within the fabric. They also include guidelines for authenticating FC entities, setting up session keys, negotiating the parameters required to ensure frame-by-frame integrity and confidentiality, and establishing and distributing policies across an FC fabric.

LUN masking and zoning, security in FC switch port, switch-wide and fabric-wide access control, and logical partitioning of a fabric (Virtual SAN) are the most commonly used SAN security methods. A stronger variant of LUN masking may sometimes be offered whereby masking can be done on the basis of source FC address. It offers a mechanism to lock down the FC address of a given node port to its WWN.

## Securing FC Switch Ports

- Port binding
  - ▶ Restricts devices that can attach to a particular switch port
    - ▶ Allows only the corresponding switch port to connect to a node for fabric access
- Port lockdown and port lockout
  - ▶ Restricts a switch port's type of initialization
- Persistent port disable
  - ▶ Prevents a switch port from being enabled even after a switch reboot

Apart from zoning and LUN masking, additional security mechanisms, such as port binding, port lockdown, port lockout, and persistent port disable, can be implemented on switch ports.

*Port binding* : Limits the devices that can attach to a particular switch port and allows only the corresponding switch port to connect to a node for fabric access. Port binding mitigates but does not eliminate WWPN spoofing.

*Port lockdown and port lockout*: Restrict a switch port's type of initialization. Typical variants of port lockout ensure that the switch port cannot function as an E-Port and cannot be used to create an ISL, such as a rogue switch. Some variants ensure that the port role is restricted to only F-Port, E-Port, or a combination of these.

*Persistent port disable*: Prevents a switch port from being enabled even after a switch reboot.

## Switch-wide and Fabric-wide Access Control

- Access control lists (ACLs)
  - ▶ Include device connection and switch connection control policies
    - ▶ Device connection control policy specifies which HBAs, storage ports can be connected to a particular switch
    - ▶ Switch connection control policy prevents unauthorized switches to join a particular switch
- Fabric Binding
  - ▶ Prevents unauthorized switch from joining a fabric
- Role-based access control (RBAC)
  - ▶ Enables assigning roles to users that explicitly specify access rights

As organizations grow their SANs locally or over longer distances, there is a greater need to effectively manage SAN security. Network security can be configured on the FC switch by using access control lists (ACLs) and on the fabric by using fabric binding.

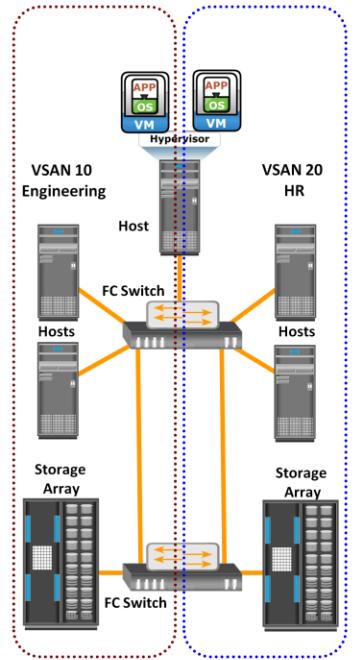
ACLs incorporate the device connection control and switch connection control policies. The device connection control policy specifies which HBAs and storage ports can be connected to a particular switch, preventing unauthorized devices from accessing it. Similarly, the switch connection control policy specifies which switches are allowed to be connected to a particular switch, preventing unauthorized switches from joining it.

Fabric binding prevents an unauthorized switch from joining any existing switch in the fabric. It ensures that authorized membership data exists on every switch and any attempt to connect any switch in the fabric by using an ISL causes the fabric to segment.

Role-based access control provides additional security to a SAN by preventing unauthorized activity on the fabric for management operations. It enables the security administrator to assign roles to users that explicitly specify privileges or access rights after logging into the fabric. For example, the zone admin role can modify the zones on the fabric, whereas a basic user may view only fabric-related information, such as port types and logged-in nodes.

## Logical Partitioning of a Fabric: VSAN

- Enables the creation of multiple logical SANs over a common physical SAN
- Fabric events in one VSAN are not propagated to the others
- Zoning should be configured for each VSAN



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

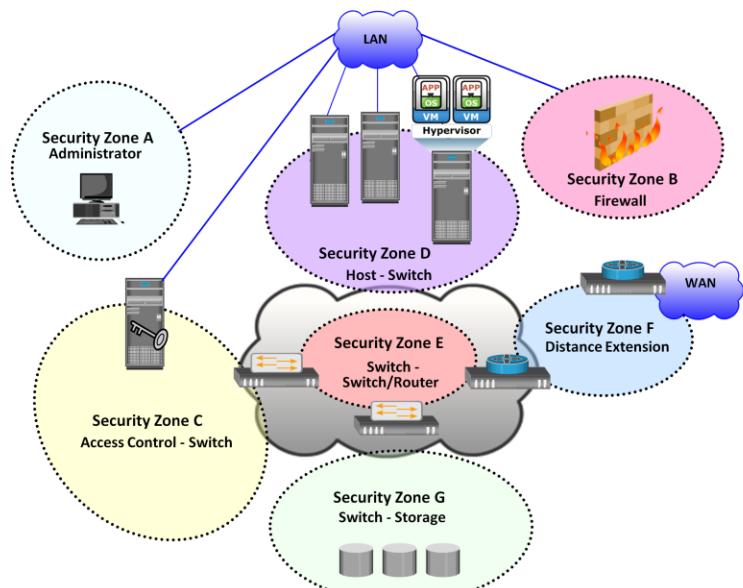
Module 14: Securing the Storage Infrastructure 24

VSANs enable the creation of multiple logical SANs over a common physical SAN. They provide the capability to build larger consolidated fabrics and still maintain the required security and isolation between them. Figure on the slide depicts logical partitioning of a fabric using VSAN.

The SAN administrator can create distinct VSANs by populating each of them with switch ports. In the example, the switch ports are distributed over two VSANs: 10 and 20—for the Engineering and HR divisions, respectively. Although they share physical switching gear with other divisions, they can be managed individually as standalone fabrics. Zoning should be done for each VSAN to secure the entire physical SAN. Each managed VSAN can have only one active zone set at a time.

VSANs minimize the impact of fabric wide disruptive events because management and control traffic on the SAN—which may include RSCNs, zone set activation events, and more—does not traverse VSAN boundaries. Therefore, VSANs are a cost-effective alternative for building isolated physical fabrics. They contribute to information availability and security by isolating fabric events and providing authorization control within a single fabric.

## SAN Security Architecture: Defense-in-Depth



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 25

Storage networking environments are a potential target for unauthorized access, theft, and misuse because of the vastness and complexity of these environments. Therefore, security strategies are based on the *defense in depth* concept, which recommends multiple integrated layers of security. This ensures that the failure of one security control will not compromise the assets under protection. Figure on the slide illustrates various levels (zones) of a storage networking environment that must be secured. FC SANs not only suffer from certain risks and vulnerabilities that are unique, but also share common security problems associated with physical security and remote administrative access. In addition to implementing SAN-specific security measures, organizations must simultaneously leverage other security implementations in the enterprise.

Comprehensive list of protection strategies that must be implemented in various security zones are listed below:

### **Zone A (Authentication at the Management Console):**

- (a) Restrict management LAN access to authorized users (lock down MAC addresses)
- (b) implement VPN tunneling for secure remote access to the management LAN
- (c) use two-factor authentication for network access

### **Zone B (Firewall)**

Block inappropriate traffic by (a) filtering out addresses that should not be allowed on your LAN, and (b) screening for allowable protocols block ports that are not in use

Cont..

### **Zone C (Access Control-Switch)**

Authenticate users/administrators of FC switches using Remote Authentication Dial In User Service (RADIUS) and DH-CHAP (Diffie-Hellman Challenge Handshake Authentication Protocol)

### **Zone D (Host to switch)**

Restrict Fabric access to legitimate hosts by implementing (a) ACLs: Known HBAs can connect on specific switch ports only; and (b) a secure zoning method, such as port zoning (also known as hard zoning)

### **Zone E (Switch to Switch/Switch to Router)**

Protect traffic on fabric by (a) using E\_Port authentication; (b) encrypting the traffic in transit; and (c) implementing FC switch controls and port controls

### **Zone F (Distance Extension)**

Implement encryption for in-flight data (a) FC-SP for long-distance FC extension, and (b) IPSec for SAN extension via FCIP

### **Zone G (Switch to Storage)**

Protect the storage arrays on your SAN via (a) WWPN-based LUN masking and (b) S\_ID locking: masking based on source Fibre Channel address

## Security Implementation in NAS

- Permissions and ACLs
  - ▶ Protection to NAS resources by restricting access
- Other authentication and authorization mechanisms
  - ▶ Kerberos and Directory services
    - ▶ Implemented to verify the identity of network users and define their privileges
  - ▶ Firewalls
    - ▶ To protect the storage infrastructure from unauthorized access and malicious attacks

NAS is open to multiple exploits, including viruses, worms, unauthorized access, snooping, and data tampering. Various security mechanisms are implemented in NAS to secure data and the storage networking infrastructure.

Permissions and ACLs form the first level of protection to NAS resources by restricting accessibility and sharing. These permissions are deployed over and above the default behaviors and attributes associated with files and folders. In addition, various other authentication and authorization mechanisms, such as Kerberos and directory services, are implemented to verify the identity of network users and define their privileges. Similarly, firewalls protect the storage infrastructure from unauthorized access and malicious attacks.

## NAS File Sharing: Windows ACLs

- Types of ACLs
  - ▶ Discretionary access control lists (DACL)
    - ▶ Commonly referred to as ACL and used to determine access control
  - ▶ System access control lists (SACL)
    - ▶ Determine what access needs to be audited if auditing is enabled
- Object Ownership
  - ▶ Object owner has hard-coded rights to that object
  - ▶ Child objects within a parent object automatically inherit the ACLs of parent object
- Security identifiers (SIDs)
  - ▶ SIDs uniquely identify a user or a user group
  - ▶ ACLs use SIDs to control access to the objects

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 28

Windows supports two types of ACLs: *discretionary access control lists* (DACLs) and *system access control lists* (SACLs). The DACL, commonly referred to as the ACL, that determines access control. The SACL determines what accesses need to be audited if auditing is enabled.

In addition to these ACLs, Windows also supports the concept of object ownership. The owner of an object has hard-coded rights to that object, and these rights do not need to be explicitly granted in the SACL. The owner, SACL, and DACL are all statically held as attributes of each object. Windows also offers the functionality to inherit permissions, which allows the child objects existing within a parent object to automatically inherit the ACLs of the parent object.

ACLs are also applied to directory objects known as security identifiers (SIDs). These are automatically generated by a Windows server or domain when a user or group is created, and they are abstracted from the user. In this way, though a user may identify his login ID as "User1," it is simply a textual representation of the true SID, which is used by the underlying operating system. Internal processes in Windows refer to an account's SID rather than the account's username or group name while granting access to an object. ACLs are set by using the standard Windows Explorer GUI but can also be configured with CLI commands or other third-party tools.

## NAS File Sharing: UNIX Permissions

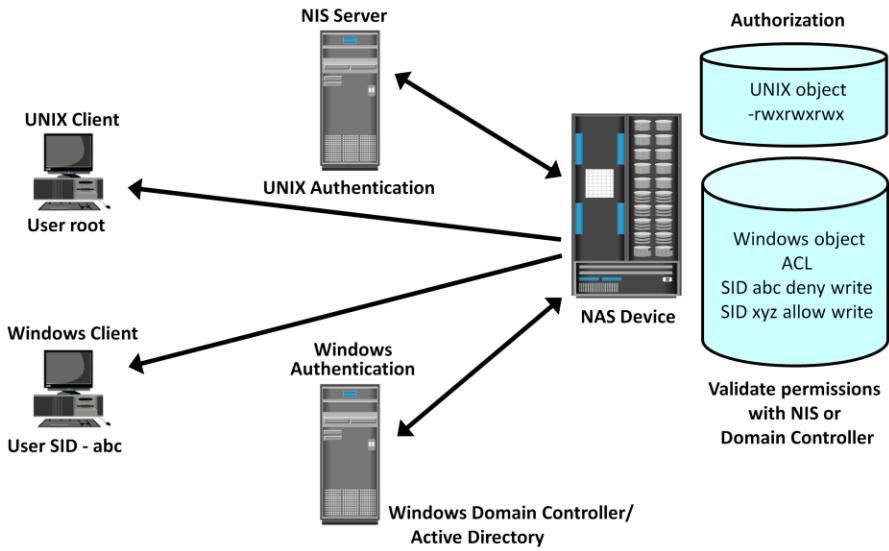
- UNIX permissions specify what can be done to a file and by whom
  - ▶ Common permissions: Read/Write/Execute
- Every file and directory (folder) has three ownership relations:
  - ▶ Rights for the file owner
  - ▶ Rights for the group the user belongs to
  - ▶ Rights for all other users

For the UNIX operating system, a *user* is an abstraction that denotes a logical entity for assignment of ownership and operation privileges for the system. A user can be either a person or a system operation. A UNIX system is only aware of the privileges of the user to perform specific operations on the system and identifies each user by a user ID (UID) and a username, regardless of whether it is a person, a system operation, or a device.

In UNIX, users can be organized into one or more groups. The concept of group serves the purpose to assign sets of privileges for a given resource and sharing them among many users that need them. For example, a group of people working on one project may need the same permissions for a set of files.

UNIX permissions specify the operations that can be performed by any ownership relation with respect to a file. In simpler terms, these permissions specify what the owner can do, what the owner group can do, and what everyone else can do with the file. For any given ownership relation, three bits are used to specify access permissions. The first bit denotes read (r) access, the second bit denotes write (w) access, and the third bit denotes execute (x) access. Because UNIX defines three ownership relations (Owner, Group, and All), a triplet (defining the access permission) is required for each ownership relationship, resulting in nine bits. Each bit can be either set or clear. When displayed, a set bit is marked by its corresponding operation letter (r, w, or x), a clear bit is denoted by a dash (-), and all are put in a row, such as rwxr-xr-x. In this example, the owner can do anything with the file, but group owners and the rest of the world can read or execute only. When displayed, a character denoting the mode of the file may precede this nine-bit pattern. For example, if the file is a directory, it is denoted as “d”; and if it is a link, it is denoted as “l.”

## Authentication and Authorization



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 30

In a file-sharing environment, NAS devices use standard file-sharing protocols, NFS and CIFS. Therefore, authentication and authorization are implemented and supported on NAS devices in the same way as in a UNIX or Windows file-sharing environment.

Authentication requires verifying the identity of a network user and therefore involves a login credential lookup on a Network Information System (NIS) server in a UNIX environment. Similarly, a Windows client is authenticated by a Windows domain controller that houses the Active Directory. The Active Directory uses LDAP to access information about network objects in the directory and Kerberos for network security. NAS devices use the same authentication techniques to validate network user credentials. Figure on the slide depicts the authentication process in a NAS environment.

Authorization defines user privileges in a network. The authorization techniques for UNIX users and Windows users are quite different. UNIX files use mode bits to define access rights granted to owners, groups, and other users, whereas Windows uses an ACL to allow or deny specific rights to a particular user for a particular file.

Although NAS devices support both of these methodologies for UNIX and Windows users, complexities arise when UNIX and Windows users access and share the same data. If the NAS device supports multiple protocols, the integrity of both permission methodologies must be maintained. NAS device vendors provide a method of mapping UNIX permissions to Windows and vice versa, so a multiprotocol environment can be supported.

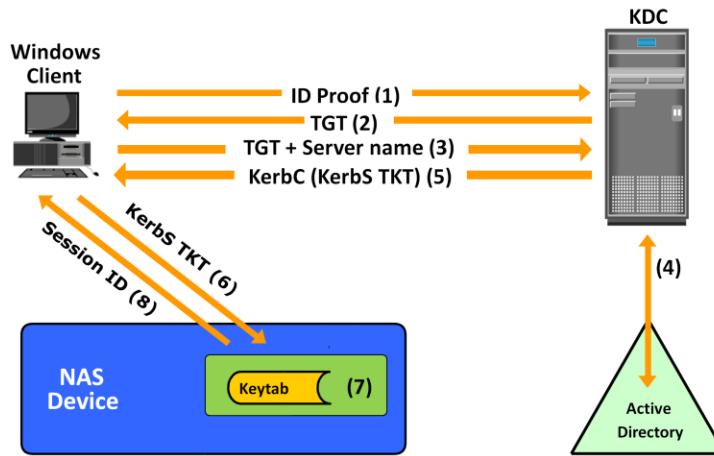
## Kerberos – Network Authentication Protocol

- Uses secret-key cryptography
- A client can prove its identity to a server (and vice versa) across an insecure network connection
- Kerberos client
  - ▶ An entity that gets a service ticket for a Kerberos service
- Kerberos server
  - ▶ Refers to the Key Distribution Center (KDC)
  - ▶ Implements the Authentication Service (AS) and the Ticket Granting Service (TGS)

Kerberos is a network authentication protocol, which is designed to provide strong authentication for client/server applications by using secret-key cryptography. It uses cryptography so that a client and server can prove their identity to each other across an insecure network connection. After the client and server have proven their identities, they can choose to encrypt all their communications to ensure privacy and data integrity.

In Kerberos, authentications occur between clients and servers. The client gets a ticket for a service and the server decrypts this ticket by using its secret key. Any entity, user, or host that gets a service ticket for a Kerberos service is called a *Kerberos client*. The term *Kerberos server* generally refers to the Key Distribution Center (KDC). The KDC implements the Authentication Service (AS) and the Ticket Granting Service (TGS). The KDC has a copy of every password associated with every principal, so it is absolutely vital that the KDC remain secure. In Kerberos, users and servers for which a secret key is stored in the KDC database are known as *principals*.

## Kerberos Authorization



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

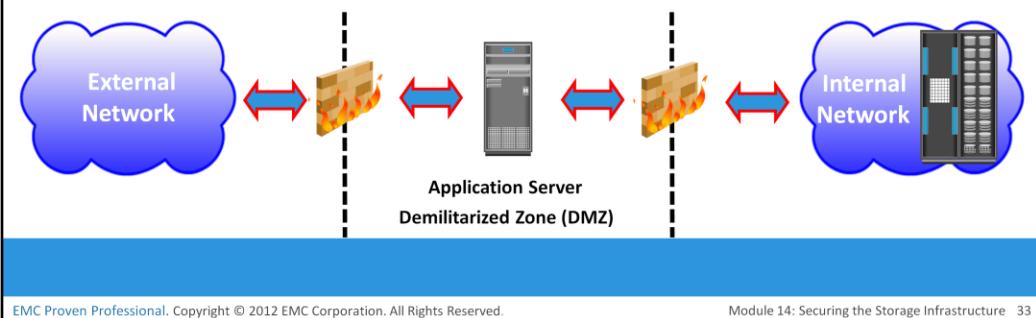
Module 14: Securing the Storage Infrastructure 32

The Kerberos authorization process shown in figure on the slide includes the following steps:

1. The user logs on to the workstation in the Active Directory domain (or forest) using an ID and a password. The client computer sends a request to the AS running on the KDC for a Kerberos ticket. The KDC verifies the user's login information from Active Directory.
2. The KDC responds with an encrypted Ticket Granting Ticket (TGT) and an encrypted session key. TGT has a limited validity period. TGT can be decrypted only by the KDC, and the client can decrypt only the session key.
3. When the client requests a service from a server, it sends a request, consisting of the previously generated TGT, encrypted with the session key and the resource information to the KDC.
4. The KDC checks the permissions in Active Directory and ensures that the user is authorized to use that service.
5. The KDC returns a service ticket to the client. This service ticket contains fields addressed to the client and to the server hosting the service.
6. The client then sends the service ticket to the server that houses the required resources.
7. The server, in this case the NAS device, decrypts the server portion of the ticket and stores the information in a keytab file. As long as the client's Kerberos ticket is valid, this authorization process does not need to be repeated. The server automatically allows the client to access the appropriate resources.
8. A client-server session is now established. The server returns a session ID to the client, which tracks the client activity, such as file locking, as long as the session is active.

## Network Layer Firewalls

- Firewalls are implemented in NAS environments
  - ▶ To protect against security threats in IP network
  - ▶ To examine network packets and compare them to a set of configured security rules
    - ▶ Packets that are not authorized by a security rule are dropped
- Demilitarized Zone (DMZ)
  - ▶ To secure internal assets while allowing Internet-based access to various resources



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

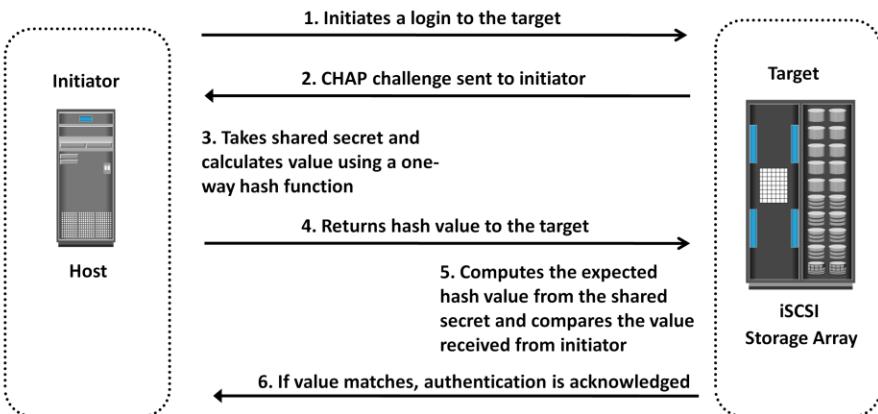
Module 14: Securing the Storage Infrastructure 33

Because NAS devices utilize the IP protocol stack, they are vulnerable to various attacks initiated through the public IP network. Network layer firewalls are implemented in NAS environments to protect the NAS devices from these security threats. These network-layer firewalls can examine network packets and compare them to a set of configured security rules. Packets that are not authorized by a security rule are dropped and not allowed to continue to the destination. Rules can be established based on a source address (network or host), a destination address (network or host), a port, or a combination of those factors (source IP, destination IP, and port number). The effectiveness of a firewall depends on how robust and extensive the security rules are. A loosely defined rule set can increase the probability of a security breach.

A demilitarized zone (DMZ) is commonly used in networking environments. A DMZ provides a means to secure internal assets while allowing Internet-based access to various resources. In a DMZ environment, servers that need to be accessed through the Internet are placed between two sets of firewalls. Application-specific ports, such as HTTP or FTP, are allowed through the firewall to the DMZ servers. However, no Internet-based traffic is allowed to penetrate the second set of firewalls and gain access to the internal network. The servers in the DMZ may or may not be allowed to communicate with internal resources. In such a setup, the server in the DMZ is an Internet-facing web application accessing data stored on a NAS device, which may be located on the internal private network. A secure design would serve only data to internal and external applications through the DMZ.

## Security Implementation in IP SAN: CHAP

- Challenge-Handshake Authentication Protocol (CHAP)
  - ▶ Provides a method for initiators and targets to authenticate each other by utilizing a secret code



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

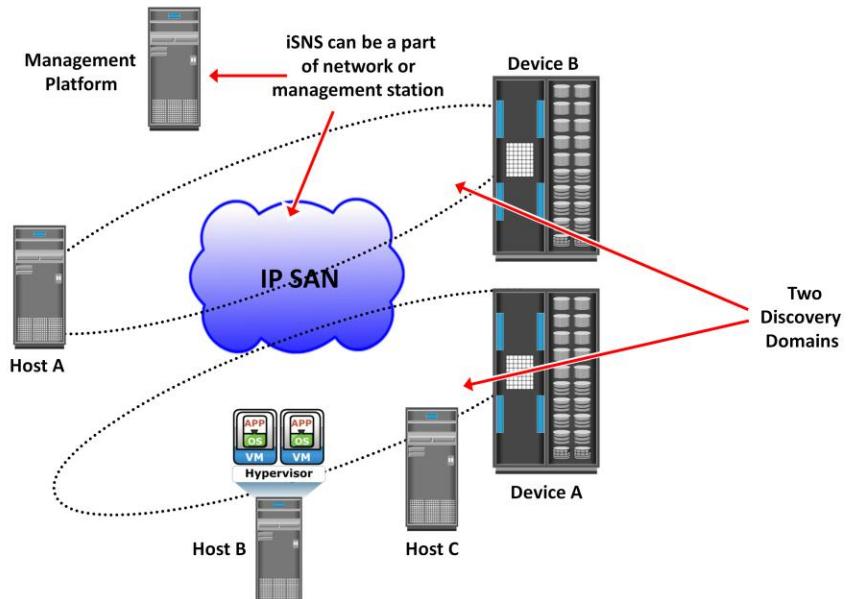
Module 14: Securing the Storage Infrastructure 34

The *Challenge-Handshake Authentication Protocol* (CHAP) is a basic authentication mechanism that has been widely adopted by network devices and hosts. CHAP provides a method for initiators and targets to authenticate each other by utilizing a secret code or password. CHAP secrets are usually random secrets of 12 to 128 characters. The secret is never exchanged directly over the communication channel; rather, a one-way hash function converts it into a hash value, which is then exchanged. A hash function, using the MD5 algorithm, transforms data in such a way that the result is unique and cannot be changed back to its original form. Figure on the slide depicts the CHAP authentication process.

If the initiator requires reverse CHAP authentication, the initiator authenticates the target by using the same procedure. The CHAP secret must be configured on the initiator and the target. A CHAP entry, composed of the name of a node and the secret associated with the node, is maintained by the target and the initiator.

The same steps are executed in a two-way CHAP authentication scenario. After these steps are completed, the initiator authenticates the target. If both authentication steps succeed, then data access is allowed. CHAP is often used because it is a fairly simple protocol to implement and can be implemented across a number of disparate systems.

## Securing IPSAN with iSNS Discovery Domains



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 14: Securing the Storage Infrastructure 35

*iSNS discovery domains* function in the same way as FC zones. Discovery domains provide functional groupings of devices in an IP-SAN. For devices to communicate with one another, they must be configured in the same discovery domain. State change notifications (SCNs) inform the iSNS server when devices are added to or removed from a discovery domain.

## Module 14: Securing the Storage Infrastructure

### Lesson 4: Security in Virtualized and Cloud Environments

During this lesson the following topics are covered:

- Security concerns
- Security measures

This lesson covers an overview of security in virtualized and cloud environment. This lesson also covers security concerns and measures in virtualized and cloud environment.

## Security in Virtualized and Cloud Environments

- These environments have additional threats due to multitenancy and lack of control over the cloud resources
- Virtualization-specific security concerns are common for all cloud models
- In public clouds, there are additional security concerns, which demand specific countermeasures
  - ▶ Clients have less control to enforce security measures in public clouds
  - ▶ Difficult for cloud service provider(CSP) to meet the security needs of all the clients

This module, so far, focused only on the security threats and measures in a traditional data center. These threats and measures are also applicable to information storage in virtualized and cloud environments. However, virtualized and cloud computing environments pose additional threats to an organization's data due to multitenancy and lack of control over the cloud resources. A public cloud has more security concerns compared to a private cloud and demands additional counter measures. This is because in a public cloud, cloud users (consumers) usually have limited control over resources, and therefore, enforcement of security mechanisms for consumers is comparatively difficult. From a security perspective, both consumers and cloud service providers (CSP) have several security concerns and face multiple threats.

## Security Concerns

- Multitenancy
  - ▶ Enables multiple independent tenants to be serviced using the same set of storage resources
    - ▶ Co-location of multiple VMs in a single server and sharing the same resources increase the attack surface
- Velocity of attack
  - ▶ Any existing security threat in the cloud spreads more rapidly and has larger impact than that in the traditional data center
- Information assurance and data privacy

Organizations are rapidly adopting virtualization and cloud computing, however they have some security concerns. These key security concerns are multitenancy, velocity of attack, information assurance, and data privacy.

*Multitenancy*, by virtue of virtualization, enables multiple independent tenants to be serviced using the same set of storage resources. In spite of the benefits offered by multitenancy, it is still a key security concern for users and service providers. Colocation of multiple VMs in a single server and sharing the same resources increase the attack surface. It may happen that business critical data of one tenant is accessed by other competing tenants who run applications using the same resources.

*Velocity-of-attack* refers to a situation in which any existing security threat in the cloud spreads more rapidly and has a larger impact than that in the traditional data center environments. *Information assurance* for users ensures confidentiality, integrity, and availability of data in the cloud. Also the cloud user needs assurance that all the users operating on the cloud are genuine and access the data only with legitimate rights and scope.

*Data privacy* is also a major concern in a virtualized and cloud environment. A CSP needs to ensure that Personally Identifiable Information (PII) about its clients is legally protected from any unauthorized disclosure.

## Security Measures

- Securing compute
  - ▶ Securing physical server, VMs, and hypervisor
- Securing network
  - ▶ Virtual firewall
    - ▶ Provides packet filtering and monitoring of the VM-to-VM traffic
  - ▶ DMZ and data encryption
- Securing storage
  - ▶ Access control and data encryption
  - ▶ Use separate LUNs for VM configuration files and VM data
  - ▶ Segregate VM traffic from management traffic

Major threats to storage systems in virtualized and cloud environments arise due to compromises at compute, network, and physical security levels. This is because access to storage systems is provided by using compute and network infrastructure. Therefore, adequate security measures should be in place at the compute and network levels to ensure storage security.

Securing a compute infrastructure includes enforcing the security of the physical server, hypervisor, VM, and guest OS (OS running within a virtual machine). Physical server security involves implementing user authentication and authorization mechanisms. These mechanisms identify users and provide access privileges on the server. To minimize the attack surface on the server, unused hardware components, such as NICs, USB ports, or drives, should be removed or disabled.

A *hypervisor* is a single point of security failure for all the VMs running on it. Rootkits and malware installed on a hypervisor make detection difficult for the antivirus software installed on the guest OS. To protect against attacks, security-critical hypervisor updates should be installed regularly. Further, the hypervisor management system must also be protected. Malicious attacks and infiltration to the management system can impact all the existing VMs and allow attackers to create new VMs. Access to the management system should be restricted to authorized administrators. Furthermore, there must be a separate firewall installed between the management system and the rest of the network.

Cont..

*VM isolation* and *hardening* are some of the common security mechanisms to effectively safeguard a VM from an attack. VM isolation helps to prevent a compromised guest OS from impacting other guest OSs. VM isolation is implemented at the hypervisor level. Apart from isolation, VMs should be hardened against security threats. Hardening is a process to change the default configuration to achieve greater security.

The key security measures that minimize vulnerabilities at the network layer are firewall, intrusion detection, demilitarized zone (DMZ), and encryption of data-in-flight. In a virtualized and cloud environment, a firewall can also protect hypervisors and VMs. For example, if remote administration is enabled on a hypervisor, access to all the remote administration interfaces should be restricted by a firewall. A firewall also secure VM-to-VM traffic. This firewall service can be provided using a *Virtual Firewall* (VF) running on the hypervisor. VF gives visibility and control over the VM traffic and enforces policies at the VM level. DMZ and data encryption are also deployed as security measures in the virtualized and cloud environments. However, these deployments work in the same way as in the traditional data center.

Common security mechanisms that protect storage include the following:

- Access control methods to regulate which users and processes access the data on the storage systems
- Zoning and LUN masking
- Encryption of data-at-rest (on the storage system) and data-in-transit. Data encryption should also include encrypting backups and storing encryption keys separately from the data.
- Data shredding that removes the traces of the deleted data

Apart from these mechanisms, isolation of different types of traffic using VSANs further enhances the security of storage systems. In the case of storage utilized by hypervisors, additional security steps are required to protect the storage. Storage for hypervisors using clustered file systems supporting multiple VMs may require separate LUNs for VM components and VM data.

## Module 14: Securing the Storage Infrastructure

### Concept in Practice

- RSA security products
- VMware vShield

The concepts in practice section covers various security products of RSA and VMware. The product includes RSA SecureID, RSA Identity and Access Management, RSA Data Protection Manager, and VMware vShield.

## RSA Security Products

- RSA SecurID
  - ▶ Provides two-factor authentication
    - ▶ Based on something a user knows (a password or PIN) and something a user has (an authenticator device)
    - ▶ Authenticator device automatically changes passwords every 60 seconds
- RSA Identity and Access Management
  - ▶ Provides identity, security, and access-control management for physical, virtual, and cloud-based environments
- RSA Data Protection Manager
  - ▶ Enables deployment of encryption, tokenization, and enterprise key management

RSA SecurID two-factor authentication provides an added layer of security to ensure that only valid users have access to systems and data. RSA SecurID is based on something a user knows (a password or PIN) and something a user has (an authenticator device). It generates a new one-time password code every 60 seconds, making it difficult for anyone other than the genuine user to input the correct token code at any given time. To access their resources, users combine their secret Personal Identification Number (PIN) with the token code that appears on their SecurID authenticator display at that given time. The result is a unique, one-time password used to assure a user's identity.

The RSA Identity and Access Management product provides identity, security, and access-control management for physical, virtual, and cloud-based environments through access management. It enables trusted identities to freely and securely interact with systems and access. The RSA Identity and Access Management family has two products: *RSA Access Manager* and *RSA Federated Identity Manager*. RSA Access Manager enables organizations to centrally manage authentication and authorization policies for a large number of users, online web portals, and application resources. RSA Federated Identity Manager enables end users to collaborate with business partners, outsourced service providers, and supply-chain partners or across multiple offices or agencies all with a single identity and logon.

RSA Data Protection Manager enables deployment of encryption, tokenization, and enterprise key management simply and affordably. RSA Data Protection Manager family is composed of two products: *Application Encryption and Tokenization* and *Enterprise Key Management*. Application Encryption and Tokenization with RSA Data Protection Manager helps to achieve compliance with regulations related to PII. Enterprise key management is an easy-to-use management tool for encrypting keys at the database, file server and storage layers.

## VMware vShield

- VMware vShield family includes three products
  - ▶ vShield App
    - ▶ Hypervisor-based application-aware firewall solution
    - ▶ Observes network activity between virtual machines
  - ▶ vShield Edge
    - ▶ Provides comprehensive perimeter network security
    - ▶ Deployed as a virtual appliance and serves as a network security gateway for all the hosts
    - ▶ Provides many services including firewall, VPN, and DHCP
  - ▶ vShield Endpoint
    - ▶ Consists of a hardened special security VM with a third party antivirus software

The VMware vShield family includes three products: *vShield App*, *vShield Edge*, and *vShield Endpoint*.

VMware vShield App is a hypervisor-based application-aware firewall solution. It protects applications in a virtualized environment from network-based threats by providing visibility into network communications and enforcing granular policies with security groups. VMware vShield App observes network activity between virtual machines to define and refine firewall policies and secure business processes through detailed reporting of application traffic.

VMware vShield Edge provides comprehensive perimeter network security for a virtualized environment. It is deployed as a virtual appliance and serves as a network security gateway for all the hosts within the virtualized environment. It provides many services including firewall, VPN, and Dynamic Host Configuration Protocol (DHCP) services.

VMware vShield Endpoint consists of a hardened special security VM with a third party antivirus software. VMware vShield Endpoint streamlines and accelerates antivirus and antimalware deployment because antivirus engine and signature files are updated only within the special security VM. VMware vShield Endpoint improves VM performance by offloading file scanning and other tasks from VMs to the security VM. It prevents antivirus storms and bottlenecks associated with multiple simultaneous antivirus and antimalware scans and updates. It also satisfies audit requirements with detailed logging of antivirus and antimalware activities.

## Module 14: Summary

Key points covered in this module:

- Information security framework
- Storage security domains
- Controls that can be deployed against identified threats in each domain
- SAN security architecture
- Protection mechanisms in SAN, NAS, and IP SAN environments
- Security in virtualized and cloud environments

This module covered information security framework and various storage security domains. This module also covered security implementation in SAN, NAS, and IP SAN. Further this module covered security in virtualized and cloud environment.

The basic security framework for storage is built around the four primary services of security: confidentiality, integrity, availability, and accountability.

To identify the threats that apply to a storage network, access paths to data storage can be categorized into three security domains: *application access, management access, and backup, replication, and archive*.

Storage networking environments are potential target for unauthorized access, theft, and misuse because of the vastness and complexity of these environments. Therefore, security strategies are based on the *defense-in-depth* concept, which recommends multiple integrated layers of security.

LUN masking and zoning, security in FC switch port, switch-wide and fabric-wide access control, and logical partitioning of a fabric (Virtual SAN) are the most commonly used SAN security methods. Permissions and ACLs, Kerberos and directory services, and firewalls are the common security mechanisms implemented in NAS. CHAP and iSNS discovery domain are the security measures implemented in IP SAN environment.

Virtualized and cloud computing environments pose additional threats to an organization's data due to multitenancy and lack of control over the cloud resources.

## Check Your Knowledge – 1

- Which attack involves performing an action and eliminating evidence that could prove the identity of the attacker?
  - A. Denial of service
  - B. Eavesdropping
  - C. Repudiation
  - D. Snooping
- What is a role of Active Directory in Kerberos authentication?
  - A. Implements the authentication service and ticket granting service
  - B. Verifies the session ID when the client-server session is established
  - C. Verifies user's login information
  - D. Maintains the security key for the servers

## Check Your Knowledge – 2

- How vulnerabilities can be reduced in an IT environment?
  - A. By maximizing attack surface and minimizing work factor
  - B. By minimizing attack surface and maximizing work factor
  - C. By maximizing both attack surface and work factor
  - D. By minimizing both attack surface and work factor
- Which SAN security mechanism prevents a switch port from being enabled even after a switch reboot?
  - A. Port lockdown
  - B. Persistent port disable
  - C. Port binding
  - D. Port zoning

## Check Your Knowledge – 3

- Which security mechanism ensures that a port can only be initialized with a specific port type?
  - A. Port lockdown
  - B. Persistent port disable
  - C. Port binding
  - D. Port zoning

This slide intentionally left blank.

# Module – 15

# Managing the Storage Infrastructure



## Module 15: Managing the Storage Infrastructure

Upon completion of this module, you should be able to:

- List the key storage infrastructure components that are monitored
- List the key monitoring parameters
- Describe the storage management activities
- Describe the storage infrastructure management challenges and their solutions
- Describe the information lifecycle management (ILM) strategy

This module focuses on management of storage infrastructure. This module lists the key storage infrastructure components that are monitored. The module also lists the monitoring parameters. Further, the module describes the storage management activities. It also describes the storage infrastructure management challenges and their solutions. Finally, this module describes the information lifecycle management strategy.

# Module 15: Managing the Storage Infrastructure

## Lesson 1: Monitoring the Storage Infrastructure

During this lesson the following topics are covered:

- Key storage infrastructure components that are monitored
- Monitoring parameters
- Types of alerts

This lesson covers the storage infrastructure components that are monitored. It also covers the monitoring parameters for storage infrastructure components. Further, it details the three types of alerts. Finally, the module provides monitoring examples.

## Storage Infrastructure Management

- Managing storage infrastructure is key to ensure continuity of business
  - ▶ Establishing management processes and implementing appropriate tools are essential to meet service levels
  - ▶ Virtualization have changed the storage infrastructure management paradigm
- Monitoring is the most important aspect that forms the basis for storage management

Unprecedented growth of information, proliferation of applications, complexity of business processes, and requirements of 24x7 availability of information have put increasingly higher demands on the IT infrastructure.

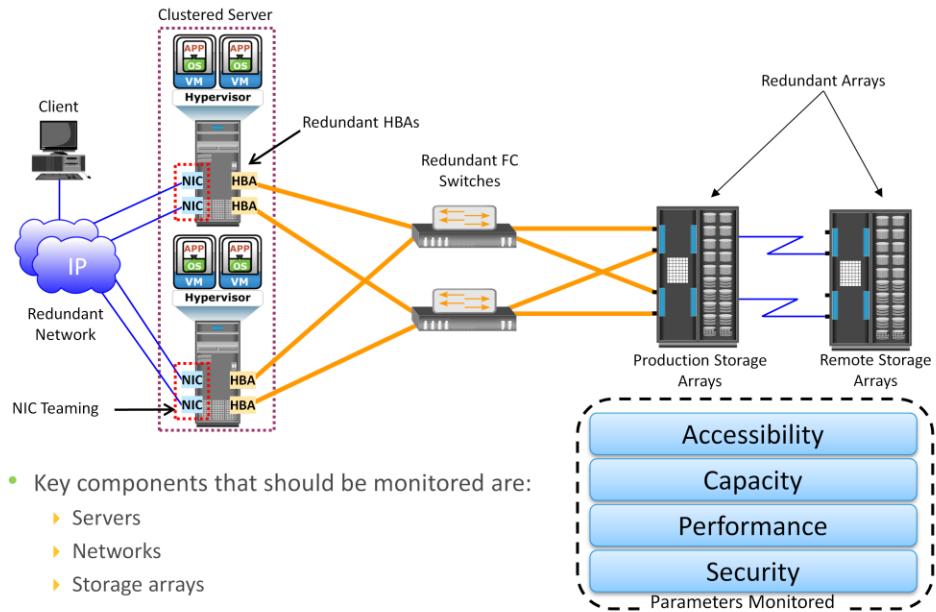
Managing storage infrastructure efficiently is a key that enables organizations to address these challenges and ensures continuity of business.

Comprehensive storage infrastructure management requires the implementation of intelligent tools and robust processes to meet the required service levels. These tools enable performance tuning, data protection, access control, centralized auditing, and meeting compliance requirements. They also ensure the consolidation and better utilization of existing resources, thereby limiting the need for excessive ongoing investment on infrastructure. The management process defines procedures for efficient handling of various operations, such as incident, problem, and change requests. It is imperative to manage not just the individual components, but also the infrastructure end to end due to the components' interdependency.

Storage infrastructure management is also composed of strategies, such as *information lifecycle management* (ILM) that optimizes the storage cost while meeting the service levels. ILM helps to manage information based on its value to the business.

Managing the storage infrastructure requires performing various activities, including accessibility, capacity, performance, and security management. All of these activities are interrelated and should be considered to maximize the return on investment. Virtualization technologies have dramatically changed the storage infrastructure management paradigm.

## Monitoring Storage Infrastructure



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 5

Monitoring is one of the most important aspects that forms the basis for managing storage infrastructure resources. Monitoring provides the performance and accessibility status of various components. It also enables administrators to perform essential management activities. Monitoring also helps to analyze the utilization and consumption of various storage infrastructure resources. This analysis facilitates capacity planning, forecasting, and optimal use of these resources. Storage infrastructure environment parameters such as heating, ventilating and air-conditioning (HVAC) are also monitored.

The key storage infrastructure components that should be monitored are:

- Servers
- Network
- Storage arrays

These components could be physical or virtualized. Each of these components should be monitored for accessibility, capacity, performance and security.

## Monitoring Parameters

- Accessibility
  - ▶ To identify failure of any component that may lead to service unavailability or degraded performance
- Capacity
  - ▶ To ensure availability of adequate amount of resources and prevent service unavailability or degraded performance
- Performance
  - ▶ To evaluate efficiency and utilization of components and identify bottlenecks
- Security
  - ▶ To ensure confidentiality, integrity, and availability of storage infrastructure

**Accessibility** refers to the availability of a component to perform its desired operation during a specified time period. Monitoring the accessibility of hardware components (for example, a port, an HBA, or a disk drive) or software component (for example, a database instance) involves checking their availability status by reviewing the alerts generated from the system. For example, a port failure might result in a chain of availability alerts.

A storage infrastructure uses redundant components to avoid a single point of failure. Failure of a component might cause an outage that affects application availability, or it might cause performance degradation even though accessibility is not compromised. Continuously monitoring for expected accessibility of each component and reporting any deviation helps the administrator to identify failing components and plan corrective action to maintain SLA requirements.

Cont..

*Capacity* refers to the amount of storage infrastructure resources available. Examples of capacity monitoring include examining the free space available on a file system or a RAID group, the mailbox quota allocated to users, or the numbers of ports available on a switch. Inadequate capacity leads to degraded performance or even application/service unavailability. Capacity monitoring ensures uninterrupted data availability and scalability by averting outages before they occur. For example, if 90 percent of the ports are utilized in a particular SAN fabric, this could indicate that a new switch might be required if more arrays and servers need to be installed on the same fabric. Capacity monitoring usually leverages analytical tools to perform trend analysis. These trends help to understand future resource requirements and provide an estimation of the time required to deploy them.

*Performance monitoring* evaluates how efficiently different storage infrastructure components are performing and helps to identify bottlenecks. Performance monitoring measures and analyzes behavior in terms of response time or the capability to perform at a certain predefined level. It also deals with the utilization of resources, which affects the way resources behave and respond. Performance measurement is a complex task that involves assessing various components on several interrelated parameters. The number of I/Os performed by a disk, application response time, network utilization, and server-CPU utilization are examples of performance parameters that are monitored.

Monitoring a storage infrastructure for security helps to track and prevent unauthorized access, whether accidental or malicious. *Security monitoring* helps to track unauthorized configuration changes of storage infrastructure elements. For example, security monitoring tracks and reports the initial zoning configuration performed and all the subsequent changes. Security monitoring also detects unavailability of information to authorized users due to security breach. Physical security of a storage infrastructure can also be continuously monitored using badge readers, biometric scans, or video cameras.

## Components Monitored – Host

- Accessibility
  - ▶ Hardware components such as HBAs, NICs, and internal disks
  - ▶ Status of various processes/applications
- Capacity
  - ▶ File system utilization
  - ▶ Database: table space/log space utilization
  - ▶ User quota
- Performance
  - ▶ CPU and memory utilization
  - ▶ Transaction response time
- Security
  - ▶ Authentication and authorization
  - ▶ Physical security (data center access)

Hosts, networks, and storage are the components within the storage environment that should be monitored for accessibility, capacity, performance, and security.

The accessibility of a host depends on the availability status of the hardware components and the software processes running on it. For example, a host's NIC failure might cause inaccessibility of the host to its user. Server clustering is a mechanism that provides high availability if a server failure occurs. In a server virtualization environment, multiple virtual machines (VMs) share a pool of resources. These resources are dynamically reallocated, which ensures better accessibility and utilization of the resources.

Monitoring the file system utilization of a host is important to ensure that sufficient storage capacity is available to the applications. Running out of file system space disrupts application availability. Monitoring helps estimate the file system's growth rate and predict when it will reach 100 percent. Accordingly, the administrator can extend (manually or automatically) the file system's space proactively to prevent application outage. Use of virtual provisioning technology enables efficient management of storage capacity requirements but is highly dependent on capacity monitoring.

Cont..

Host performance monitoring mainly involves a status check on the utilization of various server resources, such as CPU and memory. For example, if a server running an application is experiencing 80 percent of CPU utilization continuously, it suggests that the server may be running out of processing power, which can lead to degraded performance and slower response time. Administrators can take several actions to correct the problem, such as upgrading or adding more processors, shifting the workload to different servers, and restricting the number of simultaneous client access. In a virtualized environment, additional CPU and memory may be allocated dynamically from the pool, if available, to meet performance requirements.

Security monitoring on servers involves tracking of login failures and execution of unauthorized applications or software processes. Proactive measures against unauthorized access to the servers are based on the threat identified. For example, an administrator can block user access if multiple login failures are logged.

## Components Monitored – Network

- Accessibility
  - ▶ Physical components such as switches and ports
  - ▶ Logical components such as zones
- Capacity
  - ▶ Interswitch links and port utilization
- Performance
  - ▶ Assess individual component performance and help to identify network bottlenecks
    - ▶ Monitoring port performance and link utilization
- Security
  - ▶ Unauthorized changes to the fabric
  - ▶ Login failures

Storage networks need to be monitored to ensure proper communication between the server and the storage array. Uninterrupted access to data over the storage network depends on the accessibility of the physical and logical components in the storage network. The physical components of a storage network include switches, ports, cables, and power supplies. The logical components include constructs, such as zones. Any failure in the physical or logical components causes data unavailability. For example, errors in zoning, such as specifying the wrong WWN of a port, result in failure to access that port, which potentially prevents access from a host to its storage.

*Capacity monitoring* in a storage network involves monitoring the availability of ports on a switch, the number of available ports in the entire fabric, the utilization of the interswitch links, individual ports, and each interconnect device in the fabric. Capacity monitoring provides all the required inputs for future planning and optimization of fabric resources.

Monitoring the performance of the storage network is useful in assessing individual component performance and helps to identify network bottlenecks. For example, monitoring port performance involves measuring the receive or transmit link utilization metrics, which indicates how busy the switch port is. Heavily used ports can cause queuing of I/Os on the server, which results in poor performance.

For IP networks, monitoring the performance includes monitoring network latency, packet loss, bandwidth utilization for I/O, network errors, packet retransmission rates and collisions.

Storage network security monitoring provides information about any unauthorized change to the configuration of the fabric—for example, changes to the zone policies that can affect data security. Login failures and unauthorized access to switches for performing administrative changes should be logged and monitored continuously.

## Components Monitored – Storage

- Accessibility
  - ▶ Hardware components such as controllers and disks
  - ▶ Software processes such as replication processes
- Capacity
  - ▶ Capacity utilization and consumption trend
- Performance
  - ▶ Utilization rates of storage array components
  - ▶ I/O response time, cache utilization
- Security
  - ▶ Access control and physical security (data center access)

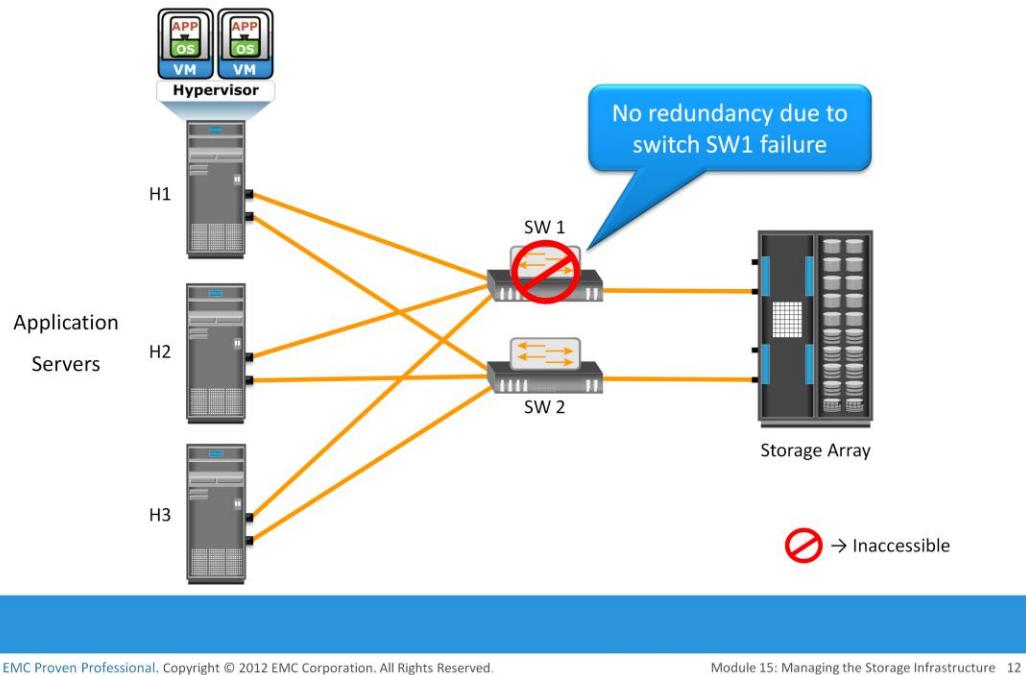
The accessibility of the storage array should be monitored for its hardware components and various processes. Storage arrays are typically configured with redundant components and therefore individual component failure does not usually affect their accessibility. However, failure of any process in the storage array can disrupt or compromise business continuity operations. For example, the failure of a replication task affects disaster recovery capabilities. Some storage arrays provide the capability to send messages to the vendor's support center if hardware or process failures occur, referred to as a *call home*.

Capacity monitoring of a storage array enables the administrator to respond to storage needs preemptively based on capacity utilization and consumption trends. Information about unconfigured and unallocated storage space is also required to decide whether a new server can be allocated storage capacity from the array.

A storage array can be monitored using a number of performance metrics, such as utilization rates of the various storage array components, I/O response time, and cache utilization. For example, an over utilized storage array component might lead to performance degradation.

A storage array is usually a shared resource, which may be exposed to security threats. Monitoring security helps to track unauthorized configuration of the storage array and ensures that only authorized users are allowed to access it.

## Accessibility Monitoring Example: Switch Failure



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 12

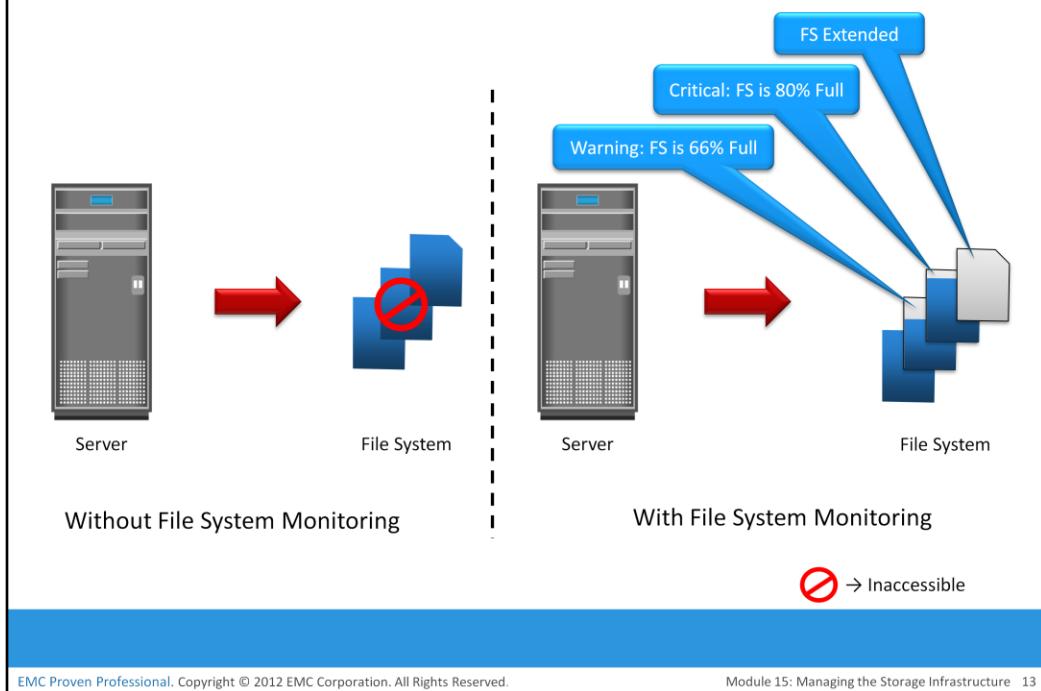
A storage infrastructure requires implementation of an end-to-end solution to actively monitor all the parameters of its critical components. Early detection and preemptive alerting ensure uninterrupted services from critical assets. In addition, the monitoring tool should analyze the impact of a failure and deduce the root cause of symptoms. Next few slides illustrates examples of monitoring storage infrastructure components for accessibility, capacity, performance, and security.

Failure of any component might affect the accessibility of one or more components due to their interconnections and dependencies. Consider an implementation in a storage infrastructure with three servers: H1, H2, and H3. All the servers are configured with two HBAs, each connected to the production storage array through two switches, SW1 and SW2, as shown on slide. All the servers share two storage ports on the storage array. Multipathing software has also been installed on all the three servers.

If one of the switches, SW1 fails, the multipathing software initiates a path failover, and all the servers continue to access data through the other switch, SW2. However, due to absence of redundant switch, a second switch failure could result in inaccessibility of the array. Monitoring for accessibility enables detecting the switch failure and helps administrator to take corrective action before another failure occurs.

In most cases, the administrator receives symptom alerts for a failing component and can initiate actions before the component fails.

## Capacity Monitoring Example: File System Space

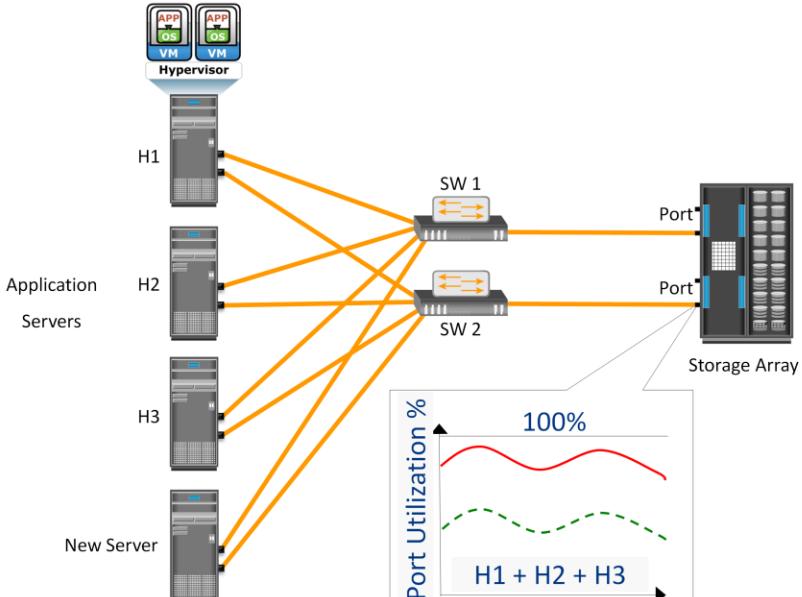


EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 13

This example illustrates the importance of monitoring the file system capacity on file servers. Figure on the slide illustrates the environment of a file system when full and that results in application outage when no capacity monitoring is implemented. Monitoring can be configured to issue a message when thresholds are reached on the file system capacity. For example, when the file system reaches 66 percent of its capacity, a warning message is issued, and a critical message is issued when the file system reaches 80 percent of its capacity. This enables the administrator to take action to extend the file system before it runs out of capacity. Proactively monitoring the file system can prevent application outages caused due to lack of file system space.

## Performance Monitoring Example: Array Port Utilization



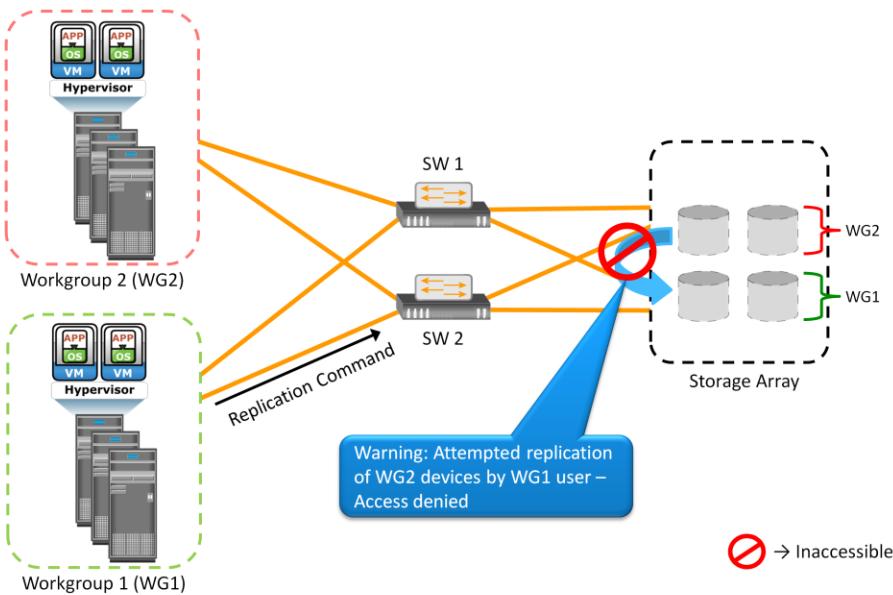
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 14

This example illustrates the importance of monitoring performance on storage arrays. In this example, servers H1, H2, and H3 (with two HBAs each) are connected to the storage array through switch SW1. The three servers share the same storage ports on the storage array to access LUNs. A new server running an application with a high work load must be deployed to share the same storage port as H1, H2, and H3.

Monitoring array port utilization ensures that the new server does not adversely affect the performance of the other servers. In this example, utilization of the shared storage port is shown by the solid and dotted lines in the graph. If the port utilization prior to deploying the new server is close to 100 percent, then deploying the new server is not recommended because it might impact the performance of the other servers. However, if the utilization of the port prior to deploying the new server is closer to the dotted line, then there is room to add a new server.

## Security Monitoring Example: Storage Array



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 15

This slide illustrates the importance of monitoring security in a storage array. In this example, the storage array is shared between two workgroups, WG1 and WG2. The data of WG1 should not be accessible by WG2 and vice versa. A user from WG1 might try to make a local replica of the data that belongs to WG2. If this action is not monitored or recorded, it is difficult to track such a violation of security protocols. Conversely, if this action is monitored, a warning message can be sent to prompt a corrective action or at least enable discovery as part of regular auditing operations.

## Alerts

- It is an integral part of monitoring
- It keeps administrators informed on the status of components and processes
- Monitoring tools enable administrators to assign different severity levels for different events
  - ▶ Classified as information, warning, and fatal alerts

Type of alerts	Description	Example
Information	<ul style="list-style-type: none"><li>• Provide useful information</li><li>• Does not require administrator intervention</li></ul>	<ul style="list-style-type: none"><li>• Creation of zone or LUN</li></ul>
Warning	<ul style="list-style-type: none"><li>• Require administrative attention</li></ul>	<ul style="list-style-type: none"><li>• FS becoming full</li><li>• Soft media errors</li></ul>
Fatal	<ul style="list-style-type: none"><li>• Require immediate attention</li></ul>	<ul style="list-style-type: none"><li>• Component failure</li><li>• Power failure</li></ul>

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 16

Alerting of events is an integral part of monitoring. Alerting keeps administrators informed about the status of various components and processes—for example, conditions such as failure of power, disks, memory, or switches, which can impact the availability of services and require immediate administrative attention. Other conditions, such as a file system reaching a capacity threshold or a soft media error on disks, are considered warning signs and may also require administrative attention.

Monitoring tools enable administrators to assign different severity levels for different conditions based on the impact of the alerted condition. Whenever a condition with a particular severity level occurs, an alert is sent to the administrator, a script is triggered, or an incident ticket is opened to initiate a corrective action. Alert classifications can range from information alerts to fatal alerts. *Information alerts* provide useful information but do not require any intervention by the administrator. The creation of a zone or LUN is an example of an information alert. *Warning alerts* require administrative attention so that the alerted condition is contained and does not affect accessibility. For example, if an alert indicates that the number of soft media errors on a disk is approaching a predefined threshold value, the administrator can decide whether the disk needs to be replaced. *Fatal alerts* require immediate attention because the condition might affect overall performance or availability. For example, if a disk fails, the administrator must ensure that it is replaced quickly.

Continuous monitoring, with automated alerting, enables administrators to respond to failures quickly and proactively. Alerting provides information that helps administrators prioritize their response to events.

# Module 15: Managing the Storage Infrastructure

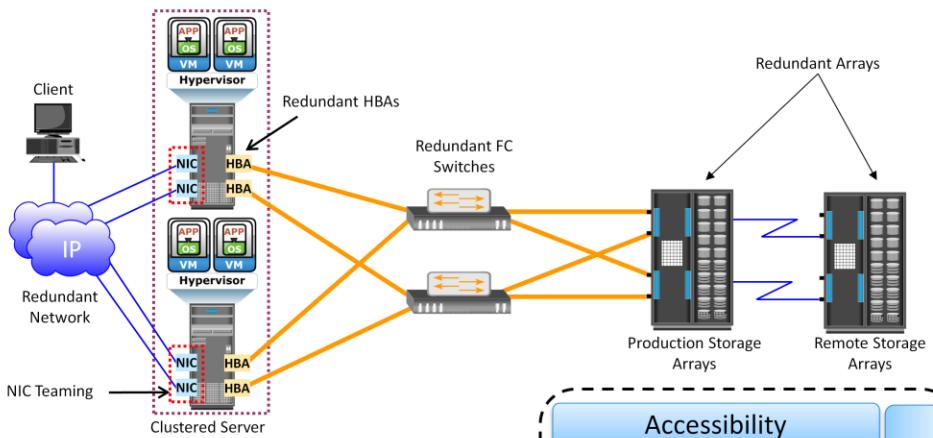
## Lesson 2: Managing the Storage Infrastructure

During this lesson the following topics are covered:

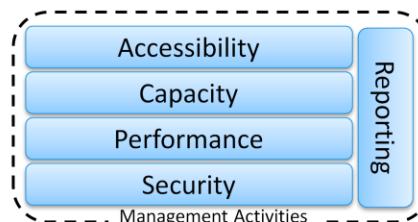
- Storage infrastructure management activities
- Storage infrastructure management in virtualized environment
- Storage infrastructure management challenges
- Ideal solution for storage infrastructure management

This lesson details storage infrastructure management activities, and management in virtualized environment. It covers storage management examples and challenges. Finally, it describes the ideal solution for storage infrastructure management.

## Storage Infrastructure Management Activities



- Key components that should be managed are:
  - ▶ Servers
  - ▶ Networks
  - ▶ Storage arrays



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 18

The pace of information growth, proliferation of applications, heterogeneous infrastructure, and stringent service-level requirements have resulted in increased complexity of managing storage infrastructures. However, the emergence of storage virtualization and other technologies, such as data deduplication and compression, thin provisioning, federated storage access, and storage tiering, have enabled administrators efficiently manage storage resources.

The key storage infrastructure management activities performed in a data center can be broadly categorized into availability management, capacity management, performance management, security management, and reporting.

# Availability Management

## Availability Management

The goal of availability management is to ensure that the availability requirements of all the components and services are constantly met.

- Key activity is to provision redundancy at all levels
- Example of availability management tasks are:
  - ▶ Installing two or more HBAs per server
  - ▶ Installing multipathing software
  - ▶ Deploying clustered server
  - ▶ Configuring RAID
  - ▶ Deploying redundant fabric
  - ▶ Configuring data backup and replication

A critical task in availability management is establishing a proper guideline based on defined service levels to ensure availability. *Availability management* involves all availability-related issues for components or services to ensure that service levels are met. A key activity in availability management is to provision redundancy at all levels, including components, data, or even site levels. For example, when a server is deployed to support a critical business function, it requires high availability. This is generally accomplished by deploying two or more HBAs, multipathing software, and server clustering. The server must be connected to the storage array using at least two independent fabrics and switches that have built-in redundancy. Provision RAID-protected LUNs to the server using at least two front-end ports. In addition, the storage arrays should have built-in redundancy for various components and should support local and remote replication.

# Capacity Management

## Capacity Management

The goal of capacity management is to ensure adequate availability of resources, based on their service level requirements.

- Example of capacity management activities are:
  - ▶ Storage provisioning
  - ▶ Enforcing capacity quota for users
  - ▶ Capacity consumption trend analysis
- Technologies such as data deduplication, compression, and virtual provisioning help to manage storage capacity efficiently

The goal of *capacity management* is to ensure adequate availability of resources based on their service level requirements. Capacity management also involves optimization of capacity based on the cost and future needs. Capacity management provides capacity analysis that compares allocated storage to forecasted storage on a regular basis. It also provides trend analysis based on the rate of consumption, which must be rationalized against storage acquisition and deployment timetables. Storage provisioning is an example of capacity management. It involves activities, such as creating RAID sets and LUNs, and allocating them to the host. Enforcing capacity quotas for users is another example of capacity management. Provisioning a fixed amount of space for their files restricts users from exceeding the allocated capacity.

Technologies, such as data deduplication and compression, have reduced the amount of data to be backed up and thereby reduced the amount of storage capacity to be managed.

# Performance Management

## Performance Management

The goal of performance management is to ensure the optimal operational efficiency of all components.

- Key activities are:
  - ▶ Fine tuning for performance enhancement
  - ▶ Identifying performance bottlenecks
- Example of performance management activities are:
  - ▶ Configuring multiple paths
  - ▶ Choosing appropriate RAID type and cache configuration

*Performance management* ensures the optimal operational efficiency of all components. Performance analysis is an important activity that helps to identify the performance of storage infrastructure components. This analysis provides information on whether a component meets expected performance levels.

Several performance management activities need to be performed when deploying a new application or server in the existing storage infrastructure. Every component must be validated for adequate performance capabilities as defined by the service levels. For example, to optimize the expected performance levels, activities on the server, such as the volume configuration, database design or application layout, configuration of multiple HBAs, and intelligent multipathing software, must be fine-tuned. The performance management tasks on a SAN include designing and implementing sufficient ISLs in a multiswitch fabric with adequate bandwidth to support the required performance levels. The storage array configuration tasks include selecting the appropriate RAID type, LUN layout, front-end ports, back-end ports, and cache configuration, when considering the end-to-end performance.

# Security Management

## Security Management

The goal of security management is to ensure confidentiality, integrity, and availability of information.

- It prevents unauthorized access and configuration of storage infrastructure components
- Examples of security management tasks are:
  - ▶ Managing user accounts
  - ▶ Configuring zoning and LUN masking
  - ▶ Configuring encryption services
  - ▶ Installing antivirus and firewalls
  - ▶ Auditing of event logs

The key objective of the *security management* activity is to ensure confidentiality, integrity, and availability of information in both virtualized and nonvirtualized environments. Security management prevents unauthorized access and configuration of storage infrastructure components. For example, while deploying an application or a server, the security management tasks include managing the user accounts and access policies that authorize users to perform role-based activities. The security management tasks in a SAN environment include configuration of zoning to restrict an unauthorized HBA from accessing specific storage array ports. Similarly, the security management task on a storage array includes LUN masking that restricts a host from accessing a defined set of LUNs.

## Reporting

- Involves gathering information from various components or processes and generating reports
- Commonly used reports are:
  - ▶ Capacity planning report
  - ▶ Configuration and asset management reports
  - ▶ Performance report
  - ▶ Chargeback report

*Reporting* on a storage infrastructure involves keeping track and gathering information from various components/processes. This information is compiled to generate reports for trend analysis, capacity planning, chargeback, and performance. Capacity planning reports contain current and historic information about the utilization of storage, file systems, database tablespace, ports, and so on. Configuration and asset management reports include details about device allocation, local or remote replicas, and fabric configuration. This report also lists all the equipment, with details, such as their purchase date, lease status, and maintenance records. Chargeback reports contain information about the allocation or utilization of storage infrastructure components by various departments or user groups. Performance reports provide details about the performance of various storage infrastructure components.

## Storage Infrastructure Management in Virtualized Environment

- Virtualization technology has enabled managing storage infrastructure efficiently
  - ▶ Virtualization at storage layer
    - ▶ Example: virtual provisioning of LUN
  - ▶ Virtualization at network layer
    - ▶ Example: VLAN and VSAN
  - ▶ Virtualization at compute layer
    - ▶ Example: Virtual machines, memory virtualization

Virtualization technology has dramatically changed the complexity of storage infrastructure management. In fact, flexibility and ease of management is one of the key drivers for wide adoption of virtualization at all layers of the IT infrastructure.

At the storage layer, storage virtualization has enabled dynamic migration of data and extension of storage volumes. Due to dynamic extension, storage volumes can be expanded nondisruptively to meet both capacity and performance requirements. Because virtualization breaks the bond between the storage volumes presented to the host and its physical storage, data can be migrated both within and across data centers without any downtime. This has made the administrators' tasks easier while reconfiguring the physical environment. Virtual storage provisioning is another tool that has changed the infrastructure management cost and complexity scenario. In conventional provisioning, storage capacity is provisioned upfront in anticipation of future growth. Because growth is uneven, some users or applications find themselves running out of capacity, whereas others have excess capacity that remains underutilized. Use of virtual provisioning can address this challenge and make capacity management less challenging. In virtual provisioning, storage is allocated from the shared pool to hosts on-demand. This improves the available capacity utilization, and thereby reduces capacity management complexities. Virtualization has also contributed to network management efficiency. VSANs and VLANs made the administrators' job easier, by isolating different networks logically using management tools rather than physically separating them. Disparate virtual networks can be created on a single physical network, and reconfiguration of nodes can be done quickly without any physical changes. It has also addressed some of the security issues that might exist in a conventional environment. On the host side, compute virtualization has made host deployment, reconfiguration, and migration easier than physical environment. Compute, application, and memory virtualization have not only improved the provisioning, but also contributed to the high availability of resources.

## Storage Multitenancy

- Enables multiple tenants to share the same storage resources provided by a single landlord (resource provider)
- Security and service level assurance are key concerns in a multitenant storage environment
- Secure storage multitenant environment should follow the four pillars of multitenancy
  - ▶ Secure separation
  - ▶ Service assurance
  - ▶ Availability
  - ▶ Management

Multiple tenants sharing the same resources provided by a single landlord (resource provider) is called multitenancy. Two common examples of multitenancy are multiple virtual machines sharing the same server hardware through the use of hypervisor running on the server, and multiple user applications using the same storage platform. Multitenancy is not a new concept; however, it has become a topic of much discussion due to the rise in popularity of cloud deployments as shared infrastructure is a core component of any cloud strategy.

As with any shared services, security and service level assurance are key concern in a multitenant storage environment. Secure multitenancy means that no tenant can access another tenant's data. To achieve this, any storage deployment should follow the four pillars of multitenancy:

- **Secure Separation:** This enables data path separation across various tenants in a multitenant environment. At the storage layer, this pillar can be divided into four basic requirements: separation of data at rest, address space separation, authentication and name service separation, and separation of data access.
- **Service Assurance:** Consistent and reliable service levels are integral to storage multitenancy. Service assurance plays an important role in providing service levels that can be unique to each tenant.
- **Availability:** High availability ensures a resilient architecture that provides fault tolerance and redundancy. This is even more critical when IT infrastructure is shared by multiple tenants – as the impact of any outage is magnified.
- **Management:** This includes provisions that allow a landlord to manage basic infrastructure while delegating management responsibilities to tenants for the resources that they interact with day to day. This concept is known as balancing the provider (landlord) in-control with the tenant in-control capabilities.

## Storage Management Example 1 – Storage Allocation to a New Server

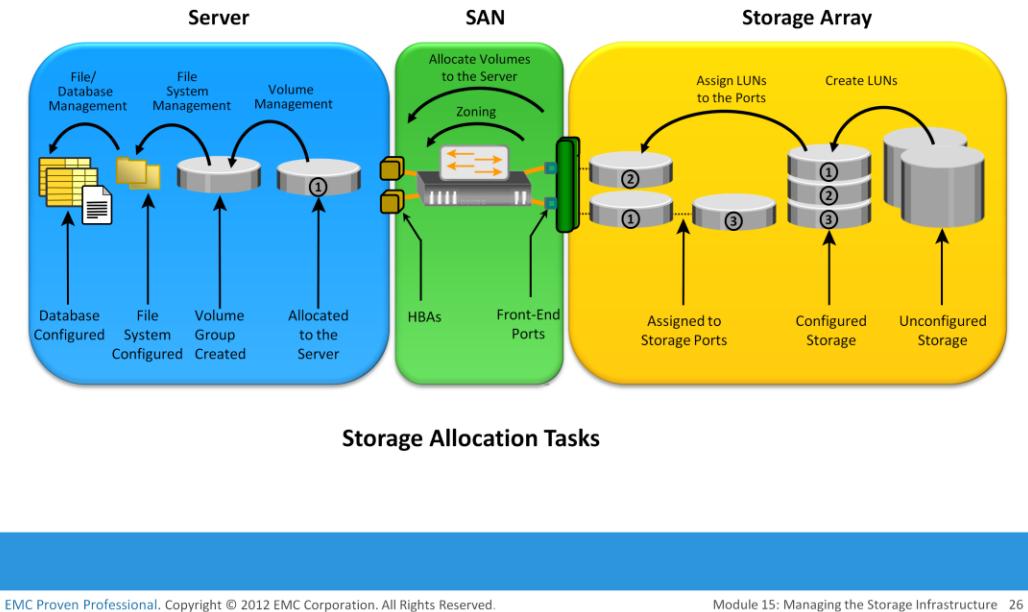


Figure on the slide illustrates the activities performed on server, SAN, and storage array while allocating storage to a new server.

Consider the deployment of a new RDBMS server to the existing nonvirtualized storage infrastructure. As a part of storage management activities, first, the administrator needs to install and configure the HBAs and device drivers on the server before it is physically connected to the SAN. Optionally, multipathing software can be installed on the server, which might require additional configuration. Further, storage array ports should be connected to the SAN.

As the next step, the administrator needs to perform zoning on the SAN switches to allow the new server access to the storage array ports via its HBAs. To ensure redundant paths between the server and the storage array, the HBAs of the new server should be connected to different switches and zoned with different array ports.

Further, the administrator needs to configure LUNs on the array and assign these LUNs to the storage array front-end ports. In addition, LUN masking configuration is performed on the storage array, which restricts access to LUNs by a specific server.

Cont..

The server then discovers the LUNs assigned to it by either a *bus rescan* process or sometimes through a server reboot, depending upon the operating system installed. A volume manager may be used to configure the logical volumes and file systems on the host. The number of logical volumes or file systems to be created depends on how a database or an application is expected to use the storage.

On the application side, the administrator's task includes installation of a database or an application on the logical volumes or file systems that were created. The last step is to make the database or application capable of using the new file system space.

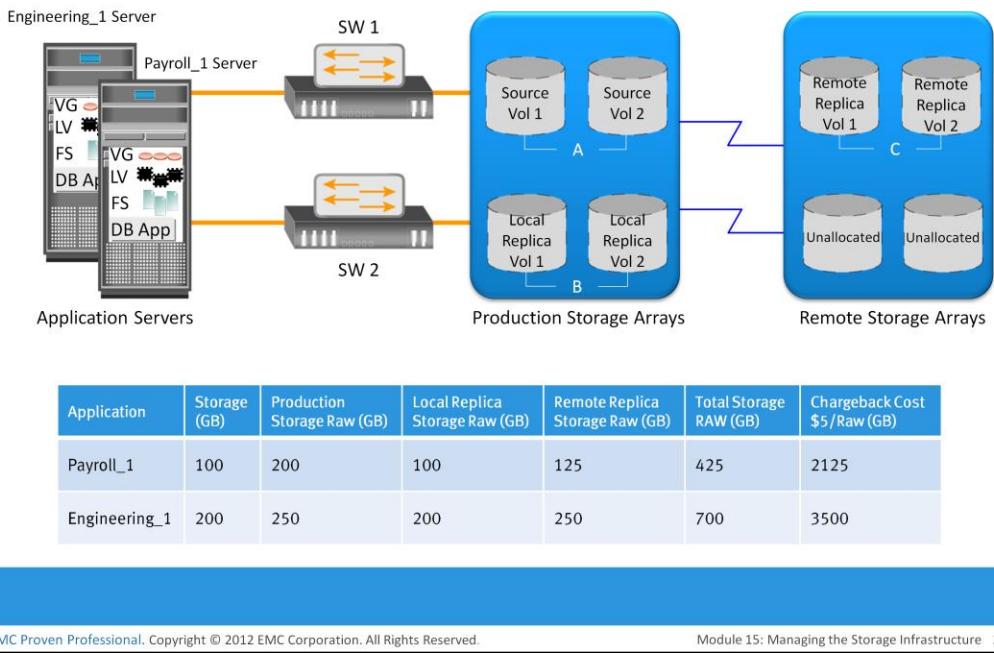
In a virtualized environment, provisioning storage to a VM that runs an RDBMS requires different administrative tasks.

Similar to a nonvirtualized environment, a physical connection must be established between the physical server, which hosts the VMs, and the storage array through the SAN. At the SAN level, a VSAN can be configured to transfer data between the physical server and the storage array. The VSAN isolates this storage traffic from any other traffic in the SAN. Further, the administrator can configure zoning within the VSAN.

At the storage side, administrators need to create thin LUNs from the shared storage pool and assign these thin LUNs to the storage array front-end ports. Similar to a physical environment, LUN masking needs to be carried out on the storage array.

At the physical server side, the hypervisor discovers the assigned LUNs. The hypervisor creates a logical volume and file system to store and manage VM files. Then, the administrator creates a VM and installs the OS and RDBMS on the VM. While creating the VM, the hypervisor creates a virtual disk file and other VM files in the hypervisor file system. The virtual disk file appears to the VM as a SCSI disk and is used to store the RDBMS data. Alternatively, the hypervisor enables virtual provisioning to create a thin virtual disk and assigns it to the VM. Hypervisors usually have native multipathing capabilities. Optionally, a third-party multipathing software may be installed on the hypervisor.

## Storage Management Example 2 – Chargeback Report



This example explores the storage infrastructure management tasks necessary to create a chargeback report. Figure on the slide shows a configuration deployed in a storage infrastructure. Three servers with two HBAs each connect to a storage array via two switches, SW1 and SW2. Individual departmental applications run on each of the servers. Array replication technology is used to create local and remote replicas. The production volume is represented as A, the local replica volume as B, and the remote replica volume as C.

A report documenting the exact amount of storage resources used by each application is created using a chargeback analysis for each department. If the unit for billing is based on the amount of raw storage (usable capacity plus protection provided) configured for an application used by a department, the exact amount of raw space configured must be reported for each application. Slide shows a sample report. The report shows the information for two applications, Payroll\_1 and Engineering\_1.

The first step to determine chargeback costs is to correlate the application with the exact amount of raw storage configured for that application. The Payroll\_1 application storage space is traced from file systems to logical volumes to volume groups and to the LUNs on the array. When the applications are replicated, the storage space used for local replication and remote replication is also identified. In the example shown, the application is using Source Vol 1 and Vol 2 (in the production array). The replication volumes are Local Replica Vol 1 and Vol 2 (in the production array) and Remote Replica Vol 1 and Vol 2 (in the remote array).

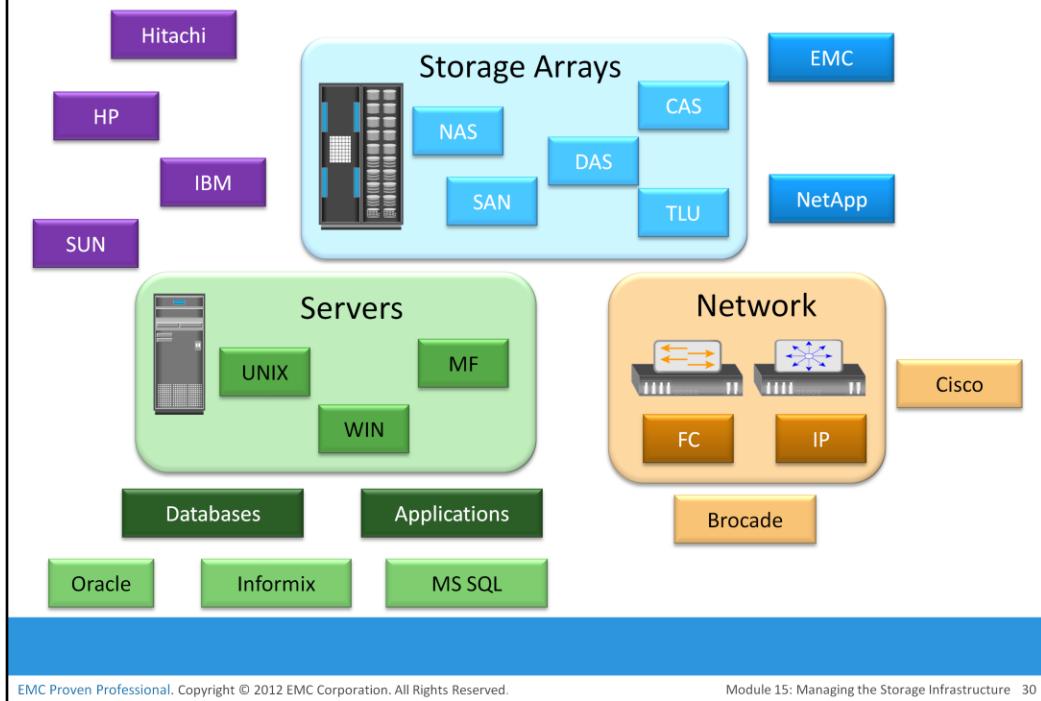
Cont..

The amount of storage allocated to the application can be easily computed after the array devices are identified. In this example, consider that Source Vol 1 and Vol 2 are each 50 GB in size, the storage allocated to the application is 100 GB (50 + 50). The allocated storage for replication is 100 GB for local replication and 100 GB for remote replication. From the allocated storage, the raw storage configured for the application is determined based on the RAID protection that is used for various array devices.

If the Payroll\_1 application's production volumes are RAID 1-protected, the raw space used by the production volumes is 200 GB. Assume that the local replicas are on unprotected volumes, and the remote replicas are protected with a RAID 5 configuration, then 100 GB of raw space is used by the local replica and 125 GB by the remote replica. Therefore, the total raw capacity used by the Payroll\_1 application is 425 GB. The total cost of storage provisioned for Payroll\_1 application will be \$2,125 (assume cost per GB of storage is \$5). This exercise must be repeated for each application in the enterprise to generate the chargeback report.

Chargeback reports can be extended to include a pre-established cost of other resources, such as the number of switch ports, HBAs, and array ports in the configuration. Chargeback reports are used by data center administrators to ensure that storage consumers are well aware of the costs of the service levels they have requested.

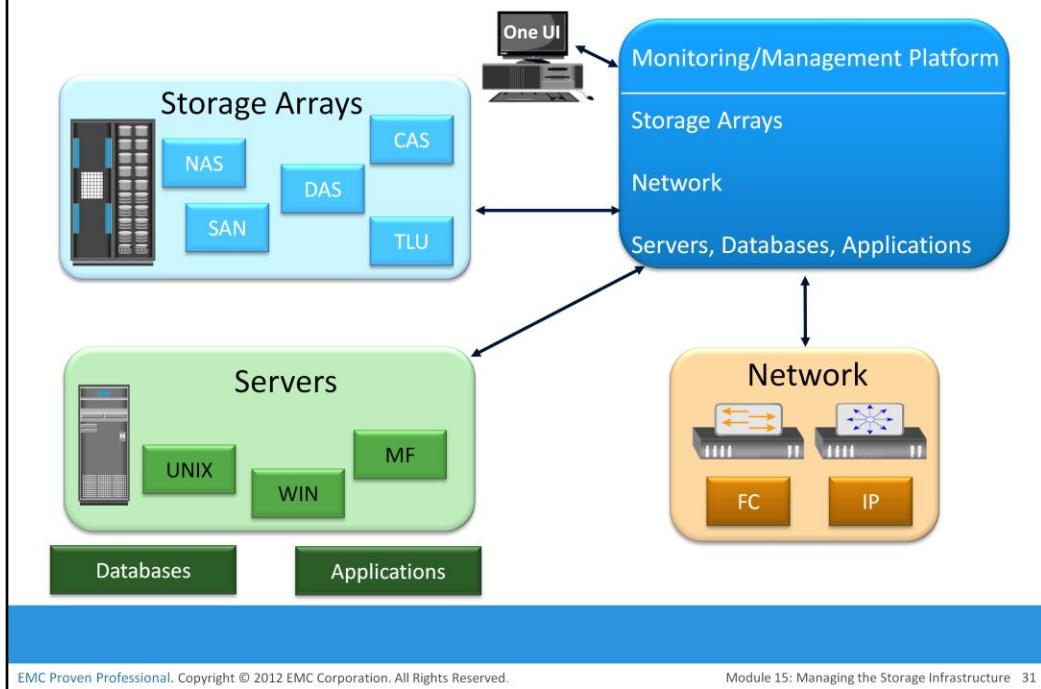
## Storage Infrastructure Management Challenges



Monitoring and managing today's complex storage infrastructure is challenging. This is due to the heterogeneity of storage arrays, networks, servers, databases, and applications in the environment. For example, heterogeneous storage arrays vary in their capacity, performance, protection, and architectures.

Each of the components in a data center typically comes with vendor-specific tools for management. An environment with multiple tools makes understanding the overall status of the environment challenging because the tools might not be interoperable. Ideally, management tools should correlate information from all components in one place. Such tools provide an end-to-end view of the environment, and a quicker root cause analysis for faster resolution to alerts.

## Developing an Ideal Solution



EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 31

An ideal solution should offer meaningful insight into the status of the overall infrastructure and provide root cause analysis for each failure. This solution should also provide central monitoring and management in a multivendor storage environment and create an end-to-end view of the storage infrastructure.

The benefit of end-to-end monitoring is the ability to correlate one component's behavior with the other. This is helpful to debug or analyze a problem, when looking at each component individually might not be sufficient to reveal the actual cause of the problem. The central monitoring and management system should gather information from all the components and manage them through a single-user interface. In addition, it must provide a mechanism to notify administrators about various events using methods, such as e-mail and Simple Network Management Protocol (SNMP) traps. It should also have the capability to generate monitoring reports and run automated scripts for task automation.

The ideal solution must be based on industry standards, by leveraging common APIs, data model terminology, and taxonomy. This enables the implementation of policy-based management across heterogeneous devices, services, applications, and deployed topologies.

Traditionally, SNMP protocol was the standard used to manage multivendor SAN environments. However, SNMP was primarily a network management protocol and was inadequate for providing the detailed information required to manage the SAN environment. The unavailability of automatic discovery functions and weak modeling constructs are some inadequacies of SNMP in a SAN environment. Even with these limitations, SNMP still holds a predominant role in SAN management, although newer open storage SAN management standards have emerged to monitor and manage storage environments more effectively.

## Storage Management Initiative (SMI)

- SNIA has been engaged in an initiative to develop a common storage management interface
- SNIA developed a specification called Storage Management Initiative-Specification (SMI-S)
  - ▶ It forms a normalized, abstracted model to which a storage infrastructure components can be mapped
  - ▶ Enables standardized and end-to-end control of resources
  - ▶ SMI-S compliant products are easier and faster to deploy
  - ▶ Eliminates the need for development of vendor-proprietary management interfaces

The Storage Networking Industry Association (SNIA) has been engaged in an initiative to develop a common storage management interface. SNIA has developed a specification called Storage Management Initiative-Specification (SMI-S). This specification is based on the Web-Based Enterprise Management (WBEM) technology, and Distributed Management Task Force's (DMTF) Common Information Model (CIM). The initiative was formally created to enable broad interoperability and management among heterogeneous storage and SAN components. For more information, see [www.snia.org](http://www.snia.org).

SMI-S offers substantial benefits to users and vendors. It forms a normalized, abstracted model to which a storage infrastructure's physical and logical components can be mapped. This model is used by management applications, such as storage resource management, device management, and data management, for standardized, end-to-end control of storage resources.

Using SMI-S, device software developers have a unified object model with details about managing the breadth of storage and SAN components. SMI-S-compliant products lead to easier, faster deployment and accelerated adoption of policy-based storage management frameworks. Moreover, SMI-S eliminates the need for the development of vendor-proprietary management interfaces and enables vendors to focus on value-added features.

## Enterprise Management Platform

- Suite of applications for managing and monitoring storage infrastructure
  - ▶ Provides end-to-end management of physical and virtual resources
  - ▶ Generates alerts to inform the status of components and processes
  - ▶ Schedules operations such as provisioning and configuration management

An enterprise management platform (EMP) is a suite of applications that provides an integrated solution for managing and monitoring an enterprise storage infrastructure. These applications have powerful, flexible, unified frameworks that provide end-to-end management of both physical and virtual resources. EMP provides a centrally managed, single point of control for resources throughout the storage environment.

These applications can proactively monitor storage infrastructure components and alert users about events. These alerts are either shown on a console depicting the faulty component in a different color, or they can be configured to send the alert by e-mail. In addition to monitoring, an EMP provides the necessary management functionality, which can be natively implemented into the EMP or can launch the proprietary management utility supplied by the component manufacturer.

An EMP also enables easy scheduling of operations that must be performed regularly, such as the provisioning of resources, configuration management, and fault investigation. These platforms also provide extensive analytical, remedial, and reporting capabilities to ease storage infrastructure management. EMC ControlCenter and EMC Prosphere described in this chapter are examples of an EMP.

# Module 15: Managing the Storage Infrastructure

## Lesson 3: Information Lifecycle Management

During this lesson the following topics are covered:

- Challenges of managing information
- Information lifecycle management
- Storage tiering

This lesson describes the challenges of managing information. It also describes information lifecycle management strategy. Finally, the lesson explains storage tiering.

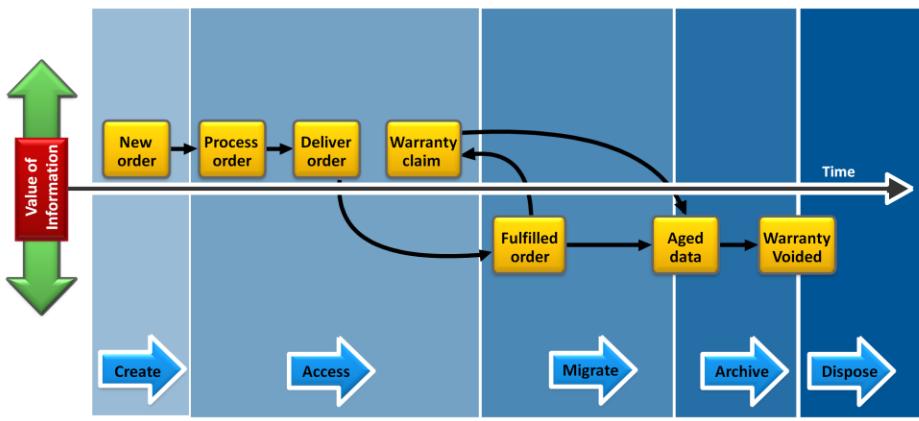
## Challenges in Managing Information

- Exploding digital universe
  - ▶ Multifold increase of information
- Increasing dependency on information
  - ▶ Strategic use of information plays an important role in determining the success of a business
- Changing value of information
  - ▶ Information that is valuable today may become less important in future

In both traditional data center and virtualized environments, managing information can be expensive if not managed appropriately. Along with the tools, an effective management strategy is also required to manage information efficiently. This strategy should address the following key challenges that exist in today's data centers:

- Exploding digital universe: The rate of information growth is increasing exponentially. Creating copies of data to ensure high availability and repurposing has contributed to the multifold increase of information growth.
- Increasing dependency on information: The strategic use of information plays an important role in determining the success of a business and provides competitive advantages in the marketplace.
- Changing value of information: Information that is valuable today might become less important tomorrow. The value of information often changes over time.

## Information Lifecycle Management



*A proactive strategy that enables an IT organization  
to effectively manage the information throughout its lifecycle*

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 36

Framing a strategy to meet these challenges involves understanding the value of information over its life cycle. When information is first created, it often has the highest value and is accessed frequently. As the information ages, it is accessed less frequently and is of less value to the organization. Understanding the value of information helps to deploy the appropriate infrastructure according to the changing value of information.

For example, in a sales order application, the value of the information (customer data) changes from the time the order is placed until the time that the warranty becomes void. The value of the information is highest when a company receives a new sales order and processes it to deliver the product. After the order fulfillment, the customer data does not need to be available for real-time access. The company can transfer this data to less expensive secondary storage with lower performance until a warranty claim or another event triggers its need. After the warranty becomes void, the company can dispose of the information.

*Information lifecycle management (ILM)* is a proactive strategy that enables an IT organization to effectively manage information throughout its life cycle based on predefined business policies. From data creation to data deletion, ILM aligns the business requirements and processes with service levels in an automated fashion. This allows an IT organization to optimize the storage infrastructure for maximum return on investment.

## Benefits of ILM

- Provides lower total cost of ownership
  - ▶ By aligning the infrastructure and management costs with changing value of information
- Provides simplified management
  - ▶ By automating the processes
- Maintains compliance
  - ▶ Knowledge of what data needs to be protected for what length of time
- Optimized utilization
  - ▶ By deploying storage tiering

Implementing an ILM strategy has the following key benefits that directly address the challenges of information management:

- Simplified management: By integrating process steps and interfaces with individual tools and by increasing automation.
- Maintaining compliance: By knowing what data needs to be protected for what length of time.
- Lower Total Cost of Ownership (TCO): By aligning the infrastructure and management costs with information value. As a result, resources are not wasted, and complexity is not introduced by managing low-value data at the expense of high-value data.
- Optimized utilization: By deploying storage tiering.

## Storage Tiering

### Storage Tiering

It is a technique of establishing a hierarchy of storage types and identifying the candidate data to relocate to the appropriate storage type to meet service level requirements at a minimal cost.

- Each tier has different levels of protection, performance, and cost
- Efficient storage tiering requires defining tiering policies
- Storage tiering implementations are:
  - ▶ Manual storage tiering
  - ▶ Automated storage tiering
- Data movement occurs between tiers
  - ▶ Within a storage array (Intra-array)
  - ▶ Between storage arrays (Inter-array)

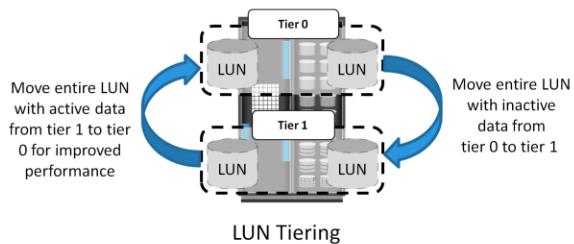
*Storage tiering* is a technique of establishing a hierarchy of different storage types (tiers). This enables storing the right data to the right tier, based on service level requirements, at a minimal cost. Each tier has different levels of protection, performance, and cost. For example, high performance solid-state drives (SSDs) or FC drives can be configured as tier 1 storage to keep frequently accessed data and low cost SATA drives as tier 2 storage to keep the less frequently accessed data. Keeping frequently used data in SSD or FC improves application performance. Moving less-frequently accessed data to SATA can free up storage capacity in high performance drives and reduce the cost of storage. This movement of data happens based on defined tiering policies. The tiering policy might be based on parameters, such as file type, size, frequency of access, and so on. For example, if a policy states “Move the files that are not accessed for the last 30 days to the lower tier,” then all the files matching this condition are moved to the lower tier.

Storage tiering can be implemented as a manual or an automated process. *Manual storage tiering* is the traditional method where the storage administrator monitors the storage workloads periodically and moves the data between the tiers. Manual storage tiering is complex and time-consuming. *Automated storage tiering* automates the storage tiering process, in which data movement between the tiers is performed nondisruptively. In automated storage tiering, the application workload is proactively monitored; the active data is automatically moved to a higher performance tier and inactive data to higher capacity, lower performance tier. Data movements between various tiers can happen within (intra-array) or between (inter-array) storage arrays.

## Intra-array Storage Tiering

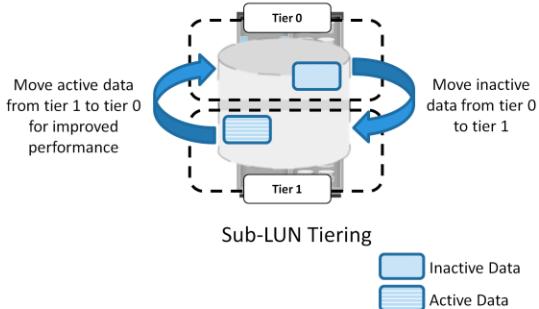
- LUN tiering

- ▶ Moves entire LUN from one tier to another
- ▶ Does not give effective cost and performance benefits



- Sub-LUN tiering

- ▶ A LUN is broken down into smaller segments and tiered at that level
- ▶ Provides effective cost and performance benefits



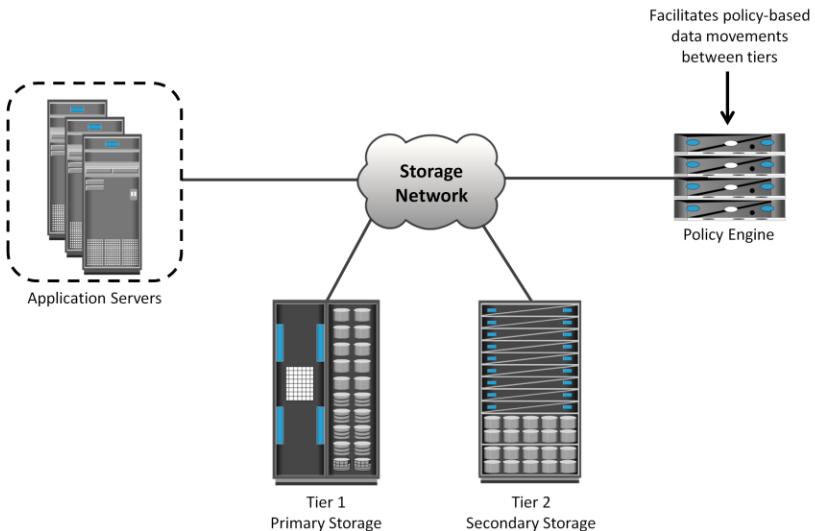
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 39

The process of storage tiering within a storage array is called *intra-array storage tiering*. It enables the efficient use of SSD, FC, and SATA drives within an array and provides performance and cost optimization. The goal is to keep the SSDs busy by storing the most frequently accessed data on them, while moving out the less frequently accessed data to the SATA drives. Data movements executed between tiers can be performed at the LUN level or at the sub-LUN level. The performance can be further improved by implementing tiered cache.

Traditionally, storage tiering is operated at the LUN level that moves an entire LUN from one tier of storage to another. This movement includes both active and inactive data in that LUN. This method does not give effective cost and performance benefits. Today, storage tiering can be implemented at the sub-LUN level. In sub-LUN level tiering, a LUN is broken down into smaller segments and tiered at that level. Movement of data with much finer granularity, for example 8 MB, greatly enhances the value proposition of automated storage tiering. Tiering at the sub-LUN level effectively moves active data to faster drives and less active data to slower drives.

## Inter-array Storage Tiering



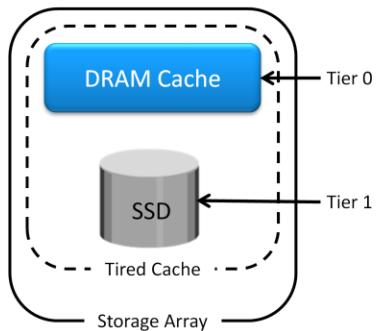
EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 40

The process of storage tiering between storage arrays is called *inter-array storage tiering*. Inter-array storage tiering automates the identification of active or inactive data to relocate them to different performance or capacity tiers between the arrays. Figure on the slide illustrates an example of a two-tiered storage environment. This environment optimizes the primary storage for performance and the secondary storage for capacity and cost. The policy engine, which can be software or hardware where policies are configured, facilitates moving inactive or infrequently accessed data from the primary to the secondary storage. Some prevalent reasons to tier data across arrays is archival or to meet compliance requirements. As an example, the policy engine might be configured to relocate all the files in the primary storage that have not been accessed in one month and archive those files to the secondary storage. For each archived file, the policy engine creates a small space-saving stub file in the primary storage that points to the data on the secondary storage. When a user tries to access the file at its original location on the primary storage, the user is transparently provided with the actual file from the secondary storage.

## Cache Tiering

- Enables creation of a large capacity secondary cache using SSDs
- Enables tiering between DRAM cache and SSDs (secondary cache)
- Most reads are served directly from high performance tiered cache



### Benefits

- Enhances performance during peak workload
- Non-disruptive and transparent to applications

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 41

Tiering is also implemented at the cache level. A large cache in a storage array improves performance by retaining large amount of frequently accessed data in a cache, so most reads are served directly from the cache. However, configuring a large cache in the storage array involves more cost. An alternative way to increase the size of the cache is by utilizing the SSDs on the storage array. In cache tiering, SSDs are used to create a large capacity secondary cache and to enable tiering between DRAM (primary cache) and SSDs (secondary cache). Server flash-caching is another tier of cache in which flash-cache card is installed in the server to further enhance the application performance.

## Module 15: Managing the Storage Infrastructure

### Concept in Practice

- EMC ControlCenter
- EMC Prosphere
- EMC Unisphere
- EMC Unified Infrastructure Management (UIM)

The Concept in Practice covers the product example of storage infrastructure management software. It covers four products: EMC ControlCenter, EMC Prosphere, EMC Unisphere, and EMC Unified Infrastructure Management.

## EMC ControlCenter

- An enterprise management platform
  - ▶ Contains storage resource management applications to manage a multi-vendor storage infrastructure
- Provides an end-to-end view of the entire networked storage infrastructure including virtualized environment
- Enables implementation of ILM strategy by providing management of tiered storage infrastructure
- Has built-in security features that provide access control, data confidentiality, data integrity, logging, and auditing

EMC ControlCenter is a family of storage resource management (SRM) applications that provide a unified solution to manage a multivendor storage infrastructure. It helps address the challenges to manage a large, complex storage environment that includes hosts, storage networks, storage, and virtualization across all the layers. ControlCenter provides capabilities, such as storage planning, provisioning, monitoring, and reporting. It enables implementing an ILM strategy by providing comprehensive management of tiered storage infrastructure. It also provides an end-to-end view of the entire networked storage infrastructure that includes SAN, NAS, and host storage resources, including virtualized environment. It provides a central administrative console, discovery of new components, quota management, event management, root cause analysis, and chargeback. ControlCenter comes with built-in security features that provide access control, data confidentiality, data integrity, logging, and auditing. It offers an intuitive, easy-to-use interface that provides insight into the complex relationships of the environment. ControlCenter uses an agent to discover the components in the environment.

## EMC ProSphere

- It is a storage resource management software, built to meet the demands of the new cloud computing era
- Following are the key capabilities:
  - ▶ Provides end-to-end visibility of all objects
  - ▶ Enables multi-site management from a single console
  - ▶ Improves productivity in virtualized environments with “Smart Groups”
  - ▶ Enables fast, easy, and efficient deployment
    - ▶ Agent-less discovery of objects
    - ▶ ProSphere is packaged as virtual appliance

EMC ProSphere is also storage resource management software built to meet the demands of the new cloud computing era. EMC ProSphere improves productivity and service levels in the virtualized and cloud environment. ProSphere includes the following key capabilities:

- End-to-end visibility: It offers an intuitive, easy-to-use interface that provides insight into the complex relationships between objects in large, virtualized environments.
- Multi-site management: From a single console, ProSphere’s federated architecture aggregates information from across sites and simplifies information management between data centers. ProSphere is managed from a web browser to allow easy access over the Internet for remote management.
- Improved productivity in growing virtualized environments: ProSphere introduces an innovative technology called Smart Groups, which groups objects with similar characteristics into a user-defined group for performing management tasks. This enables IT to take a policy-based approach to manage objects or to set data collection policies.
- Fast, easy, and efficient deployment: Agent-less discovery eliminates the burden of deploying and managing host agents. ProSphere is packaged as a virtual appliance that can be installed in a short time.
- Delivery of IT as a service: With ProSphere, service levels can now be monitored from host-to-storage layers. This allows organizations to maintain consistent service levels at an optimal price-performance ratio to meet business objectives to delivering IT-as-a-service.

## EMC Unisphere

- It is a unified storage management platform for managing:
  - ▶ EMC VNX and VNXe
  - ▶ EMC RecoverPoint/SE
- Some of the key capabilities offered by Unisphere are:
  - ▶ Provides unified management for file-based, block-based, and object-based storage
  - ▶ Supports automated storage tiering
  - ▶ Provides management of both physical and virtual components

EMC Unisphere is a unified storage management platform that provides intuitive user interfaces for managing EMC RecoverPoint SE, and EMC VNX and VNXe storage arrays. Unisphere is web-enabled and supports remote management of storage arrays. Some of the key capabilities offered by Unisphere follow:

- Provides unified management for file, block, and object storage
- Provides single sign-on for all devices in a management domain
- Supports automated storage tiering and ensures that data is stored in the correct tier to meet performance and cost
- Provides management of both physical and virtual components

## EMC Unified Infrastructure Manager (UIM)

- Unified management solution for Vblocks
- Enables configuring Vblock resources and activating services
- Provides a dashboard showing Vblock infrastructure configuration and resource utilization
- Provides a topology view of Vblock infrastructure
- Provides an alerts console that lists alerts against adversely affected resources and services
- Performs compliance check during resource configuration

EMC Unified Infrastructure Manager is a unified management solution for Vblocks. (Vblock is covered in module 13.) It enables configuring the Vblock infrastructure resources and activating cloud services. It provides a single user interface to manage multiple Vblocks and eliminates the need for configuring compute, network, and storage separately using different virtual infrastructure management tools.

UIM provides a dashboard that shows how the Vblock infrastructure is configured and how the resources are used. This enables an administrator to monitor the configuration and utilization of the Vblock infrastructure resources and to plan for capacity requirements. UIM also provides a topology or a map view of the Vblock infrastructure, which enables an administrator to quickly locate and understand the interconnections of the Vblock infrastructure components and services. It provides an alerts console, which allows an administrator to see the alerts against the Vblock infrastructure resources and the associated services affected by problems. UIM performs a compliance check during resource configuration. It validates compliance with configuration best practices. It also prevents conflicting resource identity assignments, for example, accidentally assigning a MAC address to more than one virtual NIC.

## Module 15: Summary

Key points covered in this module:

- Key storage infrastructure components that are monitored
- Key monitoring parameters
- Storage management activities
- Storage infrastructure management challenges
- Enterprise management platform
- Information lifecycle management
- Storage tiering

This module covered the key storage infrastructure management components that should be monitored such as servers, network, storage arrays, and environmental controls.

These key components should be monitored for accessibility, capacity, performance, and security.

This module also covered the key management activities such as availability management, capacity management, performance management, security management, and reporting.

This module also covered the key infrastructure management challenges and the ideal solution.

Further, this module covered information lifecycle management (ILM) which is a proactive strategy that enables an IT organization to effectively manage information throughout its lifecycle, based on predefined business policies.

Finally, this module detailed storage tiering techniques which identifies the candidate data and relocate them to the appropriate storage type to meet service level requirements at a minimal cost.

## Check Your Knowledge – 1

- Which type of alert is generated if soft media errors on a disk drive approaches its pre-defined threshold value?
  - A. Fatal
  - B. Warning
  - C. Information
  - D. Watermark
- What is a purpose of chargeback report?
  - A. Reports investment in managing infrastructure
  - B. Reports cost of decommissioning infrastructure components
  - C. Reports utilization of infrastructure components by various users
  - D. Reports charges for SLA breach

## Check Your Knowledge – 2

- Which monitoring parameter helps ensuring availability of adequate amount of resources and prevents service unavailability?
  - A. Availability
  - B. Capacity
  - C. Performance
  - D. Security
- Which pillar of multitenancy ensures consistent and reliable service levels in a multitenant storage environment?
  - A. Secure separation
  - B. Service assurance
  - C. Service availability
  - D. Storage tiering

## Check Your Knowledge – 3

- What best describes the SMI specification?
  - A. Restricts vendors to build new features and functions to manage storage subsystems
  - B. Eliminates the need for development of vendor-proprietary management interfaces
  - C. Prevents interoperability among multi-vendor resources
  - D. Enable deploying ILM supported infrastructure

## Course Summary

Key points covered in this course:

- Key elements of data center environment
- Key components of intelligent storage systems and RAID technology
- Storage networking technologies
- Business continuity solutions
- Cloud computing technology
- Security and management of storage infrastructure

This course covered the importance of information in our daily lives. The increasing dependency of information to the businesses has amplified the challenges in storing, protecting, and managing data.

The course also detailed the five core elements of data center that are essential for its functionality such as application, database management system , host or compute, network, and storage.

The course also covered RAID which is a technique of combining multiple disk drives into a logical unit and provide protection, performance, or both.

This course also covered intelligent storage systems which are feature-rich RAID arrays that provide highly-optimized I/O processing capabilities.

Further, this course detailed storage networking technologies such as FC SAN, IP SAN, and FCoE that provide block access to storage. The course also covered NAS which is a dedicated, high-performance file sharing and storage device that enables its clients to share files over an IP network. This module covered object-based and unified storage.

Besides this, the course also covered business continuity solutions such as backup and replication. It also covered data archiving solutions that enable to meet compliance.

It also covered cloud computing which is a next generation style of computing that provides highly scalable and flexible computing that is available on demand.

Finally, this module covered security of storage infrastructure that is essential to ensure confidentiality, integrity, and availability. It also covered the management activities in storage environment.



# Thank You!

EMC Proven Professional. Copyright © 2012 EMC Corporation. All Rights Reserved.

Module 15: Managing the Storage Infrastructure 52

This concludes the training. Thank you for your participation.

Please remember to complete the course evaluation available from your instructor.