DCUCD

Designing Cisco Data Center Unified Computing

Volume 1

Version 5.0

Student Guide

Text Part Number: 97-3176-01

https://t.me/learningnets



Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED “AS IS.” CISCO MAKES AND YOU RECEIVE NO WARRANTIES IN CONNECTION WITH THE CONTENT PROVIDED HEREUNDER, EXPRESS, IMPLIED, STATUTORY OR IN ANY OTHER PROVISION OF THIS CONTENT OR COMMUNICATION BETWEEN CISCO AND YOU. CISCO SPECIFICALLY DISCLAIMS ALL IMPLIED WARRANTIES, INCLUDING WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE, OR ARISING FROM A COURSE OF DEALING, USAGE OR TRADE PRACTICE. This learning product may contain early release content, and while Cisco believes it to be accurate, it falls subject to the disclaimer above.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Table of Contents

Volume 1

Course Introduction ***1***

Overview 1
Learner Skills and Knowledge 2
Course Goal and Objectives 3
Course Flow 4
Additional References 5
Cisco Glossary of Terms 6
Your Training Curriculum 7
Additional Resources 10
Introductions 12

Cisco Data Center Solution Architecture and Components ***1-1***

Overview 1-1
Module Objectives 1-1

Identifying Data Center Solutions **1-3**

Overview 1-3
Objectives 1-3
Data Center Overview 1-4
Data Center Trends: Consolidation 1-10
Data Center Trends: Virtualization 1-14
Data Center Business Challenges 1-21
Data Center Environmental Challenges 1-27
Data Center Technical Challenges 1-34
Summary 1-45
References 1-45

Identifying Data Center Applications **1-47**

Overview 1-47
Objectives 1-47
Common Data Center Applications 1-48
Server Virtualization Overview 1-66
Desktop Virtualization Overview 1-81
Desktop Virtualization Components 1-81
Summary 1-93
References 1-93

Identifying Cloud Computing **1-95**

Overview 1-95
Objectives 1-95
Cloud Computing Overview 1-96
Cloud Computing Models 1-100
Cloud Computing Service Categories 1-106
Cloud Computing Aspects 1-109
Summary 1-113
References 1-113

https://t.me/learningnets

Identifying Cisco Data Center Architecture and Components 1-115
Overview 1-115
Objectives 1-115
Cisco Data Center Architecture Overview 1-116
Cisco Data Center Architecture Unified Fabric 1-128
Cisco Data Center Network Equipment 1-150
Cisco Data Center Compute Architecture 1-165
Cisco Validated Designs 1-174
Summary 1-179
References 1-179
Module Summary 1-181
Module Self-Check 1-183
Module Self-Check Answer Key 1-185

Assess Data Center Computing Requirements 2-1

Overview 2-1
Module Objectives 2-1
Defining a Cisco Unified Computing System Solution Design 2-3
Overview 2-3
Objectives 2-3
Design Process 2-4
Design Process Phases 2-13
Design Deliverables 2-23
Summary 2-28
References 2-28
Analyzing Computing Solutions Characteristics 2-29
Overview 2-29
Objectives 2-29
Performance Characteristics 2-30
Assess Server Virtualization Characteristics 2-38
Assess Desktop Virtualization Performance Characteristics 2-48
Assess Small vSphere Deployment Requirements 2-53
Assess Small Hyper-V Deployment Requirements 2-59
Assess VMware View VDI Deployment Requirements 2-64
Design Workshop Output 2-65
Summary 2-72
References 2-72
Employing Data Center Analysis Tools 2-73
Overview 2-73
Objectives 2-73
Reconnaissance and Analysis Tools 2-74
Use Reconnaissance and Analysis Tools 2-79
Employ VMware Capacity Planner 2-85
Employ VMware CapacityIQ 2-97
Employ MAP Toolkit 2-104
Employ Cisco UCS TOC/ROI Advisor 2-113
Summary 2-118
References 2-118
Module Summary 2-119
Module Self-Check 2-121
Module Self-Check Answer Key 2-123

 © 2012 Cisco Systems, Inc

https://t.me/learningnets

Size Cisco Unified Computing Solutions *3-1*

Overview 3-1
Module Objectives 3-1

Sizing the Cisco UCS C-Series Server Solution **3-3**

Overview 3-3
Objectives 3-3
Size the Cisco UCS C-Series Solution 3-4
Cisco UCS C-Series Integration with UCS Manager 3-12
Size the Small VMware vSphere Solution—Plan 3-15
Size the Small Hyper-V vSphere Solution—Plan 3-19
Summary 3-22

Sizing the Cisco UCS B-Series Server Solution **3-23**

Overview 3-23
Objectives 3-23
Size the Cisco UCS B-Series Solution 3-24
Size the Desktop Virtualization Solution—Plan 3-34
Summary 3-40

Planning Unified Computing Deployment **3-41**

Overview 3-41
Objectives 3-41
Cisco UCS Power Calculator Tool 3-42
Create a Physical Deployment Plan 3-47
Summary 3-54
Module Summary 3-55
References 3-55
Module Self-Check 3-57
Module Self-Check Answer Key 3-58

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Course Introduction

Overview

Designing Cisco Data Center Unified Computing (DCUCD) v5.0 is a four-day course that teaches you how to design a Cisco Unified Computing System (UCS) solution for the data center.

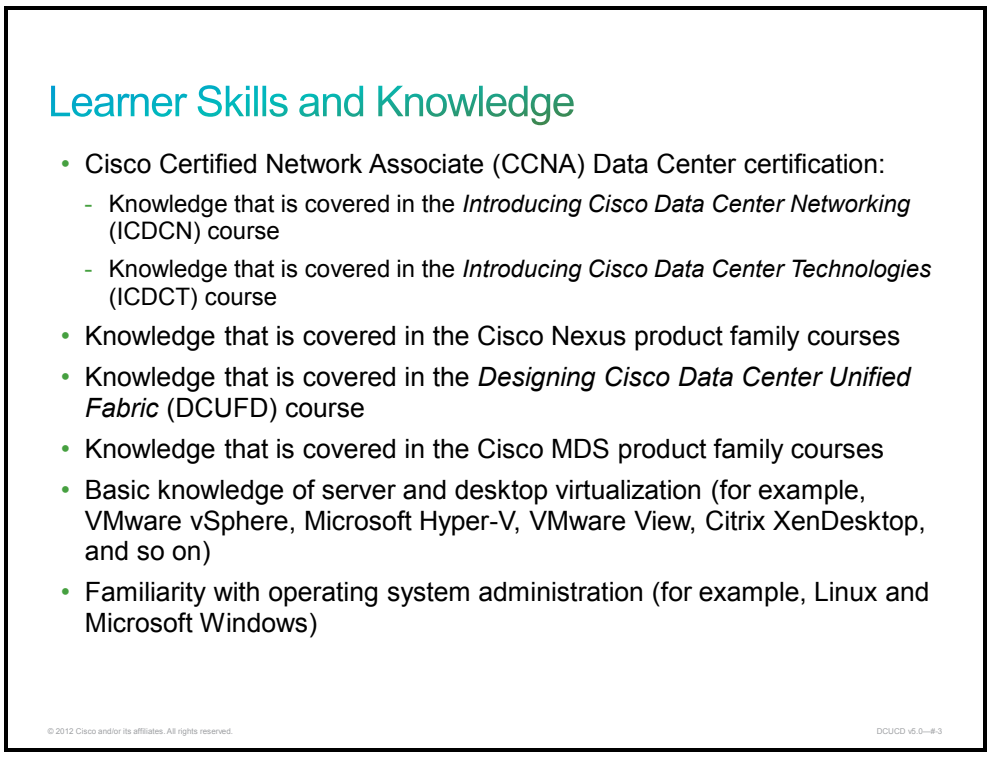
The primary focus of the course is on the next-generation data center platform: the Cisco UCS. The course also includes information about Cisco Nexus Family Switches, Cisco Multilayer Director Switches (MDSs), server and desktop virtualization, distributed applications, and more, which are all part of a Cisco UCS solution.

The course describes the design-related aspects, which include how to evaluate the hardware components and the sizing process, define the server deployment model, address the management and environmental aspects, and design the network and storage perspectives of the Cisco UCS solution.

https://t.me/learningnets

Learner Skills and Knowledge

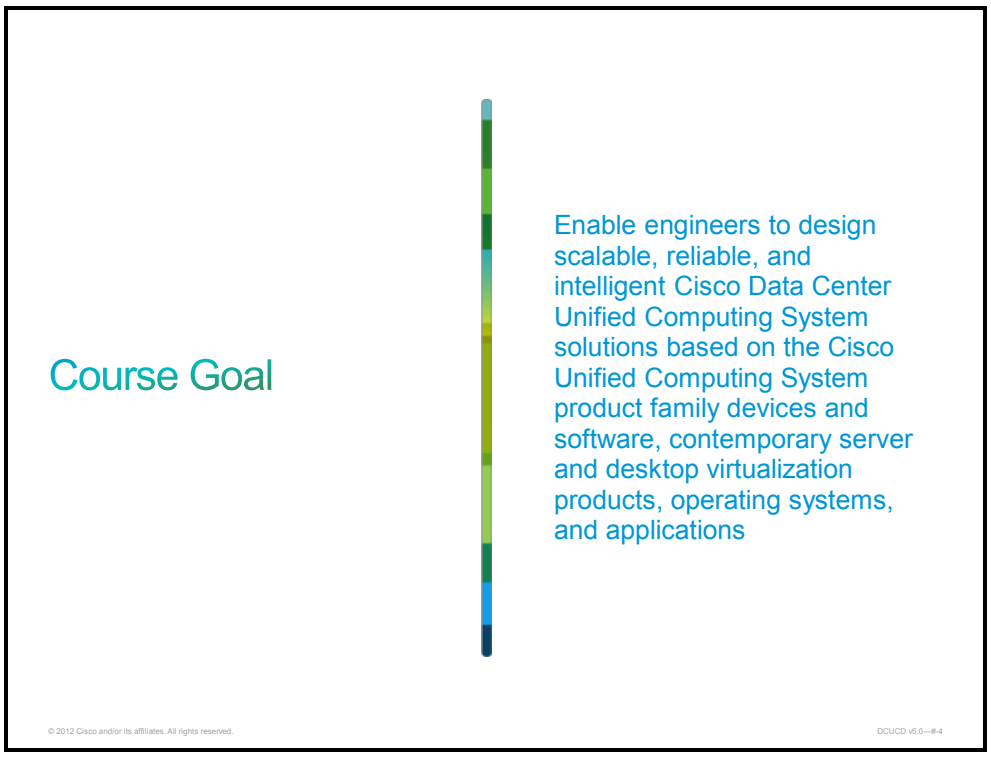
This subtopic lists the skills and knowledge that learners must possess to benefit fully from the course. The subtopic also includes recommended Cisco learning offerings that learners should first complete to benefit fully from this course.



https://t.me/learningnets

Course Goal and Objectives

This topic describes the course goal and objectives.



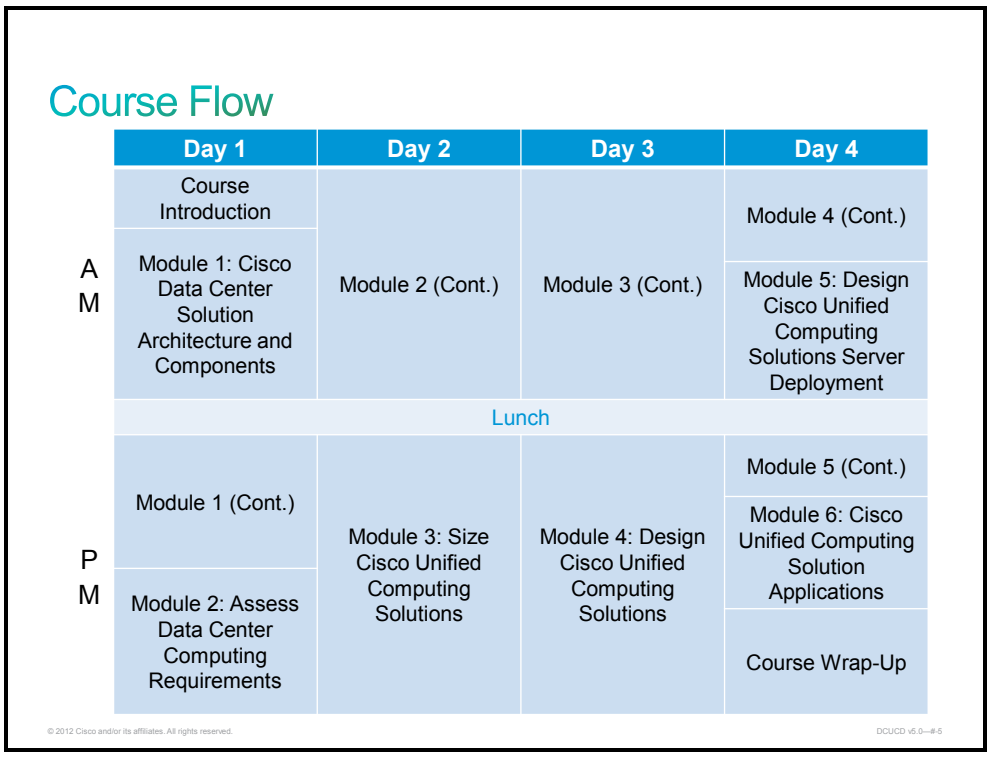
Upon completing this course, you will be able to meet these objectives:

- Evaluate the Cisco UCS solution design process in regard to the contemporary data center challenges, Cisco Data Center architectural framework, and components
- Use the reconnaissance and analysis tools to assess computing solution performance characteristics and requirements
- Identify the hardware components of Cisco UCS C-Series and B-Series and select proper hardware for a given set of requirements
- Design the Cisco UCS solution LAN and SAN connectivity
- Identify the Cisco UCS server deployment model and design a deployment model with correct naming, addressing, and management for a given set of requirements
- Identify the typical data center applications where the Cisco UCS solution is used

https://t.me/learningnets

Course Flow

This topic presents the suggested flow of the course materials.

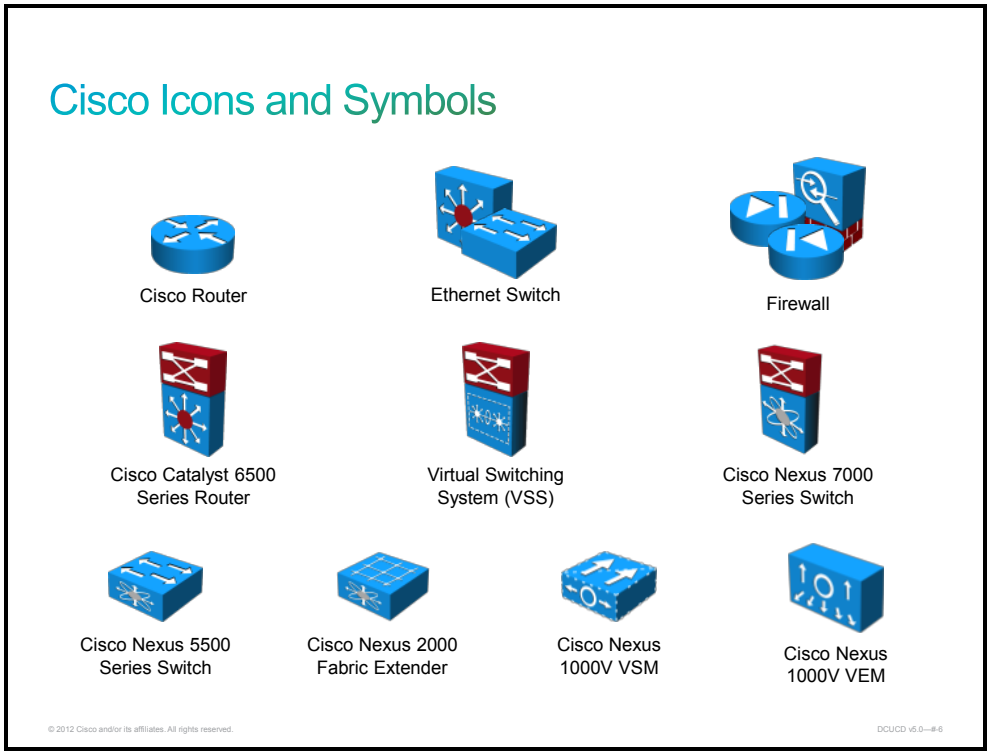


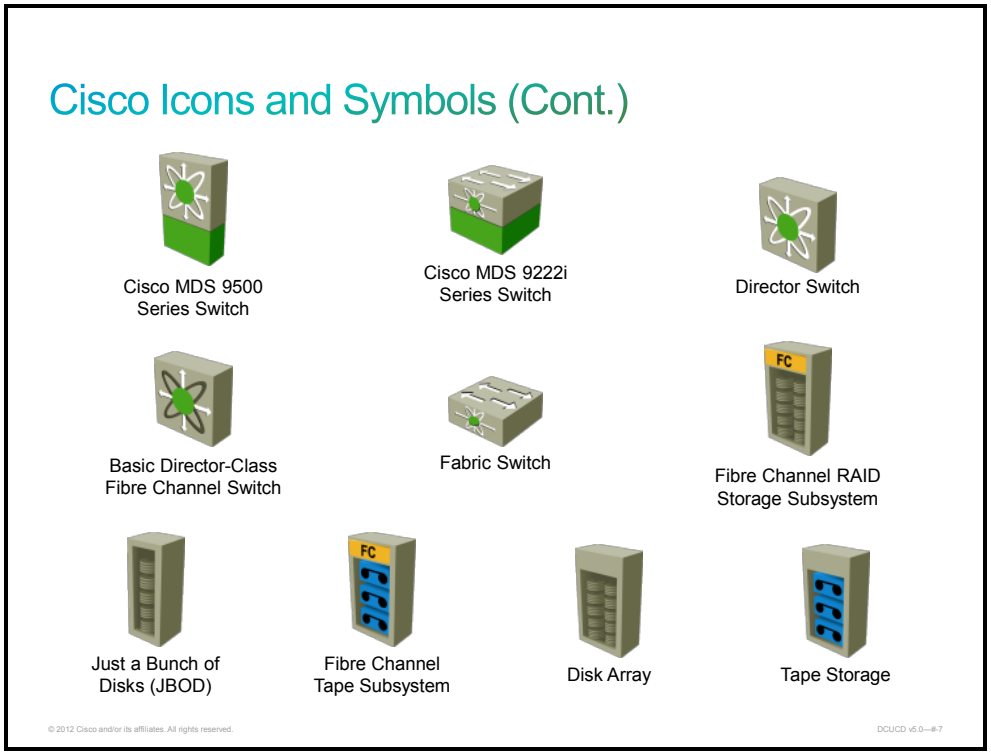
The schedule reflects the recommended structure for this course. This structure allows enough time for the instructor to present the course information and for you to work through the lab activities. The exact timing of the subject materials and labs depends on the pace of your specific class.

https://t.me/learningnets

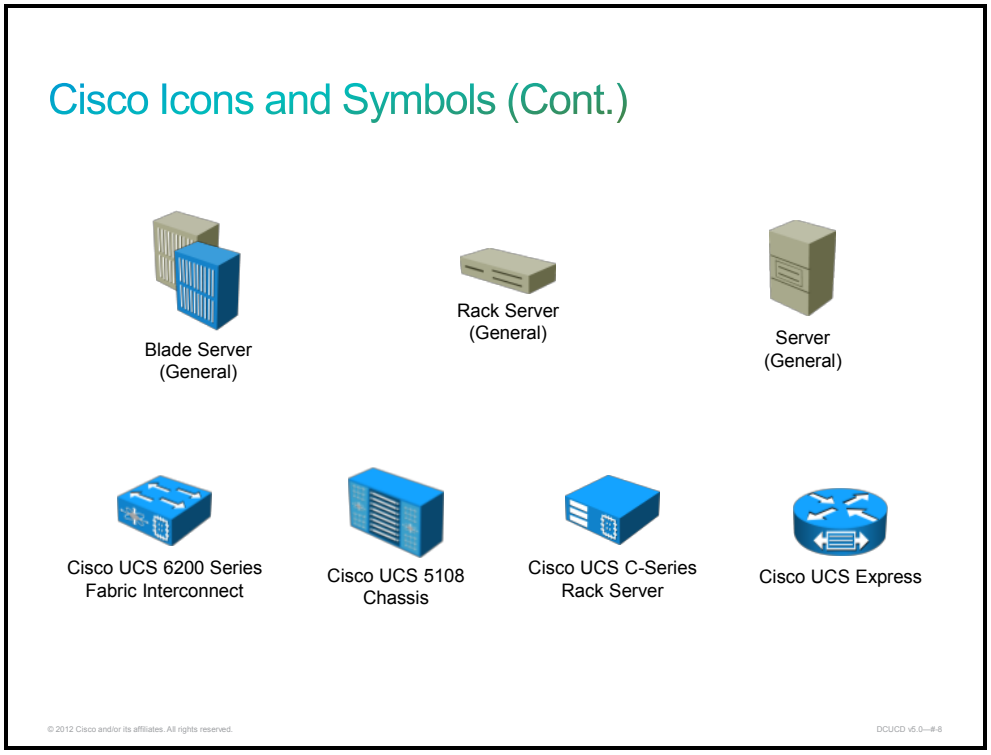
Additional References

This topic presents the Cisco icons and symbols that are used in this course, as well as information on where to find additional technical references.





https://t.me/learningnets



Cisco Glossary of Terms

For additional information on Cisco terminology, refer to the *Cisco Internetworking Terms and Acronyms* glossary of terms at http://docwiki.cisco.com/wiki/Category:Internetworking_Terms_and_Acronyms_(ITA).

https://t.me/learningnets

Your Training Curriculum

This topic presents the training curriculum for this course.

Cisco Learning Network

Developing a world of talent through collaboration, social learning, online assessment, and mentoring

https://learningnetwork.cisco.com



© 2012 Cisco and/or its affiliates. All rights reserved.

DCUCD v5.0—#-10

To prepare and learn more about IT certifications and technology tracks, visit the Cisco Learning Network, which is the home of Cisco Certifications.

https://t.me/learningnets

Cisco Career Certifications: Cisco Certified Network Professional (CCNP) Data Center

Expand Your Professional Options and Advance Your Career

CCNP-level recognition in data center

http://www.cisco.com/go/certifications



Cisco CCNP Data Center
Designing Cisco Data Center Unified Computing (DCUCD)
Implementing Cisco Data Center Unified Computing (DCUCI)
Troubleshooting Data Center Unified Computing (DCUCT)
Designing Cisco Data Center Unified Fabric (DCUFD)
Implementing Cisco Data Center Unified Fabric (DCUFI)
Troubleshooting Cisco Data Center Unified Fabric (DCUFT)

© 2012 Cisco and/or its affiliates. All rights reserved.

DCUCD v5.0—#-11

You are encouraged to join the Cisco Certification Community, a discussion forum open to anyone holding a valid Cisco Career Certification:

- Cisco CCDE®
- Cisco CCIE®
- Cisco CCDP®
- Cisco CCNP®
- Cisco CCNP® Data Center
- Cisco CCNP® Security
- Cisco CCNP® Service Provider
- Cisco CCNP® Service Provider Operations
- Cisco CCNP® Voice
- Cisco CCNP® Wireless
- Cisco CCDA®
- Cisco CCNA®
- Cisco CCNA® Data Center
- Cisco CCNA® Security
- Cisco CCNA® Service Provider
- Cisco CCNA® Service Provider Operations
- Cisco CCNA® Voice
- Cisco CCNA® Wireless

https://t.me/learningnets

It provides a gathering place for Cisco certified professionals to share questions, suggestions, and information about Cisco Career Certification programs and other certification-related topics. For more information, visit http://www.cisco.com/go/certifications.

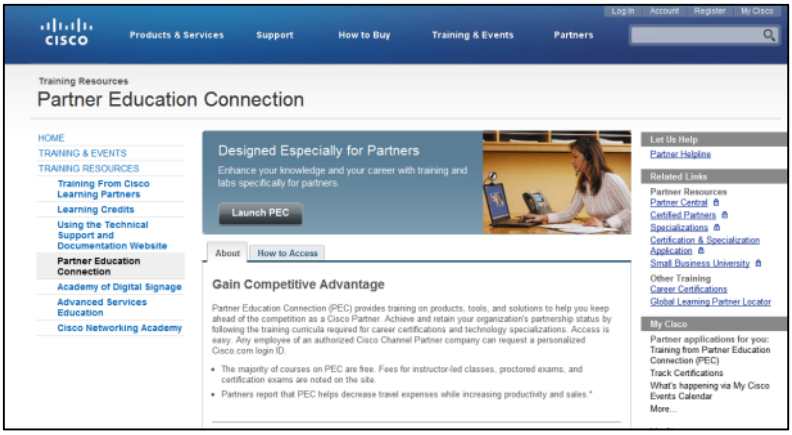
https://t.me/learningnets

Additional Resources

For additional information about Cisco technologies, solutions, and products, refer to the information available at the following pages.

Cisco Partner Education Center

http://www.cisco.com/go/pec

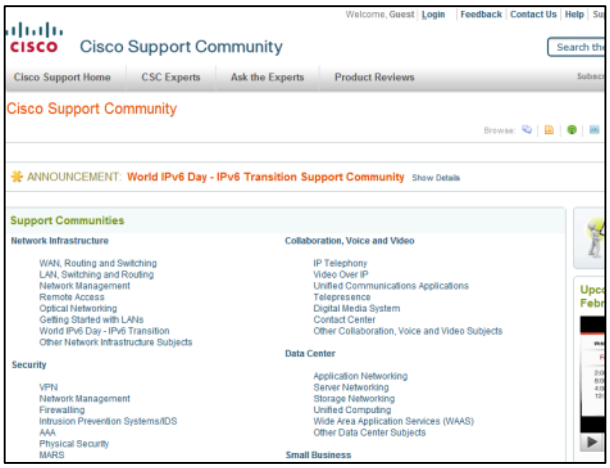


© 2012 Cisco and/or its affiliates. All rights reserved.

DCUCD v5.0—#-12

Cisco Support Community

https://supportforums.cisco.com/index.jspa



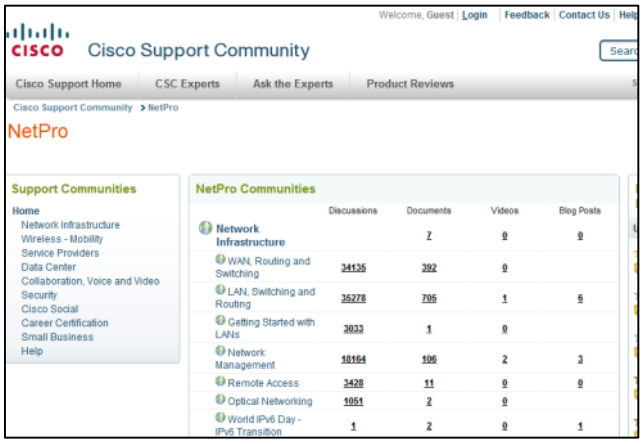
© 2012 Cisco and/or its affiliates. All rights reserved.

DCUCD v5.0—#-13

https://t.me/learningnets

Cisco NetPro Forums

https://supportforums.cisco.com/community/netpro



© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Introductions

Please use this time to introduce yourself to your classmates.

General Administration

Class-related:

- Sign-in sheet
- Length and times
- Break and lunchroom locations
- Attire
- Cell phones and pagers

Facilities-related:

- Participant materials
- Site emergency procedures
- Restrooms
- Telephones and faxes

© 2012 Cisco and/or its affiliates. All rights reserved.

Student Introductions

- Your name
- Your company
- Prerequisite skills
- Brief history
- Objective

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Module 1

Cisco Data Center Solution Architecture and Components

Overview

Modern data centers operate in high availability and are the foundation for business processes. Additionally, the cloud computing model is emerging and data centers provide the infrastructure that is needed to support various cloud computing deployments.

Cisco offers a comprehensive set of technologies and devices that are used to implement data centers, including switches, servers, security appliances, virtual appliances, and so on.

This module describes data centers, and identifies technologies and design processes to successfully design a data center.

Module Objectives

Upon completing this module, you will be able to evaluate the data center solution design process, including data center challenges, architecture, and components. This ability includes being able to meet these objectives:

- Identify data center components and trends, and understand the relation between the business, technical, and environmental challenges and goals of data center solutions
- Identify the data center applications
- Describe cloud computing, including deployment models and service categories
- Provide a high-level overview of the Cisco Data Center architectural framework and components within the solution

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Lesson 1

Identifying Data Center Solutions

Overview

Data centers are the core of any IT environment—they host the applications and data that users need. Data centers must be well-tuned to business and technical requirements—both dictating the trends and presenting the challenges to the IT staff.

This lesson describes the business, technical, and environmental challenges and goals of contemporary data center solutions.

Objectives

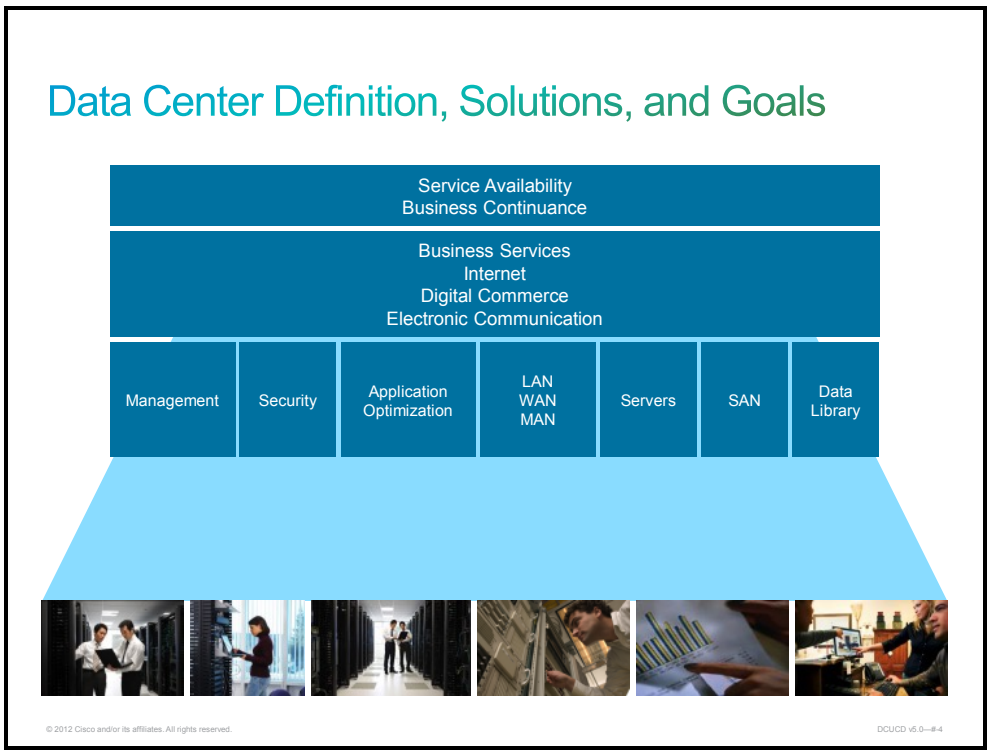
Upon completing this lesson, you will be able to identify the data center components and trends, and understand the relation between the business, technical, and environmental challenges and goals of contemporary data center solutions. This ability includes being able to meet these objectives:

- Recognize the elements of data center computing solutions
- Identify consolidation as a relevant data center trend
- Identify virtualization as a relevant data center trend
- Evaluate the business challenges of the contemporary data center solutions
- Evaluate the environmental challenges of the contemporary data center solutions
- Describe the technical challenges of the contemporary data center solutions

https://t.me/learningnets

Data Center Overview

This topic describes the elements of data center computing solutions.



Data Center Definition

A data center is a centralized or geographically distributed group of departments that houses the computing systems and their related storage equipment or data libraries. A data center has controlled centralized management that enables an enterprise to operate according to business needs.

A data center infrastructure is an essential component that supports Internet services, digital commerce, electronic communications, and other business services and solutions.

Data Center Goals

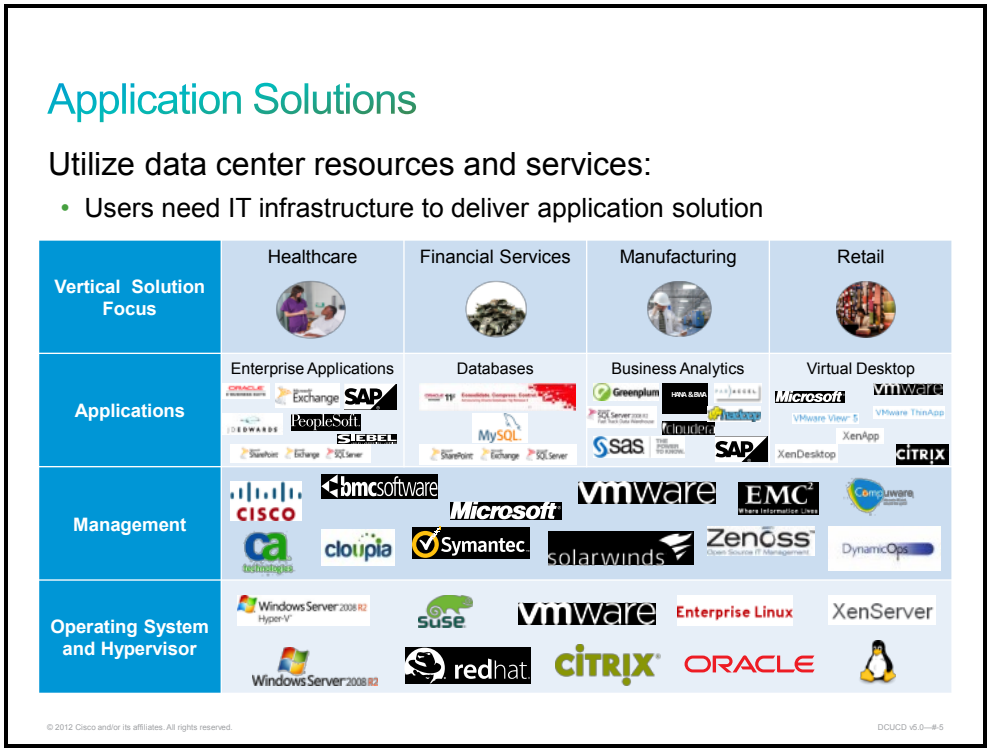
A data center goal is to sustain the business functions and operations, and to provide flexibility for future data center changes. A data center network must be flexible and support nondisruptive scalability of applications and computing resources to support the infrastructure for future business needs.

Business Continuance Definition

Business continuity is the ability to adapt and respond to risks and opportunities in order to maintain continuous business operations. There are four primary aspects of business continuity:

- High availability (disaster tolerance)
- Continuous operations
- Disaster recovery
- Disaster tolerance

https://t.me/learningnets

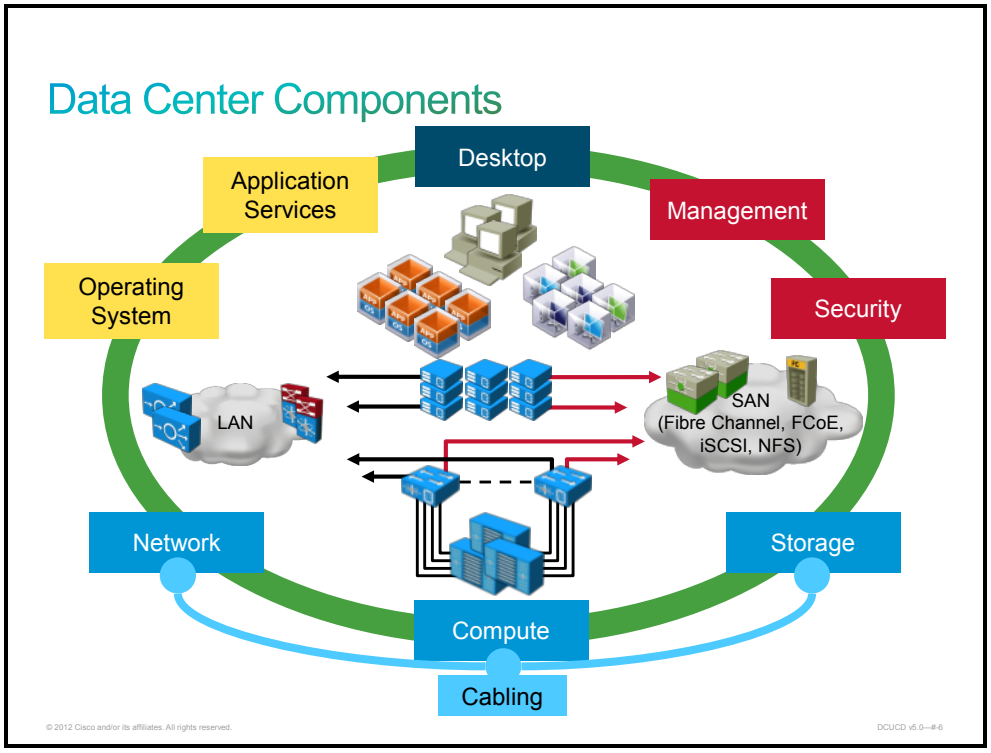


Any organization, whether it is commercial, nonprofit, or public sector (including healthcare), has applications that are mission-critical to its operations and survival. In all cases, some form of data center operations and infrastructure supports those applications.

Applications are critical for the functioning of the organization. The applications must be run on a reliable, cost-effective, flexible solution: a data center.

The primary goal of a data center is to deliver adequate resources for the applications. Users need the data center infrastructure to access applications. The data centers must be tailored as application solutions.

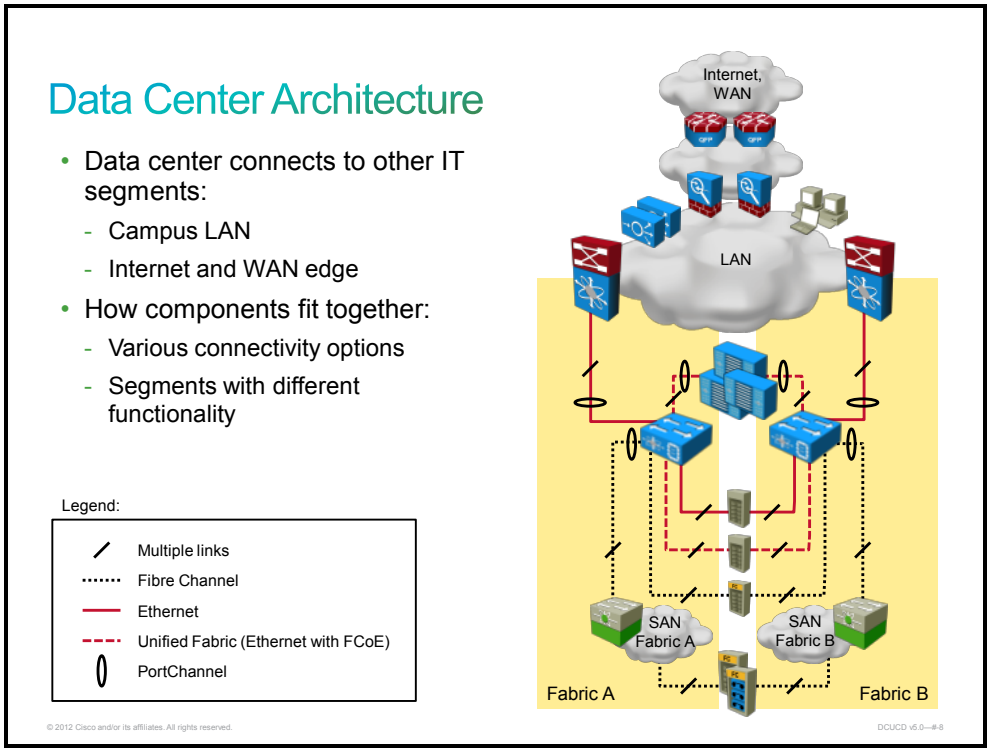
https://t.me/learningnets



Contemporary data center computing solutions encompass multiple aspects, technologies, and components:

- Network technologies and equipment, such as intelligent switches, multilayer and converged devices, high availability mechanisms, Layer 2 and Layer 3 protocols
- Storage solutions and equipment that include technologies ranging from Fibre Channel, Internet Small Computer Systems Interface (iSCSI), Network File System (NFS), Fibre Channel over Ethernet (FCoE), storage network equipment, and storage devices such as disk arrays and tape libraries
- Computing technologies and equipment, including general purpose and specialized servers. The Cisco Unified Computing System (UCS) consolidates the LAN and SAN in the management and access layers into a common infrastructure.
- Operating system and server virtualization technologies
- Application services technologies and products such as load balancers and session enhancement devices
- Management systems that are used to manage network, storage, and computing resources, operating systems, server virtualization, applications, and security aspects of the solution
- Security technologies and equipment that are employed to ensure confidentiality and security to sensitive data and systems
- Desktop virtualization solutions and access clients
- Physical cabling that connects all physical, virtual, and logical components of the data center

https://t.me/learningnets



Data center architecture is the blueprint of how components and elements of the data center are connected. The components need to correctly interact in order to deliver application services.

The data center, as one of the components of the IT infrastructure, needs to connect to other segments to deliver application services and enable users to access and use them. Such segments include the Internet, WAN edge, campus LAN, and various demilitarized zone (DMZ) segments hosting public or semi-public services.

The scheme depicts the general data center blueprint with the computing component, the Cisco UCS, as the centerpiece of the architecture. Internally, the data center is connected by LAN, SAN, or unified fabric to provide communication paths between the components. Various protocols and mechanisms are used to implement the internal architecture, with Ethernet as the key technology, accompanied by various scaling mechanisms.

https://t.me/learningnets

Other Data Center Aspects

Physical facility:

- Architectural and mechanical specifications
- Physical security
- Environmental conditions
- Power and cooling

IT organization:

- Organizational hierarchy
- Responsibilities and demarcation



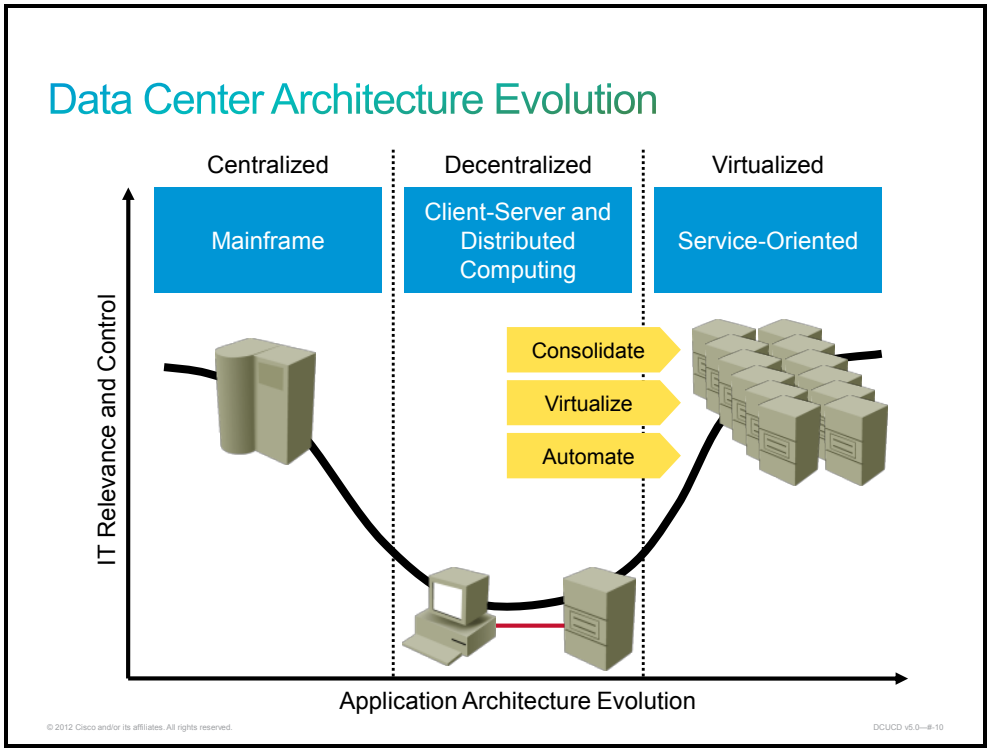


© 2012 Cisco and/or its affiliates. All rights reserved.

Apart from the already-mentioned aspects and components of the data center solution, there are two important components that influence how the solution is used and scaled:

- **Physical facility:** The physical facility includes the characteristics of the data center facility that affect the data center infrastructure, such as available power, cooling capacity, physical space and racks, physical security, fire prevention systems, and so on.
- **IT organization:** The IT organization includes the IT departments and how they interact in order to offer IT services to business users. This organization can be in the form of a single department that takes care of all IT aspects (typically, with the help of external IT partners), or, in large companies, in the form of multiple departments, with each department taking care of a subset of the data center infrastructure.

https://t.me/learningnets



Data centers have changed and evolved over time. At first, data centers were monolithic and centralized, employing mainframes and terminals that users accessed to perform their work. The mainframes are still used in the finance sector because they are an advantageous solution in terms of availability, resilience, and service level agreements (SLAs).

The second era of data center computing was characterized by pure client-server and distributed computing, with applications being designed in such a way that the users used client software to access an application, and the services were distributed due to poor computing ability and high link costs. The mainframes were too expensive.

Today, with the computing infrastructure being cheaper and with increased computing capacities, data centers are being consolidated, because the distributed approach is expensive in the long term. The new solution is equipment virtualization, making the utilization of servers more common than in the distributed approach. This solution also provides significant gains in terms of return on investment (ROI) and the total cost of ownership (TCO).

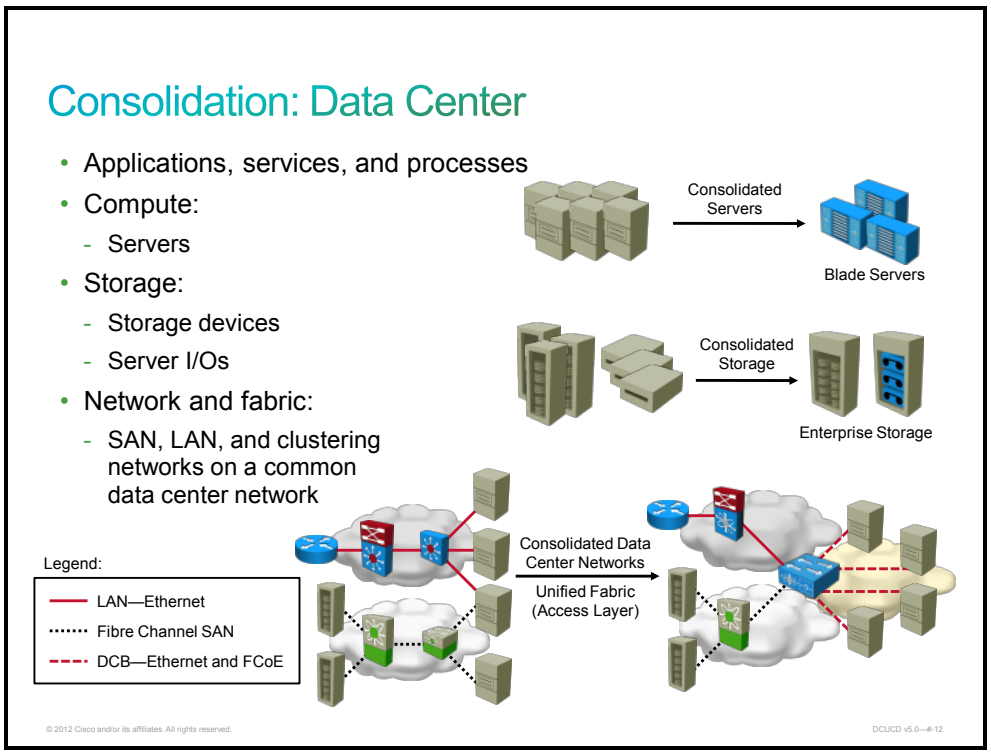
Latest data center designs and implementations have three things in common:

- Consolidation is used to unify various resources.
- Virtualization is used to ease deployment of new applications and services, and improve scalability and utilization of resources.
- The task of management is to simplify the data center design, with automation as the ultimate goal.

https://t.me/learningnets

Data Center Trends: Consolidation

This topic discusses consolidation as a relevant data center trend.

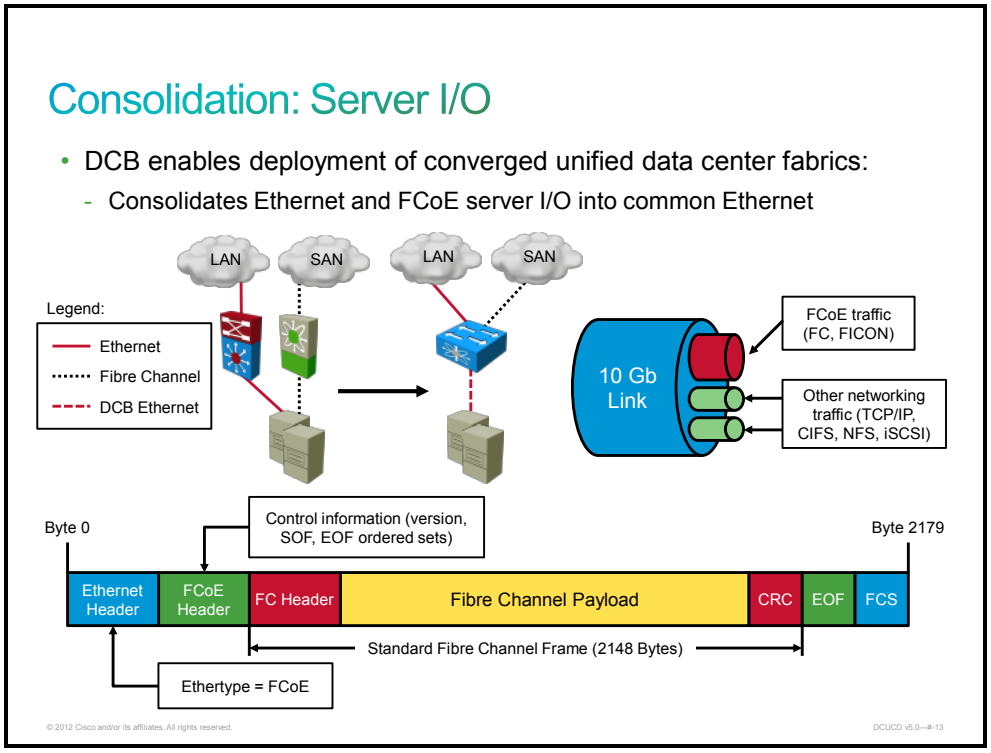


Consolidation is defined as the process of bringing together disconnected parts to make a single and complete whole. In the data center, it means replacing several small devices with a few highly capable pieces of equipment to provide simplicity.

The primary reason for consolidation is to prevent the sprawl of equipment and processes that are required to manage the equipment. It is important to understand the functions of each piece of equipment before consolidating it. There are various reasons for server, storage, server I/O, network, application, and process consolidation:

- Reduced number of servers, storage devices, networks, cables, and so on
- Increased usage of resources using resource pools (of storage and computing resources)
- Reduced centralized management
- Reduced expenses due to a smaller number of equipment pieces needed
- Increased service reliability

https://t.me/learningnets



Server I/O consolidation has been attempted several times in the past with the introduction of Fibre Channel and iSCSI protocols that carry storage, data, and clustering I/Os across the same channel.

Enhanced Ethernet is a new way of consolidating a network, using a converged network protocol that is designed to transport unified data and storage I/Os. Primary enabling technologies are PCI Express (PCIe) and 10 Gigabit Ethernet.

A growing demand for network storage is influencing demands for network bandwidth. Server virtualization allows the consolidation of multiple applications on a server, therefore influencing the server bandwidth requirement of 10 Gb/s.

10-Gb/s Data Center Bridging (DCB) uses copper and twinax cables with short distances (32.8 feet [10 meters]), but with lower cost, lower latency, and lower power requirements than 10BASE-T. FCoE and classical Ethernet can be multiplexed across the common physical DCB connection.

With the growing throughput demand, the links within the data center are becoming faster. Today, the common speed is 10 Gb/s, and in the future, 40 or even 100 Gb/s will be common.

https://t.me/learningnets

Consolidation: Server and Server Farm

Business Imperatives	IT Imperatives	Server Farm Imperatives
Cost containment	Consolidation: • Data center • Server • Storage • Operating system • Application	• Improved resource utilization—server, storage, and network • Automated infrastructure • Investment protection
Business continuance	High system and application availability	• Highly available and automated network infrastructure • Highly secure application delivery infrastructure
Agility	Improve service velocity	• Flexible and scalable network foundation to rapidly enable new applications



© 2012 Cisco and/or its affiliates. All rights reserved.

Data center server farms have diverse requirements from the perspective of integration, performance, and services. The more complex the IT infrastructure becomes, the more issues are raised:

- High operational costs, proliferation of disparate computer platforms across multiple data centers and branches, mainframe, UNIX, Windows
- Low average CPU utilization: 5 to 15 percent in the Windows operating system, 10 to 30 percent in UNIX and Linux operating systems
- Significant investment in mainframe
- Need for lower-cost, high-performance data analysis and storage resources
- Need for faster server and application deployment, new applications and services, development environments, surges in demand
- High density servers cause problems with cooling, server I/O, network connectivity

Many organizations look toward server consolidation that standardizes on blade centers or industry-standard servers (sizes from 1 rack unit [RU] to 4 RUs) that can process information much faster and lead significant traffic volumes at line rate across Gigabit Ethernet and at significant fractions of 10 Gigabit Ethernet. Organizations are also adopting, or planning to adopt, virtual server techniques, such as VMware, that further increase server densities, although at a virtual level. Additionally, these same organizations must maintain heterogeneous environments that have different applications and server platforms (blade servers, midrange servers, mainframes, and so on) that need to be factored into the data center design.

https://t.me/learningnets

Standalone Server Form Factor

When a client-server model was introduced, the data center evolved from a centralized mainframe-oriented deployment to a decentralized multiple-server deployment design. The client-server design spurred deployment of many standalone servers and the data centers evolved into a “server colony” because of using a scale-out server deployment approach to increase computing capacity and throughput. Although initially such an approach is less costly if compared to a centralized mainframe deployment, through time it became clear that space was not being efficiently used because standalone or tower servers use a considerable amount of space. Furthermore, such a standalone server deployment model is not optimal for all application types and requirements.

Rack-Optimized Server Form Factor

The next trend was to use the servers in a rack-optimized form with a better computing-to-space ratio—more servers could be deployed in less space. The rack-optimized form factor tries to address the need to optimize size, airflow, and connections, and to rationalize deployment. Although a single server unit is optimized because of power and cooling limitations, it is typically difficult, if not impossible, to deploy an entire 42-RU rack cabinet with servers. A typical data center rack cabinet has 5 to 10 kW of power and cooling capacity available for 10 to 20 servers per rack. This rack-optimized server solution still lacks cabling rationalization, serviceability simplification, and power and cooling efficiency. (Each server still has its own power supplies and fans.)

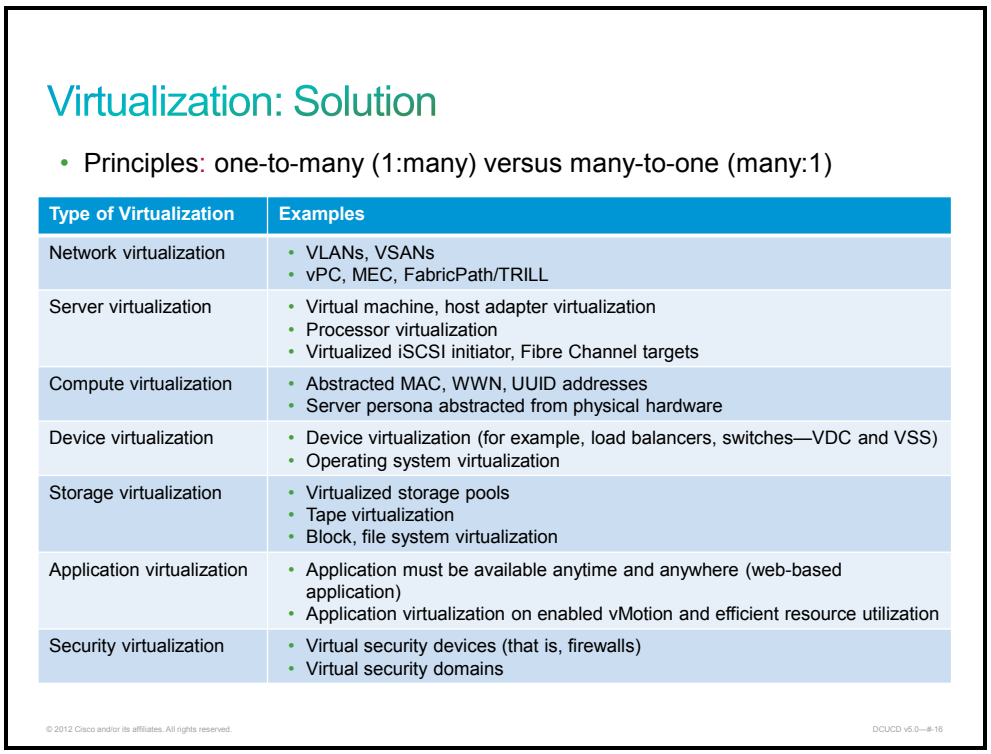
Blade Server Form Factor

The number of applications and services that data centers host has increased significantly, which has led to the popularity of the blade server form factor. The blade server form factor offers an even higher density of servers in a rack—the blade enclosure is 6 to 12 RUs high and can host from 6 to 14 blade servers.

https://t.me/learningnets

Data Center Trends: Virtualization

This topic describes virtualization as a relevant data center trend.



Virtualization offers flexibility in designing and building data center solutions. It enables enterprises with diverse networking needs to separate a single user group or data center resources from the rest of the network.

Common Goals

There are some common goals of virtualization techniques:

- **Affect utilization and reduce overprovisioning:** The main goal is to reduce operating costs of maintaining equipment that is not really needed or is not fully utilized. Overprovisioning has been used to provide for a safety margin, but with virtualization, a lower overprovisioning percentage can be used because systems are more flexible.
- **Isolation:** Security must be effective enough to prevent any undesired access across the virtual entities that share a common physical infrastructure. Performance (quality of service [QoS] and SLA) must be provided at the desired level, independently for each virtual entity. Faults must be contained.
- **Management:** Flexibly managing a virtual resource generally requires no hardware change. Administration for each virtual entity can be deployed using role-based access control (RBAC).

https://t.me/learningnets

Virtualization Types

Two virtualization types exist:

- **One-to-many:** One-to-many virtualization applies to server or network virtualization. For device virtualization, this type is used in virtual contexts on the Cisco ASA adaptive security appliance, Cisco Firewall Services Module (FWSM), and the virtual device context (VDC) on the Cisco Nexus 7000 Series Switches. In both cases, one physical device is divided into many logical devices.
- **Many-to-one:** Examples of many-to-one virtualization include storage and network system virtualization. An example of device virtualization is the Virtual Switching System (VSS) for Cisco Catalyst 6500 Series Switches, where two or more physical switches are combined into a single network element. Another example of the many-to-one type is the combination of two stackable switches into a stack.

Network Virtualization

Network virtualization can address the problem of separation. Network virtualization also provides other types of benefits such as increasing network availability, better security, consolidation of multiple networks, segmentation of networks, and increased network availability. Examples of network virtualization are VLANs and virtual SANs (VSANs) in Fibre Channel SANs. A VLAN virtualizes Layer 2 segments, making them independent of the physical topology. This virtualization gives the ability to connect two servers to the same physical switch, though they participate in different logical broadcast domains, or VLANs. A similar concept is presented by a VSAN in Fibre Channel SANs.

Server Virtualization

Server virtualization enables physical consolidation of servers on the common physical infrastructure. Deployment of a virtual server is easy because there is no need to buy a new adapter and a new server. For a virtual server to be enabled, software needs to be activated and configured properly. Server virtualization simplifies server deployment, reduces the cost of management, and increases server utilization. VMware and Microsoft are examples of companies that support server virtualization technologies.

Device Virtualization

Cisco Nexus 7000 and Cisco Catalyst 6500 Series Switches support device virtualization or Cisco Nexus Operating System (Cisco NX-OS) virtualization. A VDC represents the ability of the switch to enable multiple virtual switches on the common physical switch. This feature provides various benefits to the application services, such as higher service availability, fault isolation, separation of logical networking infrastructure that is based on traffic service types, and flexible and scalable data center design.

Storage Virtualization

Storage virtualization is the ability to pool storage on diverse and independent devices into a single view. Features such as copy services, data migration, and multiprotocol and multivendor integration can benefit from storage virtualization.

Application Virtualization

The web-based application must be available anytime and anywhere and it should be able to use unused remote server CPU resources, which implies an extended Layer 2 domain. Application virtualization enables VMware VMotion and efficient resource utilization.

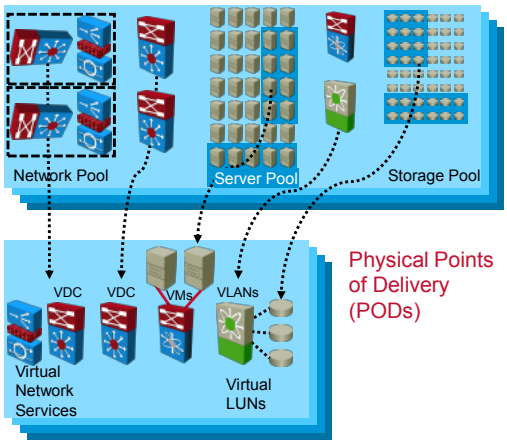
https://t.me/learningnets

Virtualization: Service and Device

Logical and Physical Data Center View

Virtualized data center POD:

- Logical instantiation of entire data center network infrastructure using VDC, VLANs, VSAN, and virtual services
- Fault isolation, high reliability
- Efficient resource pool utilization
- Centralized management
- Scalability



© 2012 Cisco and/or its affiliates. All rights reserved.

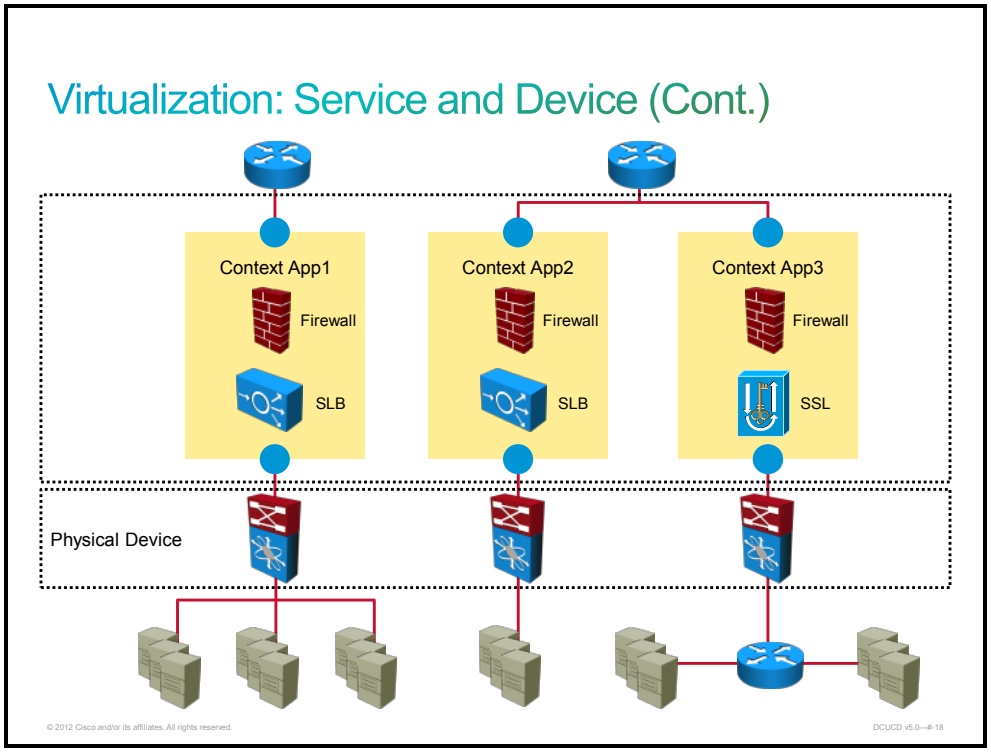
Virtualizing data center network services has changed the logical and physical data center network topology view.

Services virtualization enables higher service density by eliminating the need to deploy separate appliances for each application. There are a number of benefits of higher service density:

- Less power consumption
- Less rack space
- Reduced ports and cabling
- Simplified operational management
- Lower maintenance costs

The figure shows how virtual services can be created from the physical infrastructure, using features such as VDC, VLANs, and VSANs. Virtual network services include virtual firewalls with the Cisco adaptive security appliances (ASA or ASA-SM) or Cisco FWSM, and virtual server load-balancing contexts with the Cisco Application Control Engine (ACE) and virtual intrusion detection system (IDS).

https://t.me/learningnets



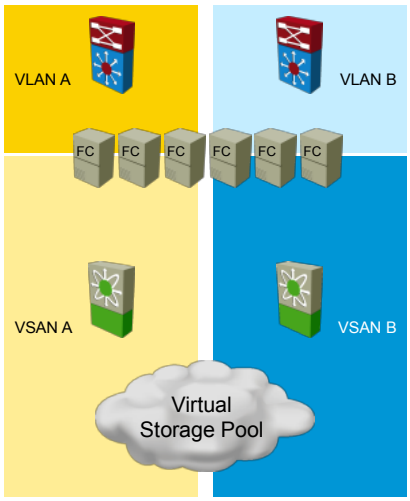
The figure shows one physical service module that is logically partitioned into several virtual service modules, and a physical switch that is logically partitioned into several virtual device contexts. This partitioning reduces the number of physical devices that must be deployed and managed, but still provides the same functionality that each device could provide.

Every device supports some kind of virtualization. Firewalls and server load-balancers support context-based virtualization, switches support VDCs, and servers use host virtualization techniques such as VMware, Microsoft Hyper-V, Citrix Xen, and so on.

https://t.me/learningnets

Virtualization: Storage and SAN

- Storage device abstracts hosts from physical disks
- SANs allow storage sharing and consolidation across multiple servers:
  - Leads to SAN proliferation (SAN islands) and poor utilization
  - Virtual SANs (VSANs) allow further consolidation of SAN islands:
    - Increased utilization
    - Easier to manage
- Storage device virtualization coupled with VSANs enable dynamic storage allocation.

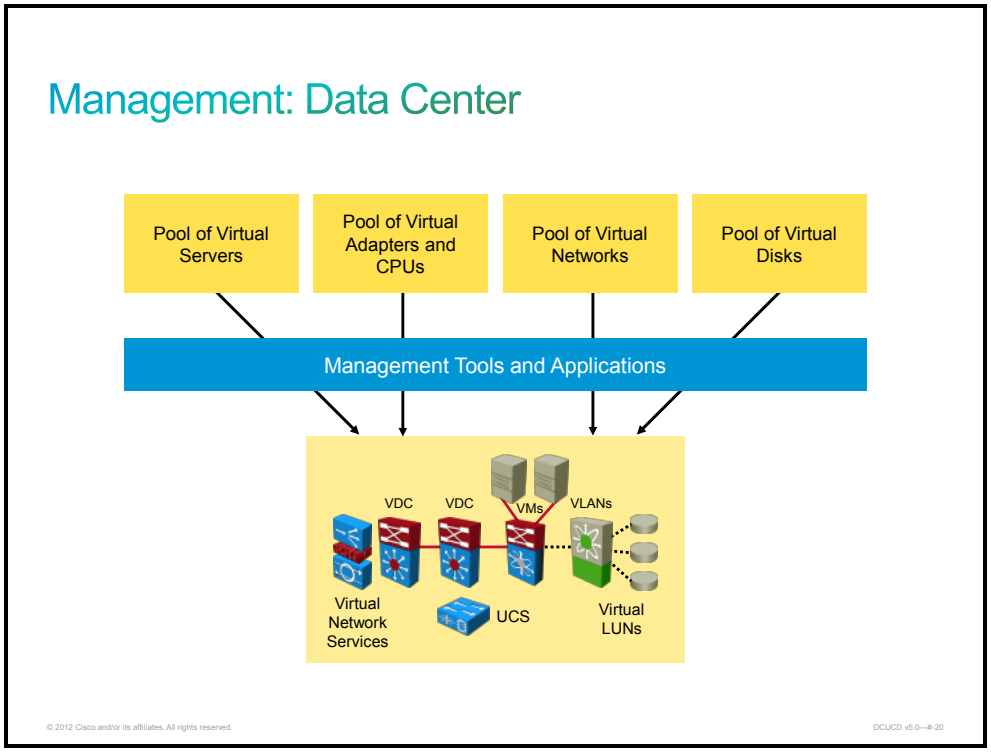


© 2012 Cisco and/or its affiliates. All rights reserved.

Data center storage virtualization starts with Cisco VSAN technology. Traditionally, SAN islands have been used within the data center to separate traffic on different physical infrastructures, providing security and separation from both a management and traffic perspective. To provide virtualization facilities, VSANs are used within the data center SAN environment to consolidate SAN islands onto one physical infrastructure, while, from the perspective of management and traffic, maintaining the separation.

Storage virtualization also involves virtualizing the storage devices themselves. Coupled with VSANs, storage device virtualization enables dynamic allocation of storage. Taking a similar approach to the integration of network services directly into data center switching platforms, the Cisco MDS 9000 platform supports third-party storage virtualization applications on an MDS 9000 services module, reducing operational costs by consolidating management processes.

https://t.me/learningnets



There are several advantages from pooling and virtualizing computing, storage, and networking resources:

- Simplified management and troubleshooting
- Increased resource utilization that results in reduced cost
- Data center service automation, which makes deployment simpler and quicker

Data center management tasks and operations can be automated based on consolidated pools of virtualized storage, computing, and networking resources. Virtualization and consolidation enable the creation of virtual server pools, virtual pools of adapters, pools of virtual processing units, virtual network pools, and pools of virtual disks.

An appliance that is attached to an existing data center that monitors application processing, computing, storage, and networking resource utilization can therefore detect the missing processing power or the lack of application storage resources, and can automatically react to it. An appliance can configure a virtual server, activate a virtual adapter, configure a server I/O channel, connect the channel across a virtual network to the dynamically allocated virtual disk, and then start the application on the newly allocated infrastructure.

In addition to policy-based resource provisioning, or the ability to automatically increase the capacity of an existing application, data center service automation also provides the ability to roll out new applications that are critical to the success of many enterprises:

- E-commerce
- Customer relationship management (CRM)
- Supply chain management (SCM)
- Human resources (HR)

https://t.me/learningnets

Policy-based provisioning is about providing an end-to-end IT service that is dynamically linked to business policies, allowing the ability to adapt to changing business conditions. For example, if a customer order entry application suddenly experiences a surge in load, just allocating more CPU might not be enough. The application might also need additional storage, more network capacity, and even additional servers and new users to process the increased activity. All of these changes must be orchestrated so that the dynamic allocation of multiple resource elements occurs seamlessly.

For servers, there are virtual server solutions such as Microsoft, Citrix, VMware, and physical server solutions (that is, modular blade server systems). For storage, there are various virtualization solutions. Almost every enterprise-class storage product supports some form of virtualization. Fabric virtualization is based on VLANs, VSANs, and VDCs.

By reducing complexity, consolidation reduces management overhead and operational expenses (OpEx). At the same time, the ability to use resources is increased because resources are no longer locked up in silos, resulting in lower capital expenditures (CapEx).

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Data Center Business Challenges

This topic describes the business challenges of the contemporary data center solutions.

Data Center Demands and Limitations

- Growing business demands:
  - Greater collaboration
  - Quicker application and information access
  - Global availability
  - Regulatory compliance
  - Organizational changes
  - Fast application deployment
- Operational limitations:
  - Power, cooling, and physical space
  - Resource utilization, provisioning, and repurposing
  - Security threats
  - Business continuance
  - Scalability limitations



© 2012 Cisco and/or its affiliates. All rights reserved.

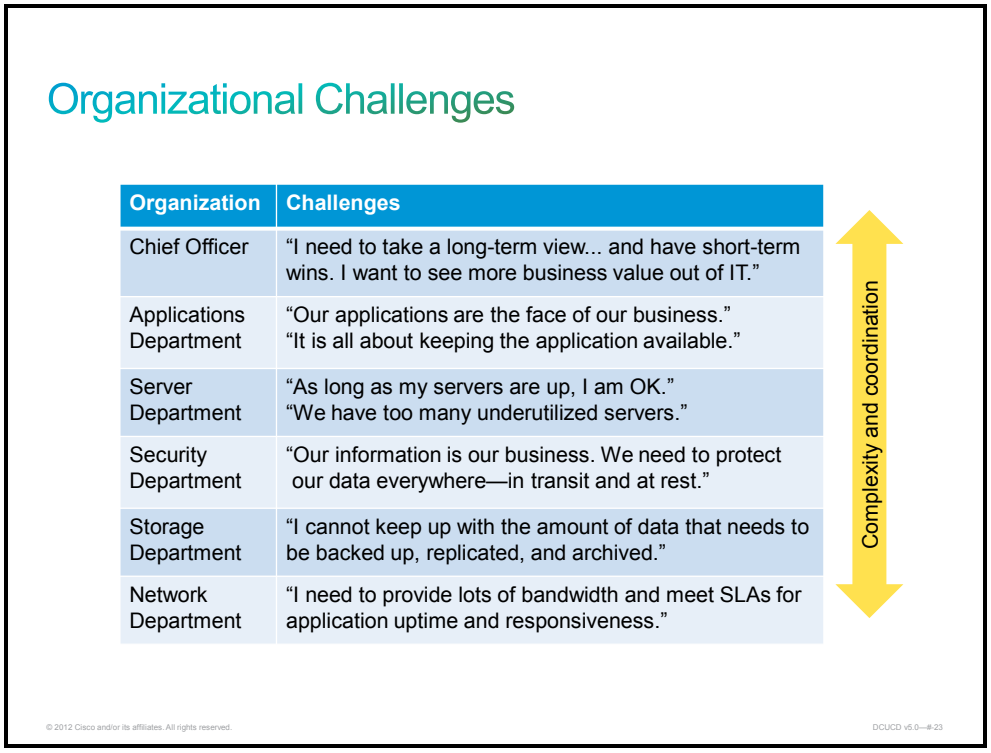
The modern enterprise is being changed by shifting business pressures and operational limitations. While enterprises prepare to meet demands for greater collaboration, quicker access to applications and information, and ever-stricter regulatory compliance, they are also being pressured by issues relating to power and cooling, efficient asset utilization, escalating security and provisioning needs, and business continuance. All of these concerns are central to data centers.

Modern data center technologies, such as multicore CPU servers and blade servers, require more power and generate more heat than older technologies, and moving to new technologies can significantly affect data center power and cooling budgets.

The importance of security is rising as well, because more services are concentrated in a single data center. If an attack were to occur in such a condensed environment, many people could be put out of work, resulting in lost time and revenue. As a result, thorough traffic inspection is required for inbound data center traffic.

Security concerns and business continuance must be considered in any data center solution. A data center should be able to provide services if an outage occurs because of a cyber-attack or because of physical conditions such as floods, fires, earthquakes, and hurricanes.

https://t.me/learningnets



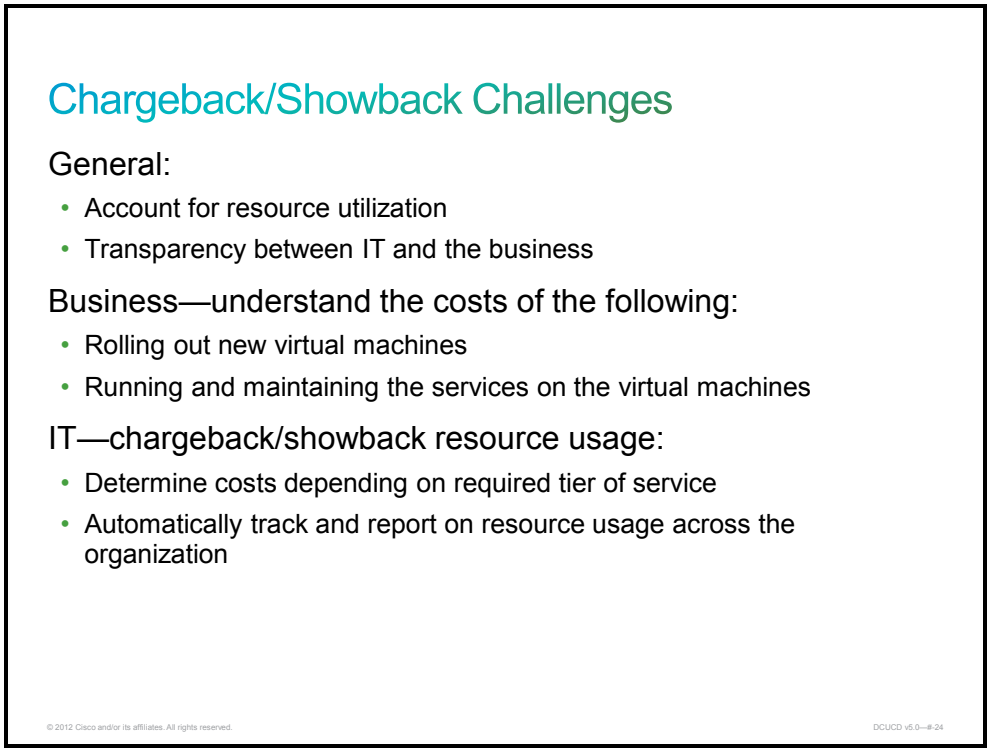
The data center is viewed from different perspectives, depending on the organization or the viewer.

Depending on which IT team you are speaking with, you will find different requirements. You have the opportunity to talk on all levels because of your strategic position and the fact that you interact with all of the different components in the data center.

Selling into the data center involves multiple stakeholders with different agendas and priorities. The traditional network contacts might get you in, but they might not be the people who make the decisions that ultimately determine how the network evolves.

The organization might be run in silos, where each silo has its own budget and power base. Conversely, many next-generation solutions involve multiple groups.

https://t.me/learningnets



IT and IT infrastructure enables business users and application owners to perform the following activities:

- Request services in a simple manner
- Specify the desired service levels
- Consume (and pay) for those services

These activities are provided with a high degree of reliability and the users do not need to understand the underlying infrastructure.

The goal is to hide the complexity of the infrastructure from the users—end users request services and IT delivers service levels with a dynamic, flexible, and reliable IT infrastructure. Such an approach transforms IT into a "service provider" with the ability to interact with end users. It also requires the IT to clearly understand and manage user expectations and ensure that the infrastructure meets user needs. Being in a role of a service provider, the IT must understand and transparently meter, report, and sometimes charge for services that are delivered.

The responsibility of IT varies from ensuring that a particular server or other piece of infrastructure is operating correctly to delivering what the business needs (for example, reliable email service, a responsive customer relationship management (CRM) system, or an e-commerce site that supports peak shopping periods).

Chargeback and Showback Rationale

IT administrators need to meter resource usage in their infrastructure, whether physical or virtual (for example, server resources, network traffic, public IP addresses, and services such as DHCP, Network Address Translation [NAT], firewalling, and so on).

To account for the operational costs that are involved in providing and maintaining an IT infrastructure, including the costs for IT services and applications (for example, licenses for the Microsoft Windows operating system or the VMware vSphere infrastructure, and so on), a chargeback or showback model needs to be defined.

https://t.me/learningnets

To address this challenge, the IT administrators can deploy and utilize the chargeback and showback tools, such as VMware Chargeback, used in VMware vSphere environments, or VKernel vOps, which can be used in other environments such as Microsoft SQL Server, Microsoft Exchange, and Active Directory.

Chargeback and Showback Aspects

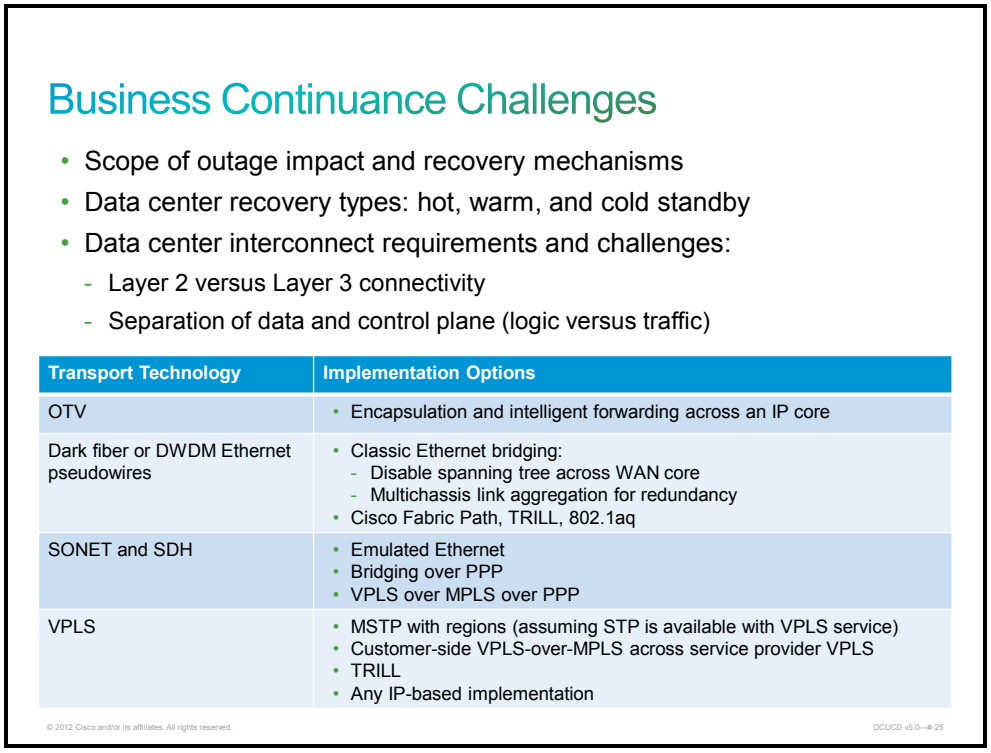
The metering tool and policy need to address the following:

- Provide accurate visibility into the true costs and usage of workloads, to aid in improving resource utilization
- Provide business owners with complete transparency and accountability for self-service resource requests
- Enable setting up an infrastructure cost model in a flexible way, including organizational processes and policies

The metering tool should have the following functions:

- **Precise cost and usage reporting:** The tool should take into account many different factors, ranging from hardware costs (CPU, memory, storage, and so on) to any additional elements such as power and cooling. It should be able to incorporate these variables to provide comprehensive information for cost and usage, enabling chargeback or showback to individual business units and the business as a whole, including the following:
  - Knowledge of the workload cost and usage
  - Proper allocation of resource costs and usage across organization units
  - Comprehensive reporting
- **Ability to customize resource cost and usage models and metrics:** The tool should enable IT administrators to enter resource cost and usage information and tune calculations, based on specific requirements, including the following:
  - Ability to add reservation- and utilization-based cost and usage policy
  - Ability to enter granular resource cost and usage policy structures (that is, base cost and usage model, fixed cost and usage, and multiple rates) to calculate proper resource cost and usage
  - Ability to export the information, create reports, and import any existing cost and usage policies
- **Simplify billing and usage reporting:** The tool should automatically create detailed billings and usage reports that can be submitted to business units within an organization to provide them with a clear view of the resources that are consumed, and their associated costs, if necessary.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Business Continuance

Business continuance is one of the main reasons to implement data center interconnections, and may dictate the use of a disaster recovery site. You should always try to lower the probability of a disaster scenario by migrating the workload before an anticipated disaster.

Business needs may also dictate that you use an active/active data center, where multiple data centers are active at the same time. The same application runs concurrently in multiple data centers, which provides the optimal use of resources.

The table shows Layer 2 Cisco Data Center Interconnect (DCI) transport technologies and their implementation options.

Note The Unified Fabric and related technologies such as DCI are discussed in detail in the DCUFD course.

Technology Challenges

Technology requirements when interconnecting data centers may require that you replicate storage to the disaster recovery site. For this to be possible, you may need WAN connectivity at the disaster recovery site. You should always try to lower the probability of disaster by adjusting the application load, WAN connectivity, and load balancing.

From the technology perspective, several challenges exist:

- Control and data plane separation (that is, logic versus traffic node active role)—for example, in active/standby solutions, questions arise about the location of the active/standby firewall and how data should flow.
- Which technology to use to connect two or more data center locations? There are several options, as indicated in the table.

https://t.me/learningnets

- How to address active/active solutions—you need to use global load balancing to handle requests and traffic flows between data centers.

Note	The Cisco global server load balancing (GSLB) solution is the Cisco ACE Global Site Selector.

Outage Impact and Recovery Types

There are different types of outages that might affect the data center functions and operation. An outage in data center operations can occur and can damage the data center at any level.

Typically, the types of outages are classified based on the scope of the outage impact:

- **Outage with an impact at the data center level:** An outage of this type is an outage of a system or a component such as hardware or software. These types of outages can be recovered using reliable, resilient, and redundant data center components, using fast routing and switching reconvergence, and stateful module and process failovers.
- **Outage with an impact at the campus level:** This type of outage affects a building or an entire campus. Fire or loss of electricity can cause damage at the campus level and can be recovered using redundant components such as power supplies and fans, or by using the secondary data center site or Power over Ethernet (PoE).
- **Outage with an impact at the regional level:** This type of outage affects a region, such as earthquakes, flooding, or tornados. Such outages can be recovered using geographically dispersed, standby data centers that use global site selection and redirection protocols to seamlessly redirect user requests to the secondary site.
- **Data center recovery types:** Different types of data center recovery provide different levels of service and data protection, such as cold standby, warm standby, hot standby, immediate recovery, continuous availability, continuous operation, gradual recovery, and back-out plan.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Data Center Environmental Challenges

This topic evaluates the environmental challenges for contemporary data center solutions.

Facility Challenges

- Architectural and mechanical specifications:
  - Space available
  - Load capacity
  - Power (electrical) capacity
  - Cooling capacity
  - Cabling infrastructure
- Limited capacities
- Compliance and regulations
- New versus existing solution or facility
- Physical security:
  - Access to the premises
  - Fire suppression
- Environmental conditions:
  - Operating temperature
  - Humidity level



© 2012 Cisco and/or its affiliates. All rights reserved.

The data center facility has multiple aspects that need to be addressed when the facility is being planned, designed, and built, because the facility capacities are limited and need to be correctly designed.

The companies must also address regulatory issues, enable business resilience, and comply with environmental requirements. Data centers need infrastructures that can protect and recover applications, communications, and information, and that can provide uninterrupted access.

In building a reliable data center and maximizing an investment, the design must be considered early in the building development process and should include coordinated efforts that cut across several areas of expertise, including telecommunications, power, architectural components, and heating, ventilating, and air conditioning (HVAC) systems.

Each of the components of the data center and its supporting systems must be planned, designed, and implemented to work together to ensure reliable access while supporting future requirements. Neglecting any aspect of the design can render the data center vulnerable to cost failures, early obsolescence, and intolerable levels of availability. There is no substitute for careful planning and following the guidelines for data center physical design.

Architectural and Mechanical Specifications

The architectural and mechanical facility specifications must consider the following:

- The amount of available space
- The load that a floor can bear
- The power capacity that is available for data center deployment

https://t.me/learningnets

- Available cooling capacity
- The cabling infrastructure type and management

In addition, the facility must meet certain environmental conditions: the types of data center devices define the operating temperatures and humidity levels that must be maintained.

Physical Security

Physical security is vital because the data center typically houses data that should not be available to third parties, so access to the premises must be well controlled. Protection from third parties is important, as well as protection of the equipment and data from certain disasters. Fire suppression equipment and alarm systems to protect against fires should be in place.

Space

The space aspect involves the physical footprint of the data center—how to size the data center, where to locate servers within a multipurpose building, how to make it adaptable for future needs and growth, and how to construct the data center to effectively protect the valuable equipment inside.

The data center space defines the number of racks that can be used and thus the equipment that can be installed. That is not the only parameter—equally important is the floor-loading capability, which determines which and how much equipment can be installed into a certain rack and thus what the rack weight should be. The placement of current and future equipment must be very carefully considered so that the data center physical infrastructure and support is optimally deployed.

Although sometimes neglected, the size of the data center has a great influence on cost, lifespan, and flexibility. Determining the proper size of the data center is a challenging and essential task that should be done correctly and must take into account several variables:

- The number of people supporting the data center
- The number and type of servers and the storage and networking equipment that is used
- The sizes of the server, storage, or network areas, which depend on how the passive infrastructure is deployed

A data center that is too small will not adequately meet server, storage, and network requirements and will thus inhibit the productivity and will incur additional costs for upgrades or expansions.

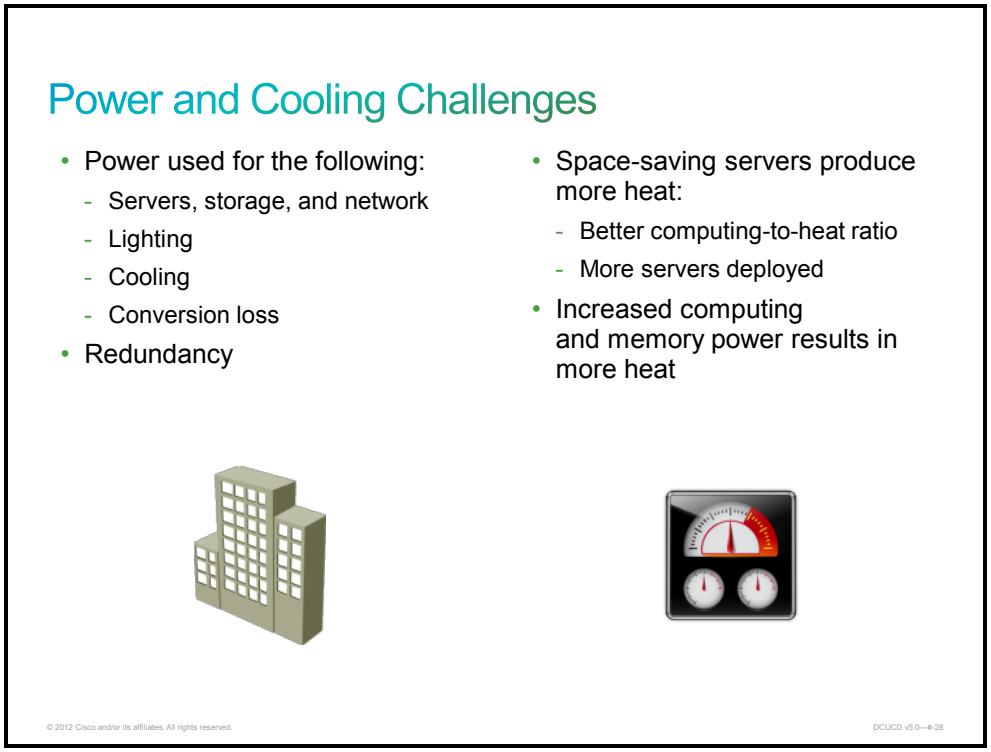
Alternatively, a data center that is too spacious is a waste of money, not only from the initial construction cost but also from the perspective of ongoing operational expenses.

Correctly sized data center facilities also take into account the placement of equipment. The data center facility should be able to grow, when needed. Otherwise, costly upgrades or relocations must be performed.

Cabinets and racks are part of the space requirements and other aspects must be considered:

- Loading, which determines what and how many devices can be installed
- The weight of the rack and equipment that is installed
- Heat that is produced by the equipment that is installed
- Power that is consumed by the equipment that is installed

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Power

The power in the data center facility is used to power servers, storage, network equipment, lighting, and cooling devices (which take up most of the energy). Some power is also lost upon conversion.

The variability of usage is difficult to predict when determining power requirements for the equipment in the data center. For the server environment, the power usage depends on the computing load. If the server must work harder, more power has to be drawn from the power supply and there is greater heat output that needs to be dissipated.

Power requirements are based on the desired reliability and may include two or more power feeds from the utility, an uninterruptible power supply (UPS), multiple circuits to systems and equipment, and on-site generators. Determining power requirements requires careful planning.

Estimating power needs involves determining the power that is required for all existing devices and for devices that are anticipated in the future. Power requirements must also be estimated for all support equipment such as the UPS, generators, conditioning electronics, HVAC system, lighting, and so on. The power estimation must include required redundancy and future growth.

The facility electrical system must not only power data center equipment (servers, storage, network equipment, and so on) but must also insulate the equipment against surges, utility power failures, and other potential electrical problems (thus addressing the redundancy requirements).

The power system must physically accommodate electrical infrastructure elements such as power distribution units (PDUs), circuit breaker panels, electrical conduits, wiring, and so on.

https://t.me/learningnets

Cooling

The temperature and humidity conditions must be controlled and considered by deploying probes to measure temperature fluctuations, data center hotspots, and relative humidity, and by using smoke detectors.

Overheating is an equipment issue with high-density computing:

- More heat overall
- Hotspots
- High heat and humidity, which threaten equipment life spans
- Computing power and memory requirements, which demand more power and generate more heat
- Data center demand for space-saving servers: density = heat. 3 kilowatts (kWs) per chassis is not a problem for one chassis, but five or six chassis per rack = 20 kW
- Humidity levels that affect static electricity and condensation—maintaining a 40 to 55 percent relative humidity level is recommended

The facilities must have airflow to reduce the amount of heat that is generated by concentrated equipment. Adequate cooling equipment must be available for flexible cooling. Additionally, the cabinets and racks should be arranged in an alternating pattern to create “hot” and “cold” aisles. In the cold aisle, equipment racks are arranged face-to-face. In the hot aisle, the equipment racks are arranged back-to-back. Perforated tiles in the raised floor of the cold aisles allow cold air to be drawn into the face of the equipment. This cold air washes over the equipment and is expelled out of the back into the hot aisle. In the hot aisle, there are no perforated tiles. This fact keeps the hot air from mingling with the cold air.

Because not every active piece of equipment exhausts heat out of the back, other considerations for cooling include the following:

- Increasing airflow by blocking unnecessary air escapes or by increasing the height of the raised floor
- Spreading equipment out over unused portions of the raised floor, if space permits
- Using open racks instead of cabinets when security ID is not a concern, or using cabinets with mesh fronts and backs
- Using perforated tiles with larger openings

Helpful Conversions

One watt is equal to 3.41214 British thermal units (BTUs). This is a generally used value for converting electrical values to BTUs, and vice versa. Many manufacturers publish kW, kilovolt-ampere (kVA), and BTU measurements in their equipment specifications. Sometimes, dividing the BTU value by 3.41214 does not equal the published wattage. Where the information is provided by the manufacturer, use it. Where it is not provided, this formula can be helpful.

Increasing Heat Production

Although the blade server deployment optimizes the computing-to-heat ratio, the heat that is produced actually increases because the blade servers are space-optimized and allow more servers to be deployed in a single rack. High-density equipment produces much heat.

In addition, the increased computing and memory power of a single server results in higher heat production. Therefore, the blade server deployment results in more heat being produced, which

© 2012 Cisco Systems, Inc

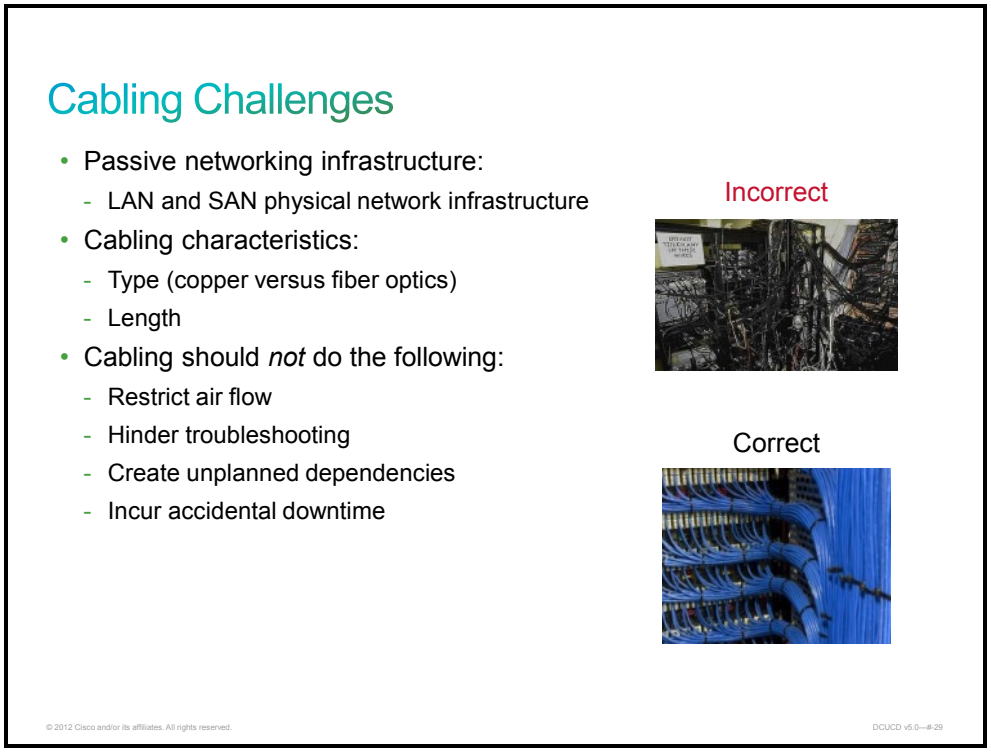
https://t.me/learningnets

requires appropriate cooling capacity and good data center design. Solutions that address the increasing heat requirements must be considered when blade servers are deployed within the data center. The design must take into consideration the cooling that is required for the current sizing of the data center servers, but the design must also anticipate future growth, thus also taking into account future heat production.

If cooling is not properly addressed, the result is a shortened equipment life span. Cooling solutions include the following:

- Increasing the space between the racks and rows
- Increasing the number of HVAC units
- Increasing the airflow through the devices
- Using new technologies such as water-cooled racks

https://t.me/learningnets



Cabling Infrastructure

The data center cabling (the passive infrastructure) is equally important for proper data center operation. The infrastructure needs to be a well-organized physical hierarchy that aids the data center operation. While the electrical infrastructure is crucial for keeping server, storage, and network devices operating, the physical network—the cabling, which runs and terminates between devices—dictates if and how these devices communicate with each other and the outside world.

The cabling infrastructure also governs the physical connector and the media type of the connector. Two options are widely used today—copper-based cabling and fiber optics-based cabling.

Fiber optics-based cabling is less susceptible to external interferences and covers greater distances, while copper-based cabling is common and less costly. The cabling must be abundant to provide ample connectivity and must employ various media types to accommodate different connectivity requirements, but it must remain well organized for the passive infrastructure to be simple to manage and easy to maintain (no one wants a data center where the cables are on the floor, creating a health and safety hazard). Typically, the cabling needs to be deployed in tight spaces, terminating at various devices.

Cabling Aspects

Cabling usability and simplicity are affected by the following:

- Media selection
- Number of connections provided
- Type of cabling termination organizers

These parameters must be addressed during the initial facility design, and the server, storage, and network components and all the technologies to be used must be considered.

https://t.me/learningnets

Problems to avoid with the cabling infrastructure include the following:

- Improper cooling due to restricted airflow
- Difficult-to-implement troubleshooting
- Unplanned dependencies that result in more downtime upon single component replacement
- Downtimes due to accidental disconnects

For example, with under-floor cabling, airflow is restricted by the power and data cables. Raised flooring is a difficult environment in which to manage cables because cable changes mean lifting floor panels and potentially having to move equipment racks.

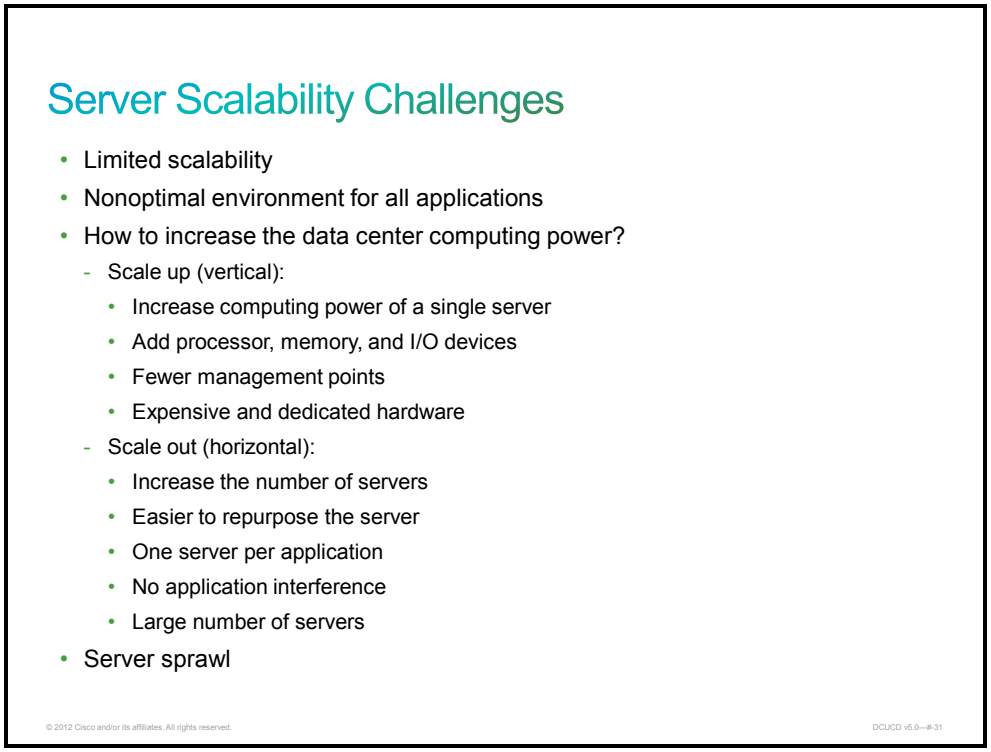
The solution is a cable management system that consists of integrated channels for connectivity that are located above the rack. Cables should be located in the front or rear of the rack for easy access. Typically, cabling is located in the front of the rack in service provider environments.

When data center cabling is deployed, the space constraints and presence of operating devices (namely servers, storage, and networking equipment) make the cabling infrastructure reconfiguration very difficult. Thus, scalable cabling is crucial for good data center operation and lifespan. Conversely, poorly designed cabling will incur downtime due to reconfiguration or expansion requirements that were not considered by the original cabling infrastructure shortcomings. The designer of a data center should work with the facilities team that installs and maintains the data center cabling in order to understand the implications of a new or reconfigured environment in the data center.

https://t.me/learningnets

Data Center Technical Challenges

This topic describes technical challenges for contemporary data center solutions.



Each server has a limited amount of resources in terms of processor, memory, or I/O throughput. Depending on resource availability, a server can run either one or several applications.

Regardless of the server power and resource capacity, these limitations are eventually reached. Furthermore, not every server type is optimal for different applications.

There are two approaches used to scale the computing power:

- **Scale-up (vertical) approach:** The scale-up approach scales computing power vertically by adding resources to a single server, adding processors, memory, or I/O devices, and so on. The benefits of this approach are as follows:
  - It enables you to leverage server virtualization technologies more effectively by providing more resources to the hosted operating system and applications.
  - There are few management points, thus the management overhead is controlled.
  - If combined with server virtualization, the solution is more power- and cooling-efficient.

  There are also drawbacks to this approach:
  - Use of dedicated and more expensive hardware
  - Limited number of operating systems
  - Hard limitation of the total amount of load per server

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

- **Scale-out (horizontal) approach:** A scale-out approach scales the computing power horizontally by increasing the number of servers. The benefits of this scaling approach are as follows:
  - Easy repurposing of a server
  - Uses less-expensive commoditized servers
  - Applications do not interfere with each other because applications can be deployed per-server

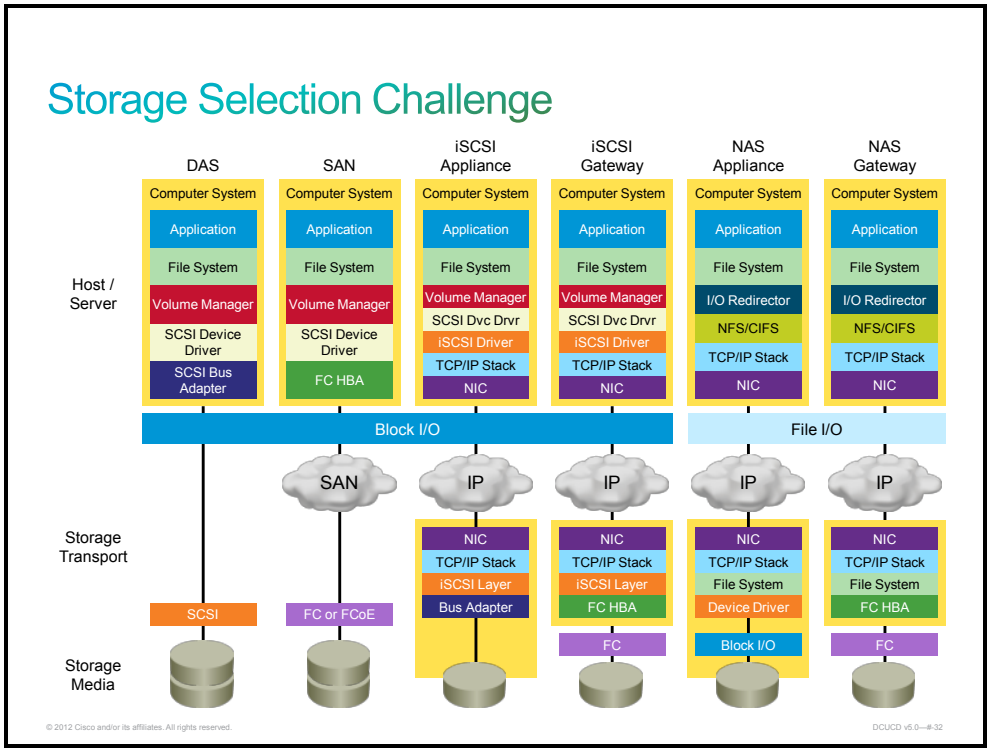
  The major drawback of the scale-out approach is that it results in a large number of servers, thus increasing management complexity.

Server Sprawl

Although the evolution of the server form-factor from a standalone (tower) server, to rack-optimized server, to blade server has led to better space usage, it also brings new challenges. A denser server deployment results in increased floor loading, which must be taken into account when deploying the solution. The greater challenge comes from the application deployment approach—one application per server leads to an even higher number of servers deployed. Even if server virtualization is used, challenges remain.

When a large number of servers are deployed, the total power consumption of the data center increases, and the data center topology, physical cabling, and overall management become more complex and difficult. Also, the scalability of the solution is limited.

https://t.me/learningnets



The storage component of the data center architecture encompasses various protocols and device options.

Network-Attached Storage

Network-attached storage (NAS) is file-level computer data storage that is connected to a computer network, and provides data access to heterogeneous clients. NAS devices enable users to attach scalable, file-based storage directly to existing LANs, based on IP and Ethernet, which provides easy installation and maintenance.

NAS systems are networked appliances that contain one or more hard drives, often arranged into logical, redundant storage containers or Redundant Array of Independent Disks (RAID) arrays. NAS removes the responsibility of file serving from other servers on the network, which typically provide access to files using network file sharing protocols such as Network File System (NFS). A NAS unit is a computer that is connected to a network that only provides file-based data storage services to other devices on the network.

Internet Small Computer Systems Interface

Internet Small Computer Systems Interface (iSCSI) is an IP-based storage networking standard for linking data storage facilities. By carrying SCSI commands over IP networks, iSCSI is used to make data transfers possible over intranets and to manage storage over long distances. iSCSI can be used to transmit data over LANs and can enable location-independent data storage and retrieval. The protocol allows initiators to send SCSI commands to SCSI storage devices (targets) on remote servers. It allows organizations to consolidate storage into data center storage arrays while providing hosts (such as database and web servers) with the illusion of locally attached disks. Unlike traditional Fibre Channel, which requires special-purpose cabling, iSCSI can be run over long distances using existing network infrastructure.

https://t.me/learningnets

Storage Area Network

A SAN is a dedicated storage network that provides access to consolidated, block-level storage. SANs are primarily used to make storage devices accessible to servers so that the devices appear as being locally attached to the operating system. A SAN typically has its own network of storage devices that are generally not accessible through the regular network by regular devices. A SAN alone does not provide the “file” abstraction but provides only block-level operations.

Storage Options Comparison

The key difference between direct-attached storage (DAS) and NAS is that DAS is simply an extension to an existing server and is not necessarily networked. NAS is designed as an easy and self-contained solution for sharing files over the network.

When both are served over the network, NAS may have better performance than DAS because the NAS device can be tuned precisely for file serving, which is less likely to happen on a server that is responsible for other processing. Both NAS and DAS can have a different amount of cache memory, which can affect performance. When you compare the use of NAS with the use of local (non-networked) DAS, the performance of NAS depends mainly on the speed of the network and congestion on the network.

Despite their differences, SAN and NAS are not mutually exclusive. They may be combined as a SAN-NAS hybrid, which offers both file-level protocols (it serves up a file) and block-level protocols (it gives you a disk drive) from the same system.

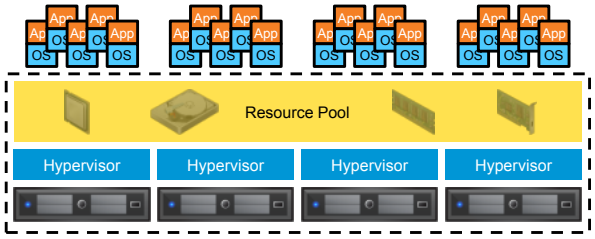
Many data centers use Ethernet for TCP/IP networks and Fibre Channel for SANs. With FCoE, Fibre Channel becomes another network protocol running on Ethernet alongside traditional IP traffic. FCoE operates directly above Ethernet in the network protocol stack, in contrast to iSCSI, which runs in addition to TCP and IP.

Because classic Ethernet has no flow control, unlike Fibre Channel, FCoE requires enhancements to the Ethernet standard to support a flow control mechanism, which prevents frame loss.

https://t.me/learningnets

Virtualization Challenges

- Virtual domains are growing fast and becoming larger
- Network administrators are involved in virtual infrastructure deployments:
  - The network access layer must support consolidation and mobility
  - Higher network (LAN and SAN) attach rates are required:
    - Multicore processor deployment affects virtualization and networking requirements
- Virtualization solution infrastructure management:
  - Where is the management demarcation?



© 2012 Cisco and/or its affiliates. All rights reserved.

Virtualization, though being the "promised solution" for server-, network-, and space-related problems, presents a few challenges.

- **The complexity factor:** The leveraging of high-density technologies, such as multicore processors, unified fabrics, higher density memory formats, and so on, increases complexity of equipment and networks.
- **Support efficiency:** Trained personnel are required to support such networks and the support burden is heavier. However, new generation management tools ease these tasks.
- **The challenges of virtualized infrastructure:** These challenges involve management, common policies, security aspects, and adaptation of organizational processes and structures.

All these aspects require higher integration and collaboration from the personnel of the various service teams.

LAN and SAN Implications

The important challenge that server virtualization brings to the network is the loss of administrative control on the network access layer. By moving the access layer into the hosts, the network administrators have no insight into configuration or troubleshooting of the network access layer. On the other hand, when obtaining access, network administrators are faced with virtual interface deployments.

Second, by enabling mobility for virtual machines (VMs), information about the VM connection point becomes lost. If the information is lost, the configuration of the VM access port does not move with the machine.

https://t.me/learningnets

Virtualization Challenges (Cont.)

- Cluster workload size:
  - Defined by the number of virtual machines running per host
  - Server and desktop virtualization results in more than 100 virtual machines per host
- Increased infrastructure load:
  - Host CPU, memory, and I/O define scalability limits
  - Lack of resources—not only size aspect but performance
- Fault domain size for high availability (influences recovery times)



© 2012 Cisco and/or its affiliates. All rights reserved.

Third, by using virtualization, the servers, network, and storage facilities are under increased loads, and therefore they need more resources and better performance. For example, storage required to support multiple VMs running at the same time must provide sufficient I/O operations per second (IOPS).

Server virtualization results in multiple VMs being deployed on a single physical server. Though the resource utilization is increased, which is desired, this increase can result in more I/O throughput. When there is more I/O throughput, more bandwidth is required per physical server.

To solve this challenge, multiple interfaces are used to provide server connectivity.

- Multiple Gigabit Ethernet interfaces provide LAN connectivity for data traffic to flow to and from the clients or to other servers. Using multiple interfaces also ensures that the redundancy requirement is correctly addressed.
- Multiple Fibre Channel interfaces provide SAN connectivity for storage traffic to allow servers and therefore VMs to access storage on a disk array.
- Typically, a dedicated management interface is also provided to allow server management.

Virtualization thus results in a higher interface count per physical server and, with SAN and LAN infrastructures running in parallel, there are the following implications:

- The network infrastructure costs more and is less efficiently used.
- There are a higher number of adapters, cabling, and network ports, which results in higher costs.
- Multiple interfaces also cause multiple fault domains and more complex diagnostics.

The use of multiple adapters increases management complexity. More management effort is put into proper firmware deployment, driver patching, and version management.

https://t.me/learningnets

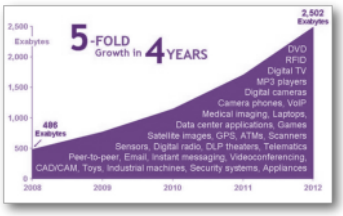
Data and Traffic Explosion

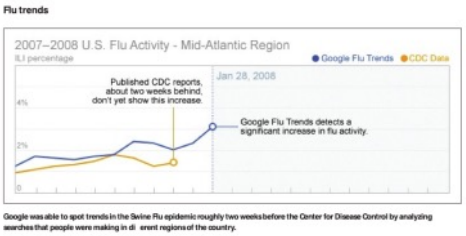
Data explosion:

- IDC estimates that 2.44 zettabytes (more than 2500 billion GB) of information will be created in 2012

Traffic explosion—bandwidth consumption:

- Social applications
- Video and voice
- Virtual desktop infrastructure





Source: http://www.storagenewsletter.com/news/miscellaneous/idc-digital-information-created

© 2012 Cisco and/or its affiliates. All rights reserved.

Data and traffic are rising at exponential rates, largely due to widespread use of social networking and applications. The growth of video and voice content is also caused by smart personal devices (for example, smart phones and tablets), which enable users to access and share information anywhere.

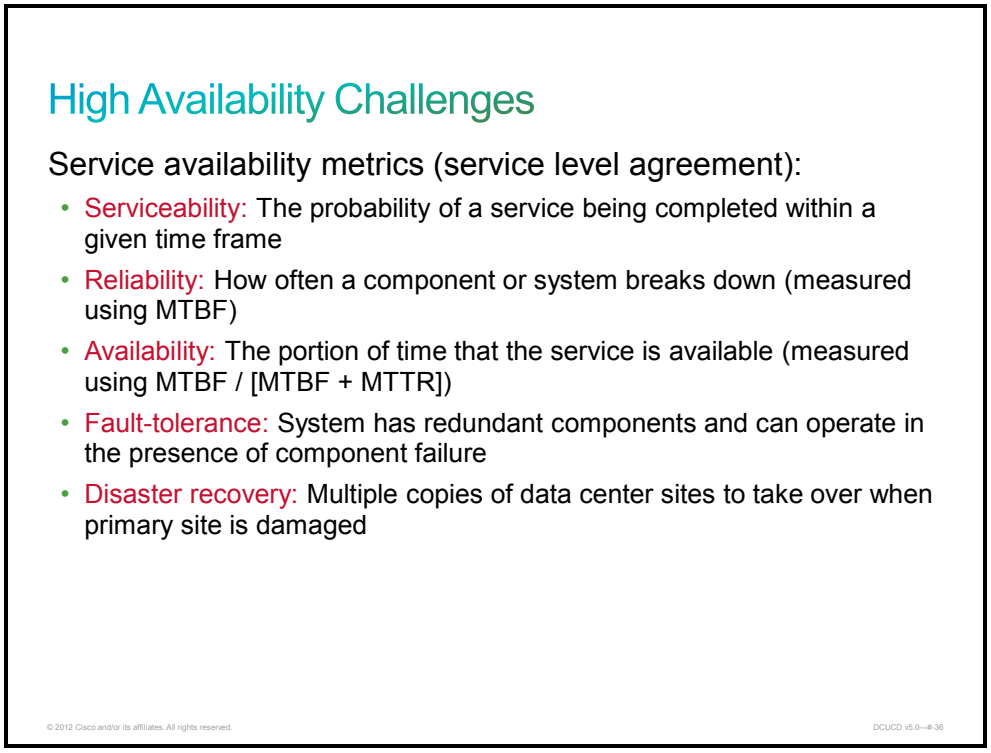
Part of the traffic growth that translates in more bandwidth being used is also a consequence of virtual desktop infrastructure (VDI) solutions, which bring many benefits but also impose more stress on links and connectivity.

In addition to traffic growth, the amount of data is exploding. The International Data Corporation (IDC), for example, estimates that 2.44 zettabytes of information will be created in 2012, which would translate into more than 2.5 billion 1-TB disk drives.

Note 1 zettabyte = 1024 exabytes = 1,048,576 petabytes = 1,073,741,824 terabytes

Infrastructure and application hardware, with the data center at the core, must be able to scale and manage the amount of data and traffic.

https://t.me/learningnets



Protecting against failure is expensive, and downtime is also expensive. It is important to identify the most serious causes of service failure and to build cost-effective safeguards against them. A high degree of component reliability and data protection with redundant disks, adapters, servers, clusters, and disaster recovery decreases the chances of service outages. Data center architecture that provides service availability must consider several different levels of high-availability features, such as the following:

- **Serviceability:** Serviceability is the probability of a service being completed within a given time. For example, if a system has serviceability of 0.98 for 3 hours, then there is a 98 percent probability that the service will be completed within 3 hours. In an ideal situation, a system can be serviced without any interruption of user support.
- **Reliability:** Reliability represents the probability of a component or system not to encounter any failure over a time span. The focus of reliability is to make the data center system components unbreakable. Reliability is a component of high availability that measures how rarely a component or system breaks down, and is expressed as the mean time between failures (MTBF). For example, a battery may have a useful life of 10 hours, but its MTBF is 50,000 hours. In a population of 50,000 batteries, this translates into one battery failure every hour during its 10-hour life span. Mean time to repair (MTTR) is the average time that is required to complete a repair action. A server with 99 percent reliability will be down for 3.65 days every year.

  For a system with 10 components, where each component can fail independently but have 99 percent reliability, the reliability of the entire system is not 99 percent. The entire system reliability is 90.44 percent (0.99 to the 10th power). This translates to 34.9 days of downtime in one year. Hardware reliability problems cause only 10 percent of the downtime. Software, human, or process-related failures comprise the other 90 percent.

https://t.me/learningnets

- **Availability:** Availability is the portion of time that an application or service is available for productive work. A more resilient system results in higher availability. An important decision to consider when building a system is the required availability level, which is a compromise between the downtime cost and the cost of the high availability configuration. Availability measures the ability of a system or group of systems to keep the application or service operating. Designing for availability assumes that the system will fail and that the system is configured to mask and recover from component-to-server failures with minimum application outage. Availability can be calculated as follows:

  Availability = MTBF / (MTBF + MTTR)

  or

  Availability = Uptime / (Uptime + Downtime)

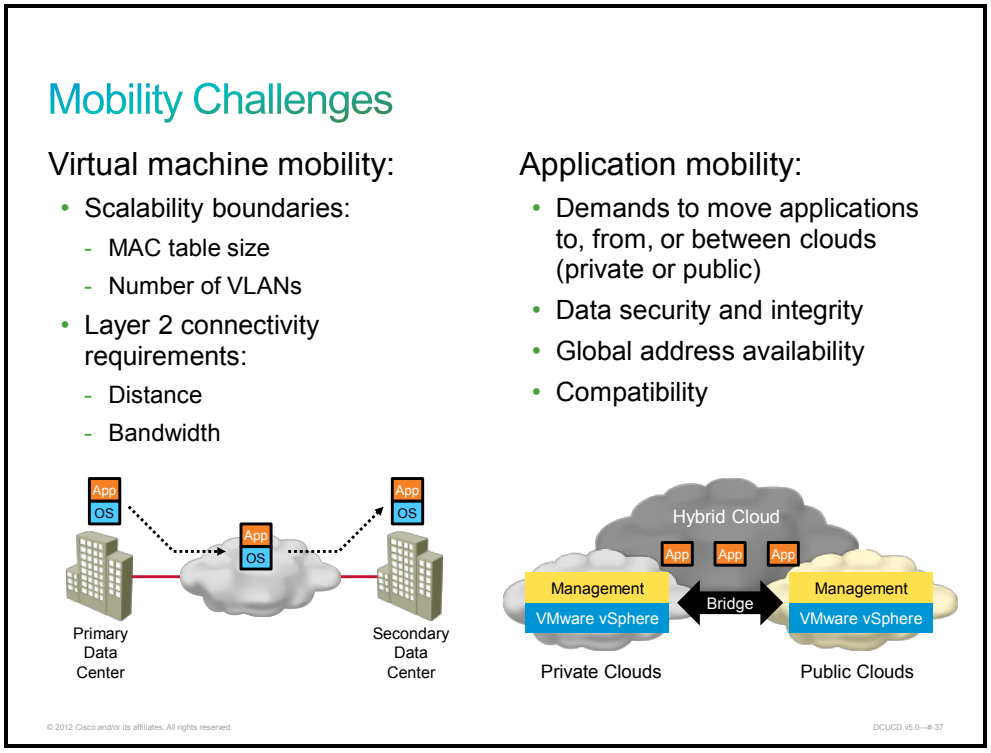
  Achieving the property of availability requires either building very reliable components (high MTBF) or designing components and systems that can rapidly recover from failure (low MTTR). As downtime approaches zero, availability approaches 100 percent.

- **Fault-tolerance:** Fault-tolerant systems are systems that have redundant hardware components and can operate in the presence of individual component failure. Several components of the system have built-in component redundancy. Clusters are also examples of a fault-tolerant system. Clusters can provide uninterrupted service despite node failure. If a node that is running on one or more applications fails, one or more nodes in the cluster takes over the applications from the failed server. A fault-tolerant server has a fully replicated hardware design that allows uninterrupted service in the event of component failure. The recovery time or performance loss that is caused by a component failure is close to zero, and information and disk content are preserved. The problem with fault-tolerant systems is that the system itself is a single point of failure.

- **Disaster recovery:** Disaster recovery is the ability to recover a data center at a different site if a disaster destroys the primary site or otherwise makes the primary site inoperable. Disaster, in the context of online applications, is an extended period of outage of mission-critical service or data that is caused by events such as fire or attacks that damage the entire facility. A disaster recovery solution requires a remote, mirrored (backup and secondary data center) site where business and mission-critical applications can be started within a reasonable period of time after the destruction of the primary site.

Setting up a new, offsite facility with duplicate hardware, software, and real-time data synchronization enables organizations to quickly recover from a disaster at the primary site. The data center infrastructure must deliver the desired Recovery Point Objective (RPO) and Recovery Time Objective (RTO). RTO determines how long it takes for a certain application to recover, and RPO determines to which point (in backup and data) the application can recover. These objectives also outline the requirements for disaster recovery and business continuity. If these requirements are not met in a deterministic way, an enterprise carries significant risk in terms of its ability to deliver on the desired service level agreements (SLAs). SLAs are fundamental to business continuity. Ultimately, SLAs define your minimum levels of data center availability and often determine what actions will be taken in the event of a serious disruption. SLAs record and prescribe the levels of service availability, serviceability, performance support, and other attributes of the service, such as billing and even penalties, in the case of violation of the SLAs. For example, SLAs can prescribe different expectations in terms of guaranteed application response time (such as 1, 0.5, or 0.1 second), guaranteed application resource allocation time (such as 1 hour or automatic), and guaranteed data center availability (such as 99.999, 99.99, or 99.9 percent). Higher levels of guaranteed availability imply higher SLA charges.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Currently, mobility is of utmost importance—everyone demands and requires it. IT infrastructure users and businesses demand to be able to access their applications and data from anywhere, imposing new challenges. At the same time, IT needs to cut costs of the infrastructure, thus more users and businesses are moving to the cloud.

Virtual Machine Mobility

VM mobility requires that data center architectures are correctly created. From the VM perspective, these considerations must be taken into account in the solution architecture:

- MAC address tables and VLAN address space present a challenge when VMs need to move outside of their own environments (for example, when moving a VM from a primary to a secondary data center, or from a private to a public IT infrastructure).
- To ensure correct VM operation, and also the operation of the application hosted by the VM, Layer 2 connectivity between the segments to which the VM is moved is commonly required, which introduces challenges such as the following:
  - Distance limitations
  - Selection of an overlay technology that enables seamless Layer 2 connectivity
  - Unwanted traffic carried between sites (broadcast, unknown unicast, and so on) that consumes bandwidth
  - Extending IP subnets and split-brain problems upon data center interconnect failure

https://t.me/learningnets

Application Mobility

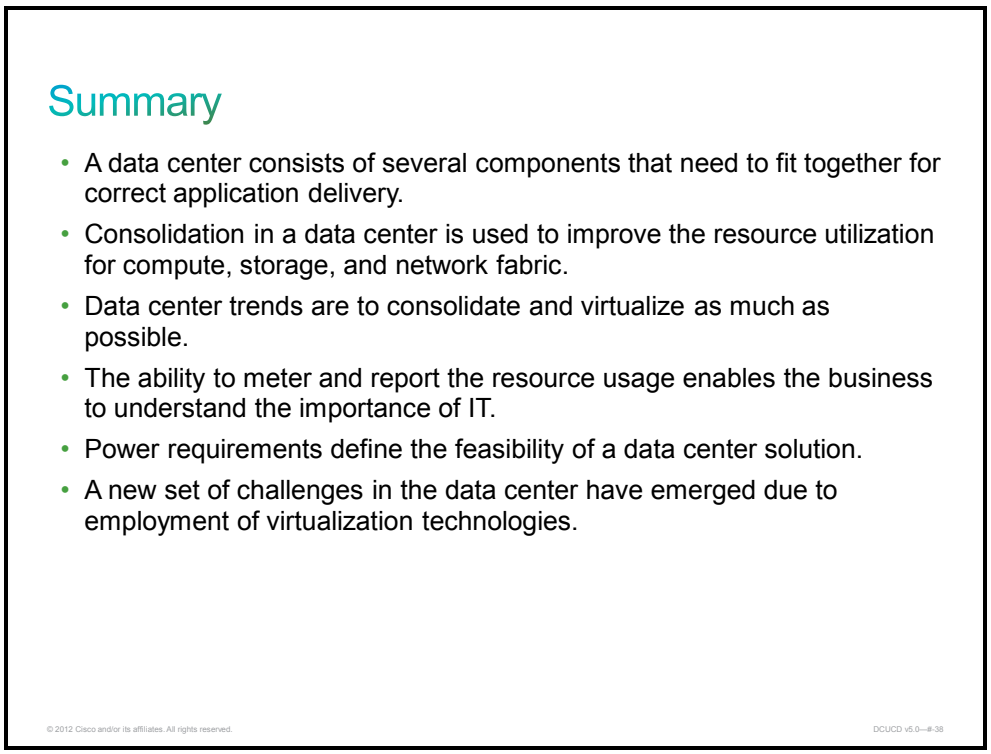
Application mobility means that users must be able to access applications from any device, but, from the IT perspective, it also includes the ability to move application load between IT infrastructures (that is, clouds). This imposes another set of challenges:

- Data security and integrity when moving the application load from its own, controlled IT infrastructure to an outsourced infrastructure (that is, a public cloud)
- Similarly to VM mobility, access to application (that is, enabling application access by the same name, regardless of its location)
- Because IT infrastructures typically do not use standardized underlying architecture, equipment, and so on, there are compatibility issues between such infrastructures, which can limit or affect application mobility by requiring the use of conversion tools. These problems can diminish seamless application mobility and limit the types of applications that can be moved (for example, critical applications do not allow downtime due to the conversion process).

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- http://www.cisco.com/en/US/netsol/ns340/ns394/ns224/index.html
- http://en.wikipedia.org/wiki/Data_center
- http://en.wikipedia.org/wiki/Virtualization

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Lesson 2

Identifying Data Center Applications

Overview

Data centers exist for applications—applications are run on the resources provided in the data centers. This lesson introduces some of the data center applications, with the emphasis on server and desktop virtualization.

Objectives

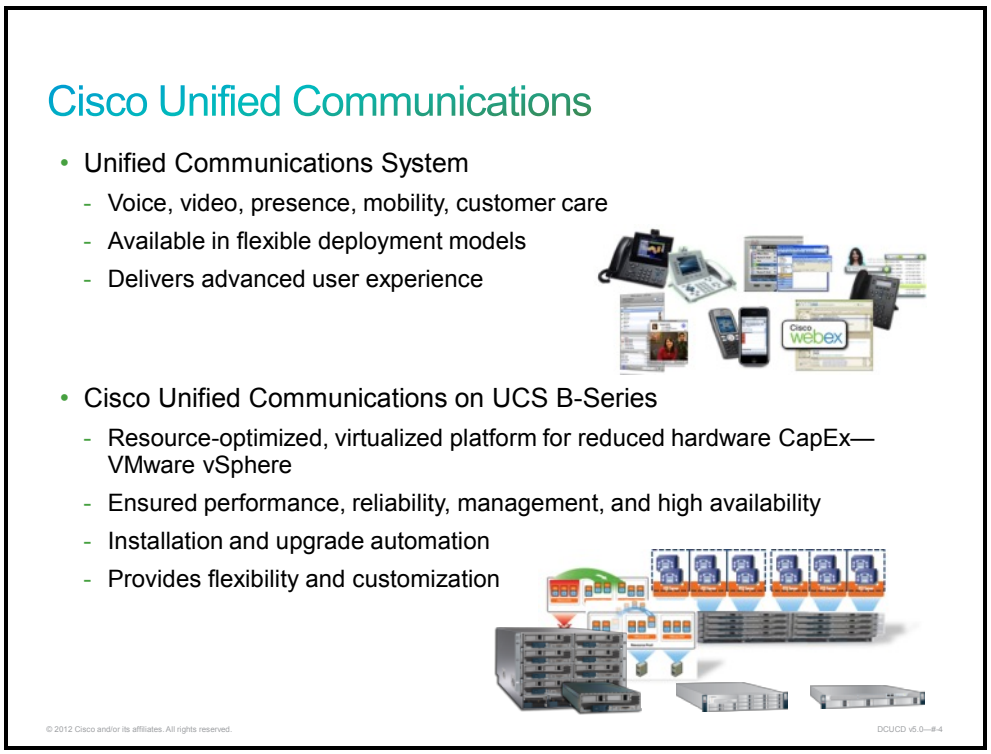
Upon completing this lesson, you will be able to identify the contemporary data center applications. This ability includes being able to meet these objectives:

- Describe common data center applications
- Describe server virtualization characteristics
- Describe desktop virtualization characteristics

https://t.me/learningnets

Common Data Center Applications

This topic describes common data center applications.



Cisco Unified Communications

Cisco Unified Communications enables you to do the following:

- Connect coworkers, partners, vendors, and customers with needed information and expertise
- Access and share video on the desktop, on the road, and on demand, as easily as making a phone call
- Facilitate better team interactions, dynamically bringing together individuals, virtual workgroups, and teams
- Make mobile devices extensions of the corporate network so that mobile workers can be productive anywhere
- Innovate across the value chain by integrating collaboration and communications into applications and business processes

Cisco Unified Communications uses the network as a platform for collaboration. Applications can be flexibly deployed onsite, on-demand, and in blended deployment models. Intercompany and intracompany collaboration is facilitated using with a wide array of market-leading solutions:

- **Conferencing:** Providing compelling, productive, and cost-effective conferencing experiences with high-quality voice, video, and web conferencing.
- **Customer care:** Getting closer to customers, while increasing satisfaction and loyalty. Proactively connecting people with information, expertise, and support.

https://t.me/learningnets

- **IP communications:** Enhancing workforce productivity by extending consistent communications services to employees in all workspaces, using a full suite of IP communications solutions and endpoints.
- **Messaging:** Communicating effectively and with a high level of security, within and between companies. Viewing real-time presence information and communicating using email, instant messaging, and voice mail.
- **Mobile applications:** Increasing mobile employee productivity and responsiveness to customers while controlling mobile costs by making mobile devices extensions of the enterprise network.

The Cisco Unified Communications features and solutions include the following:

- Cisco Unified Communications Manager
- Cisco Unified Contact Center Express
- Cisco Unified Presence
- Cisco Unity
- Cisco Unity Connection

Cisco Unified Communications on Cisco UCS

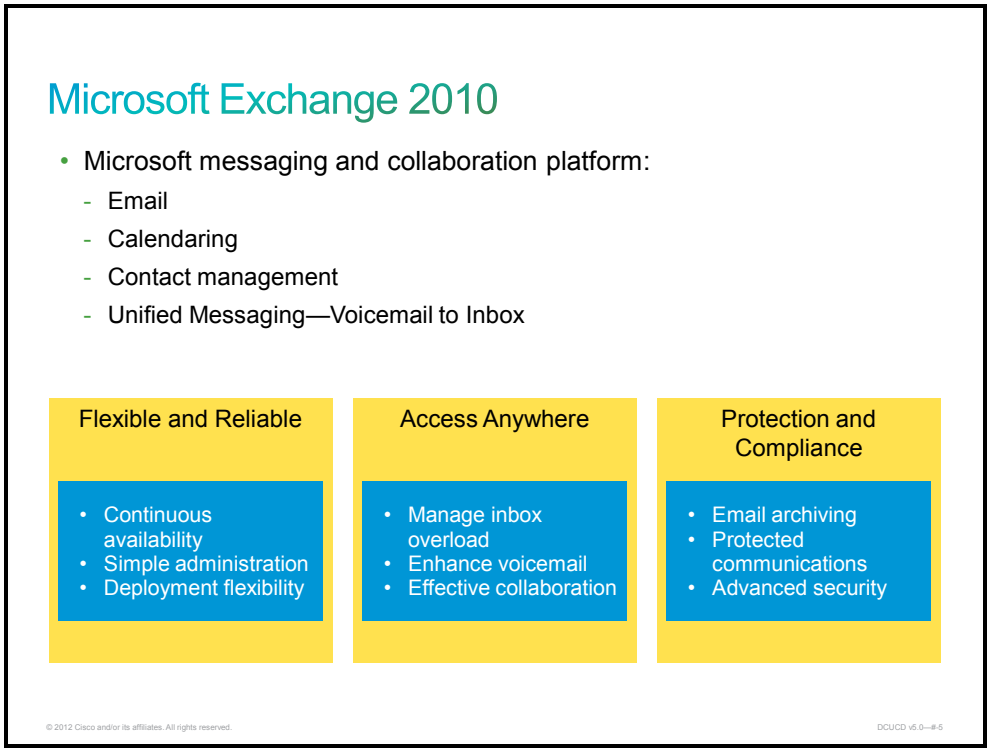
The Cisco Unified Communications on Cisco Unified Computing System (UCS) enables customers to deploy the applications in a virtualized environment.

Customers can deploy Cisco Unified Communications applications as follows:

- On supported Cisco Unified Computing System setups
- VMware vSphere ESXi server virtualization platform

This enables customers to ensure performance, reliability, management, and high availability for the Cisco Unified Communications.

https://t.me/learningnets

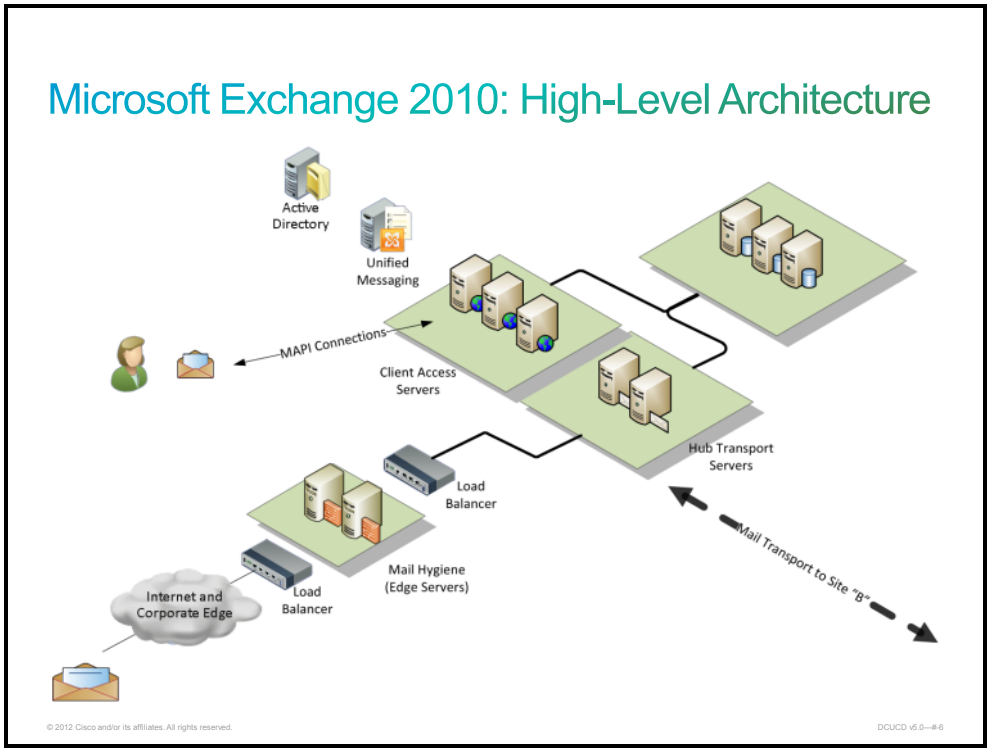


Microsoft Exchange 2010

Microsoft Exchange 2010 is a Microsoft messaging and collaboration platform that provides email, calendar, contacts, phone, and web, and keeps them connected and in sync. The platform includes these key features:

- Protection against spam attacks and phishing threats with multilayered, antispam filtering and continuous updates
- Support for large and reliable mailboxes with a variety of storage
- Support for work from anywhere, and with support for the various web browsers and devices

https://t.me/learningnets



Microsoft Exchange deployments come in different combinations and sizes.

Microsoft Exchange Server Roles

Exchange 2010 deployment encompasses the following roles:

- Hub Transport server
- Client Access server
- Mailbox server
- Edge Transport server
- Unified Messaging server

Hub Transport Server Role

For those familiar with earlier versions of Exchange Server 2007, the Hub Transport server role replaces what was formerly known as the bridgehead server in Exchange Server 2003. The function of the Hub Transport server is to intelligently route messages within an Exchange Server 2010 environment. By default, SMTP transport is inefficient at routing messages to multiple recipients because it takes a message and sends multiple copies throughout an organization. For example, if a message with a 5-MB attachment is sent to 10 recipients in an SMTP network, typically at the sendmail routing server, the 10 recipients are identified from the directory and 10 individual 5-MB messages are transmitted from the sendmail server to the mail recipients, even if all of the recipient mailboxes reside on a single server.

The Hub Transport server takes a message destined to multiple recipients, identifies the most efficient route to send the message, and keeps the message intact for multiple recipients to the most appropriate endpoint. Hence, if all of the recipients are on a single server in a remote location, only one copy of the 5-MB message is transmitted to the remote server. At that server, the message is then separated, with a copy of the message dropped into each of the recipient mailboxes at the endpoint.

https://t.me/learningnets

The Hub Transport server in Exchange Server 2010 does more than just intelligent bridgehead routing. It also acts as the policy compliance management server. Policies can be configured in Exchange Server 2010 so that, after a message is filtered for spam attacks and viruses, the message goes to the policy server so that it can be determined whether the message meets or fits into any regulated message policy, and appropriate actions are taken. The same is true for outbound messages—the messages go to the policy server, the content of the message is analyzed, and if the message is determined to meet specific message policy criteria, the message can be routed unchanged, or the message can be held or modified based on the policy. For example, an organization might want any communications referencing a specific product code name or a message that has content that appears to be private health information to be held, or encryption to be enforced on the message before it continues its route.

Client Access Server Role

The Client Access server role in Exchange Server 2010 (also in Exchange Server 2007) performs many of the tasks that were formerly performed by the Exchange Server 2003 front-end server, such as providing a connecting point for client systems. A client system can be an Office Outlook client, a Windows Mobile handheld device, a connecting point for Outlook Web Access (OWA), or a remote laptop user using Outlook Anywhere to perform an encrypted synchronization of their mailbox content.

Unlike a front-end server in Exchange Server 2003, which effectively just passed user communications to the back-end Mailbox server, the Client Access server does intelligent assessment of where a user mailbox resides and then provides the appropriate access and connectivity. Exchange Server 2010 now has replicated mailbox technology, where a user mailbox can be active on a different server in the event of a primary mailbox server failure. By allowing the Client Access server to redirect the user to the appropriate destination, Exchange Server provides more flexibility in providing redundancy and recoverability of mailbox access in the event of a system failure.

Mailbox Server Role

The Mailbox server is merely a server that holds user mailbox information. It is the server that has the Exchange Server databases. However, rather than just being a database server, the Exchange Server 2010 Mailbox server role can be configured to perform several functions that keep the mailbox data online and replicated. For organizations that want to create high availability for Exchange Server data, the Mailbox server role systems would likely be clustered, and not just a local cluster with a shared drive (and, thus, a single point of failure on the data), but rather one that uses the new Exchange Server 2010 Database Availability Groups. The Database Availability Group allows the Exchange Server to replicate data transactions between Mailbox servers within a single-site data center or across several data centers at multiple sites. In the event of a primary Mailbox server failure, the secondary data source can be activated on a redundant server with a second copy of the data intact. Downtime and loss of data can be minimized or eliminated, with the ability to replicate mailbox data on a real-time basis.

Microsoft eliminated single-copy clusters, local continuous replication, clustered continuous replication, and standby continuous replication in Exchange 2010 and substituted in their place Database Availability Group (DAG) replication technology. The DAG is effectively clustered continuous replication, but instead of a single active and single passive copy of the database, DAG provides up to 16 copies of the database and provides a staging failover of data from primary to replica copies of the mail. DAGs still use log shipping as the method of replication of information between servers. Log shipping means that the 1-MB log files that note the information written to an Exchange Server are transferred to other servers, and the logs are replayed on that server to build up the content of the replica system from data known to be accurate. If, during a replication cycle, a log file does not completely transfer to the remote

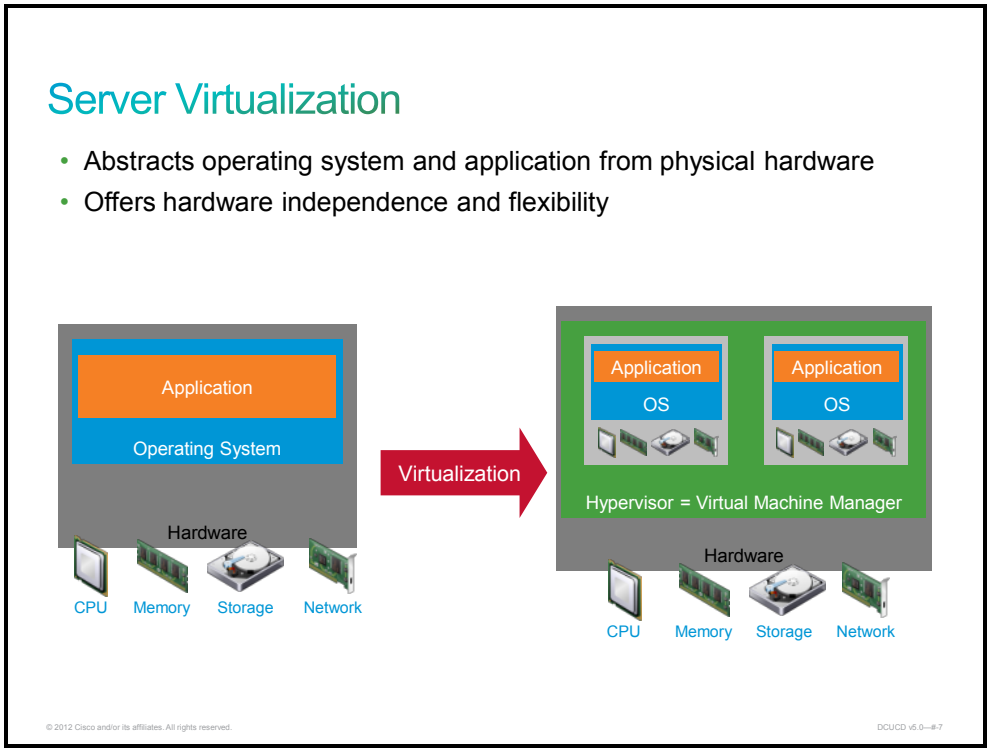
© 2012 Cisco Systems, Inc.
https://t.me/learningnets

system, individual log transactions are backed out of the replicated system and the information is re-sent.

Unlike bit-level transfers of data between source and destination, which are used in SANs and most other Exchange Server database replication solutions, if a system fails, bits do not transfer and Exchange Server has no idea what the bits were, what to request for a resend of data, or how to notify an administrator what file or content the bits referenced. The Microsoft implementation of log shipping provides organizations with a clean method of knowing what was replicated and what was not. In addition, log shipping is done with small 1-MB log files to reduce bandwidth consumption of Exchange Server 2010 replication traffic. Other uses of the DAG include staging the replication of data so that a third or fourth copy of the replica resides offline in a remote data center. Instead of having the data center actively be a failover destination, the remote location can be used to simply be the point where data is backed up to tape or a location where data can be recovered if a catastrophic enterprise environment failure occurs.

A major architecture change with Exchange Server 2010 is how Outlook clients connect to Exchange Server. In previous versions of Exchange Server, even Exchange Server 2007, Remote Procedure Call (RPC)/HTTP and RPC/HTTPS clients would initially connect to the Exchange Server front end or Client Access server to reach the Mailbox servers, while internal Messaging Application Programming Interface (MAPI) clients would connect directly to their Mailbox server. With Exchange Server 2010, all communications (initial connection and ongoing MAPI communications) go through the Client Access server, regardless of whether the user was internal or external. Therefore, architecturally, the Client Access server in Exchange Server 2010 needs to be close to the Mailbox server, and a high-speed connection should exist between the servers for optimum performance.

https://t.me/learningnets



Historically, physical servers are deployed with one application and one operating system within a single set of hardware. A single operating system is isolated to one machine, running, for example, the Windows or Linux operating system. This means that the physical server is linked with the underlying hardware, which makes migration or replacement a process that requires time and skill.

If additional applications are put on a physical server, these multiple applications start competing for resources, which typically causes problems related to performance or insufficient resources—challenges that are difficult to address and manage. Thus, a single application might run on a single server, resulting in server resource underutilization—with average utilization ranging from 5 to 10 percent.

When a new application must be deployed, such as a web service, a physical server must be deployed, racked, stacked, connected to external resources, and configured—all of which requires a substantial amount of time.

Because numerous applications are used, some of them demanding high availability as well, a data center ends up with numerous server deployments. In many cases, this causes various problems ranging from insufficient space to excessive power requirements.

Server Virtualization

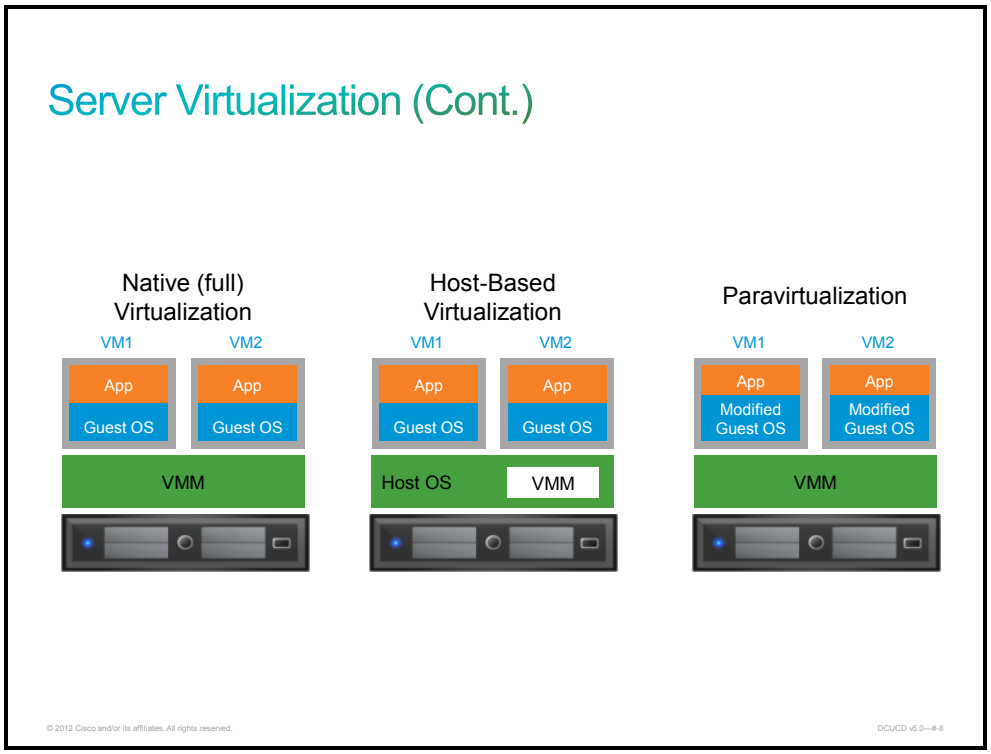
Server virtualization decouples the server from the physical hardware—this makes the server independent of the underlying physical server. The hardware is literally abstracted or separated from the operating system.

The operating system and the applications are contained in a container—a virtual machine. A single physical server with server virtualization software deployed (for example, VMware ESX) can run multiple virtual machines. While virtual machines do share the physical resources of the underlying physical server, with virtualization, tools exist to control how the resources are allocated to individual virtual machines.

The virtual machines on a physical server are isolated from each other and do not interact. In other words, they run deployed applications without affecting each other. Virtual machines can

https://t.me/learningnets

be brought online without the need for installing new server hardware, which allows rapid expansion of computing resources to support greater workloads.



The virtual machine (VM) is isolated from the underlying hardware. The physical host is the place where the hypervisor or virtual machine manager (VMM) resides. The most often used approaches in server virtualization (that is, hypervisors) are the following:

- Native or full virtualization
- Host-based virtualization
- Paravirtualization

Native (Full) Virtualization

Native virtualization has the following characteristics:

- The hypervisor runs on bare metal—that is, directly on the physical server hardware without the need for a host operating system.
- The hypervisor completely virtualizes hardware from the guest operating systems. Drivers used to access the hardware exist in the hypervisor.
- The guest operating system deployed in a VM is unmodified.

Such an approach enables almost any guest operating system deployment and allows the best scalability. The most widely used example of native virtualization is the VMware ESX hypervisor.

Host-Based Virtualization

Host-based virtualization has the following characteristics:

- The VMM runs in a host operating system that is not directly located on the physical server hardware.

https://t.me/learningnets

- Drivers used to access physical hardware are based on the host operating system kernel, whereas the hardware is still emulated by the VMM.
- The guest operating system deployed in a VM is unmodified, but must be supported by the VMM and host operating system.

Examples of host-based virtualization are Microsoft Virtual Server and VMware Server solutions. Such solutions typically have a larger footprint due to host operating system usage and the additional I/O that is used for the host operating system communication.

Microsoft Virtual Server can be deployed with Windows 7, Windows XP, Windows Vista, or Windows 2003 host operating systems, and can host Windows NT, Windows 2000, Windows 2003, and Linux as a guest operating system. The current version is Microsoft Virtual Server 2005 R2 SP1.

Hybrid Virtualization

Microsoft Hyper-V Server is a hybrid native-host virtualization solution, where a hypervisor resides on a bare metal server but requires a parent VM or partition running Windows 2008. The parent partition creates child partitions hosting guest operating systems. The virtualization stack runs on the parent partition, which has direct access to the hardware devices and provides physical hardware access to child partitions. The guest operating systems include Windows 2000, Windows 2003, Windows 2008, SUSE Linux Enterprise Server 10 SP1 or SP2, Windows Vista SP1, Windows 7, and Windows XP Professional SP2, SP3, or x64.

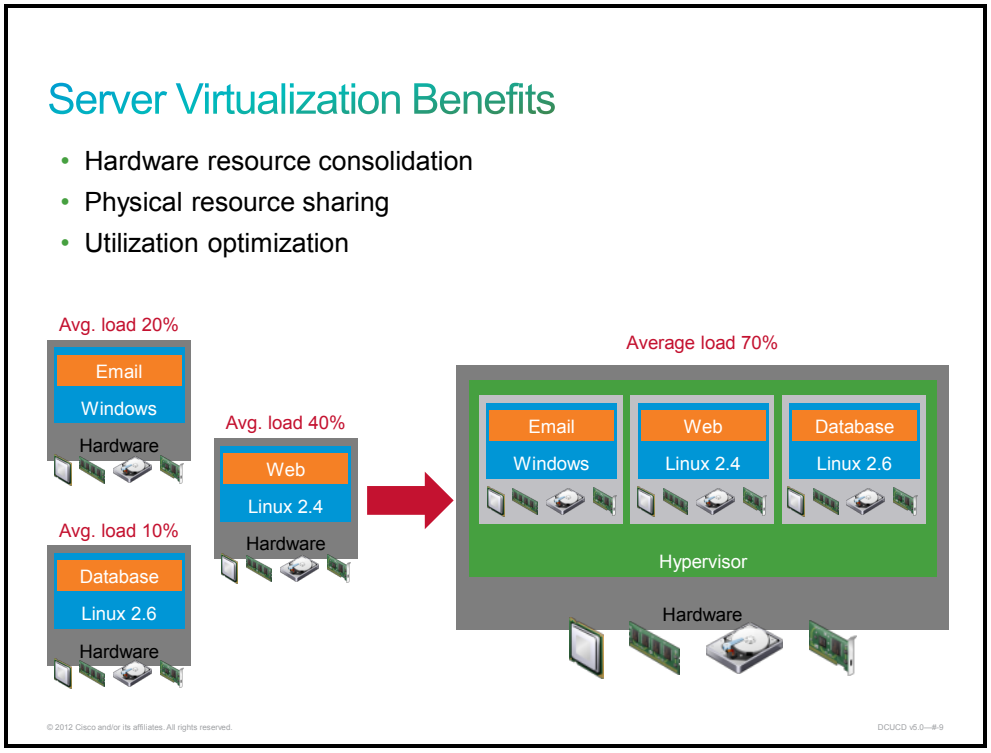
Paravirtualization

Paravirtualization has the following characteristics:

- The hypervisor runs on bare metal—that is, directly on the physical server hardware without the need for a host operating system.
- The guest operating system deployed must be modified to make calls to or receive events from the hypervisor. This typically requires a guest operating system code change.
- The application binary interface used by the application software remains intact, thus the applications do not have to be changed to run inside a VM.
- Unmodified guest operating systems can be supported if virtualization is supported in hardware architecture, for example, Intel VT-x or AMD Pacific.

An example of a paravirtualization solution is Xen, which supports guest operating systems such as Linux, NetBSD, FreeBSD, Solaris 10, and certain Windows versions.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



With physical server deployment, a single operating system is used by each server. The software and hardware are tightly coupled, which makes the solution inflexible. Because multiple applications do not typically run on a single machine due to potential conflicts, the resources are underutilized and the computing infrastructure cost is high.

Benefits

When server virtualization is used, the operating systems and applications are independent of the underlying hardware, allowing a virtual machine to be provisioned to any physical server. The operating system and applications are encapsulated in a virtual machine, so multiple virtual machines can be run on the same physical server. Thus, server virtualization offers significant benefits as compared to physical server deployment:

- Physical hardware can be consolidated.
- Resources of a physical machine can be shared among virtual machines.
- Resource utilization is improved—fewer resources are wasted.

The figure shows application deployment in a physical server environment compared to a virtualized server environment. A physical server configuration uses three servers that have a low average load ranging from 10 to 40 percent. A virtualized solution uses just one server with three virtual machines deployed—so the total physical server average load is the sum of virtual machine average loads, which is around 70 percent—significantly better than with the physical server configuration.

Abstraction of the operating system and application from the underlying physical server hardware is an important benefit of virtualization. The abstraction can be employed in disaster recovery scenarios because it intelligently addresses the traditional requirement of physical server-based disaster recovery—the need to provide identical hardware at the backup data center.

With complete abstraction, any VM can be brought online on any supported physical server without having to consider hardware or software compatibility.

https://t.me/learningnets

Desktop Virtualization

- Replacement or augmentation of the desktop using virtualization:
  - Separate physical endpoint from the logical desktop
  - Host logical desktops in a data center
  - Deliver desktop over the network
  - Best suited to specific use cases
- Virtual desktop infrastructure (VDI) is an architecture that consists of the following:
  - Multiple hardware and software components
  - Multiple vendors



© 2012 Cisco and/or its affiliates. All rights reserved.

Desktop Virtualization Challenges

The traditional desktop virtualization deployments have many challenges:

- Application-related challenges:
  - Managing updates
  - Licensing compliance
  - Security and policy compliance
  - New applications
- Operating system-related:
  - Driver compatibility
  - Integration
  - Patching
  - Upgrading
  - New installs
- Desktop (that is, hardware)-related:
  - Performance
  - Lifecycle management
  - Security
  - Mobility
  - Supportability

https://t.me/learningnets

The key challenges can be summarized in these five categories:

- Hardware costs:
  - Updates may require changes in hardware.
  - Updates can “break” a system because of incompatibility.
- Compliance and data security:
  - Lost devices can contain sensitive or secure data.
  - Compliance must be checked on each device.
- IT productivity:
  - Each device type requires different support models.
  - Different tools are needed to manage the desktop virtualization.
- Growth:
  - Provisioning new desktops can take days.
  - Refresh cycles are needed.
- Resilience:
  - It is difficult to restore lost virtualization desktops.
  - Backing up data is critical.

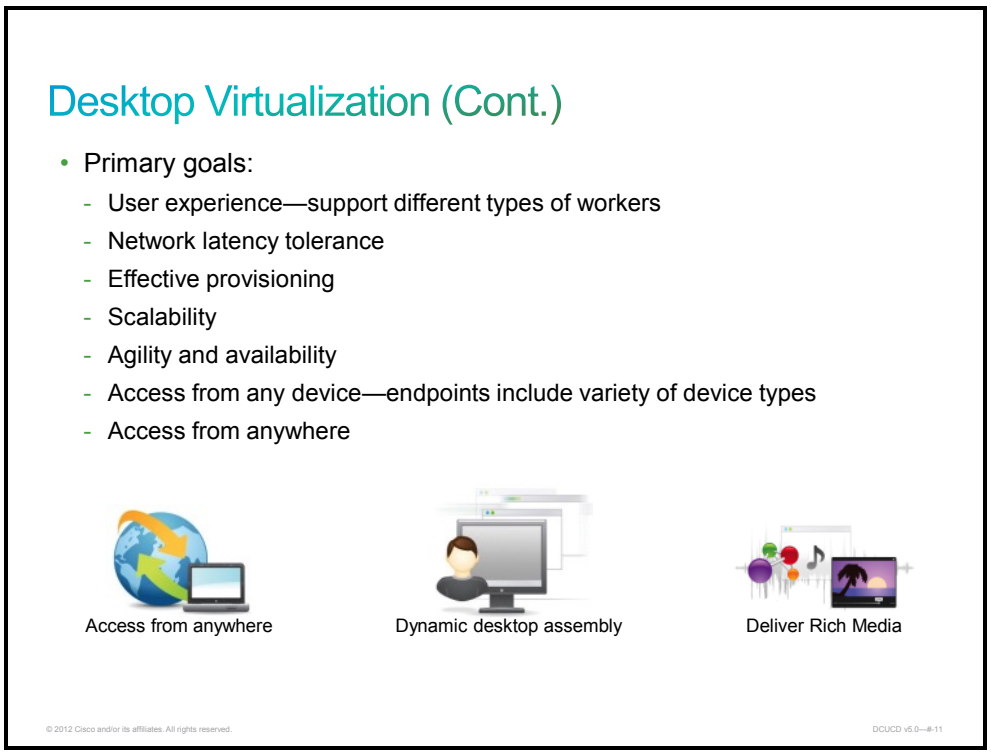
Desktop Virtualization

Desktop virtualization, as a concept, separates a PC desktop environment from a physical machine using a client-server model of computing. The model stores the resulting “virtualized” desktop on a remote central server instead of on the local storage of a remote client. Thus, when users work from their remote desktop client, all of the programs, applications, processes, and data used are kept and run centrally. This allows users to access their desktops on any capable device, such as a traditional PC, notebook computer, smart phone, or thin client.

Desktop virtualization involves encapsulating and delivering access to a remote client device in order to access the entire information system environment. The client device may use a different hardware architecture than that used by the projected desktop environment, and may also be based upon an entirely different operating system.

The desktop virtualization model allows the use of VMs to let multiple network subscribers maintain individualized desktops on a single, centrally located computer or server. The central machine may operate at a residence, business, or data center. Users may be geographically scattered, but all may be connected to the central machine by a LAN, a WAN, or the public Internet.

https://t.me/learningnets



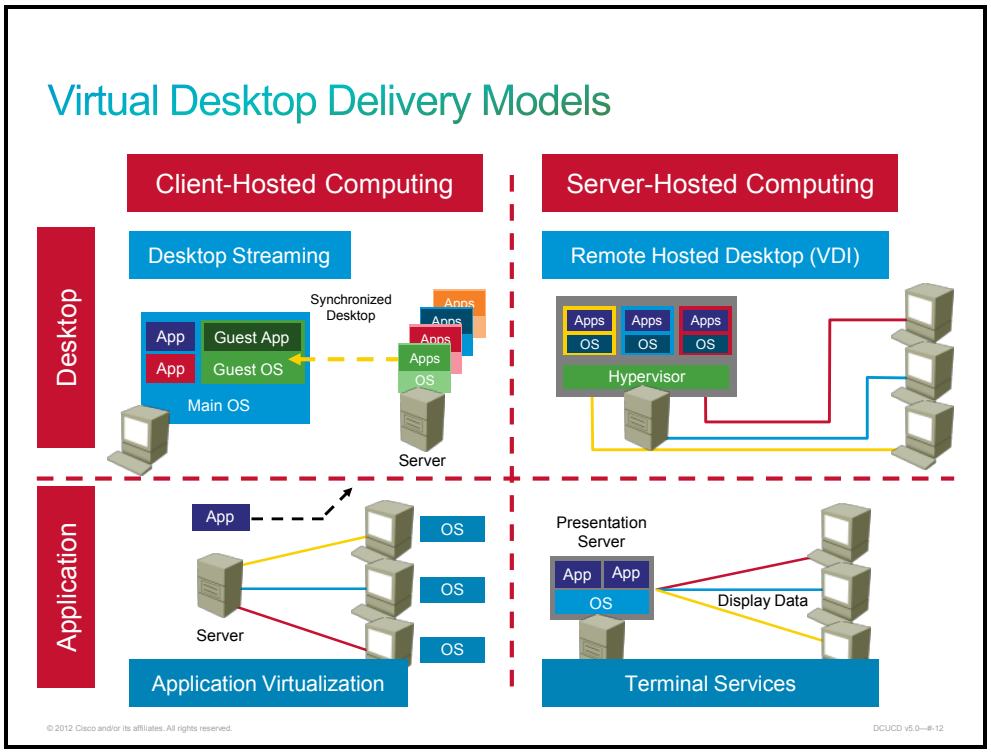
Virtual desktop infrastructure (VDI) is the practice of hosting a desktop operating system within a VM running on a centralized server or servers.

The following major reasons and goals make VDI an appealing solution:

- **User experience:**
  - Instant on with fast boot
  - Mobility: Desktop available through any device
  - Functionality: Fully functioning, personalized desktop
  - Support for peripherals: USB, network, printing, scanning, and so on
  - Support: Performance, service-level compliance, quick resolution of problems
- **Tolerance to network latency:** VDI can be used in all contemporary networks today.
- **Effective provisioning:** With VDI, it is easier to provision desktops, because all are located in the data center.
- **Scalability:** Similar to server virtualization, the VDI is easily scaled—an administrator just needs to add hardware resources.
- **Agility and availability:**
  - From the user perspective:
    - The same environment, regardless of device
    - Allows the same usage of peripherals, regardless of device
    - Allows the same personalization, regardless of device
    - Allows access to the same desktop, regardless of location
  - From the IT perspective:
    - Should be able to use any hypervisor

https://t.me/learningnets

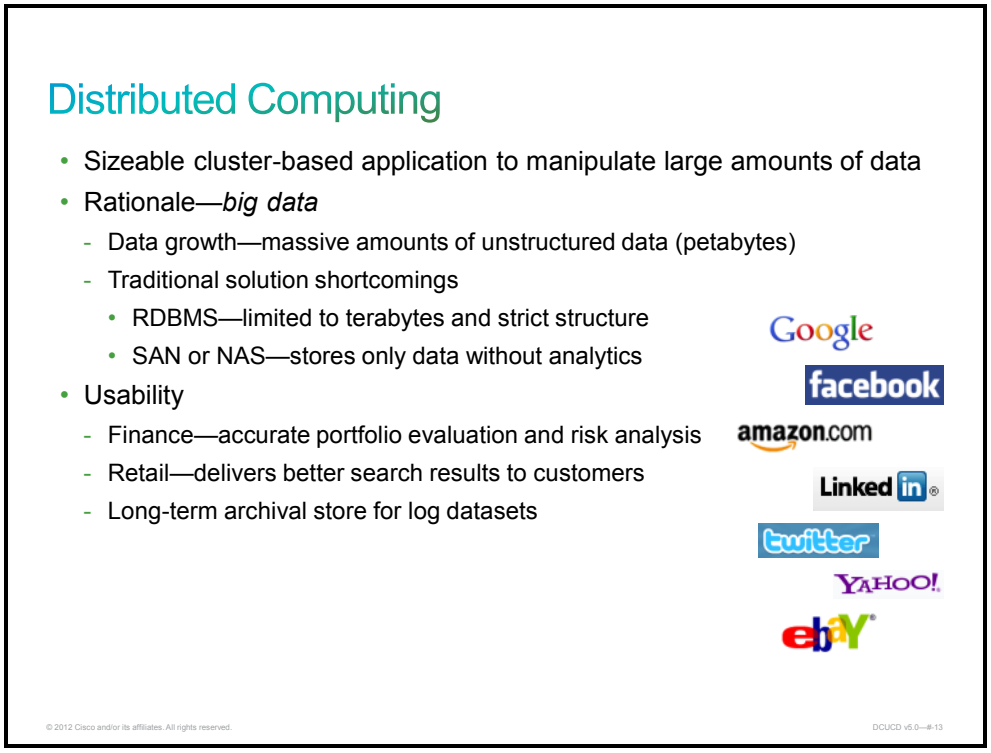
- Multiple types of virtual storage
- Supports every major operating system
- Can deliver the same applications as a physical desktop



The desktop virtualization solutions can be divided into two main categories, based on the means of delivering the virtual desktop:

- **Desktop-oriented solutions:** The user desktop can be used in these ways:
  - **Desktop streaming:** The desktop is streamed to the user client device, allowing the desktops to be centrally managed via synchronization, but still requiring the computing capacity on the end-user side (for example, Citrix XenDesktop with XenClient, VMware View with offline mode). The users can work even if the connectivity to the data center is not available (if the applications allow it).
  - **Remote hosted virtual desktop or VDI:** The desktop is hosted in the data center servers and the end user merely uses a terminal access to the desktop. For this solution to work, it is important to have connectivity between user client devices and the data center.
- **Application-oriented solutions:** The applications rather than the desktops are used in these ways:
  - **Application virtualization:** Applications are streamed similarly to the desktop streaming solution.
  - **Terminal services:** Applications are hosted on central servers and users use remote access to the central server to access the applications.

https://t.me/learningnets



Big data is a foundational element of social networking and Web 2.0-based information companies. The enormous amount of data is generated as a result of democratization and ecosystem factors such as the following:

- **Mobility trends:** Mobile devices, mobile events and sharing, and sensory integration
- **Data access and consumption:** Internet, interconnected systems, social networking, and convergent interfaces and access models (Internet, search and social networking, and messaging)
- **Ecosystem capabilities:** Major changes in the information processing model and the availability of an open source framework for general-purpose computing and unified network integration

Data generation, consumption, and analytics have provided competitive business advantages for Web 2.0 portals and Internet-centric firms that offer services to customers and service differentiation through correlation of adjacent data.

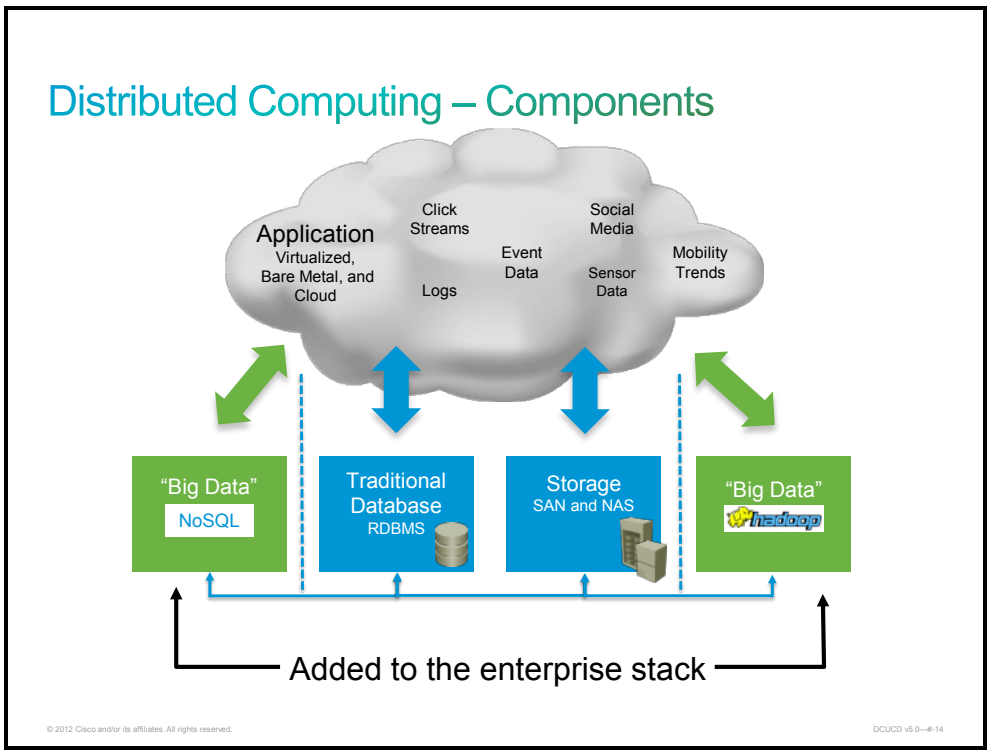
With the rise of business intelligence data mining, analytics, market research, behavioral modeling, and inference-based decision-making, data can be used to provide a competitive advantage. Here are a few use cases of big data for companies with a large Internet presence:

- Targeted marketing and advertising
- Related sale promotions
- Analysis of behavioral social patterns
- Metadata-based optimization of workload and performance management

The requirements of traditional enterprise data models for application, database, and storage resources have grown over the years, and the cost and complexity of these models has increased to meet the needs of big data. This rapid change has prompted changes in the fundamental models that describe the way that big data is stored, analyzed, and accessed. The new models are based on a scaled-out, shared-nothing architecture, bringing new challenges to enterprises to decide which technologies to use, where to use them, and how. The traditional

https://t.me/learningnets

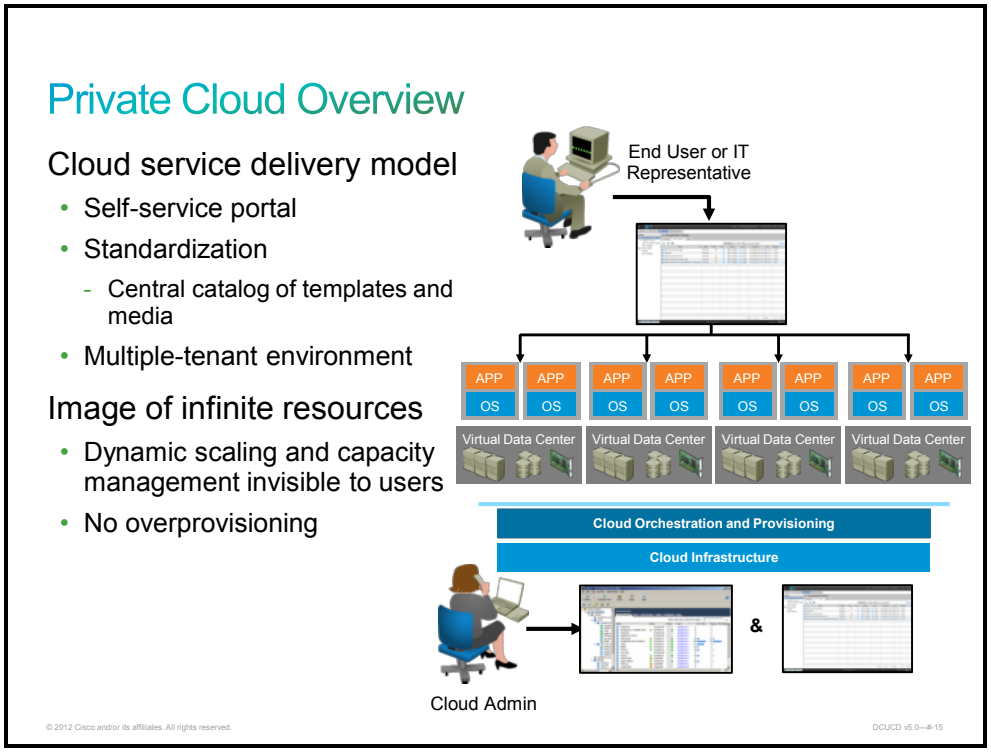
model is now being expanded to incorporate new building blocks that address the challenges of big data with new information processing frameworks purpose-built to meet requirements of the big data. These purpose-built systems must also meet the inherent requirement for integration into current business models, data strategies, and network infrastructures.



Two main building blocks are being added to the enterprise stack to accommodate significant data:

- **Hadoop:** Provides storage capability through a distributed, shared-nothing file system, and analysis capability through MapReduce.
- **NoSQL:** Provides the capability to capture, read, and update, in real time, the large influx of unstructured data and data without schemas. Examples include click streams, social media, log files, event data, mobility trends, and sensor and machine data.

https://t.me/learningnets



Private clouds provide an ideal way to solve some of your biggest business and technology challenges. A private cloud can deliver IT-as-a-Service (ITaaS), which helps reduce costs, reach new levels of efficiency, and introduce innovative new business models. Consequently, an enterprise can become more agile and efficient, while simplifying its operations and infrastructure.

Cisco Private Cloud

To deliver the full benefits of a private cloud, the Cisco cloud infrastructure integrates the Cisco Data Center architecture with intelligent, cloud-ready networks. This powerful combination helps to simplify IT operations, tighten security, increase business agility, improve data center economics, and enable consistent cloud experiences for end users.

Cisco Data Center provides a complete, highly efficient data center infrastructure. It integrates compute, network, security, and management into a fabric architecture that delivers outstanding performance and agility for the business while simplifying operations and lowering costs for IT. Its holistic approach unifies and dynamically optimizes compute, storage, and network resources, allowing them to be securely and rapidly repurposed and managed on demand. Cisco Data Center comprises three elements:

- Unified Computing System
- Unified Fabric
- Unified Management

Cisco Virtualized Multi-Tenant Data Center

The Cisco Virtualized Multi-Tenant Data Center (VMDC) architecture provides an end-to-end architecture and design for a complete private cloud, providing ITaaS capabilities. VMDC consists of several components of a cloud design, from the IT infrastructure building blocks to all the components that complete the solution, including orchestration for automation and configuration management. The building blocks are based on stacks of integrated infrastructure components that can be combined and scaled: Vblock Infrastructure Packages from the Virtual

https://t.me/learningnets

Computing Environment (VCE) coalition, developed in partnership with EMC and VMware and the Secure Multi-Tenancy (SMT) stack developed in partnership with NetApp and VMware. Workload management and infrastructure automation is achieved using BMC Cloud Lifecycle Management (CLM). Clouds built on VMDC can also be interconnected or connected to service provider clouds with Cisco Data Center Interconnect (DCI) technologies. This solution is built on a service delivery framework that can be used to host other services in addition to ITaaS on the same infrastructure, such as, for example, VDI.

VMware vCloud Solution

VMware vCloud is a system enabling seamless deployment of private cloud solutions. It enables a cloud service delivery that works as follows:

- Defines a new data center unit, a virtual data center (VDC)
- Uses a standardized, catalog-based service delivery
- Enables self-service user access with metering, monitoring, and charge-back.

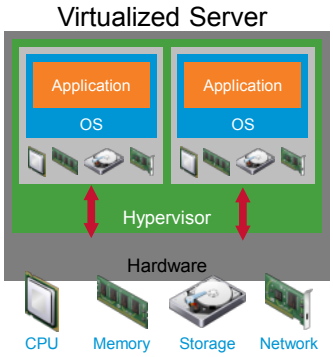
https://t.me/learningnets

Server Virtualization Overview

This topic describes server virtualization.

Hypervisor – Abstraction Layer

- Hypervisor or VMM:
  - Thin operating system between hardware and virtual machine
  - Controls and manages hardware resources
  - Manages virtual machines (creates, destroys, etc.)
  - Runs on host (physical server)
- Virtualizes hardware resources:
  - CPU process time-sharing
  - Memory span from physical memory
  - Network
  - Storage



© 2012 Cisco and/or its affiliates. All rights reserved.

A hypervisor, or virtual machine monitor (VMM), is server virtualization software that allows multiple operating systems to run concurrently on a host computer.

The hypervisor provides abstraction of the physical server hardware for the virtual machine. A thin operating system performs the following basic tasks:

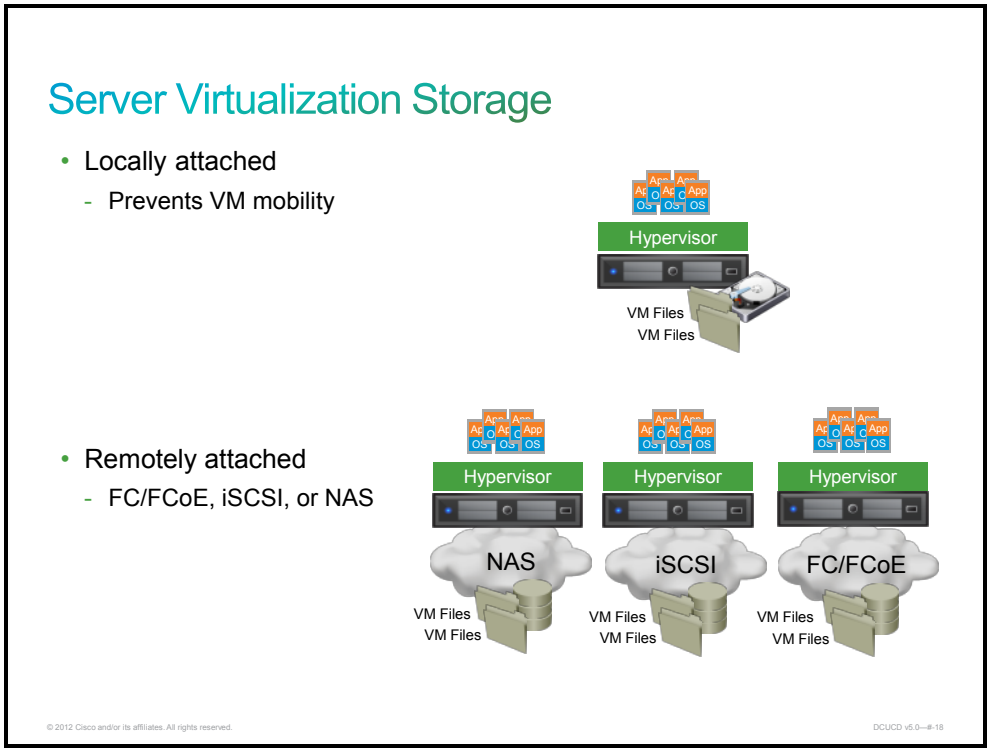
- Control and management of physical resources by assigning them to virtual machines and monitoring resource access and usage
- Control and management of virtual machines—the hypervisor creates and maintains virtual machines and, if requested, destroys the virtual machine (if the VMM is alive)

Ideally, a hypervisor abstracts all physical server components—CPU, memory, network, and storage. CPU abstraction is achieved with CPU time-sharing between virtual machines, and memory abstraction is achieved by assigning memory span from a physical memory.

A virtual server is used to enable a particular service or application, and, from the server perspective, CPU, memory, I/O, and storage resources are important.

Note When multiple virtual machines are deployed, they can oversubscribe resources. The hypervisor, therefore, must employ an intelligent mechanism to allow oversubscription without incurring performance penalties.

https://t.me/learningnets



Traditionally, storage logical unit numbers (LUNs) are presented to the hypervisor and then formatted as volumes. Each volume can contain one or more VMs, which are stored as files on the volume:

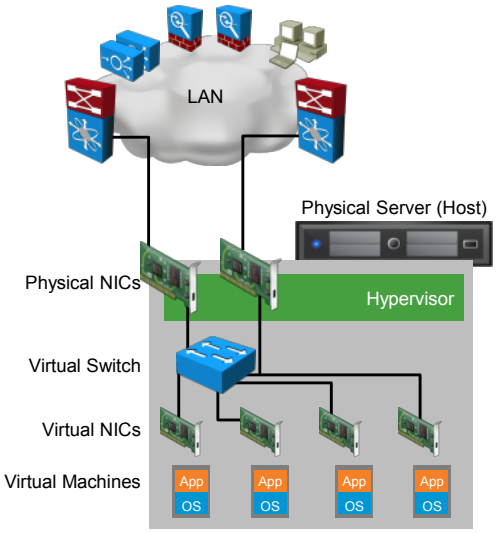
- LUNs are masked and zoned to the hypervisor, not the VM.
- LUNs are formatted by the hypervisor with the correct clustered file system.
- VMs (operating system and data) are stored as files on volumes.

Virtual disks can be presented to the VMs as Small Computer Systems Interface (SCSI) LUNs using a virtual SCSI hardware adapter.

https://t.me/learningnets

Network Virtual Access Layer

- Connects host and VMs to the network
  - Extends network into servers
- Virtual switch
  - Uplink ports: physical NICs
  - VM-facing ports: virtual NICs



© 2012 Cisco and/or its affiliates. All rights reserved.

The server virtualization solution extends the access layer into the host server with the VM networking layer. The following components are used to implement server virtualization networking:

- **Physical network:** Physical devices connecting hosts for resource sharing. Physical Ethernet switches are used to manage traffic between hosts, the same as in a regular LAN environment.
- **Virtual networks:** Virtual devices running on the same system for resource sharing.
- **Virtual Ethernet switch:** Similar to a physical switch. It maintains table of connected devices, which is used for frame forwarding. It can be connected via uplink to a physical switch via a physical network interface card (NIC). It does not provide the advanced features of a physical switch.
- **Port group:** Subset of ports on a virtual switch for VM connectivity.
- **Physical NIC:** Physical network interface card used to uplink host to the external network.

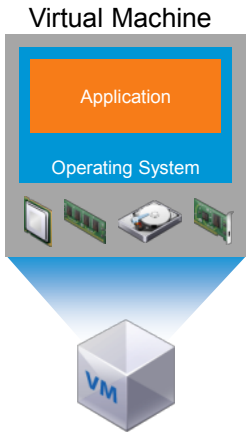
As multiple VMs are created on each physical server, virtual networks are also constructed to support the I/O needs of the VMs. These networks sit outside the boundary of standard networking controls and best practices.

Virtual networking is deployed on each host server and extends the access layer into configured physical servers—a virtual access layer. The virtual access layer does not have the same functionality as a physical access layer, typically lacking access control list (ACL) and QoS configuration options.

https://t.me/learningnets

Virtual Machine

- Virtual machine (VM) contains an operating system and application:
  - Operating system = guest operating system
  - Guest operating system does not have full control over hardware
  - Applications are isolated from each other
- VMs contain the following:
  - vMAC address
  - vIP address
  - Memory, CPU, storage space



© 2012 Cisco and/or its affiliates. All rights reserved.

A virtualized server is called a virtual machine (VM). A virtual machine is a container holding the operating system and the applications. The operating system in a VM is called the guest operating system.

A VM is defined as a representation of a physical machine by software that has its own set of virtual hardware on which an operating system and applications can be loaded. With virtualization, each virtual machine is provided with consistent virtual hardware, regardless of the underlying physical hardware that the host server runs on. A virtualized server has the same characteristics as a physical machine:

- CPU
- Memory
- Network adapters
- Disks

All the virtual server resources are virtualized. Each VM also has its own set of parameters—for example, a virtual MAC address and virtual IP address—to allow it to communicate with the external world. Therefore, a single physical server will typically have multiple MAC addresses and IP addresses—those defined and used by the VMs that it serves.

Because a VM uses virtualized resources, the guest operating system is no longer in control of hardware—this is the privilege of the hypervisor. Underlying physical machine resources are shared between different virtual machines, each running its own operating system instance.

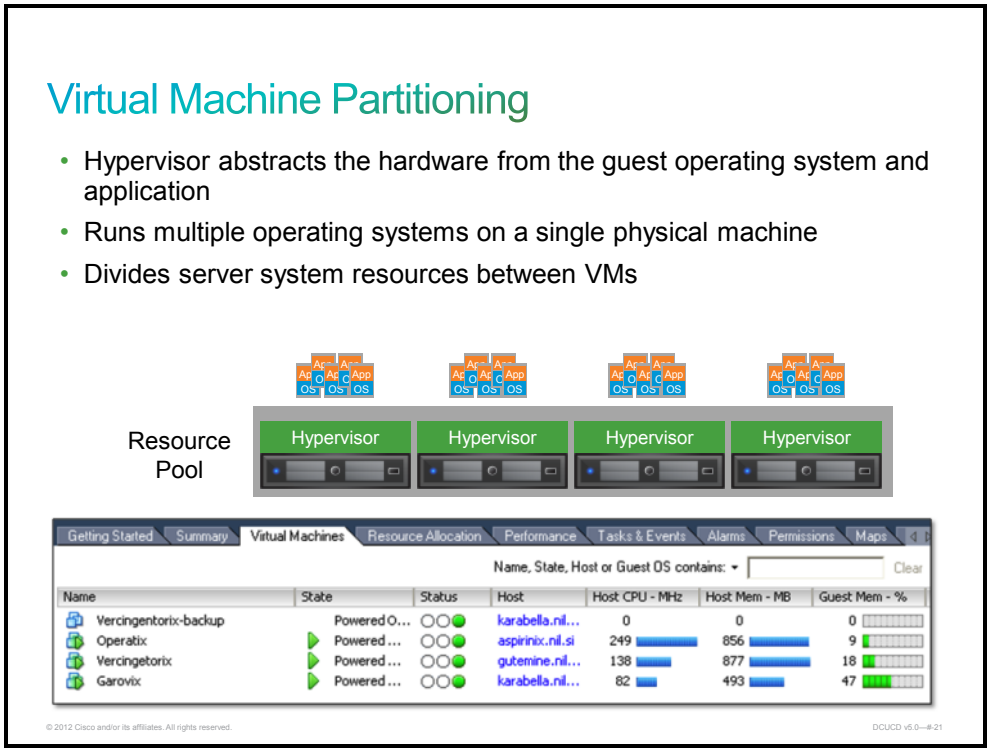
The VM resources are defined by the server administrator, which creates a VM—defines the characteristics of a VM—the CPU speed, amount of memory, storage space, network connectivity, and so on.

https://t.me/learningnets

VM Benefits

Using a VM provides four significant benefits:

- **Hardware partitioning:** Multiple virtual machines run on the same physical server at the same time.
- **VM isolation:** A VM running on the same physical server cannot affect the stability of the other VM.
- **VM encapsulation:** A VM is kept in a couple of files, which eases VM mobility.
- **Hardware abstraction:** The VM is not tied to a physical machine and can be moved according to business or administrative demand. The load can be dynamically balanced among the physical machines.



Partitioning means that a physical server (host) runs two or more operating systems with different applications installed. The VM operating system is called the guest operating system.

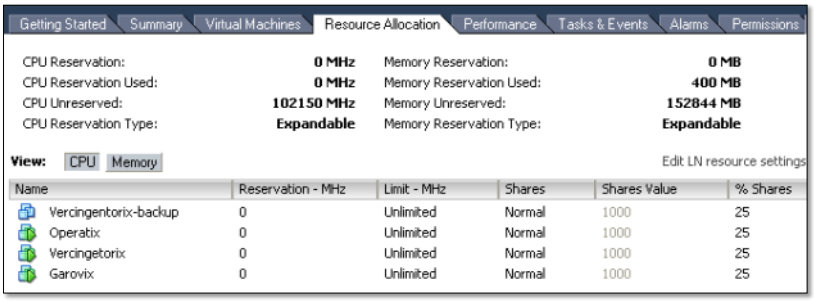
The guest operating systems might be different—hypervisors typically support different operating systems, including Windows, Linux, Solaris, NetWare, or any other vendor-specific system. None of the guest operating systems have any knowledge of others running on top of the hypervisor on the same physical host. They share the physical resources of the physical server.

The control and abstraction of the hardware and physical resources is done by the hypervisor—a thin operating system that provides the hardware abstraction.

https://t.me/learningnets

Virtual Machine Isolation

- Hardware-level fault and security isolation:
  - VMs are not aware of presence of other VMs.
  - VM failure does not affect other VMs on the same host.
- Advanced server resource control:
  - To preserve and control performance



© 2012 Cisco and/or its affiliates. All rights reserved.

A second key VM characteristic is isolation. Isolation means that VMs do not know about other VMs that might be running on the same host. They have no knowledge of any other VM.

The implication of isolation is, of course, security. Not knowing about each other, the VMs do not interfere with data from the others. Isolation also prevents any specific VM failure from affecting any other VM operation.

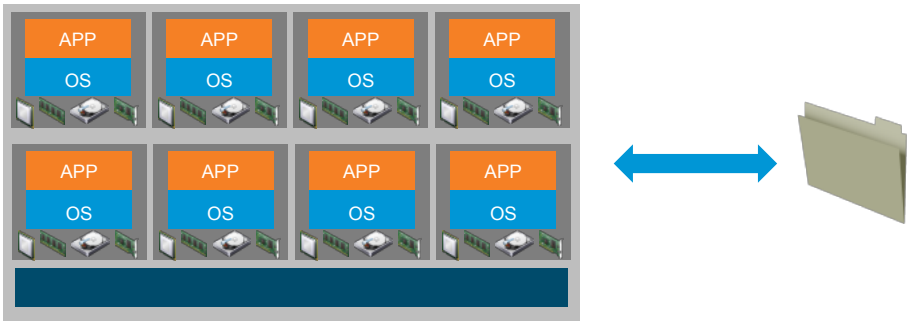
VMs on the same or different physical servers can communicate if network configuration permits it.

To ensure proper performance for a VM, the hypervisor allows advanced resource control, where certain resources can be reserved per VM, such as when the hypervisor allocates and dedicates memory.

https://t.me/learningnets

Virtual Machine Encapsulation

- Each VM has a state that can be saved to a set of known files.
- VMs can be moved or copied (cloned):
  - Simple move or copy file operation



© 2012 Cisco and/or its affiliates. All rights reserved.

A third key VM characteristic is encapsulation—a VM is a collection of files on a host operating system (the ESX storage space), which saves the VM state that contains the following information:

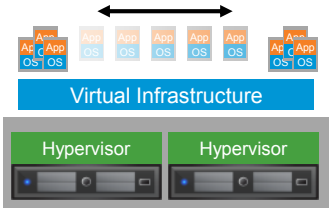
- The guest operating system and applications installed
- The VM parameters, including memory size, number of CPUs, and so on

An encapsulated VM can easily be moved or copied for backup or cloning purposes—this is just a simple move or copy operation on the host ESX system. A VM is independent of the underlying physical server, so it can be moved to and started on a different ESX server.

https://t.me/learningnets

Virtual Machine Hardware Abstraction

- Any VM can be provisioned or migrated to any other host with similar physical characteristics.
- Support for multiple operating systems:
  - Windows, Linux, Solaris, NetWare



Resource Pool



© 2012 Cisco and/or its affiliates. All rights reserved.

The fourth key characteristic is hardware abstraction. As already mentioned, this is performed by the ESX hypervisor to provide VM hardware independency.

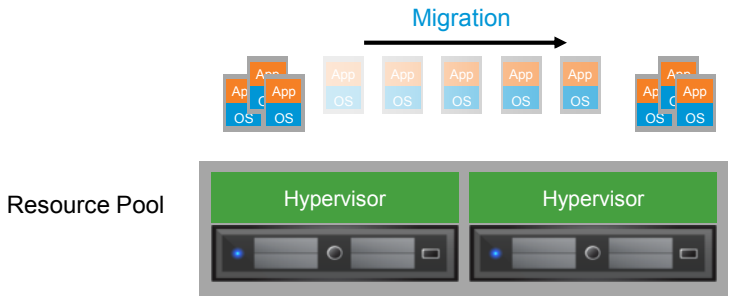
Being hardware-independent, the VM can be migrated to another ESX server to use the physical resources of that server. Mobility also provides scalable, on-demand server provisioning, server resource pool growth, and failed server replacement.

With advanced VMware mechanisms such as Dynamic Resource Scheduler (DRS), the VM can be moved to a less-used physical server, thus dynamic load balancing is provided.

https://t.me/learningnets

Virtual Machine Mobility

- Migrating live VMs:
  - Moves VMs across physical servers without interruption
  - Changes hardware resources dynamically
  - Eliminates downtime and provides continuous service
  - Balances workloads for computing resource optimization



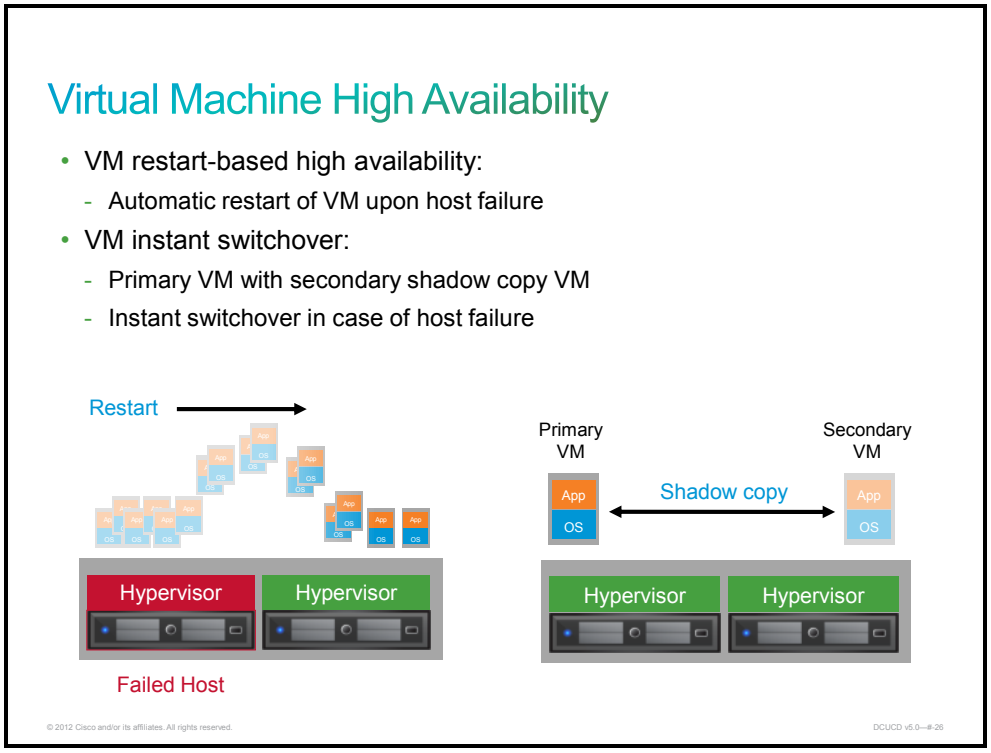
© 2012 Cisco and/or its affiliates. All rights reserved.

VM mobility is achieved with migration of live VMs (for example, VMware VMotion, Microsoft Hyper-V LiveMigration), which allows the moving of VMs across physical hosts with no interruption. During such a migration, the transactional integrity is preserved, and the VM resource requirements are dynamically shifted to the new host.

VM mobility can be used to eliminate downtime normally associated with hardware maintenance. It can also be employed to optimize server utilization by balancing virtual machine workloads across available host resources. VM mobility enables server administrators to transparently move running VMs from one physical server to another physical server across the Layer 2 network.

For example, a Cisco UCS blade needs additional memory. VM mobility could be used to migrate all running VMs off the blade, allowing the blade to be removed so that memory could be added without impact to VM applications.

https://t.me/learningnets



The VM high availability can be divided into two different mechanisms:

- VM restart-based high availability
- VM instant switchover

Restart High Availability

The restart high availability enables automatic failover of VMs upon host failure. The high availability automatic failover restarts the VM that was running on a failed host on another host that is part of the high availability cluster.

Because the VM is restarted and thus the operating system has to boot, high availability does not provide automatic service or application failover in the sense of maintaining client sessions. Upon failure, a short period of downtime occurs. The exact amount of downtime depends on the time needed to boot the VMs. Because the failover is achieved with VM restart, there is potential that some data might be lost due to host failure.

High availability typically requires the following:

- Dedicated network segment with assigned IP subnet
- Hosts configured in a high availability cluster
- Operational Domain Name System (DNS) server
- Hosts must have access to the same shared storage
- Hosts must have identical networking configuration

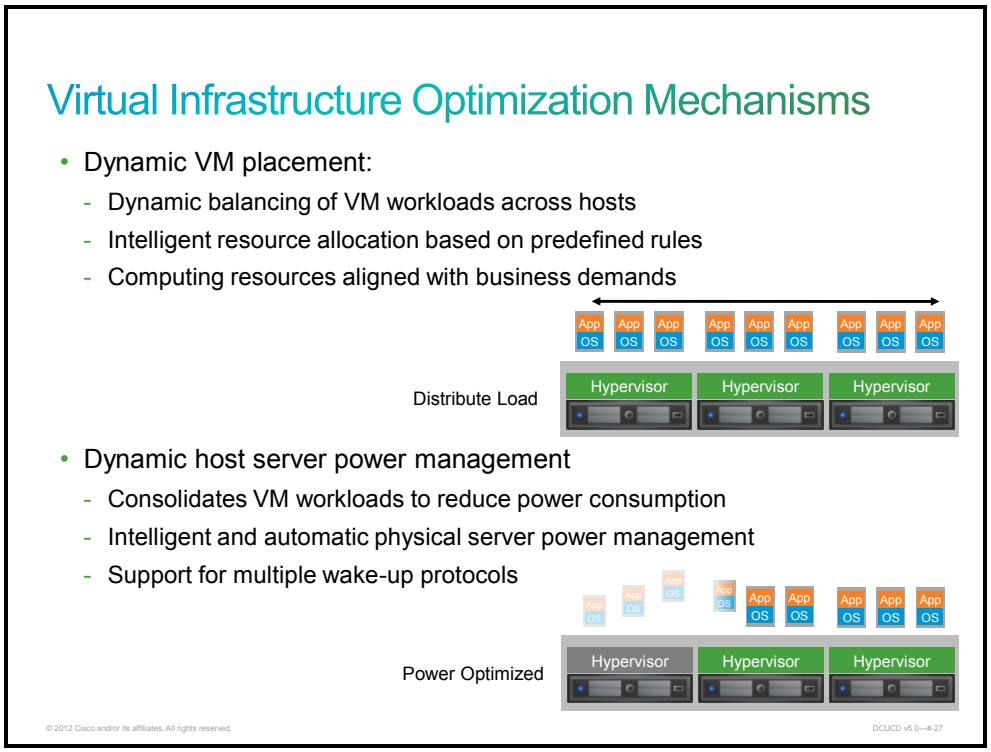
When designing high availability, it is important to observe whether the remaining hosts in a high-availability cluster will be overcommitted upon member failure.

https://t.me/learningnets

Instant Switchover

Instant switchover advances the restart high-availability functionality by enabling true zero downtime switchover time. This is achieved by running primary and secondary VMs, where the secondary is an exact copy of the primary VM. The secondary VM runs as a shadow copy and ends up in the same state as the primary VM. The difference between the two VMs is that the primary VM owns network connectivity.

Upon failure, the switchover to the secondary VM preserves the live client session because the VM is not restarted. Instant switchover is typically enabled per-VM.



Dynamic Virtual Machine Placement

A useful and interesting application is dynamic VM placement, which allows dynamic and intelligent allocation of hardware resources to ensure optimal alignment between business demands and computing resources.

When deployed, it is used to dynamically balance VMs across computing resource pools. The resource allocation decisions are made based on predefined rules, which may be defined by the administrator.

Using dynamic VM placement increases administrator productivity by automatically maintaining optimal workload balancing and avoiding situations where an individual host could become overcommitted.

Dynamic Host Power Management

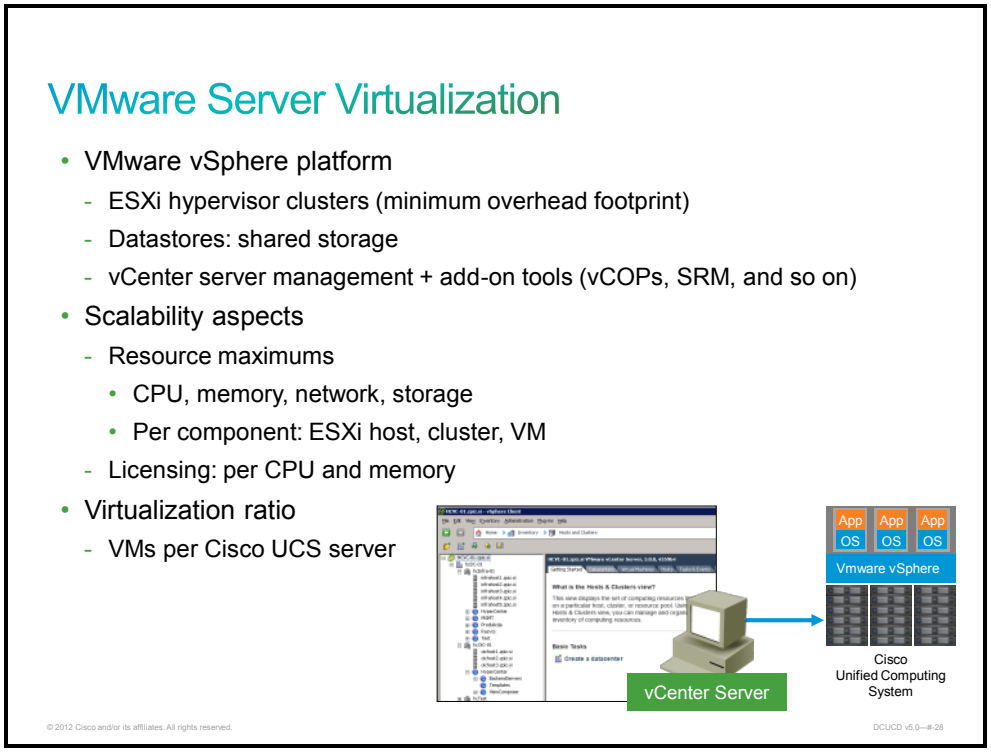
Dynamic host power management can be used to reduce the power and cooling expenses related to physical servers. It consolidates the virtual machines on a minimum number of physical servers, or hosts, by constantly monitoring resource requirements and power consumption across hosts in a cluster.

When fewer resources are required, the virtual machines are consolidated on a couple of hosts, and those that are unused are put in standby mode.

https://t.me/learningnets

If the resource utilization increases and workload requirements increase, dynamic power management brings the standby host servers back online, and then redistributes the VMs across the newly available resources.

Dynamic power management typically requires a supported power management protocol on the host, such as Intelligent Platform Management Interface (IPMI), Integrated Lights Out (ILO), and Wake on LAN (WOL).



VMware vSphere

VMware vSphere is the VMware server virtualization solution. VMware vSphere manages large collections of infrastructure—CPUs, storage, networking—as a seamless, flexible, and dynamic operating environment.

VMware vSphere comprises the following:

- Management and automation with infrastructure optimization, business continuity, desktop management, and software lifecycle tools
- Virtual infrastructure with resource management, availability, mobility, and security tools
- ESX, ESXi, virtual symmetric multiprocessing (SMP), and virtual machine file system (VMFS) virtualization platforms

https://t.me/learningnets

Among key vSphere solution tools and applications are the following:

- ESX and ESXi hypervisor
- Dynamic Resource Scheduler (DRS)
- VMFS native ESX cluster
- Distributed switch that can be native VMware or Cisco Nexus 1000V
- VMotion, Storage VMotion, high availability, fault tolerance, and data recovery availability tools
- Virtual SMP, enabling VMs to use multiple physical processors (that is, upon creation, the administrator can assign multiple virtual CPUs to a VM.)

vSphere Datastore

VMware vSphere storage virtualization allows VMs to access underlying physical storage as though it were Just a Bunch of Disks (JBOD) SCSI within the VM, regardless of the physical storage topology or protocol. In other words, a VM accesses physical storage by issuing read and write commands to what appears to be a local SCSI controller with a locally-attached SCSI drive. Either an LSILogic or BusLogic SCSI controller driver is loaded in the VM so that the guest operating system can access storage exactly as if this were a physical environment.

VMFS

The vast majority of (unclustered) VMs use encapsulated disk files stored on a VMFS volume.

VMFS is a high-performance file system that stores large, monolithic virtual disk files and is tuned for this task alone.

To understand why VMFS is used requires an understanding of VM disk files. Perhaps the closest analogy to a VM disk file is an ISO image of a CD-ROM disk, which is a single, large file containing a file system with many individual files. Through the virtualization layer, the storage blocks within this single, large file are presented to the VM as a SCSI disk drive, made possible by the file and block translations. To the VM, this file is a hard disk, with physical geometry, files, and a file system. To the storage controller, the file is a range of blocks.

Raw Device Mapping

Raw device mapping can allow a VM to access a LUN in much the same way as a nonvirtualized machine. In this scenario, where LUNs are created on a per-machine basis, the strategy of tuning a LUN for the specific application within a VM may be more appropriate.

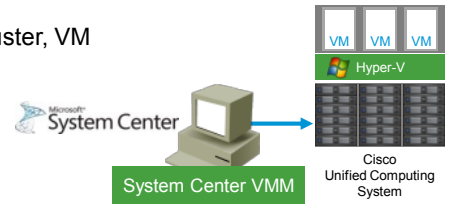
Because raw device mappings do not encapsulate the VM disk as a file within the VMFS file system, LUN access more closely resembles the native application access for which the LUN is tuned.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Microsoft Hyper-V R2 Server Virtualization

- Hyper-V = Windows 2008 R2 server role
  - Windows 2008 R2 root partition + Hyper-V hypervisor
    - Failover clustering
    - Live migration
    - Cluster Shared Volumes (CSVs)
  - System Center Virtual Machine Manager (SCVMM)
- Scalability aspects
  - Resource maximums
    - CPU, memory, network, storage
    - Per component: Hyper-V host, cluster, VM
  - Licensing: per host or CPU
- Virtualization ratio
  - VMs per Cisco UCS server



© 2012 Cisco and/or its affiliates. All rights reserved.

Hyper-V is built on the architecture of Windows Server 2008 Hyper-V and enables integration with new technologies.

Hyper-V can provide these benefits:

- Increased server consolidation
- Dynamic data center
- Virtualized centralized desktop

Hyper-V is part of Windows Server 2008 hypervisor-based architecture and is available as a new virtualization role that can be configured as a full role with local user interface or as a Windows Server Core role.

Hyper-V as a role in Windows Server 2008 provides you with the tools and services to create a virtualized server computing environment.

Hyper-V has specific requirements. Hyper-V requires an x64-based processor, hardware-assisted virtualization, and hardware data execution prevention (DEP). Hyper-V is available in x64-based versions of Windows Server 2008—specifically, the x64-based versions of Windows Server 2008 Standard, Windows Server 2008 Enterprise, and Windows Server 2008 data center.

Hyper-V delivers high levels of availability for production workloads via flexible and dynamic management while reducing overall costs through efficient server consolidation:

- Better flexibility:
  - Live migration
  - Cluster Shared Volumes
  - Hot add or remove of storage
  - Processor compatibility mode for live migration

https://t.me/learningnets

- Improved performance:
  - Improved memory management
  - TCP offload support
  - Virtual machine queue (VMQ) support
  - Improved networking
- Greater scalability:
  - x64 logical processor support
  - Enhanced green IT with core parking

Supported Guest Operating Systems

The operating systems that Microsoft supports as guest sessions under Windows 2008 Hyper-V are as follows:

- Windows Server 2008 x86 and x64
- Windows Server 2003 SP2 or higher x86 and x64
- Windows 2000 Server with SP4 and Windows 2000 Advanced Server with SP4
- Windows Vista x86 and x64
- Windows XP SP2 or later x86 and x64
- SUSE Linux Enterprise Server 10 SP1 or later x86 and x64

Failover Cluster

Failover clusters typically protect against hardware failure. Overall system failures (system unavailability) are not usually the result of server failures, but are more commonly caused by power outages, network stoppages, security issues, or misconfiguration. A redundant server will not generally protect against an unplanned outage such as lightning striking a power substation, a backhoe cutting a data link, an administrator inadvertently deleting a machine or service account, or the misapplication of a zoning update in a Fibre Channel fabric.

A failover cluster is a group of similar computers (referred to as nodes) working in a coordinated way to increase the availability of specific services or applications. You typically employ failover clusters to increase availability by protecting against the loss of a single physical server from an unanticipated hardware failure or through proactive maintenance.

Cluster Shared Volumes

Cluster Shared Volumes (CSVs) is a feature of failover clustering available in Windows Server 2008 R2 for use with the Hyper-V role. A CSV is a standard cluster disk containing an NTFS volume that is made accessible for read and write operations by all nodes within the cluster. This gives the VM complete mobility throughout the cluster, because any node can be an owner and changing owners is easy.

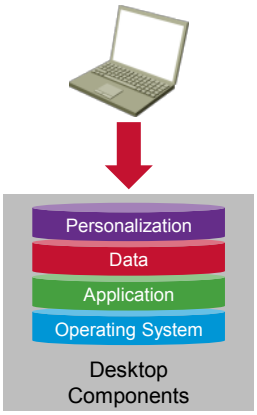
© 2012 Cisco Systems, Inc
https://t.me/learningnets

Desktop Virtualization Overview

This topic describes desktop virtualization.

Desktop Virtualization Infrastructure

- Virtual desktop (desktop VM)
- Desktop virtualization endpoints
  - Thin, thick, and zero clients
- Display protocol
  - Delivers desktop over the network
- Virtual server infrastructure
  - Running desktop and other VMs
- VDI management
- Connection broker
  - Provides connection point to desktop VM
- Infrastructure services
  - Active Directory, DNS, DHCP
- Application virtualization
  - Delivers virtualized applications for desktops, decouples application from operating system



© 2012 Cisco and/or its affiliates. All rights reserved.

Desktop Virtualization Components

The desktop virtualization solution comprises various components.

VDI Platform

The VDI platform hosts both infrastructure VMs and desktop VMs. Physical hosts can be clustered to take advantage of high availability and other features. The important part of the platform is the hypervisor. You must consider the following when you choose a hypervisor:

- Management
- High availability features
- Operating system support
- VM requirements

Services Infrastructure

A VDI solution requires certain services in order to operate. Each VDI requires a services infrastructure. Among these services are the following:

- Application servers (farms)
- Management servers
- Communication grooming
- Dynamic provisioning server
- Application profiler
- Domain controller

https://t.me/learningnets

- DNS
- RDP licensing server
- DHCP

Data and Profile Management

Clients can access a different desktop VM each time that they connect. To maintain personality and data, the profiles are used. Typically, profiles are stored on a network location, and storage of data can be directed to it and the profiles stored on the network are applied as needed.

Access Protocols

The access protocols are used to access the virtualized desktops. They should support various clients, including fat and thin clients that are directly involved in the end-user experience.

The access protocols impact user experience—the performance of the desktop—and are thus critical for providing feature-rich content to the virtual desktop.

There are different protocol options that are vendor-specific:

- **Citrix:** Independent Computing Architecture (ICA)
- **VMware:** PC over IP (PCoIP)
- **Microsoft:** Route Discovery Protocol (RDP)

Virtual desktop access can also be offered via web portals for easy access from anywhere.

Connection Broker

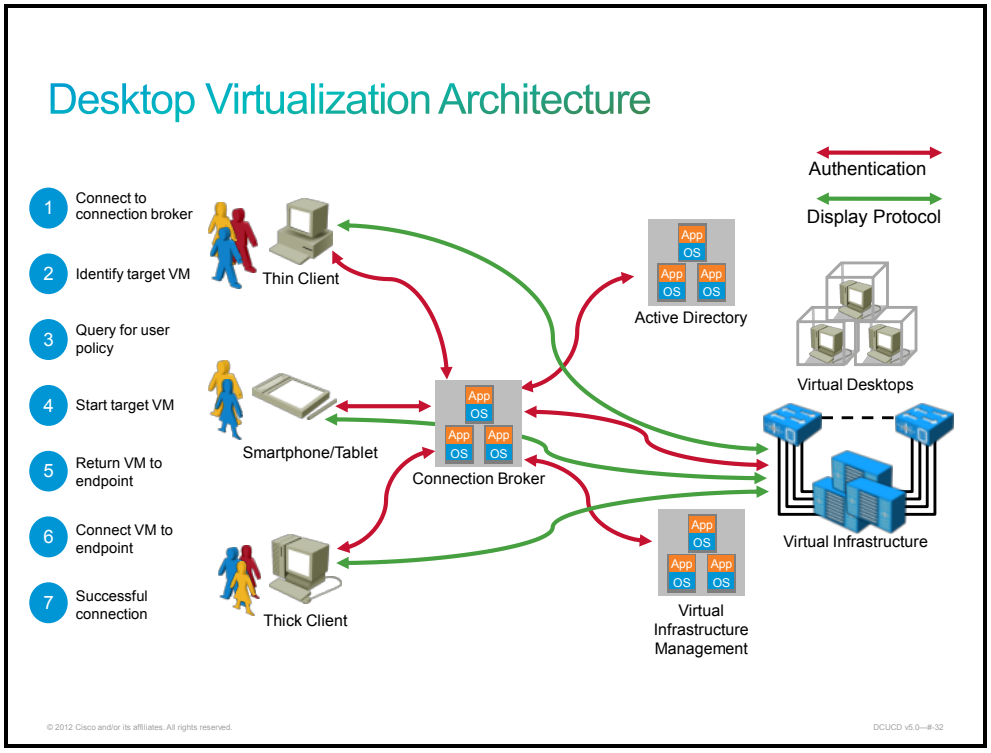
Connection broker is the component of VDI deployment that coordinates the connection to a virtual desktop. It directs users to new VM desktops or redirects clients to previous desktops and is responsible for connection distribution and management.

Desktop Components

Each desktop comprises these distinct components:

- Operating system
- One or more applications
- Data (user and other type)
- Profile (set of desktop settings, or personalization)

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



In a typical desktop virtualization environment, clients access virtual desktops through a connection broker server and are authenticated by an Active Directory server. Virtual desktops are hosted on a virtual infrastructure (for example, VMware vSphere, Microsoft Hyper-V, Citrix XenServer) with network and storage connectivity.

A desktop virtualization solution can use static or dynamic architecture:

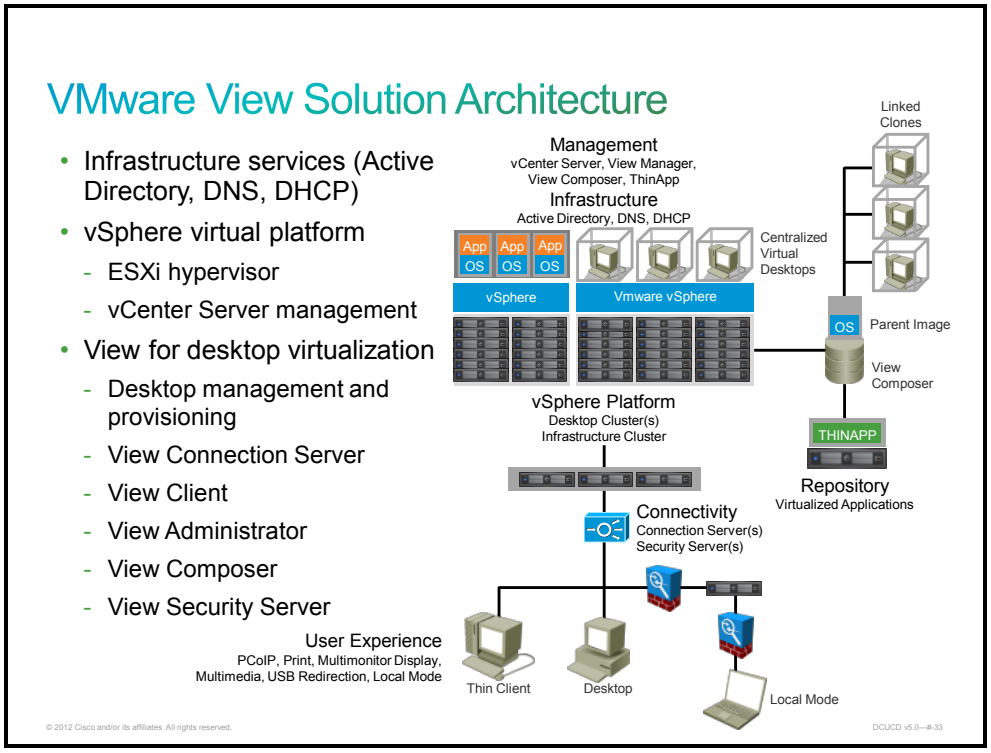
- **Static architecture:** Used where the user is mapped to the same VM upon each connection
- **Dynamic architecture:** Creates the VM each time that a user connects

VDI Advantages

The shared resources model inherent in desktop virtualization offers advantages over the traditional model, in which every computer operates as a completely self-contained unit with its own operating system, peripherals, and application programs. Overall hardware expenses may diminish as users can share resources allocated to them on an as-needed basis. Virtualization potentially improves the data integrity of user information because all data can be maintained and backed-up in the data center. Other potential advantages include the following:

- Simpler provisioning of new desktops
- Reduced downtime in the event of server or client hardware failures
- Lower cost of deploying new applications
- Desktop image-management capabilities
- Longer refresh cycle for client desktop infrastructure
- Secure remote access to an enterprise desktop environment

https://t.me/learningnets



VMware View

VMware View is a commercial desktop virtualization product developed by VMware. It enables you to deliver desktops from the data center as a secure, managed service. Built on VMware vSphere, it is the only platform designed specifically for desktop virtualization.

VMware View supports the RDP and PCoIP protocols, which accelerate the VMware View performance for remote users (such as those communicating over a slow WAN connection).

Apart from the VMware View components, the important components to the solution are vSphere virtual infrastructure and infrastructure services: Active Directory, DNS, and DHCP.

The ESXi hypervisor is the basis for the server virtualization—in this solution, it hosts the virtual desktops and virtual machines to host the server components of VMware View, including Connection Server instances, Active Directory servers, and vCenter Server instances.

A vCenter Server acts as a central administrator for VMware ESX servers that are connected on a network. It provides the central point for configuring, provisioning, and managing virtual machines in the data center.

Management

VMware View Manager is the core to the VMware View solution, providing centralized management of the desktop environment and brokering of connections to desktops in the data center.

VMware View Composer delivers storage optimizations to reduce storage requirements and simplify desktop management.

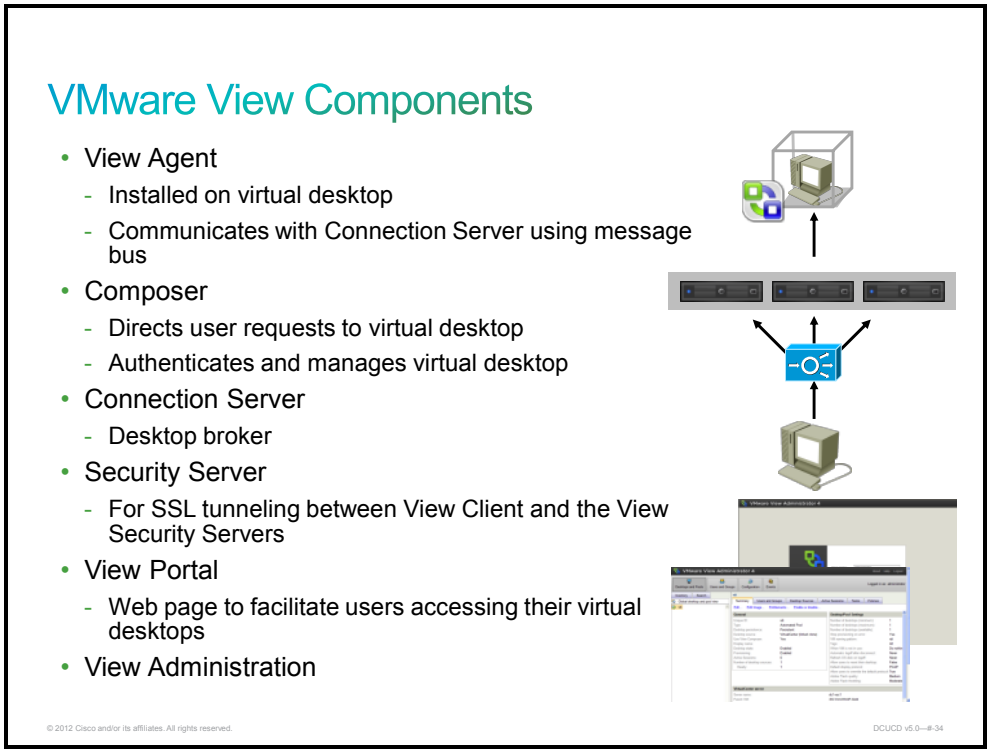
VMware ThinApp addresses the requirement for application virtualization in both virtual and physical desktop environments.

User Experience

Users can access their virtual desktops from various devices, including desktops, laptops, and thin clients. These users can be on the LAN or across the WAN.

https://t.me/learningnets

VMware View has a number of technologies addressing user experience.



View Components

VMware View consists of several components:

- View Connection Server
- View Agent
- View Client
- View Administrator
- View Composer
- View Transfer Server

View Connection Server

This is a software service that acts as a broker for client connections. The View Connection Server authenticates users through Windows Active Directory and directs the request to the appropriate VM, physical or blade PC, or Windows Terminal Services server. The Connection Server provides these services:

- Authenticates users
- Entitles users to specific desktops and pools
- Assigns applications packaged with VMware ThinApp to specific desktops and pools
- Manages local and remote desktop sessions
- Establishes secure connections between users and desktops
- Enables single sign-on
- Sets and applies policies

https://t.me/learningnets

The VDI solutions can also be used by users coming from the Internet. In such designs, the network is segmented into more- and less-trusted networks. Such segmentation influences the View design and deployment:

- Inside the corporate firewall, you install and configure a group of two or more View Connection Server instances. Their configuration data is stored in an embedded Lightweight Directory Access Protocol (LDAP) directory and is replicated among members of the group.
- Outside the corporate firewall, in the demilitarized zone (DMZ), you can install and configure View Connection Server as a security server. Security servers in the DMZ communicate with View Connection Servers inside the corporate firewall. Security servers offer a subset of functionality and are not required to be in an Active Directory domain.

View Agent

View Agent is a service installed on all virtual machines, physical systems, and Terminal Service servers that are used as sources for View desktops. This agent communicates with View Client to provide features such as connection monitoring, virtual printing, and access to locally connected USB devices.

If the desktop source is a VM, the View Agent service is first installed on that VM and then uses the VM as a template or as a parent of linked clones. When you create a pool from this virtual machine, the agent is automatically installed on every virtual desktop.

View Client

The client software for accessing View desktops runs either on a Windows or Mac PC as a native application or on a thin client with View Client for Linux.

After logging in, users select from a list of virtual desktops that they are authorized to use. Authorization can require Active Directory credentials, a User Principal Name (UPN), a smart card PIN, or an RSA SecurID token.

An administrator can configure View Client to allow end users to select a display protocol. Protocols include PCoIP, Microsoft RDP, and HP RGS. The speed and display quality of PCoIP rival that of a physical PC.

View Client with Local Mode (formerly called Offline Desktop) is a version of View Client that has been extended to allow end users to download virtual machines and use them on their local systems, regardless of whether they have a network connection.

View Administrator

This web-based application allows administrators to configure View Connection Server, deploy and manage View desktops, control user authentication, troubleshoot end-user issues, initiate and examine system events, and carry out analytical activities.

When a View Connection Server instance is installed, the View Administrator application is also installed. This application allows administrators to manage View Connection Server instances from anywhere without having to install an application on their local computer.

View Composer

You can install the View Composer software service on a vCenter Server instance to manage VMs. View Composer can then create a pool of linked clones from a specified parent VM. This strategy reduces storage costs by up to 90 percent.

Each linked clone acts like an independent desktop, with a unique hostname and IP address, yet the linked clone requires significantly less storage because it shares a base image with the parent.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Because linked-clone desktop pools share a base image, you can quickly deploy updates and patches by updating only the parent virtual machine. End-user settings, data, and applications are not affected. As of View 4.5, you can also use linked-clone technology for View desktops that you download and check out to use on local systems.

View Transfer Server

This View Transfer Server software manages and streamlines data transfers between the data center and View desktops that are checked out for use on end-user local systems. View Transfer Server is required to support desktops that run View client with Local Mode (formerly called Offline Desktop).

View Transfer Server synchronizes local desktops with the corresponding desktops in the data center by replicating user-generated changes to the data center. Replications occur at intervals that you specify in local-mode policies. It can also be initiated in View Administrator. View Transfer Server keeps local desktops up-to-date by distributing common system data from the data center to local clients. View Transfer Server downloads View Composer base images from the image repository to local desktops.

If a local computer is corrupted or lost, View Transfer Server can provision the local desktop and recover the user data by downloading the data and system image to the local desktop.

https://t.me/learningnets

VMware View Client

- Connects to View Connection Server
- Form factor depends on client device and operating system:
  - Desktop, laptop, or thin client
  - Windows, Linux, Mac, or vendor operating system
- Uses display protocol:
  - PCoIP
  - RDP



© 2012 Cisco and/or its affiliates. All rights reserved.

View Client connects to View Manager to access the virtual desktop.

A form factor depends on the client device and operating system. A form factor can be in the form of a desktop, laptop, or thin client, and can use Windows, Linux, or a vendor operating system.

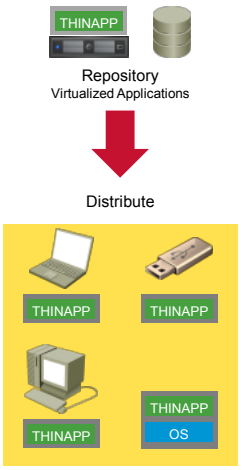
To connect to the virtual desktop, the client works with display protocols, which can be PCoIP, RDP, or Remote Graphic Service (RGS).

View Client supports extended USB redirection, virtual printing, and smart card authentication, as well as a local mode of operation.

https://t.me/learningnets

VMware ThinApp

- Agentless application virtualization solution delivering virtualized applications for virtual and physical desktops:
  - Decouple applications and data from operating system
  - Agentless architecture
  - Wide platform and application support
  - Plugs into existing application management tools
- Runs application in an isolated container:
  - Encapsulates each application and all required components
  - Stores user-specific configurations and changes in own unique container
  - Can be linked together to share interdependent components by multiple applications



© 2012 Cisco and/or its affiliates. All rights reserved.

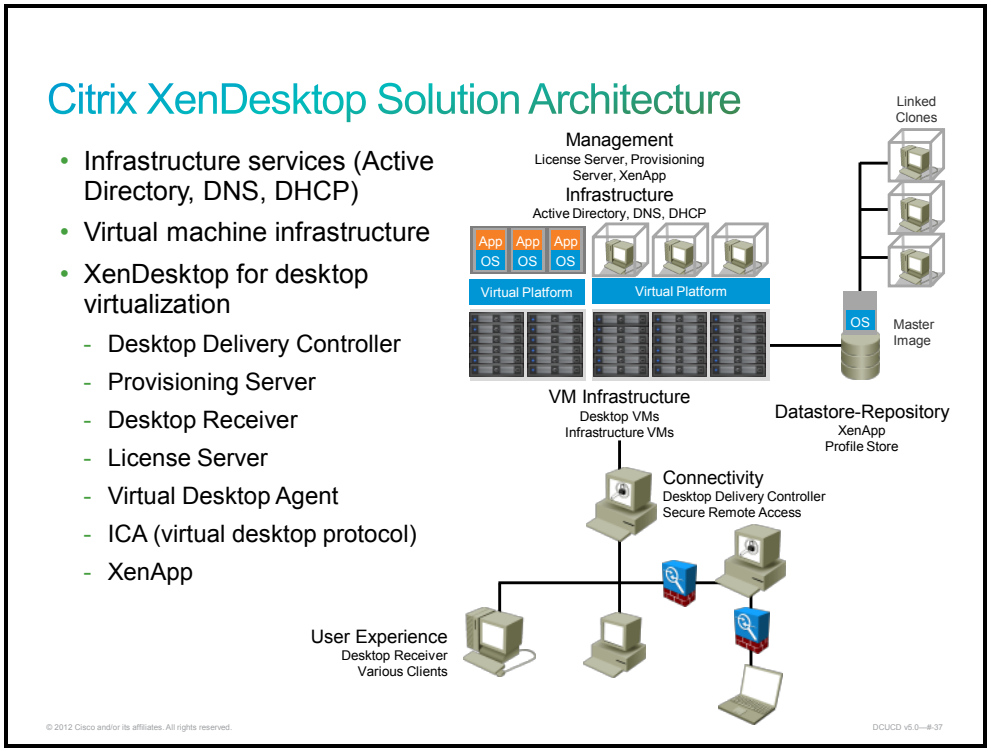
VMware ThinApp is an agentless application virtualization solution that can be used to deliver virtualized applications for virtual and physical desktops.

ThinApp does not have any complex agents to deploy or maintain, so IT is not burdened with additional agent footprints to maintain or install. It does not require additional servers or infrastructure, you simply package your applications and use your existing management frameworks to deploy and manage the ThinApp (virtualized) applications.

ThinApp enables end users to run multiple versions of the same application (such as Microsoft Office, web browsers, and so on) side by side without conflicts, because resources are unique to each application, and ThinApp reduces storage costs for VMware View:

- ThinApp enables View administrators to reuse VM templates because application-specific and user-specific configuration data (such as from the web or a network share) can be stored and streamed from a central server.
- Compared to traditional desktop deployments, where the applications are installed on every desktop, enterprises can save a significant amount of storage by delivering applications via ThinApp.
- Storage costs are significantly reduced by enabling a pool of users to leverage a single application without impacting the application or each other.
- In a 1000-desktop implementation, simply using ThinApp with Microsoft Office can save over 1 TB of shared storage. Over a three-year period, that is a savings of $30 per user. For each terabyte of shared storage that can be reduced, there is a cost saving of another $30 per user.

https://t.me/learningnets



Citrix XenDesktop transforms Windows desktops into an on-demand service that can be accessed by any user, on any device, anywhere, with simplicity and scalability. Whether you are using tablets, smart phones, laptops or thin clients, XenDesktop can quickly and securely deliver virtual desktops and applications to them with a high-definition user experience.

XenDesktop Architecture

The Citrix modular architecture provides the foundation for building a scalable virtual desktop infrastructure.

The modular architecture creates a single design for a data center, integrating all FlexCast models. The control module manages user access and virtual desktop allocation. The desktop modules integrate the aforementioned FlexCast models into the modular architecture. The imaging module provides the virtual desktops with the master desktop image. Numerous options exist for all three levels because users have different requirements and the technology must align with the user needs.

HDX Technology

Citrix HDX technology delivers a rich, complete user experience that rivals a local PC, from optimized graphics and multimedia, to high-definition webcam, broad USB device support, and high-speed printing.

HDX technology is a set of capabilities that delivers a “high definition” desktop virtualization user experience to end users for any application, device, or network. These user experience enhancements balance performance with low bandwidth – anything else becomes impractical to use and scale. HDX technology provides network and performance optimizations to deliver the best user experience over any network, including low bandwidth and high latency WAN connections.

FlexCast

Citrix FlexCast delivery technology lets you deliver desktops for any use case—from simple and standardized, to high-performance and personalized—using a single solution. With

https://t.me/learningnets

FlexCast, IT can deliver every type of virtual desktop—specifically tailored to meet the performance, security, and flexibility requirements of each user.

Citrix FlexCast consists of the following base virtual desktop models:

- **Hosted shared:** Hosted shared provides a locked down, streamlined, and standardized environment with a core set of applications, ideally suited for task workers where personalization is not needed or allowed.
- **Hosted VDI:** Hosted VDI offers a personalized Windows desktop, typically needed by office workers, which can be securely delivered over any network to any device. Hosted VDI desktops can be shared among many users or dedicated, where users have complete control to customize to suit their needs. These virtual desktops can be physical or virtual, but are connected to remotely.
- **Streamed VDI:** Streamed VDI leverages the local processing power of rich clients, while providing centralized single-image management of the desktop. These types of desktops are often used in computer labs and training facilities, and when users require local processing for certain applications or peripherals.
- **Local VM:** Local VM delivers a centrally managed desktop image to physical endpoint devices, enabling users to disconnect from the network. These types of desktops are usually required by sales, consultants, and executives.

XenDesktop Components

Citrix XenDesktop has different delivery technologies for the various desktop types:

- Pooled desktops are delivered via the Citrix Provisioning Server and Desktop Delivery Controller (connection broker)
- Assigned desktops do not use Citrix Provisioning Server, but they use Desktop Delivery Controller

Desktop Delivery Controller

Desktop Delivery Controller is a connection broker.

The XenDesktop Controller provides the link between the web interface and the XenDesktop site. The controllers authenticate users, enumerate resources for the users, and direct user launch requests to the appropriate virtual desktop. The controllers manage and maintain the state of the XenDesktop site to help control desktop startups, shutdowns, and heart beats. The controllers constantly query and update the SQL database with site status, allowing controllers to go offline without impacting user activities. It is recommended that at least two controllers be deployed per XenDesktop site to provide high availability. As the site grows, additional controllers might be required if the allocated CPU cannot service the user requests fast enough.

Provisioning Services

Provisioning Services provide images to physical and virtual desktops. Desktops use network booting to obtain the image, and only portions of the desktop images are streamed across the network, as needed. Provisioning Services use similar identity management functionality used by the Machine Creation Services. Provisioning Services require additional server resources, but these can be either physical or virtual servers, depending on the capacity requirements and hardware configuration. Also, Provisioning Services do not require the desktop to be virtualized, because Provisioning Services can deliver desktop images to physical desktops.

Citrix Provisioning Server is a streaming technology that can be used to provide multiple virtual desktops from a single copy of an operating system in the storage back-end. It provides the same image management capabilities as VMware Linked Clones, even though they use

https://t.me/learningnets

different technologies. Similar to Linked Clones, the limitation is that it cannot be used for assigned desktops.

Licensing Server

The Licensing Server is responsible for managing the licenses for all of the components of XenDesktop 5. XenDesktop has a 90-day grace period, which allows the system to function normally for 90 days if the license server becomes unavailable. This grace period offsets the complexity involved with building redundancy into the license server. This service is minimally impacted and is a prime candidate for virtualization.

Virtual Desktop

XenDesktop provides a powerful desktop computing infrastructure that is easy to manage and support. The open architecture works with your existing hypervisor, storage, Microsoft, and system management infrastructures, with complete integration and automation via the comprehensive software development kit (SDK).

Virtual Desktop Agent and Desktop Receiver

Citrix Receiver, a lightweight universal client, enables any PC, Mac, smart phone, tablet, or thin client to access corporate applications and desktops—easily and securely.

XenApp

Citrix application virtualization technology isolates applications from the underlying operating system and from other applications to increase compatibility and manageability. As a modern application delivery solution, XenApp virtualizes applications via integrated application streaming and isolation technology. This application virtualization technology enables applications to be streamed from a centralized location into an isolation environment on the target device. With XenApp, applications are not installed in the traditional sense. The application files, configuration, and settings are copied to the target device and the application execution at run time is controlled by the application virtualization layer. When executed, the application run time believes that it is interfacing directly with the operating system when, in fact, it is interfacing with a virtualization environment that proxies all requests to the operating system.

XenApp is unique in that it is a complete system for application delivery, offering both online and offline application access through a combination of application hosting and application streaming directly to user devices. When users request an application, XenApp determines if their device is compatible and capable of running the application in question. The minimum requirements of a target device are a compatible Windows operating system and appropriate Citrix client software. If the user device meets minimum requirements, then XenApp initiates application virtualization via application streaming directly into an isolated environment on the user device. If the user device is not capable of running a particular application, XenApp initiates session virtualization.

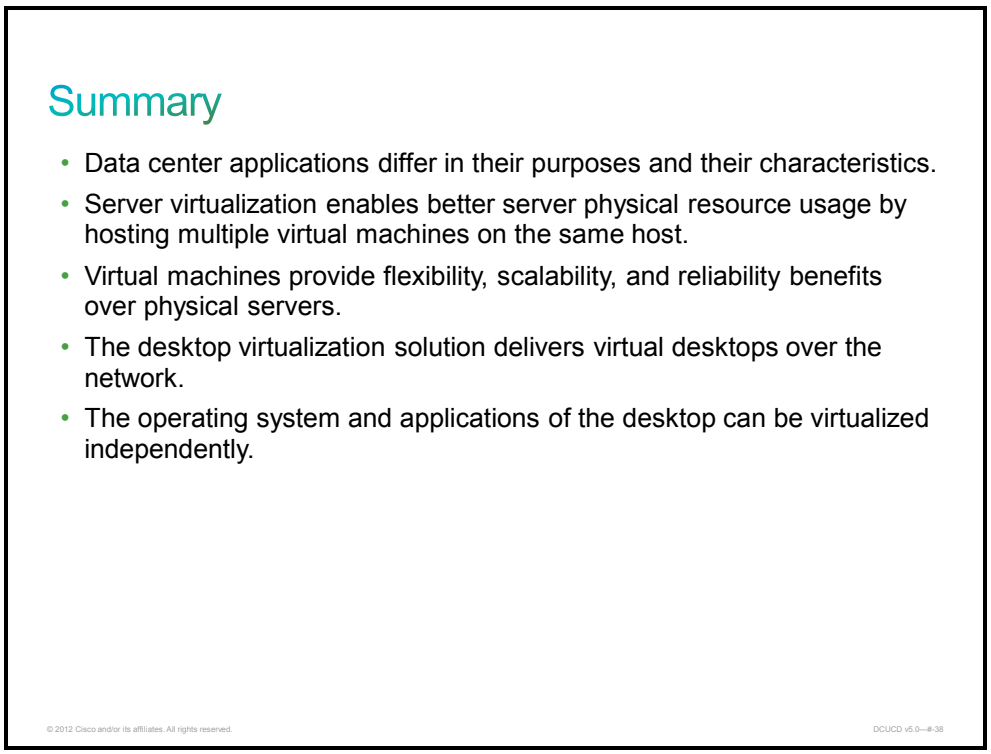
Application Streaming Profiler

The Citrix Streaming Profiler is the software component that is required to create the application packages. Just like other application virtualization solutions, this component should be installed on a clean machine where only the necessary files and settings are saved within the record process. The installation of the software component is easily done by following instructions from the wizard. When done, you are ready to create your first Application Profile. It is best practice to use a virtual workstation for these activities, so you can create a snapshot of the state before you start the profiling process. Citrix is updating this software component on a regular basis, so, when building the virtual machine, it is a good idea to check the Citrix website for the latest version.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- http://www.cisco.com/en/US/netsol/ns340/ns394/ns224/index.html
- http://www.cisco.com/en/US/netsol/ns743/networking_solutions_program_home.html
- http://en.wikipedia.org/wiki/virtualization
- http://hadoop.apache.org/
- http://www.citrix.com/lang/English/home.asp
- http://www.vmware.com/products/vsphere/mid-size-and-enterprise-business/overview.html
- http://www.vmware.com/products/view/overview.html
- http://www.microsoft.com/en-us/server-cloud/windows-server/default.aspx
- http://www.microsoft.com/en-us/server-cloud/windows-server/hyper-v.aspx

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Lesson 3

Identifying Cloud Computing

Overview

Cloud computing is a major trend in the IT industry that promises to lower the cost of IT by improving IT efficiency, lessening the time used by IT departments to maintain IT deployments, and minimizing the time required to deploy new services and applications.

This lesson describes cloud computing, the deployment models, and the service categories that can be used for the cloud computing solution, and identifies the important aspects of any cloud computing solution.

Objectives

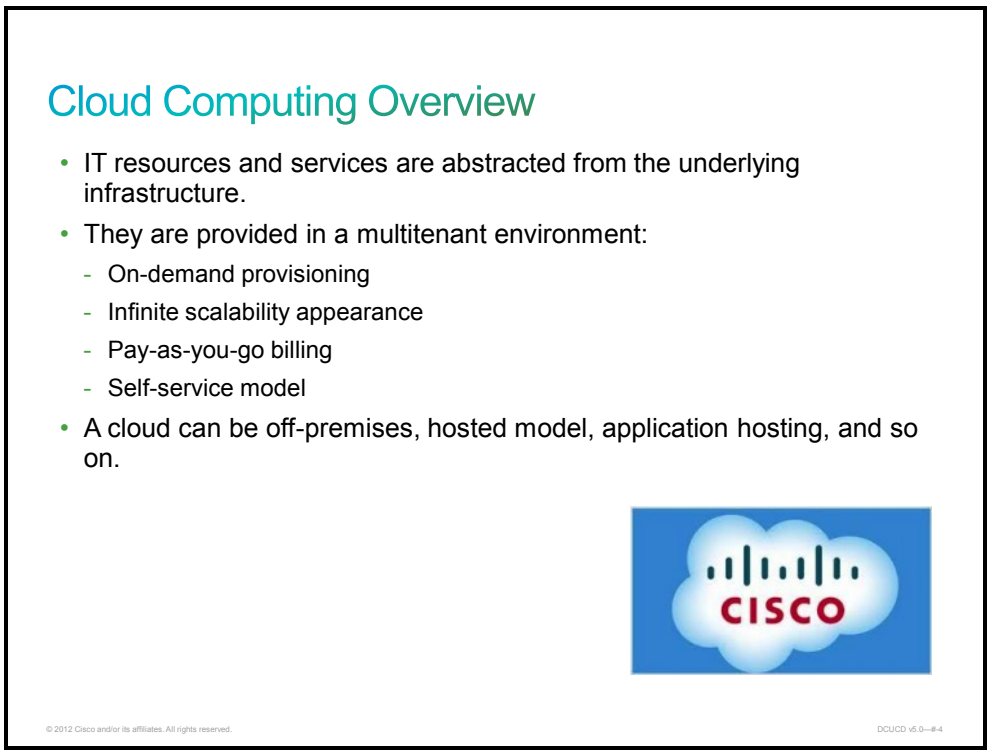
Upon completing this lesson, you will be able to identify cloud computing, deployment models, service categories, and important aspects of a cloud computing solution. This ability includes being able to meet these objectives:

- Evaluate the cloud computing solution, terms, and general characteristics
- Recognize cloud computing deployment models
- Compare cloud computing service delivery categories, the responsibilities demarcation, and their applicability
- Recognize the aspects of cloud computing services and solutions

https://t.me/learningnets

Cloud Computing Overview

This topic describes the cloud computing solution, terms, and general characteristics.



Cloud computing translates into IT resources and services that are abstracted from the underlying infrastructure and are provided “on-demand” and “at scale” in a multitenant environment. Today, clouds are associated with an off-premises model, a hosted model, application hosting, private self-service IT infrastructure, and so on.

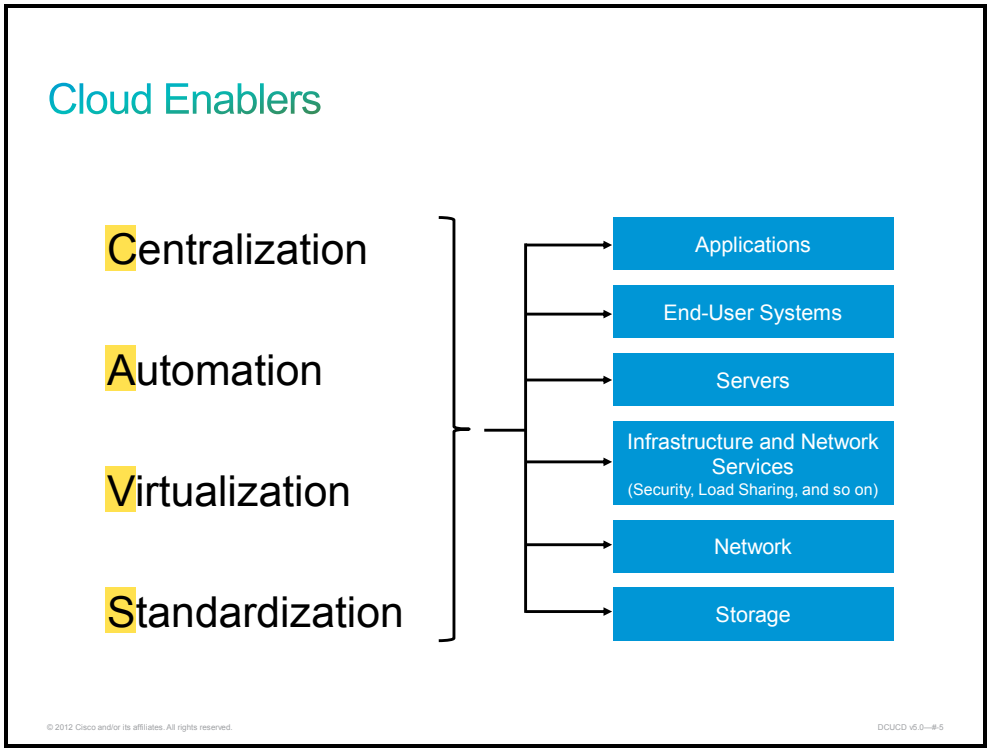
Currently, cloud computing is still quite new and there are no definite and complete standards yet.

There are many cloud solutions that are available and offered, and more are anticipated in the future.

Cloud computing can be seen as a virtualized data center. However, there are some details that differentiate a cloud computing solution from a virtualized data center.

One such detail is on-demand billing, where, typically, the resource usage is tracked in a cloud at a granular level and billed to the customer on a short interval. Another difference is the way in which associations of resources to an application are done. That is, they are more dynamic and ad hoc than in a virtual data center.

https://t.me/learningnets



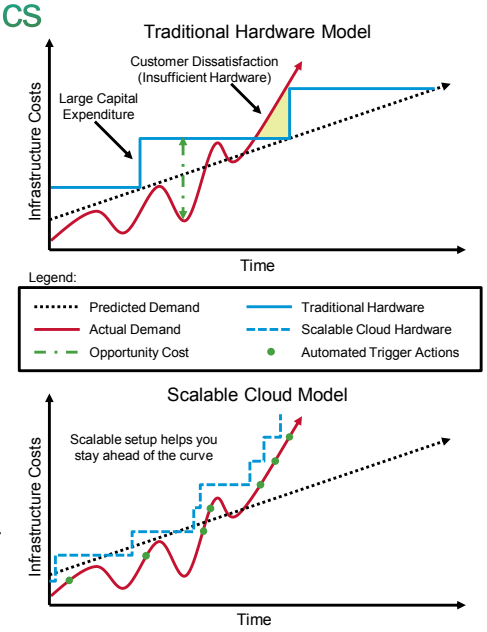
Cloud computing is possible due to fundamental principles that are used and applied in modern IT infrastructures and data centers:

- **Centralization (that is, consolidation):** Aggregating the compute, storage, network, and application resources in central locations, or data centers
- **Virtualization:** Enables seamless scalability and quick provisioning
- **Standardization:** Enables the "last but not least" principle
- **Automation:** Brings time savings and enables user-based self-provisioning of IT services

https://t.me/learningnets

Terms and Characteristics

- Cloud types:
  - Public cloud
  - Private cloud
  - Virtual private cloud
  - Hybrid cloud
  - Community cloud
- Architecture
- Characteristics:
  - Multitenancy and isolation
  - Security
  - Automation
  - Standards
  - Elasticity
- The solution depends on the vendor
- Common standards are being defined



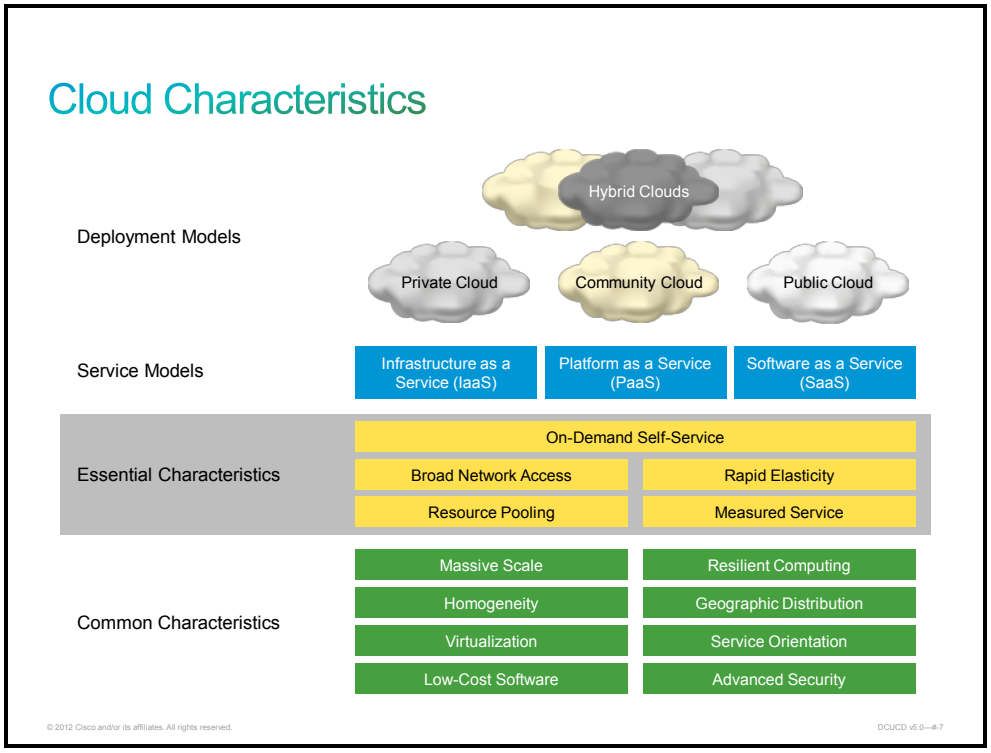
© 2012 Cisco and/or its affiliates. All rights reserved.

The effect of cloud services on IT services can be explained by the way that the cloud hides the complexity and allows control of resources, while providing the automation that tends to decrease the complexity.

The general characteristics of a cloud computing solution are multitenancy with proper security and isolation, automation for self-service and self-provisioning, standardized components, and the appearance of infinite resources.

When comparing the traditional IT model to a cloud-based solution, the cloud-based solution uses physical resources more efficiently from a customer perspective because the cloud defaults to per-use resource allocation. That is, the amount of physical resources used is almost the same as the amount of physical resources available (such as a pay-per-use model).

https://t.me/learningnets



The National Institute of Standards and Technology (NIST) is a U.S. government agency, part of the U.S. Department of Commerce, that is responsible for establishing and providing standards of all types as needed by industry or government programs.

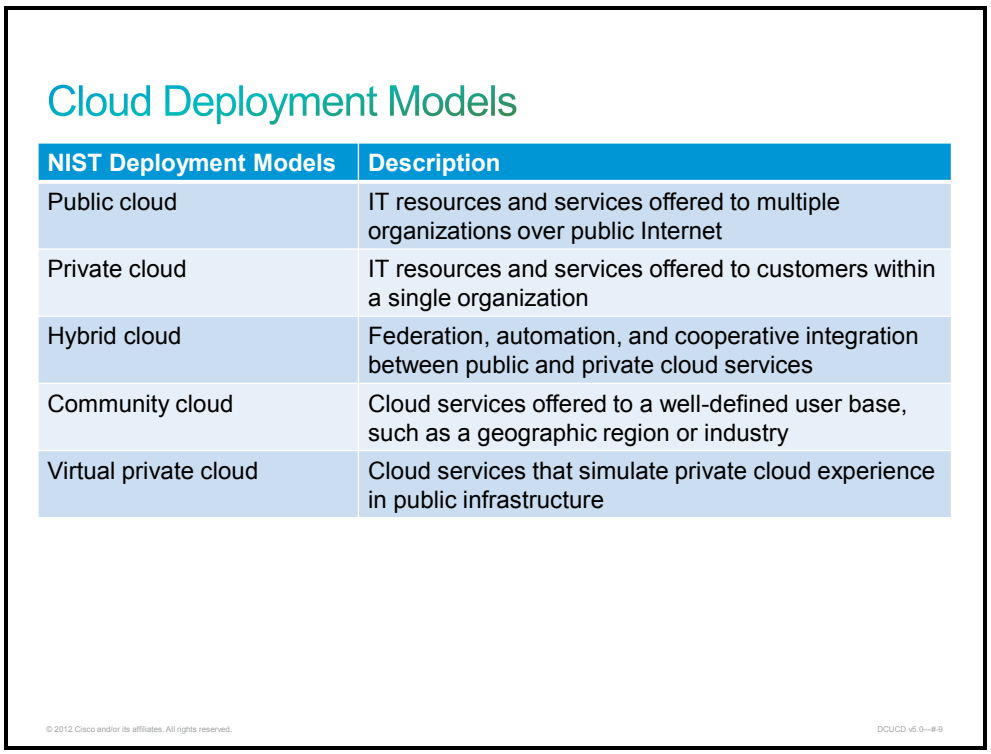
The NIST defines the cloud solution as having these characteristics:

- **Multitenancy and isolation:** This characteristic defines how multiple organizations use and share a common pool of resources (network, compute, storage, and so on) and how their applications and services that are running in the cloud are isolated from each other.
- **Security:** This feature is implicitly included in the previous characteristic, multitenancy and isolation. It defines not only the security policy, mechanisms, and technologies that are used to secure the data and applications of companies that are using cloud services, but it is also anything that secures the infrastructure of the cloud itself.
- **Automation:** This feature is an important characteristic that defines how a company can get the resources and set up its applications and services in the cloud, without too much intervention from the cloud service support staff.
- **Standards:** There should be standard interfaces for protocols, packaging, and access to cloud resources so that the companies that are using an external cloud solution (that is, a public cloud or open cloud) can easily move their applications and services between the cloud providers.
- **Elasticity:** The flexibility and elasticity allows users to scale up and down at will—utilizing the resources of all kinds (CPU, storage, server capacity, load balancing, and databases).

https://t.me/learningnets

Cloud Computing Models

This topic identifies cloud computing deployment models.



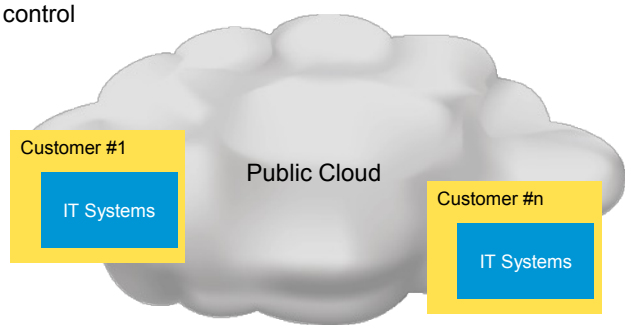
As mentioned, cloud computing is not very well-defined in terms of standards, yet there are trends and activities towards defining common cloud computing solutions.

The NIST also defines the cloud computing deployment models with five categories: public, private, hybrid, community, and virtual private.

https://t.me/learningnets

Public Cloud

- IT resources and services of customer premises sold with cloud computing qualities:
  - Self-service
  - Pay-as-you-go billing
  - On-demand provisioning
  - Appearance of infinite scalability
  - Limited customer control



© 2012 Cisco and/or its affiliates. All rights reserved.

A public cloud can offer IT resources and services that are sold with cloud computing qualities, such as self-service, pay-as-you-go billing, on-demand provisioning, and the appearance of infinite scalability.

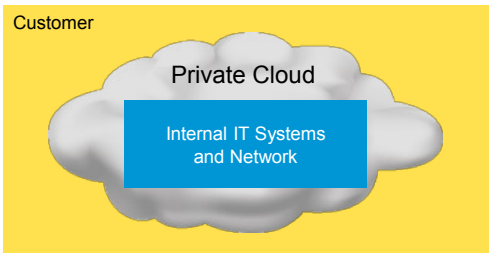
Here are some examples of cloud-based offerings:

- Cisco WebEx TelePresence
- Google services and applications such as Google Docs and Gmail
- Amazon web services and the Amazon Elastic Compute Cloud (EC2)
- Salesforce.com cloud-based customer relationship management (CRM)
- Skype voice services

https://t.me/learningnets

Private Cloud

- Enterprise IT infrastructure with cloud computing qualities:
  - Self-service
  - Resource usage showback
  - On-demand provisioning
  - Appearance of infinite scalability
- You can consolidate, virtualize, and automate data center resources with provisioning and cost-metering interfaces to enable self-service IT consumption.



© 2012 Cisco and/or its affiliates. All rights reserved.

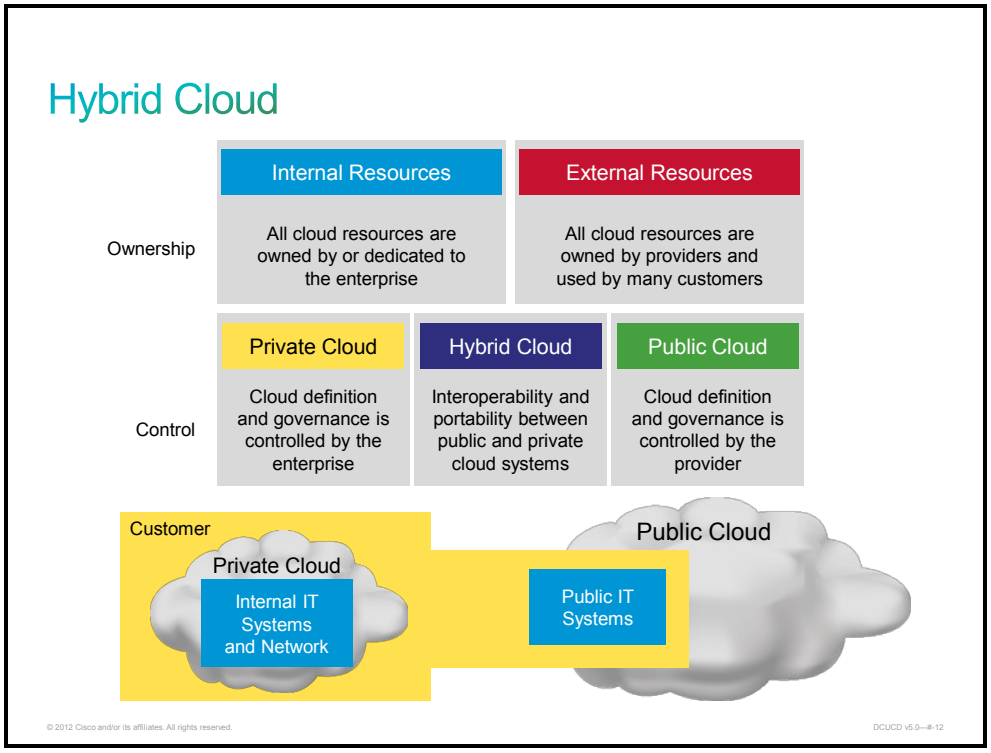
A private cloud is an enterprise IT infrastructure that is managed with cloud computing qualities, such as self-service, resource usage showback (that is, internal resources usage metering and reporting per-user), on-demand provisioning, and the appearance of infinite scalability.

The private clouds have the following characteristics:

- Consolidated, virtualized, and automated existing data center resources
- Provisioning and usage-metering interfaces to enable self-service IT consumption
- Targeted at one or two noncritical application systems

When a company decides to use a private cloud service, the private cloud scales by pooling IT resources under a single cloud operating system or management platform. It can support up to thousands of applications and services. Such solutions will enable new architectures to target very large-scale activities.

https://t.me/learningnets



The intersection between the private and public cloud is the hybrid cloud, which provides a way to interconnect the private and public cloud, running IT services and applications partially in the private and partially in the public cloud.

A hybrid cloud links disparate cloud computing infrastructures (that is, enterprise private cloud with service provider public cloud) with each other by connecting their individual management infrastructures and allowing the exchange of resources.

The hybrid cloud can enable these actions:

- Disparate cloud environments can leverage other cloud-system resources.
- The “federation” can occur across data center and organization boundaries with cloud internetworking.

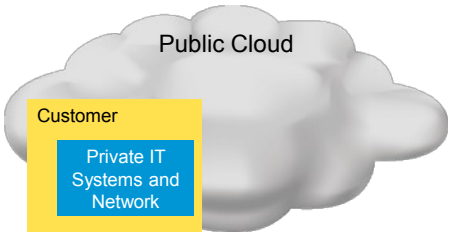
The characteristics and aspects of a hybrid cloud are those of a private and public cloud, such as limited customer control for the public part, the ability to meter and report on resource usage, and the ability to understand the cost of public resources.

https://t.me/learningnets

Virtual Private Cloud

A virtual private cloud is a service offering that allows enterprises to create and use their private clouds on infrastructure services that are provided by the service provider:

- Leverage services of third-party IaaS providers
- Virtualize trust boundaries through cloud internetworking standards and services
- Access vendor billing and management tools through a private cloud management system



© 2012 Cisco and/or its affiliates. All rights reserved.

A virtual private cloud is a service offering that allows enterprises to create their private clouds on the public infrastructure (that is, a public cloud that is provided by the service provider).

The closed-cloud service provider enables the enterprises to perform these activities:

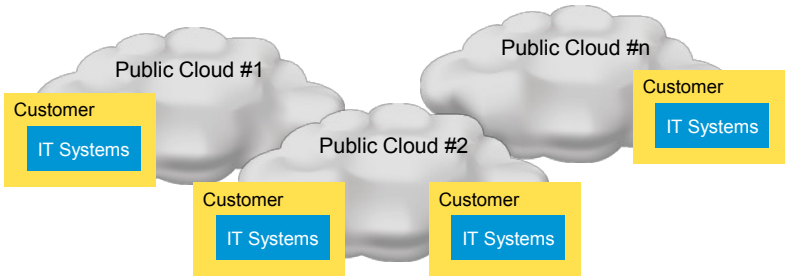
- Leverage virtual port channel (vPC) services that are offered by third-party Infrastructure as a Service (IaaS) providers
- Virtualize trust boundaries through cloud internetworking standards and services
- Access vendor billing and management tools through a private cloud management system

https://t.me/learningnets

Open Cloud

An open cloud is a federation of similar infrastructures offered by different service providers:

- Open, on-demand infrastructure services
- Intercloud standards and services
- Standards for consolidated application and service management, and billing
- Decouple one-to-one business relationships with service providers



© 2012 Cisco and/or its affiliates. All rights reserved.

The largest and most scalable cloud computing system—the open cloud—is a service provider infrastructure that allows a federation with a similar infrastructure offered by other providers. Enterprises can choose freely among participants, and service providers can leverage other provider infrastructures to manage exceptional loads on their own offerings.

A federation will link disparate cloud computing infrastructures with each other by connecting their individual management infrastructures and allowing the exchange of resources and the aggregation of management and billing streams.

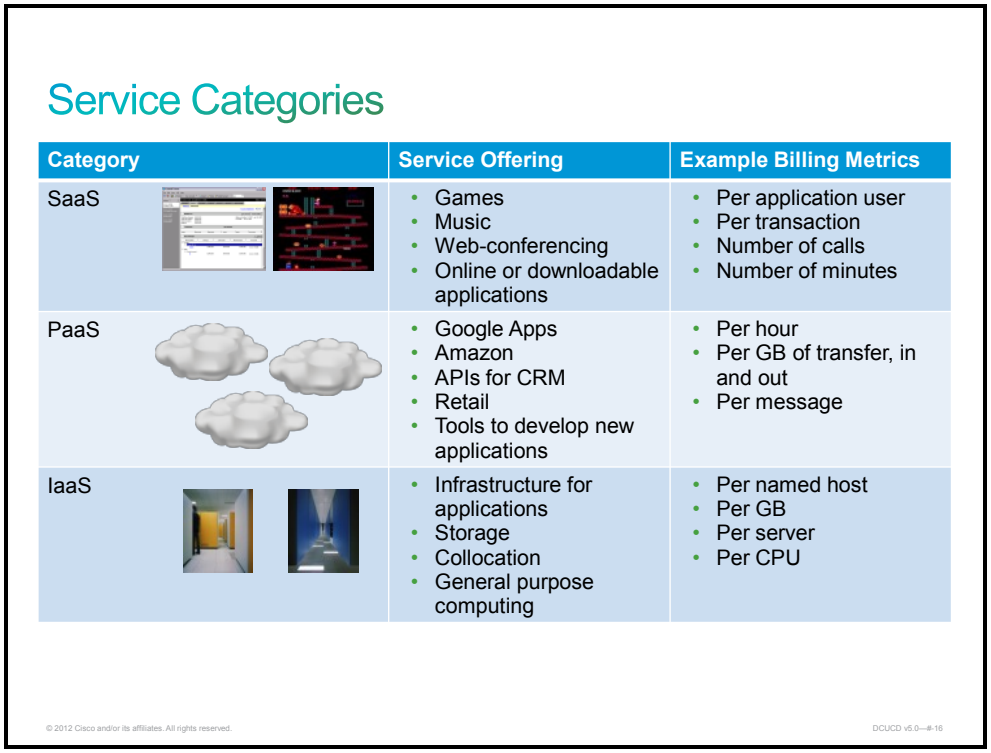
The federation can enable these options:

- Disparate cloud environments can leverage other cloud-system resources.
- The federation can occur across data center and organization boundaries with cloud internetworking.
- The federation can provide unified metering and billing and “one-stop” self-service provisioning.

https://t.me/learningnets

Cloud Computing Service Categories

This topic identifies cloud computing service delivery categories, the responsibilities demarcation, and their applicability.



There are three basic cloud computing service models:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

SaaS

SaaS is software that is deployed over the Internet or is deployed to run behind a firewall in your LAN or PC. A provider licenses an application to customers as a service on demand, through a subscription or a "pay-as-you-go" model. SaaS is also called "software on demand." SaaS vendors develop, host, and operate software for customer use.

Rather than installing software onsite, customers can access the application over the Internet. The SaaS vendor may run all or part of the application on its hardware or may download executable code to client machines as needed—disabling the code when the customer contract expires. The software can be licensed for a single user or for a group of users.

PaaS

PaaS is the delivery of a computing platform and solution stack as a service. It facilitates the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software and provisioning hosting capabilities, providing all of the facilities that are required to support the complete life cycle of building and delivering web applications and services entirely from the Internet.

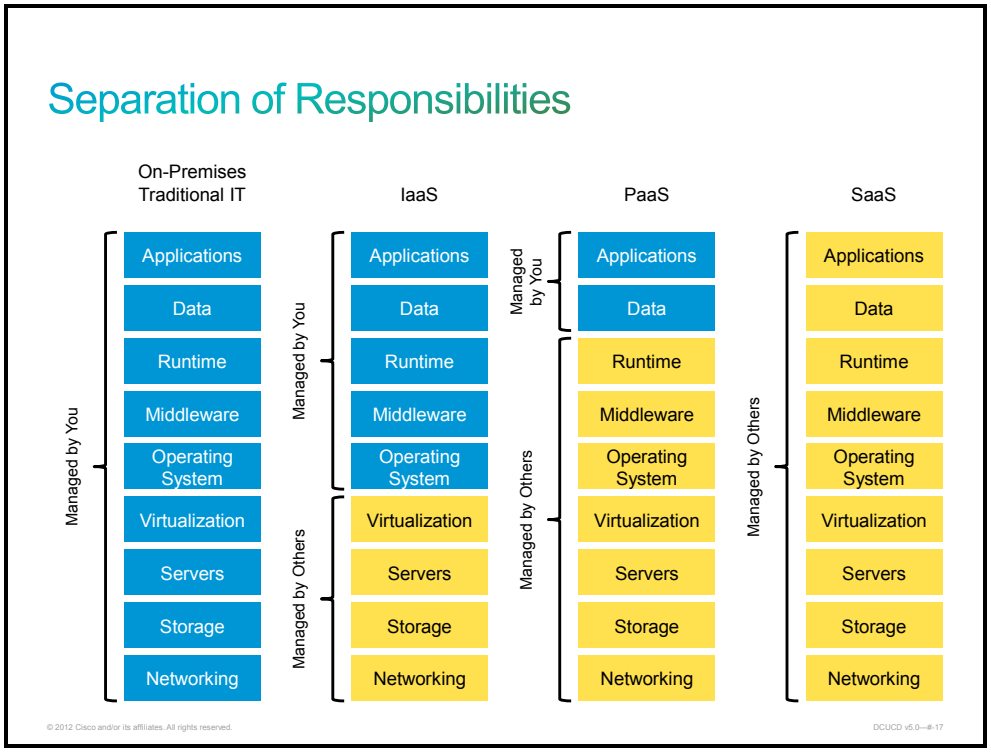
https://t.me/learningnets

The offerings may include facilities for application design, application development, testing, deployment, and hosting as well as application services such as team collaboration, web service integration and marshaling, database integration, security, scalability, storage, persistence, state management, application versioning, application instrumentation, and developer community facilitation. These services may be provisioned as an integrated solution online.

IaaS

IaaS, or cloud infrastructure services, can deliver computer infrastructure, typically a platform virtualization environment, as a service. Rather than purchasing servers, software, data center space, or network equipment, clients instead can buy those resources as a fully outsourced service.

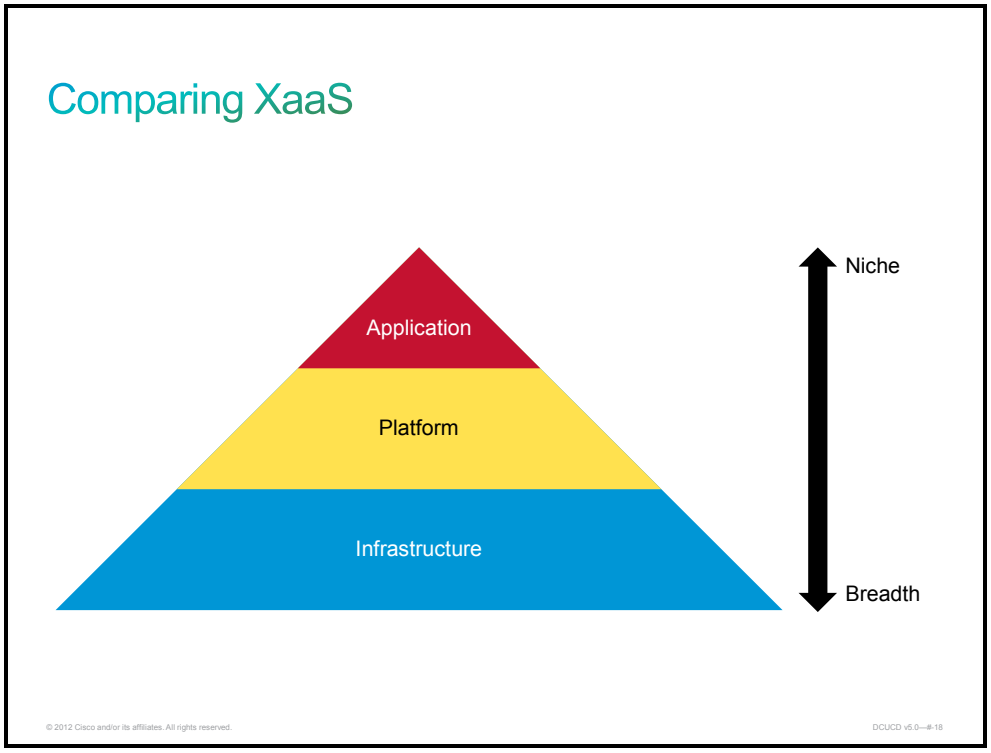
The service is typically billed on a utility computing basis and the amount of resources consumed (and therefore the cost) will typically reflect the level of activity. It is an evolution of virtual private server offerings.



The type of service category defines the demarcation point of the management responsibilities. IaaS has shared responsibilities between the customer and service provider, while with SaaS, almost all management responsibilities are on the service provider side.

This demarcation point means that, particularly with PaaS and SaaS, the service provider is invested with more trust and must have better understanding.

https://t.me/learningnets



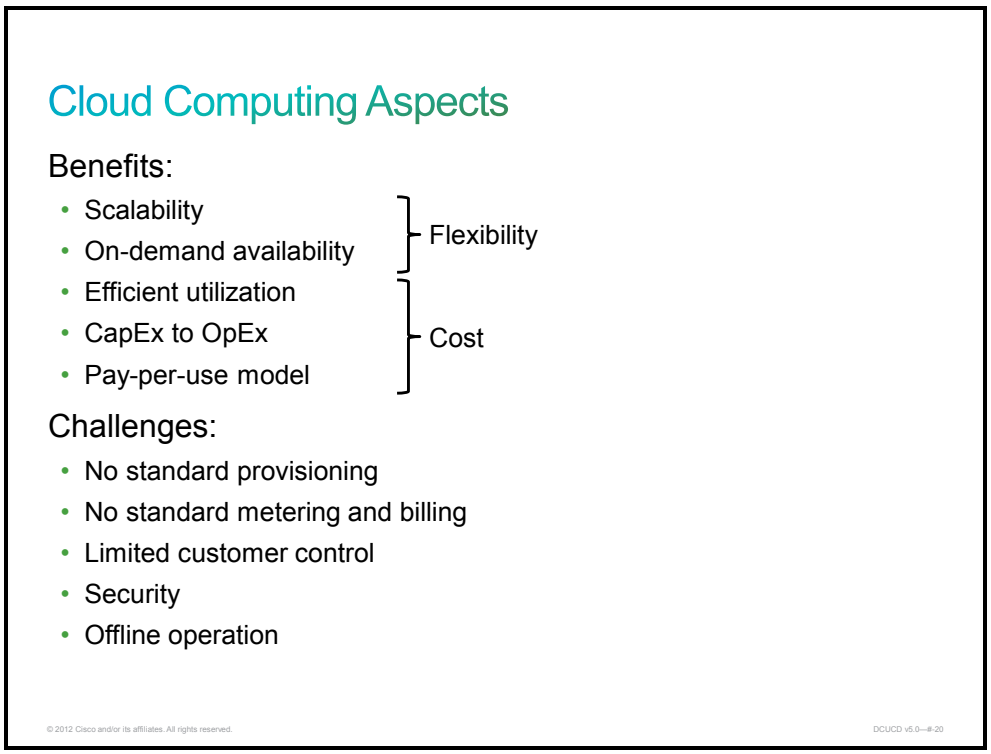
When comparing the service categories, it is obvious that the IaaS has the most breadth, because it is the most general category (that is, it does not depend on the application being used by the customer). For example, running a virtual machine (VM) on IaaS allows a customer to deploy CRM, a collaboration application or infrastructure service such as Active Directory (for example, Microsoft Exchange 2010, SAP, Microsoft Active Directory), whereas SaaS is the most “niche” service category because it defines the application type and usage. For example, Salesforce CRM is used solely for customer relationship management and cannot be used for infrastructure service.

From a user or customer type perspective, the IaaS is targeted at IT administrators and departments, the PaaS is targeted mainly at software developers, and the SaaS is targeted at the applications or software users.

https://t.me/learningnets

Cloud Computing Aspects

This topic identifies the aspects of cloud computing services and solutions.



Cloud computing offers several benefits but also introduces some challenges or drawbacks that are not present in the current IT deployment model.

The benefits are the following:

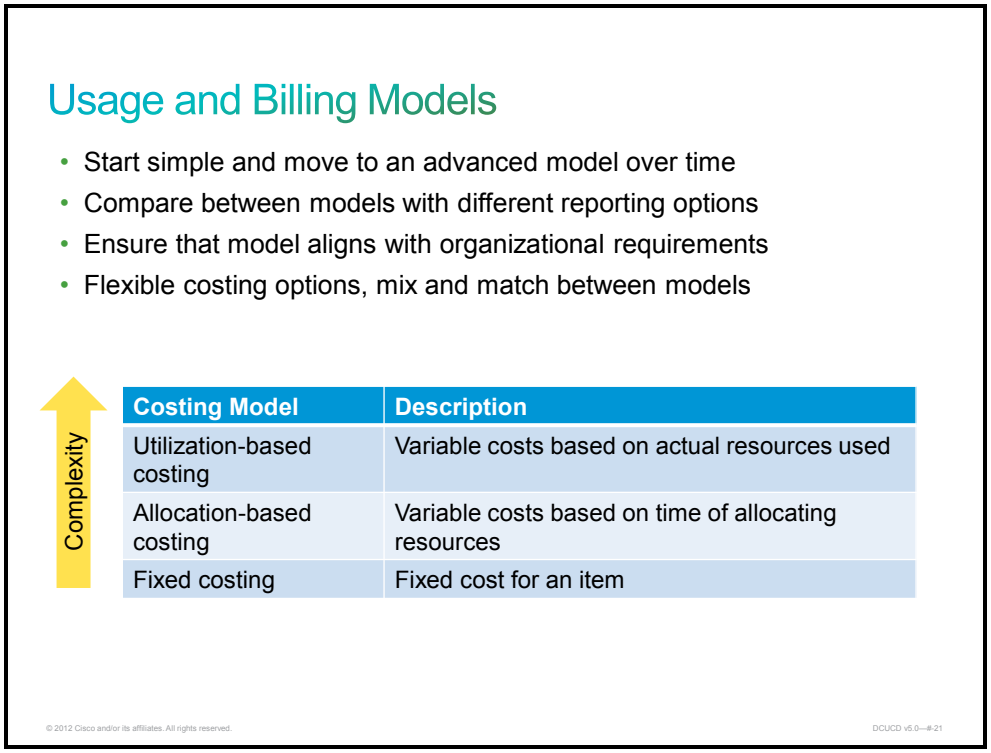
- **Scalability:** When needed, the service provider can easily add resources to the cloud.
- **On-demand availability:** When needed, the customer gets resources (if available) upon request.
- **Efficient utilization:** Sharing the physical resources among many customers and using the well-known service provider model enables even higher utilization of the infrastructure resources.
- **Capital expenditures (CapEx) to operating expenses (OpEx):** Instead of buying and owning the equipment, the customer rents the resources when needed and releases them when no longer needed. This arrangement effectively reduces the CapEx.
- **Pay-per-use model:** Because the equipment and resources are not owned by the customer, the costs are only paid for when really required. This model enables a pay-per-use and pay-as-you-grow operational model for the customers.

The current drawbacks of cloud computing are the following:

- Lack of standardized provisioning solutions.
- Lack of standardized metering and billing mechanisms and applications.
- Limited control from the customer perspective—the customer effectively does not know where the resources are and has no control over how they are used.

https://t.me/learningnets

- When applications and services are running in the cloud, the customer heavily relies on the service provider for security and confidentiality enforcement. For some customers who are bound by government directives and legislation, a public cloud is not possible.
- Having applications and services run in the cloud can increase business process dependency on network connectivity. Furthermore, for such applications and services, the customer exclusively relies on the high-availability policies of the service provider for continuous IT operation.



Service providers or internal IT have a key building block for delivering IT as a service: the infrastructure. Thus there is a need to meter resources offered and used, including broadband network traffic, public IP addresses, and other services such as DHCP, NAT, firewalling, and so on.

Service providers must create a showback or charge-back hierarchy that provides the basis for determining cost structures and delivery of reports. The main difference between charge-back and showback is that charge-back is used by service providers to charge for the cloud resources used, whereas the showback is used by the internal IT to meter and report on physical resource usage.

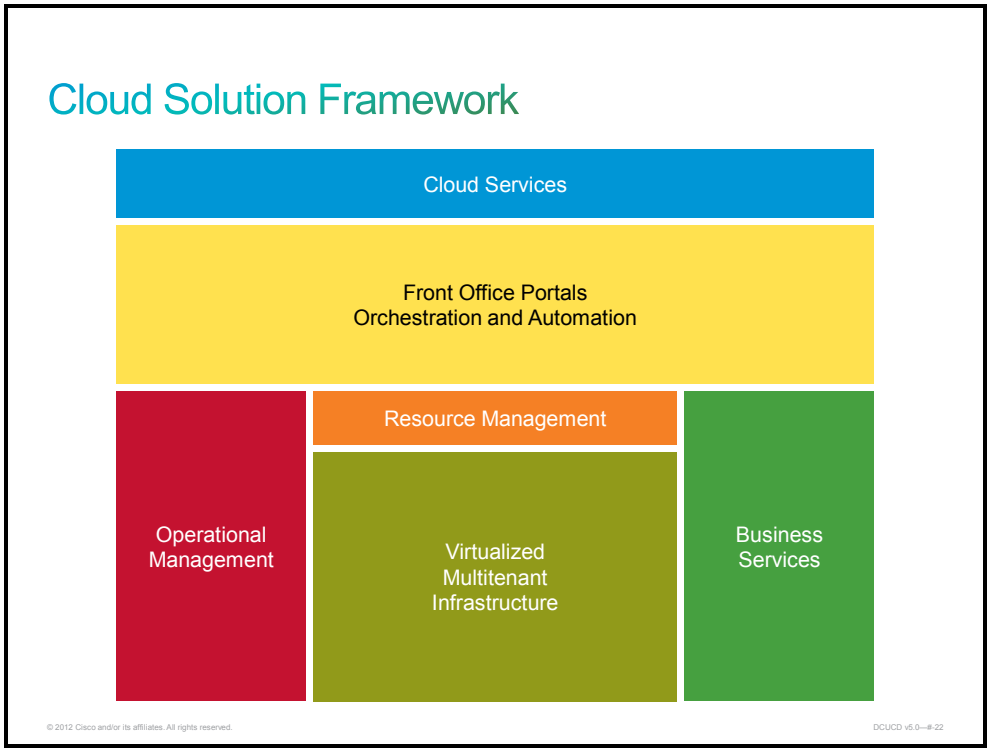
Multiple cost models provide flexibility in measuring costs.

Three basic cost models are typically used:

- **Fixed cost:** Specific per-VM instance costs, such as floor space, power and cooling, software, administrative overhead, or licenses. These aspects either cannot be metered, it is too complicated to properly meter the usage, or the usage does not change through time.
- **Allocation-based costing:** Variable costs per virtual machine based on allocated resources, such as the amount of memory, CPU, or storage allocated or reserved for the virtual machine.
- **Utilization-based costing:** Variable costs per virtual machine based on actual resources used, including average memory, disk and CPU usage, network I/O, and disk I/O.

https://t.me/learningnets

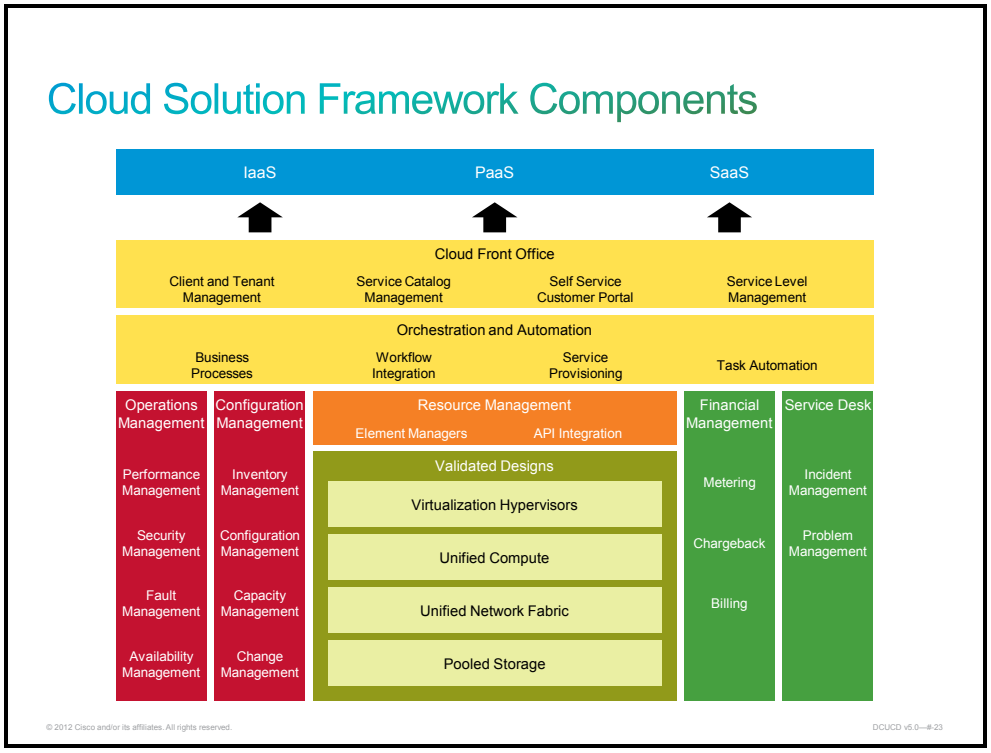
Cost models can be combined in a cost template, making it easy to start with a simple charge-back model and align with organizational requirements. Typically, the final cost comprises multiple elements.



Any cloud solution is based on the following high-level framework:

- **Multitenant virtualized infrastructure:** The basis for the solution is the resources combined in a multitenant virtualized infrastructure, which combines all the elements of the data center solution, with some extensions outside of the data center that are appropriate for the given cloud solution.
- **Operational management:** The vital part of the framework is the operational management, which enables the owner of the cloud infrastructure to internally maintain and operate it.
- **Resource management:** The resource management is the middleware between the infrastructure and the customer front-end systems. It provides the elements with which the infrastructure components and resources are managed, as well as the necessary API for the front end and the integration with customer-side management systems.
- **Business services:** The business services define the way in which the cloud services are sold and managed from a customer-facing side.
- **Front office portals, orchestration and automation:** These are the customer-facing front-end systems through which the customer, by the means of self-service and self-provisioning, uses the cloud solution, chooses among the available services from the catalog, selects the service level management, and so on.
- **Cloud services:** The cloud services are the services that the customer— IT administrator, software developer, or end user—is using.

https://t.me/learningnets

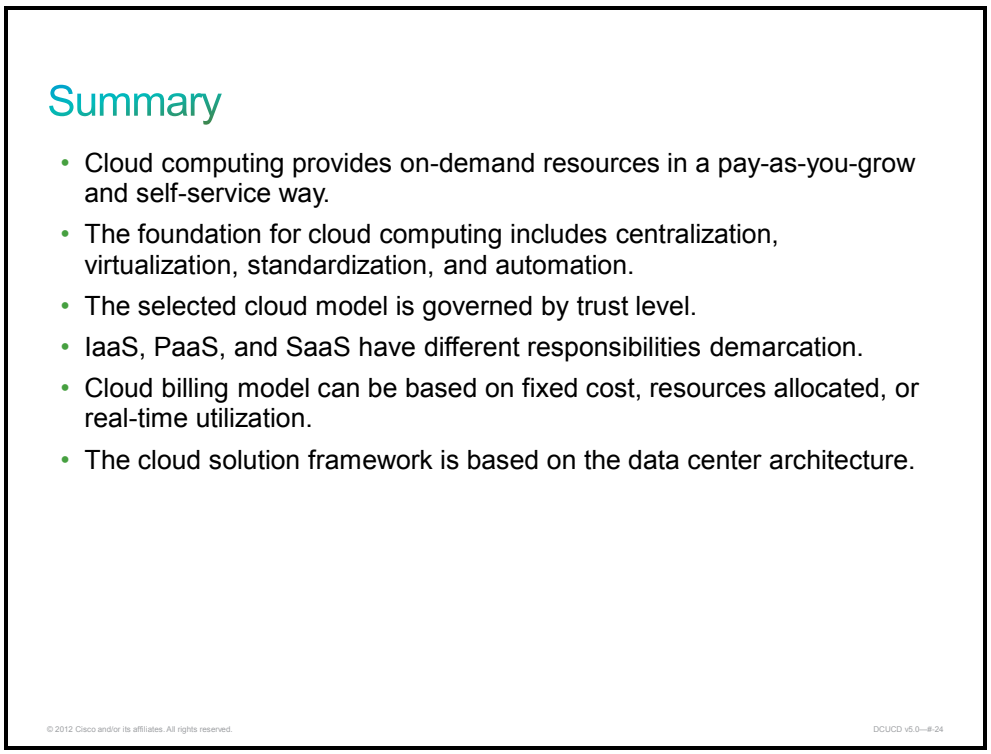


The architectural framework points to the differences between managed and cloud-based solutions. Cloud computing adds a large number of management applications and services, which include orchestration and self-provisioning portals accompanied by the service catalogs and service level management.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- http://www.cisco.com/web/solutions/trends/cloud/index.html
- http://www.cisco.com/web/solutions/trends/cloud/cloud_services.html#~CiscoCollabCloud
- http://www.nist.gov/itl/cloud/index.cfm
- http://en.wikipedia.org/wiki/Cloud_computing

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Lesson 4

Identifying Cisco Data Center Architecture and Components

Overview

Cisco Data Center is an architectural framework on which modern data center solutions are built. This lesson describes the Cisco Data Center architectural framework and the technologies, mechanisms, and equipment used in the solutions.

Objectives

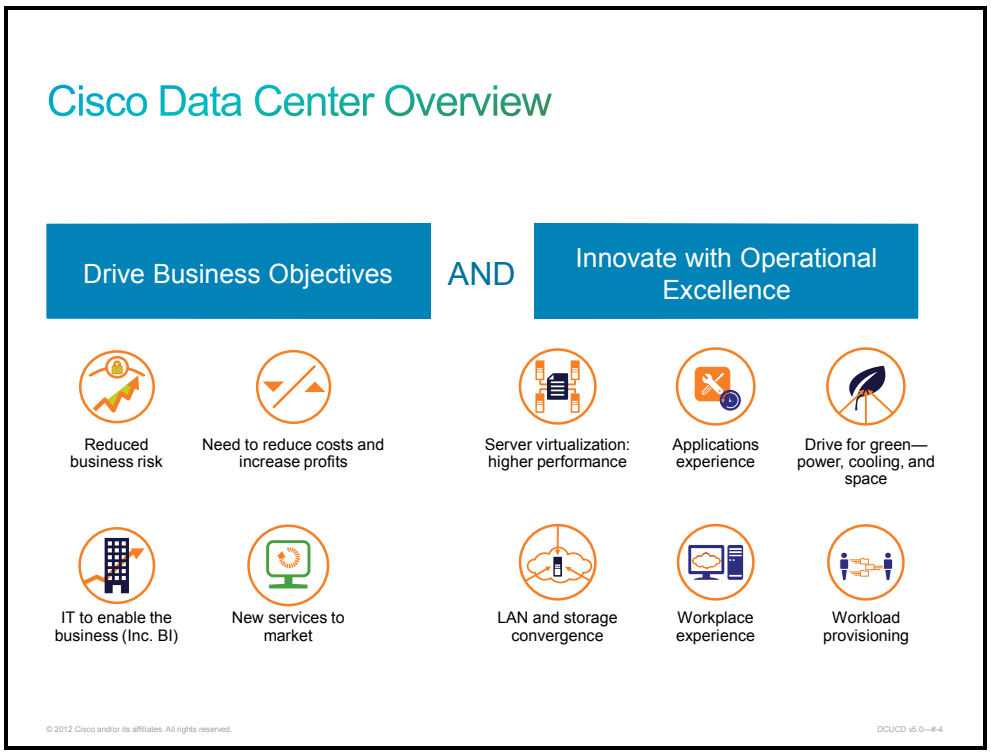
Upon completing this lesson, you will be able to identify the Cisco Data Center architectural framework and components within the solution. This ability includes being able to meet these objectives:

- Describe the Cisco Data Center architectural framework
- Describe the Cisco Data Center architectural framework unified fabric component
- Describe the Cisco Data Center network equipment
- Describe the Cisco Data Center architectural framework compute component
- Describe Cisco Validated Designs

https://t.me/learningnets

Cisco Data Center Architecture Overview

This topic describes the Cisco Data Center architectural framework.



Cisco Data Center represents a fundamental shift in the role of IT into becoming a driver of business innovation. Businesses can create services faster, become more agile, and take advantage of new revenue streams and business opportunities. Cisco Data Center increases efficiency and profitability by reducing capital, operating expenses, and complexity. It also transforms how a business approaches its market and how IT supports and aligns with the business, to help enable new and innovative business models.

https://t.me/learningnets

Cisco Data Center Benefits

Category	Business Goals (CEO Focus)	Data Center Capabilities (CIO Focus)
Growth	Market responsiveness and new markets, customers, models, acquisitions, and branches	New service creation, faster application deployment, accelerating revenue
Margin	Increased margins, customer satisfaction and retention, increased ROI on value-added IT spending	Profitability through service quality, cloud-based customer models, and converged architecture
Risk	Risk mitigation, compliance, regulatory environment, security of data, policy management, and access	New business models governance and risk, policy-based provisioning, integrated management that reduces control points

© 2012 Cisco and/or its affiliates. All rights reserved.

Businesses today are under three pressure points: business growth, margin, and risk.

For growth, the business needs to be able to respond to the market quickly and lead market reach into new geographies and branch openings. Businesses also need to gain better insight with market responsiveness (new services), maintain customer relationship management (CRM) for customer satisfaction and retention, and encourage customer expansion.

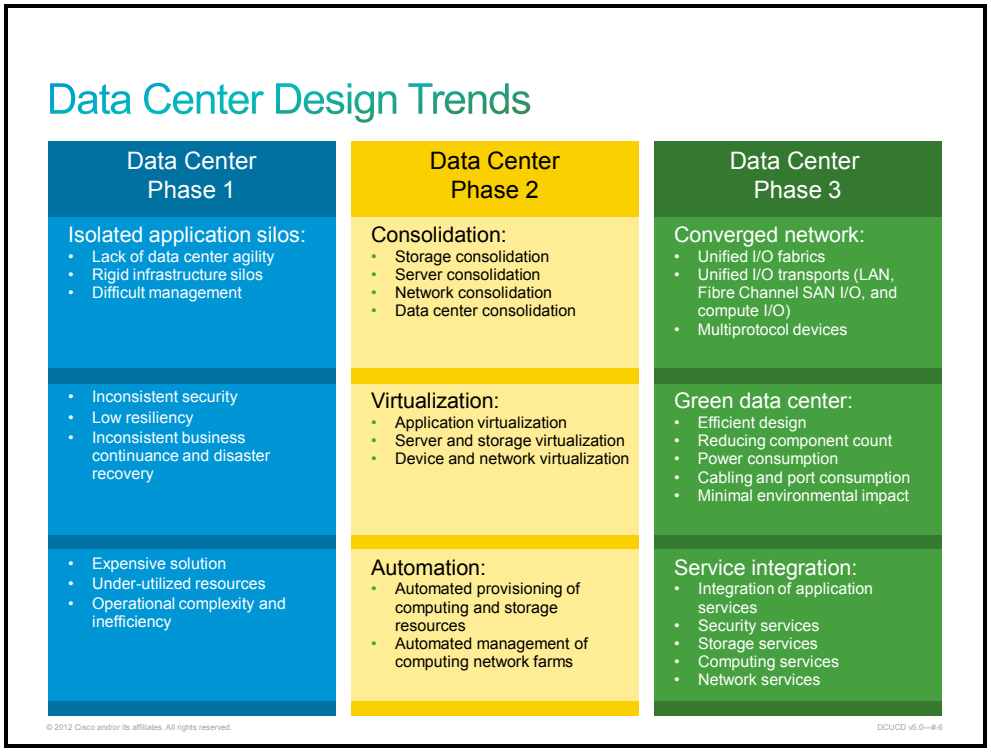
The data center capabilities can help influence growth and business. By enabling the ability to affect new service creation and faster application deployment through service profiling and rapid provision of resource pools, the business can enable service creation without spending on infrastructure, and provide increased service level agreements.

Cost cutting, margin, and efficiencies are all critical elements for businesses today in the current economic climate. When a business maintains focus on cutting costs, increasing margins through customer retention and satisfaction, and product brand awareness and loyalty, this will result in a higher return on investment (ROI). The data center works toward a robust, converged architecture to reduce costs. At the same time, the data center enhances application experience and increasing productivity through a scalable platform for collaboration tools.

The element of risk in a business must be minimized. While the business focuses on governing and monitoring changing compliance rules and a regulatory environment, it is also highly concerned with security of data, policy management, and access. The data center must ensure a consistent policy across services so that there is no compromise on the *quality* of service versus the *quantity* of service. Furthermore, the business needs the flexibility to implement and try new services quickly, while being sure they can retract them quickly if they prove unsuccessful, all with limited impact.

These areas show how the IT environment, and the data center in particular, can have a major impact on business.

https://t.me/learningnets



Data center trends that affect the data center architecture and design can be summarized by phases and stages:

- Phase 1:
  - Isolated application silos
- Phase 2:
  - Consolidation
  - Virtualization
  - Automation
- Phase 3:
  - Converged network
  - Green data center
  - Service integration

Phase 1

Isolated Application Silos

Data centers are about servers and applications. The first data centers were mostly mainframe, glass-house, raised-floor structures that housed the computer resources, as well as the intellectual capital (programmers and support staff) of the enterprise. Most data centers have evolved on an ad hoc basis. The goal was to provide the most appropriate server, storage, and networking infrastructure that supported specific applications. This strategy led to data centers with stovepipe architectures or technology islands that were difficult to manage or adapt to changing environments.

https://t.me/learningnets

There are many server platforms in current data centers, all designed to deploy a series of applications:

- IBM mainframe applications
- Email applications on Microsoft Windows servers
- Business applications on IBMAS/400 servers
- Enterprise resource planning (ERP) applications on UNIX servers
- R&D applications on Linux servers

In addition, a broad collection of storage silos exists to support these disparate server environments. These storage silos can be in the form of integrated, direct-attached storage (DAS), network-attached storage (NAS), or small SAN islands.

This silo approach has led to underutilization of resources, difficulty in managing these disparate complex environments, and difficulty in applying uniform services such as security and application optimization. It is also difficult to implement strong, consistent disaster-recovery procedures and business continuance functions.

Phase 2

Consolidation

Consolidation of storage, servers, and networks has enabled centralization of data center components, any-to-any access, simplified management, and technology convergence. Consolidation has also reduced the cost of data center deployment.

Virtualization

Virtualization is the creation of another abstraction layer that separates physical from logical characteristics and enables further automation of data center services. Almost any component of a data center can now be virtualized—storage, servers, networks, computing resources, file systems, file blocks, tape devices, and so on. Virtualization in a data center enables creation of huge resource pools, increased, efficient, and more flexible resource utilization, and automated resource provisioning, allocation, and assignment to applications. Virtualization represents the foundation for further automation of data center services.

Automation

Automation of data center services has been made possible by consolidating and virtualizing data center components. The advantages of data center automation are automated dynamic resource provisioning, automated Information Lifecycle Management (ILM), and automated data center management. Computing and networking resources can be automatically provisioned whenever needed. Other data center services can also be automated, such as data migration mirroring, and volume management functions. Monitoring data center resource utilization is a necessary condition for an automated data center environment.

Phase 3

Converged Network

Converged networks promise the unification of various networks and single all-purpose communication applications. Converged networks potentially lead to reduced IT cost and increased user productivity. Data center fabric is based on a unified I/O transport protocol, which could potentially transport SAN, LAN, WAN, and clustering I/Os. Most protocols tend to be transported across a common unified I/O channel and common hardware and software components of data center architecture.

https://t.me/learningnets

Green Data Center

A green data center is a data center in which the mechanical, lighting, electrical, and computer systems are designed for maximum energy efficiency and minimum environmental impact. The construction and operation of a green data center include advanced technologies and strategies, such as minimizing the footprints of the buildings; using low-emission building materials, sustainable landscaping, and waste recycling; and installing catalytic converters on backup generators. The use of alternative energy technologies such as photovoltaic, heat pumps, and evaporative cooling is also being considered.

Building and certifying a green data center or other facility can be initially expensive, but long-term cost savings can be realized on operations and maintenance. Another advantage is the fact that green facilities offer employees a healthy, comfortable work environment. There is growing pressure from environmentalists and, increasingly, from the general public, for governments to offer green incentives. This pressure is in terms of monetary support for the creation and maintenance of ecologically responsible technologies.

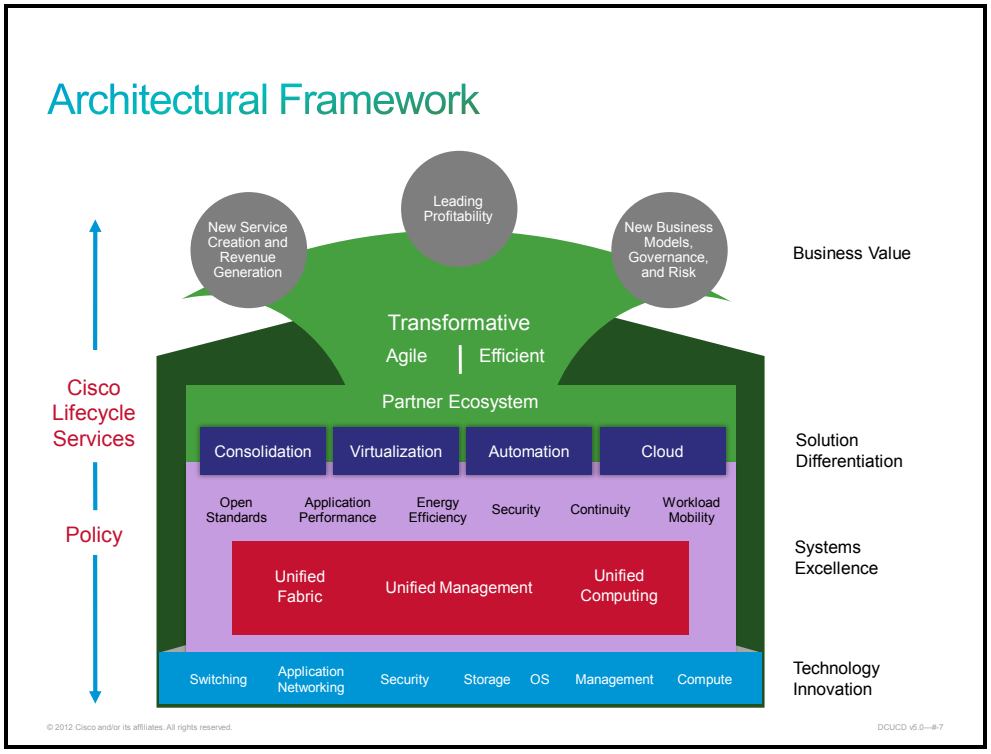
Today, green data centers provide an 85 percent power reduction using virtualized, integrated modules. Rack space is being saved with virtualized, integrated modules, and additional savings are being derived from reduced cabling, port consumption, and support cost.

Service Integration

The data center network infrastructure integrates network intelligence, including application services, security services, computing services, and storage services, among others.

© 2012 Cisco Systems, Inc

https://t.me/learningnets



Cisco Data Center is an architectural framework that connects technology innovation with business innovation. It is the foundation for the model of the dynamic networked organization and can enable the following key aspects:

- Quick and efficient innovation
- Control of data center architecture
- Freedom to choose technologies

The Cisco Data Center architectural framework is delivered as a portfolio of technologies and systems that can be adapted to meet organizational needs. You can adopt the framework in an incremental and granular fashion to control when and how you implement data center innovations. This framework allows you to easily evolve and adapt the data center to keep pace with changing organizational needs.

The Cisco approach to the data center is to provide an open and standards-based architecture. System-level benefits such as performance, energy efficiency, and resiliency are addressed, along with workload mobility and security. Cisco offers tested, preintegrated, and validated designs, providing businesses with a faster deployment model and time-to-market.

The components of the Cisco Data Center architecture are categorized into four areas: technology innovations, systems excellence, solution differentiation, and business advantage.

https://t.me/learningnets

Architectural Framework Benefits

- The architectural framework for dynamic networked organizations provides these benefits:
  - Reduced total cost of ownership
  - Accelerated business growth
  - Extension of the infrastructure life, making the data center more efficient, agile, and resilient
- Cost reduction and increased efficiency leverages
  - Infrastructure consolidation
  - Energy efficiency
  - Business continuance
  - Workforce productivity

© 2012 Cisco and/or its affiliates. All rights reserved.

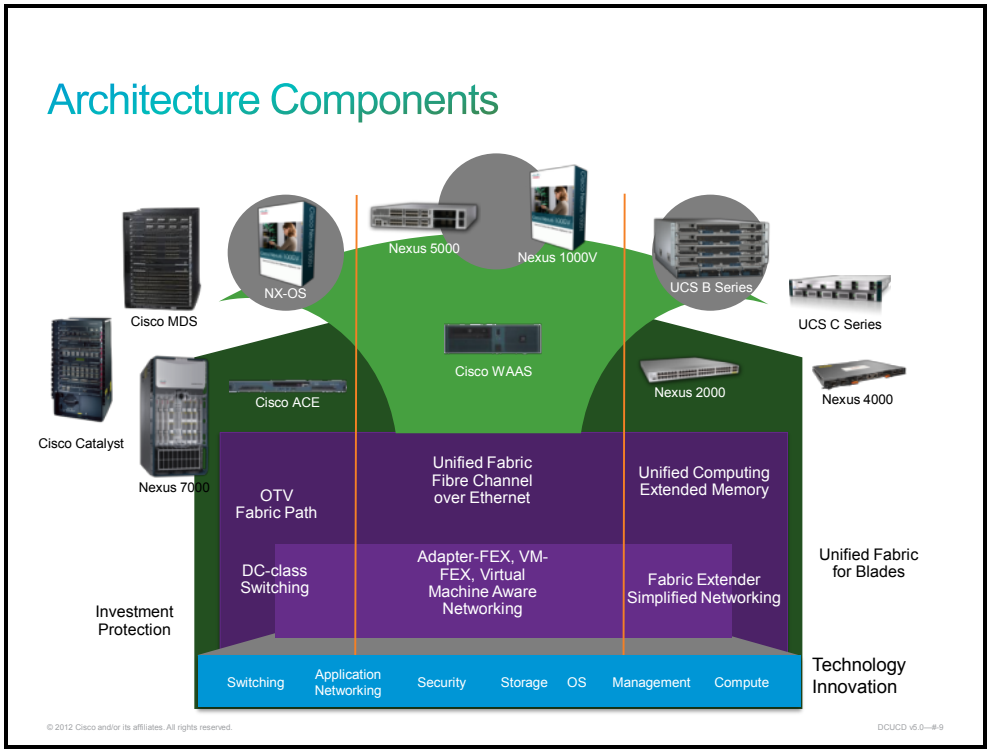
The Cisco Data Center framework is an architecture for dynamic networked organizations. The framework allows organizations to create services faster, improve profitability, and reduce the risk of implementing new business models. It can provide the following benefits:

- Delivering business value
- Providing flexibility and choice with an open ecosystem
- Providing innovative data center services

The Data Center is a portfolio of practical solutions that are designed to meet IT and business needs and can help to perform the following actions:

- Reduce total cost of ownership (TCO)
- Accelerate business growth
- Extend the life of the current infrastructure by making your data center more efficient, agile, and resilient

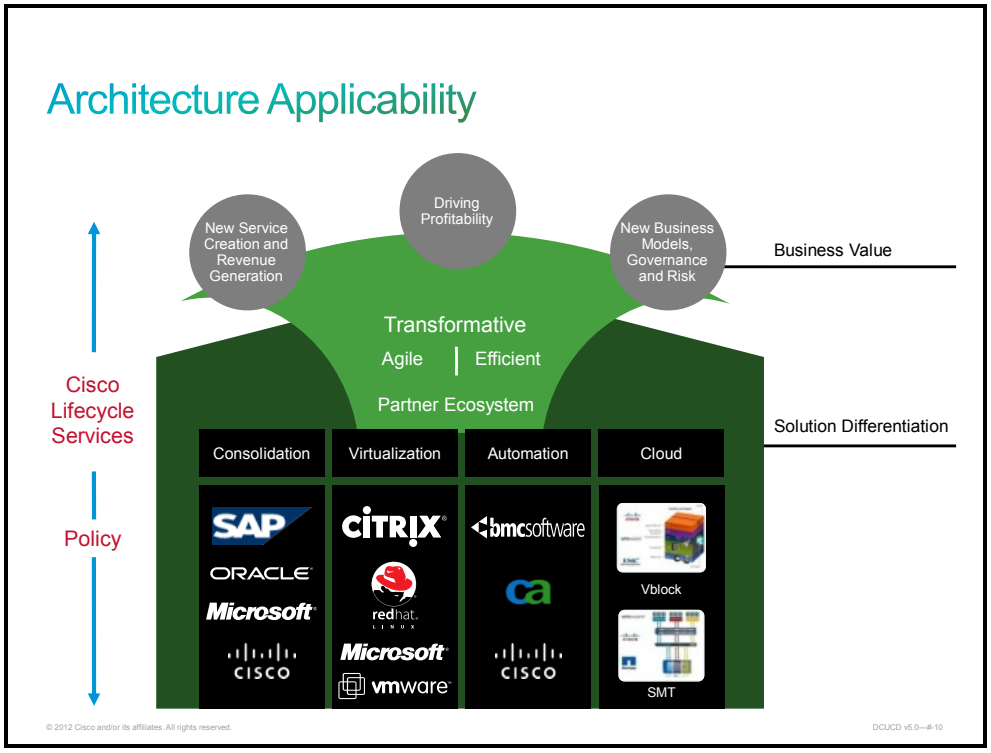
https://t.me/learningnets



The framework brings together multiple technologies:

- **Data center switching:** These next-generation virtualized data centers need a network infrastructure that delivers the complete potential of technologies such as server virtualization and unified fabric.
- **Storage networking solutions:** SANs are central to the Cisco Data Center architecture, and they provide a networking platform that helps IT departments to achieve a lower TCO, enhanced resiliency, and greater agility through Cisco Data Center storage solutions.
- **Cisco Application Networking Solutions:** You can improve application performance, availability, and security, while simplifying your data center and branch infrastructure. Cisco Application Networking Services (ANS) solutions can help you lower your TCO and improve IT flexibility.
- **Data center security:** Cisco Data Center security solutions enable you to create a trusted data center infrastructure that is based on a systems approach and that is using industry-leading security solutions.
- **Cisco Unified Computing System (UCS):** You can improve IT responsiveness to rapidly changing business demands with the Cisco Unified Computing System. This next-generation data center platform accelerates the delivery of new services simply, reliably, and securely through end-to-end provisioning and migration support.
- **Cisco Virtualization Experience Infrastructure (VXI):** Cisco VXI can deliver a superior collaboration and rich-media user experience with a best-in-class return on investment (ROI) in a fully integrated, open, and validated desktop virtualization solution.

https://t.me/learningnets



The architectural framework encompasses the network, storage, application services, security, and compute equipment.

The architectural framework is open and can integrate with other vendor solutions and products, such as VMware vSphere, VMware View, Microsoft Hyper-V, Citrix XenServer, and Citrix XenDesktop.

Starting from the top down, virtual machines (VMs) are one of the key components of the framework. VMs are entities that run an application within the client operating system, which is further virtualized and running on common hardware.

The logical server personality is defined by using management software, and it defines the properties of the server: the amount of memory, percentage of total computing power, number of network interface cards, boot image, and so on.

The network hardware for consolidated connectivity serves as one of the key technologies for fabric unification.

VLANs and VSANs provide for virtualized LAN and SAN connectivity, separating physical networks and equipment into virtual entities.

On the lowest layer of the framework is the virtualized hardware. Storage devices can be virtualized into storage pools, and network devices are virtualized by using device contexts.

The Cisco Data Center switching portfolio is built on the following common principles:

- **Design flexibility:** Modular, rack, and integrated blade switches are optimized for both Gigabit Ethernet and 10 Gigabit Ethernet environments.
- **Industry-leading switching capabilities:** Layer 2 and Layer 3 functions can build stable, secure, and scalable data centers.
- **Investment protection:** The adaptability of the Cisco Nexus and Catalyst families simplifies capacity and capability upgrades.
- **Operational consistency:** A consistent interface and consistent tools simplify management, operations, and trouble resolutions.

https://t.me/learningnets

- **Virtualization:** Cisco Data Center switches provide VM mobility support, management, and operations tools for a virtualized environment. In addition, the Cisco Catalyst 6500 Switch offers Virtual Switching System (VSS) capabilities, and the Cisco Nexus 7000 Switch offers hypervisor-like virtual switch capabilities in the form of virtual device contexts (VDCs).

The Cisco storage solution provides the following:

- **Multiprotocol storage networking:** By providing flexible options for Fibre Channel, fiber connectivity (FICON), Fibre Channel over Ethernet (FCoE), Internet Small Computer Systems Interface (iSCSI), and Fibre Channel over IP (FCIP), any business risk is reduced.
- **A unified operating system and management tools:** Operational simplicity, simple interoperability, and feature consistency can reduce operating expenses.
- **Enterprise-class storage connectivity:** Significantly larger virtualized workloads can be supported, providing availability, scalability, and performance.
- **Services-oriented SANs:** The "any network service to any device" model can be extended, regardless of protocol, speed, vendor, or location.

Cisco ANS provides the following attributes:

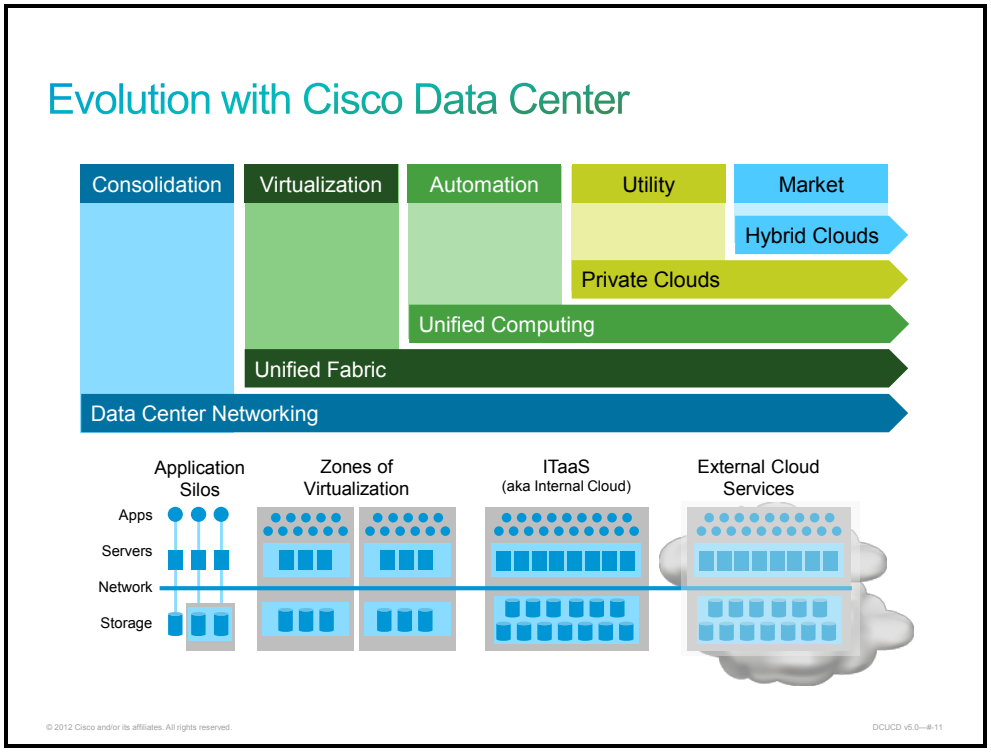
- **Application intelligence:** You can take control of applications and the user experience.
- **Cisco Unified Network Services:** You can connect any person to any resource with any device.
- **Integrated security:** There is built-in protection for access, identity, and data.
- **Nonstop communications:** Users can stay connected with a resilient infrastructure that enables business continuity.
- **Virtualization:** This feature allows simplification of the network and the ability to maximize resource utilization.
- **Operational manageability:** You can deploy services faster and automate routine tasks.

The Cisco Data Center security solutions enable businesses to create a trusted data center infrastructure that is based on a systems approach and industry-leading security solutions. These solutions enable the rapid deployment of data center technologies without compromising the ability to identify and respond to evolving threats, protect critical assets, and enforce business policies.

The Cisco UCS provides the following benefits:

- Streamlining of data center resources to reduce TCO
- Scaling of the service delivery to increase business agility
- Reducing the number of devices that require setup, management, power, cooling, and cabling

https://t.me/learningnets



The Cisco Data Center architecture brings sequential and stepwise clarity to the data center. Data center networking capabilities bring an open and extensible data center networking approach to the placement of the IT solution. This approach supports business processes, whether the processes are conducted on a factory floor in a pod or in a Tier 3 data center 200 feet (61 meters) below the ground in order to support precious metal mining. In short, it delivers location freedom.

Phase 1: Data Center Networking and Consolidation

The delivery of data center transport and network services—server networking, storage networking, application networking, and security networking—as an integrated system improves reliability and architectures.

Phase 2: Unified Fabric

The introduction of a unified fabric is the first step towards removing the network barriers to deploying any workload on any server in seconds.

The building block of the data center is not the physical server but is the VM, which provides a consistent abstraction between physical hardware and logically defined software constructs.

Phase 3: Unified Computing and Automation

Bringing the network platform that is created by the unified fabric together with the virtualization platform and the compute and storage platform introduces a new paradigm in the Cisco Data Center. Thus the architectural solution becomes simpler, more efficient, and extends the lifecycle of capital assets, but also enables the enterprise to execute business processes in the best places and in the most efficient ways, all with high availability. Cisco Unified Computing focuses on automation simplification for a predesigned virtualization platform.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Phase 4: Enterprise-Class Clouds and Utility

With the integration of cloud technologies, principles, and architectures and the Cisco Unified Computing architecture, workloads can become increasingly portable. Bringing security, control, and interoperability to standalone cloud architectures can enable enterprise-class clouds. The freedom of choice about where business processes are executed is extended across the boundaries of an organization, to include the providers as well (with no compromise on service levels).

After the process is automated, enterprise internal IT resources will be seen as a utility that is able to automate and dynamically provision the infrastructure across the network, compute, and virtualization platforms, thus simplifying the line of business and services in the enterprise.

It is an evolution that is enabled by integrating cloud computing principles, architectures, and technologies.

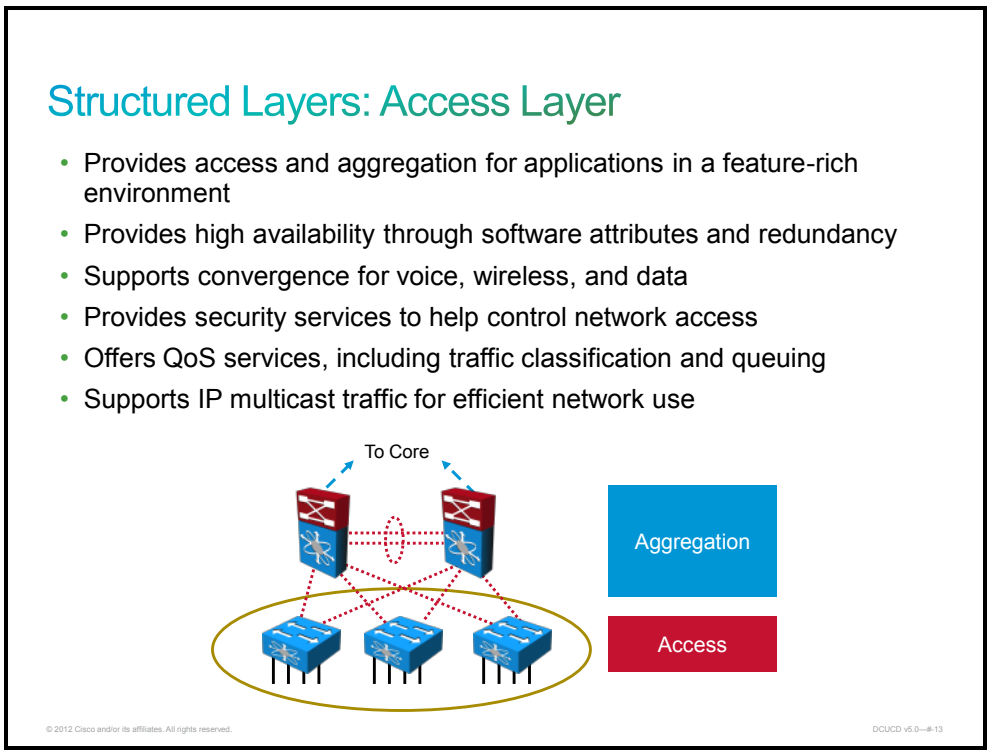
Phase 5: Inter-Cloud Solutions

The inter-cloud solutions extend from the enterprise to the provider and from the provider to another provider, based on available capacity, power cost, proximity, and so on.

https://t.me/learningnets

Cisco Data Center Architecture Unified Fabric

This topic describes the Cisco Data Center architectural framework unified fabric component.



The architectural components of the infrastructure are the access layer, the aggregation layer, and the core layer. The principal advantages of this model are its hierarchical structure and its modularity. A hierarchical design avoids the need for a fully meshed network in which all network nodes are interconnected. Modules in a layer can be put into service and taken out of service without affecting the rest of the network. This ability facilitates troubleshooting, problem isolation, and network management.

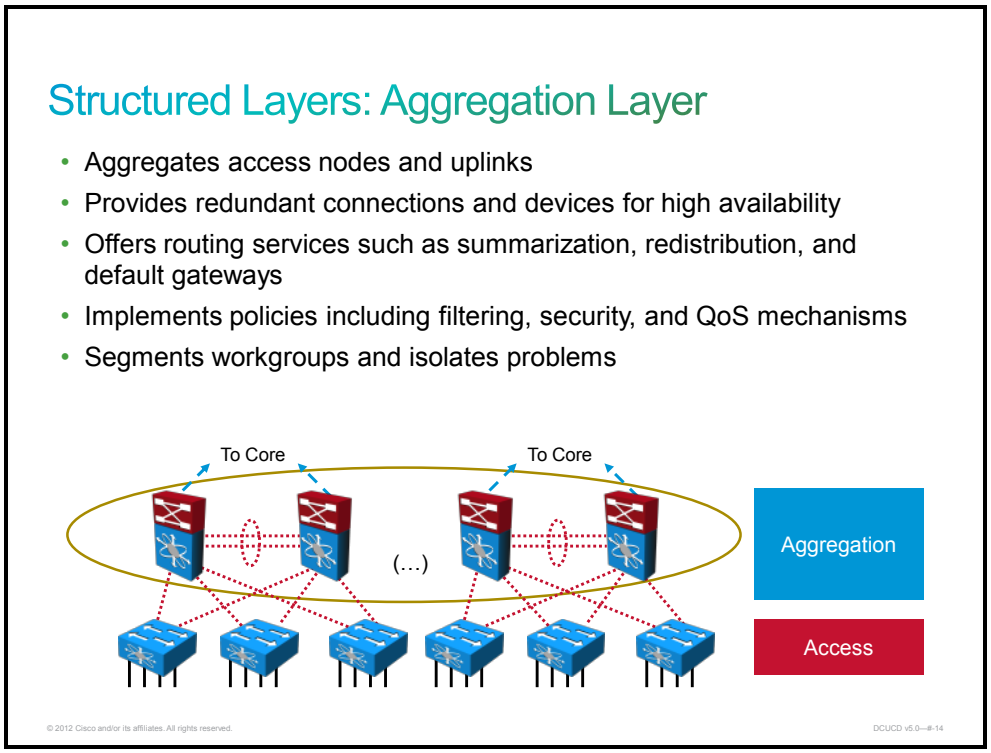
The access layer aggregates end users and provides uplinks to the aggregation layer. The access layer can be a feature-rich environment:

- **High availability:** The access layer is supported by many hardware and software attributes. This layer offers system-level redundancy by using redundant supervisor engines and redundant power supplies for crucial application groups. The layer also offers default gateway redundancy by using dual connections from access switches to redundant aggregation layer switches that use a First Hop Redundancy Protocol (FHRP), such as Hot Standby Router Protocol (HSRP).
- **Convergence:** The access layer supports inline Power over Ethernet (PoE) for IP telephony and wireless access points. This support allows customers to converge voice onto their data networks and provides roaming wireless LAN (WLAN) access for users.
- **Security:** The access layer provides services for additional security against unauthorized access to the network. This security is provided by using tools such as IEEE 802.1X, port security, DHCP snooping, Dynamic Address Resolution Protocol (ARP) Inspection (DAI), and IP Source Guard.
- **Quality of service (QoS):** The access layer allows prioritization of mission-critical network traffic by using traffic classification and queuing as close to the ingress of the network as possible. The layer supports the QoS trust boundary.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

- **IP multicast:** The access layer supports efficient network and bandwidth management by using software features such as Internet Group Management Protocol (IGMP) snooping.



Availability, load balancing, QoS, and provisioning are the important considerations at the aggregation layer. High availability is typically provided through dual paths from the aggregation layer to the core and from the access layer to the aggregation layer. Layer 3 equal-cost load sharing allows uplinks from the aggregation to the core layer to be used.

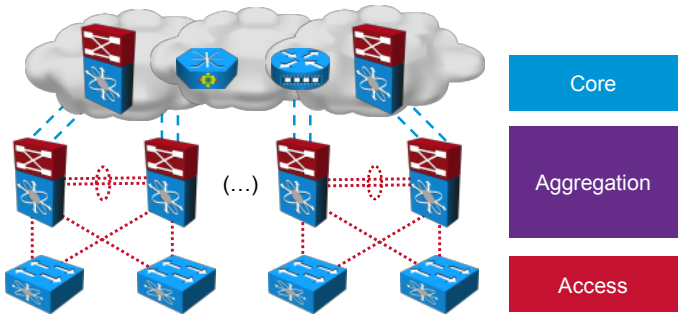
The aggregation layer is the layer in which routing and packet manipulation is performed and can be a routing boundary between the access and core layers. The aggregation layer represents a redistribution point between routing domains or the demarcation between static and dynamic routing protocols. This layer performs tasks such as controlled-routing decision-making and filtering to implement policy-based connectivity and QoS. To further improve routing protocol performance, the aggregation layer summarizes routes from the access layer. For some networks, the aggregation layer offers a default route to access layer routers and runs dynamic routing protocols when communicating with core routers.

The aggregation layer uses a combination of Layer 2 and multilayer switching to segment workgroups and to isolate network problems, so that they do not affect the core layer. This layer is commonly used to terminate VLANs from access layer switches. The aggregation layer also connects network services to the access layer and implements policies regarding QoS, security, traffic loading, and routing. In addition, this layer provides default gateway redundancy by using an FHRP such as HSRP, Gateway Load Balancing Protocol (GLBP), or Virtual Router Redundancy Protocol (VRRP).Default gateway redundancy allows for the failure or removal of one of the aggregation nodes without affecting endpoint connectivity to the default gateway.

https://t.me/learningnets

Structured Layers: Core Layer

- The core layer is a high-speed backbone and aggregation point for the enterprise.
- It provides reliability through redundancy and fast convergence.
- Separate core layer helps in scalability during future growth.

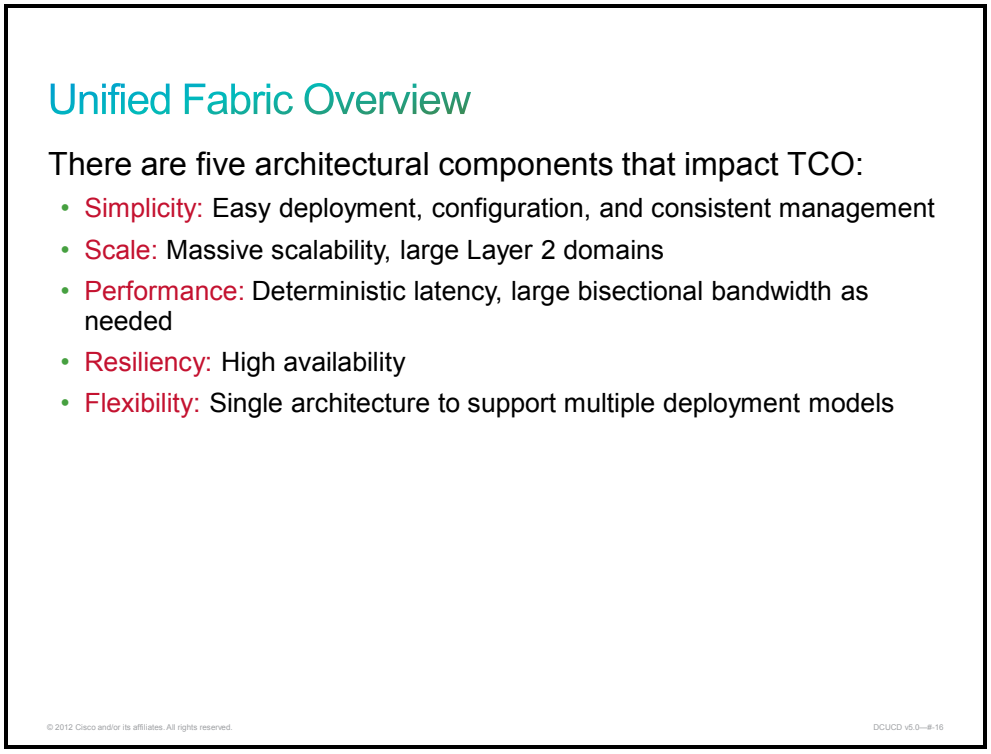


© 2012 Cisco and/or its affiliates. All rights reserved.

The core layer is the backbone for connectivity and is the aggregation point for the other layers and modules in the Cisco Data Center architecture. The core must provide a high level of redundancy and must adapt to changes very quickly. Core devices are most reliable when they can accommodate failures by rerouting traffic and can respond quickly to changes in the network topology. The core devices must be able to implement scalable protocols and technologies, alternate paths, and load balancing. The core layer helps in scalability during future growth.

The core should be a high-speed Layer 3 switching environment that uses hardware-accelerated services. For fast convergence around a link or node failure, the core uses redundant point-to-point Layer 3 interconnections in the core. That type of design yields the fastest and most deterministic convergence results. The core layer should not perform any packet manipulation, such as checking access lists and filtering, which would slow down the switching of packets.

https://t.me/learningnets



Cisco Unified Fabric delivers transparent convergence, massive three-dimensional scalability, and sophisticated intelligent services to provide the following benefits:

- Support for traditional and virtualized data centers
- Reduction in TCO
- An increase in ROI

The five architectural components that impact TCO include the following:

- **Simplicity:** Businesses require the data center to be able to provide easy deployment, configuration, and consistent management of existing and new services.
- **Scale:** Data centers need to be able to support large Layer 2 domains that can provide massive scalability without the loss of bandwidth and throughput.
- **Performance:** Data centers should be able to provide deterministic latency and large bisectional bandwidth to applications and services as needed.
- **Resiliency:** The data center infrastructure and implemented features need to provide high availability to the applications and services that they support.
- **Flexibility:** A single architecture that can support multiple deployment models to provide the flexible component of the architecture.

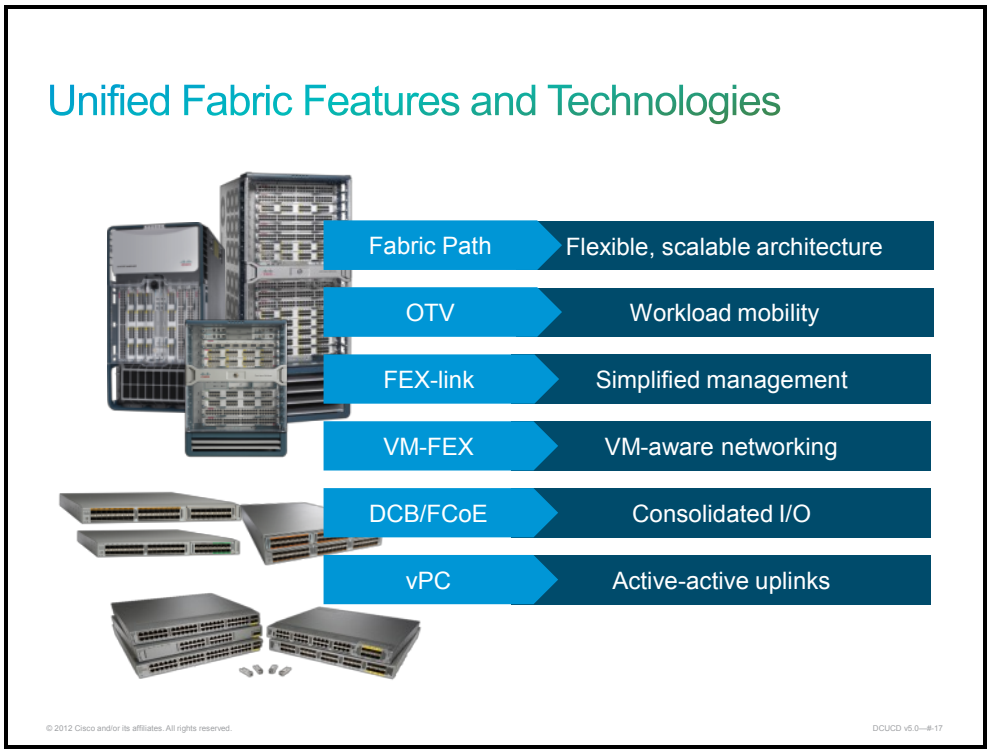
Universal I/O brings efficiency to the data center through “wire-once” deployment and protocol simplification. This efficiency, in the Cisco WebEx data center, has shown the ability to increase workload density by 30 percent in a flat power budget. In a 30-megawatt (MW) data center, this increase accounts for an annual $60 million cost deferral. Unified fabric technology enables a “wire-once” infrastructure in which there are no physical barriers in the network to redeploying applications or capacity, thus delivering hardware freedom.

The main advantage of Cisco Unified Fabric is that it offers LAN and SAN infrastructure consolidation. It is no longer necessary to plan for and maintain two completely separate

https://t.me/learningnets

infrastructures. The network comes in as a central component to the evolution of the virtualized data center and to the enablement of cloud computing.

Cisco Unified Fabric offers a low-latency and lossless connectivity solution that is fully virtualization-enabled. Unified Fabric offers you a massive reduction of cables, adapters, switches, and pass-through modules.



The Cisco Unified Fabric is a foundational pillar for the Cisco Data Center Business Advantage architectural framework. Cisco Unified Fabric complements Unified Network Services and Unified Computing to enable IT and business innovation.

- Cisco Unified Fabric convergence offers the best of both SANs and LANs by enabling users to take advantage of Ethernet economy of scale, an extensive vendor community, and future innovations.
- Cisco Unified Fabric scalability delivers performance, ports, bandwidth, and geographic span.
- Cisco Unified Fabric intelligence embeds critical policy-based intelligent functionality into the Unified Fabric for both traditional and virtualized data centers.

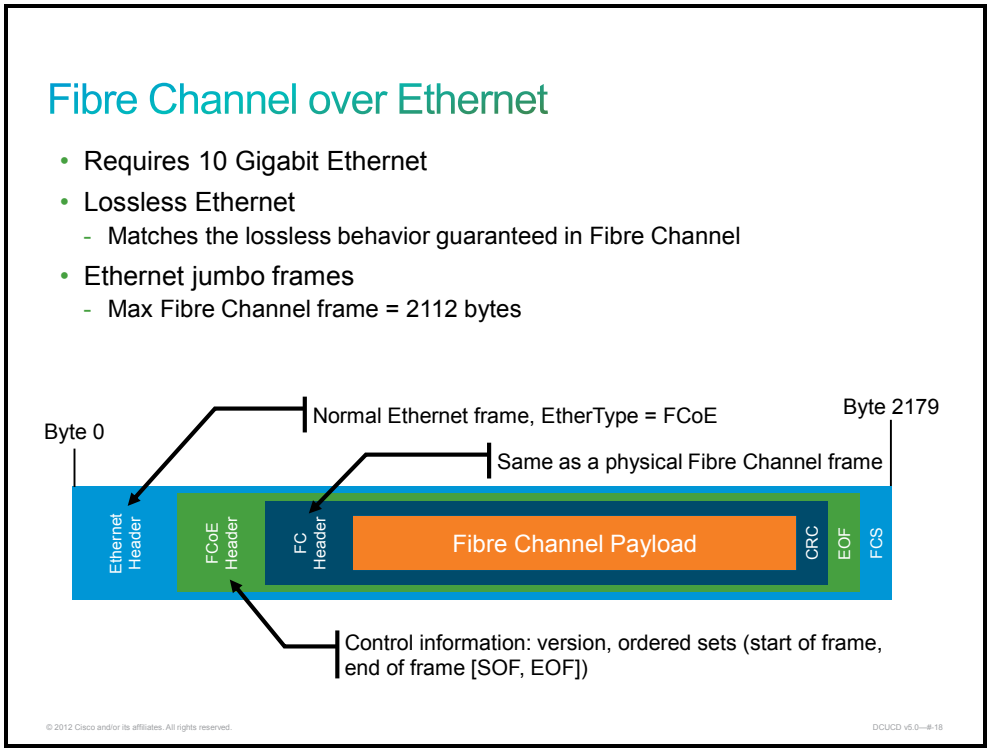
To support the five architectural attributes, the Cisco Unified Fabric evolution continues to provide architectural innovations.

- **Cisco FabricPath:** Cisco FabricPath is a set of capabilities within the Cisco Nexus Operating System (NX-OS) Software combining the “plug-and-play” simplicity of Ethernet with the reliability and scalability of Layer 3 routing. Cisco FabricPath enables companies to build highly scalable Layer 2 multipath networks without the Spanning Tree Protocol (STP). These networks are particularly suitable for large virtualization deployments, private clouds, and high-performance computing environments.

https://t.me/learningnets

- **Cisco Overlay Transport Virtualization (OTV):** Cisco OTV is an industry-first solution that significantly simplifies extending Layer 2 applications across distributed data centers. Cisco OTV allows companies to deploy virtual computing resources and clusters across geographically distributed data centers, delivering transparent workload mobility, business resiliency, and superior computing resource efficiencies.
- **Cisco FEX-Link:** Cisco FEX-Link technology enables data center architects to gain new design flexibility while simplifying cabling infrastructure and management complexity. Cisco FEX-Link uses the Cisco Nexus 2000 Series Fabric Extenders (FEXs) to extend the capacities and benefits that are offered by upstream Cisco Nexus switches.
- **Cisco VM-FEX:** Cisco VM-FEX provides advanced hypervisor switching as well as high-performance hardware switching. It is flexible, extensible, and service-enabled. Cisco VM-FEX architecture provides virtualization-aware networking and policy control.
- **Data Center Bridging (DCB) and FCoE:** Cisco Unified Fabric provides the flexibility to run Fibre Channel, IP-based storage such as network-attached storage (NAS) and Small Computer System Interface over IP, or FCoE, or a combination of these technologies, on a converged network.
- **vPC:** Cisco virtual PortChannel (vPC) technology enables the deployment of a link aggregation from a generic downstream network device to two individual and independent Cisco NX-OS devices (vPC peers). This multichassis link aggregation path provides both link redundancy and active-active link throughput, scaling high-performance failover characteristics.

https://t.me/learningnets



FCoE is a new protocol that is based upon the Fibre Channel layers that are defined by the ANSI T11 committee, and it replaces the lower layers of Fibre Channel traffic. FCoE addresses the following:

- **Jumbo frames:** An entire Fibre Channel frame (2180 bytes in length) can be carried in the payload of a single Ethernet frame.
- **Fibre Channel port:** World wide name (WWN) addresses are encapsulated in the Ethernet frames and MAC addresses are used for traffic forwarding in the converged network.
- **FCoE Initialization Protocol (FIP):** This protocol provides a login for Fibre Channel devices into the fabric.
- **Quality of service (QoS) assurance:** This ability monitors the Fibre Channel traffic with respect to lossless delivery of Fibre Channel frames and bandwidth reservations for Fibre Channel traffic.
- **Minimum 10-Gb/s Ethernet platform**

FCoE traffic consists of a Fibre Channel frame that is encapsulated within an Ethernet frame. The Fibre Channel frame payload may in turn carry SCSI messages and data or, in the future, may use FICON for mainframe traffic.

FCoE is an extension of Fibre Channel (and its operating model) onto a lossless Ethernet fabric. FCoE requires 10 Gigabit Ethernet and maintains the Fibre Channel operation model, which provides seamless connectivity between two networks.

FCoE positions Fibre Channel as the storage networking protocol of choice and extends the reach of Fibre Channel throughout the data center to all servers. Fibre Channel frames are encapsulated into Ethernet frames with no fragmentation, which eliminates the need for higher-level protocols to reassemble packets.

© 2012 Cisco Systems, Inc
https://t.me/learningnets

Fibre Channel overcomes the distance and switching limitations that are inherent in SCSI. Fibre Channel carries SCSI as its higher-level protocol. SCSI does not respond well to lost frames, which can result in significant delays when recovering from a loss. Because Fibre Channel carries SCSI, it inherits the requirement for an underlying lossless network.

FCoE transports native Fibre Channel frames over an Ethernet infrastructure, which allows existing Fibre Channel management modes to stay intact. One FCoE prerequisite is for the underlying network fabric to be lossless.

Frame size is a factor in FCoE. A typical Fibre Channel data frame has a 2112-byte payload, a header, and a frame check sequence (FCS). A classic Ethernet frame is typically 1.5 KB or less. To maintain good performance, FCoE must utilize jumbo frames (or the 2.5-KB "baby jumbo") to prevent a Fibre Channel frame from being split into two Ethernet frames.

Data Center Bridging

DCB provides a lossless data center transport layer that enables the convergence of LANs and SANs onto a single Unified Fabric.

Protocol	IEEE Standard	Description
Priority-based flow control (PFC)	802.1Qbb	Provides lossless delivery for selected classes of service (CoS)
Enhanced transmission selection (ETS)	802.1Qaz	Provides bandwidth management and priority selection
Data Center Bridging Exchange Protocol (DCBX)	802.1AB	Protocol that exchanges parameters between DCB devices and that leverages functions provided by LLDP

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Unified Fabric is a network that can transport many different protocols, such as LAN, SAN, and high-performance computing (HPC) protocols, over the same physical network.

10 Gigabit Ethernet is the basis for a new DCB protocol that, through enhanced features, provides a common platform for lossy and lossless protocols that carry LAN, SAN, and HPC data.

The IEEE 802.1 DCB is a collection of standards-based extensions to Ethernet and it can enable a Converged Enhanced Ethernet (CEE). It provides a lossless data center transport layer that enables the convergence of LANs and SANs onto a single unified fabric. In addition to supporting FCoE, DCB enhances the operation of iSCSI, NAS, and other business-critical traffic.

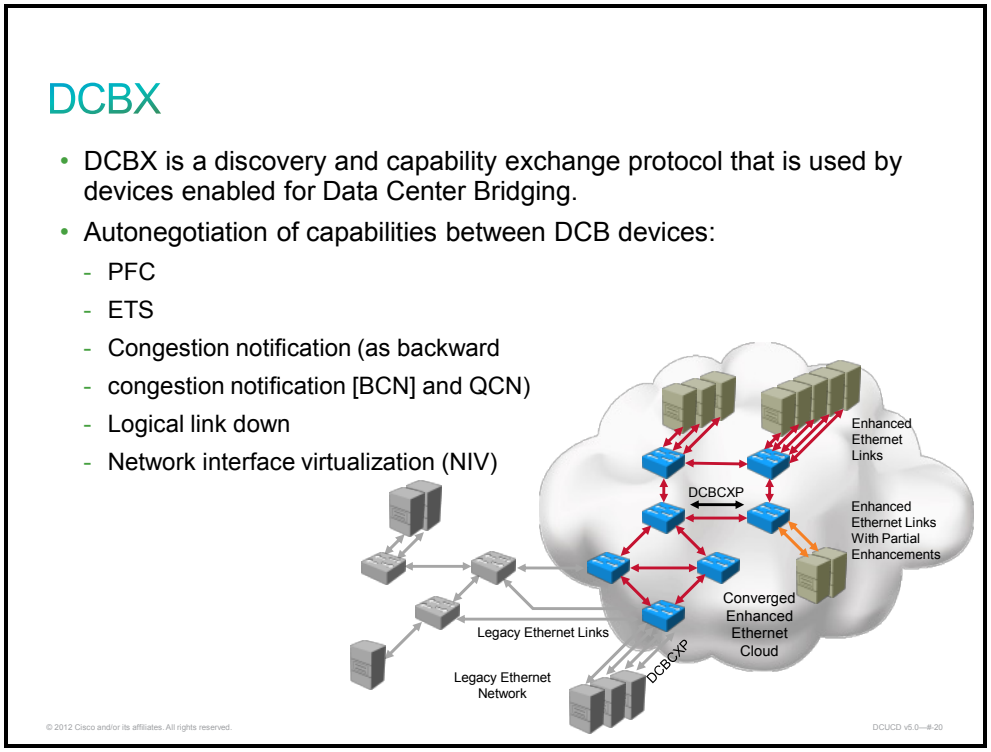
- **Priority flow control (PFC):** Provides lossless delivery for selected classes of service (CoS) (802.1Qbb)
- **Enhanced transmission selection (ETS):** Provides bandwidth management and priority selection (802.1Qaz)

https://t.me/learningnets

- **Data Center Bridging Exchange Protocol (DCBX):** Exchanges parameters between DCB devices and leverages functions that are provided by Link Layer Discovery Protocol (LLDP) (802.1AB)

Different organizations have created different names to identify the specifications. IEEE has used the term Data Center Bridging, or DCB. IEEE typically calls a standard specification by a number, such as 802.1az. IEEE did not have a way to identify the group of specifications with a standard number, so the organization grouped the specifications into DCB.

The term Converged Enhanced Ethernet was created by IBM to reflect the core group of specifications and to gain consensus among industry vendors (including Cisco) as to what a Version 0 list of the specifications would be before they all become standards.



The DCBX protocol allows each DCB device to communicate with other devices and to exchange capabilities within a unified fabric. Without DCBX, each device would not know if it could send lossless protocols such as FCoE to another device that was not capable of dealing with lossless delivery.

DCBX is a discovery and capability exchange protocol that is used by devices that are enabled for Data Center Ethernet to exchange configuration information. The following parameters of the Data Center Ethernet features can be exchanged:

- Priority groups in ETS
- PFC
- Congestion notification (as backward congestion notification [BCN] or as quantized congestion notification [QCN])
- Application types and capabilities

https://t.me/learningnets

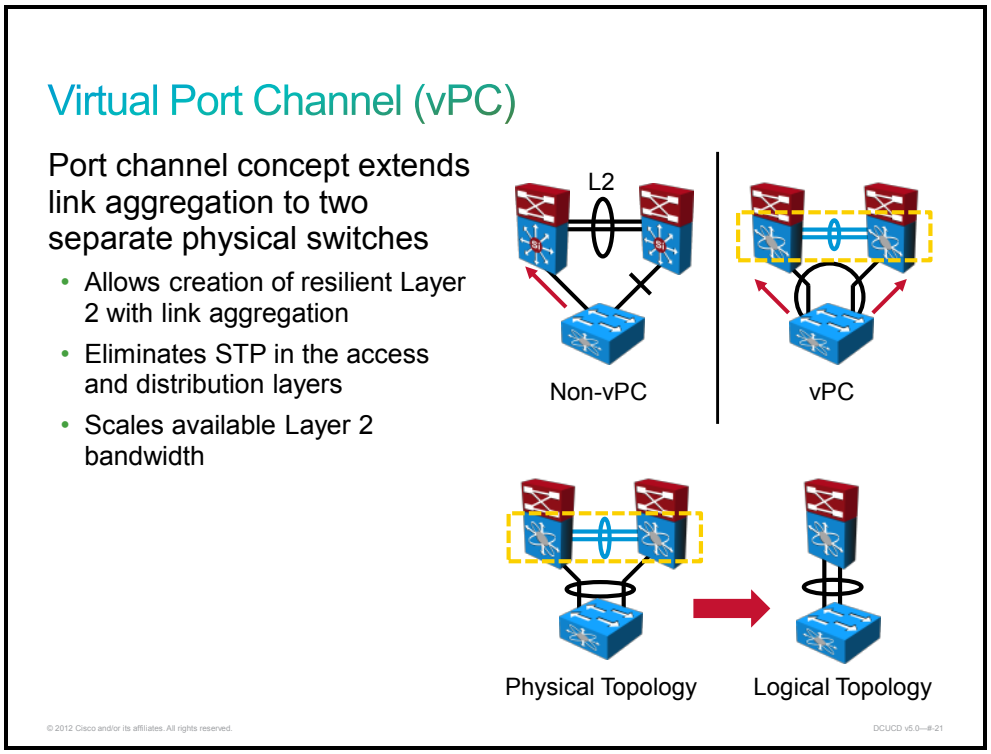
- Logical link down to signify the loss of a logical connection between devices, even though the physical link is still up
- Network interface virtualization (NIV)

Note See http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf for more details.

Devices need to discover the edge of the enhanced Ethernet cloud:

- Each edge switch needs to learn that it is connected to a legacy switch.
- Servers need to learn whether they are connected to enhanced Ethernet devices.
- Within the enhanced Ethernet cloud, devices need to discover the capabilities of peers.
- The Data Center Bridging Capability Exchange Protocol (DCBCXP) utilizes the LLDP and processes the local operational configuration for each feature.
- Link partners can choose supported features and can accept configurations from peers.

Note Details on DCBCXP can be found at http://www.intel.com/technology/eedc/index.htm.



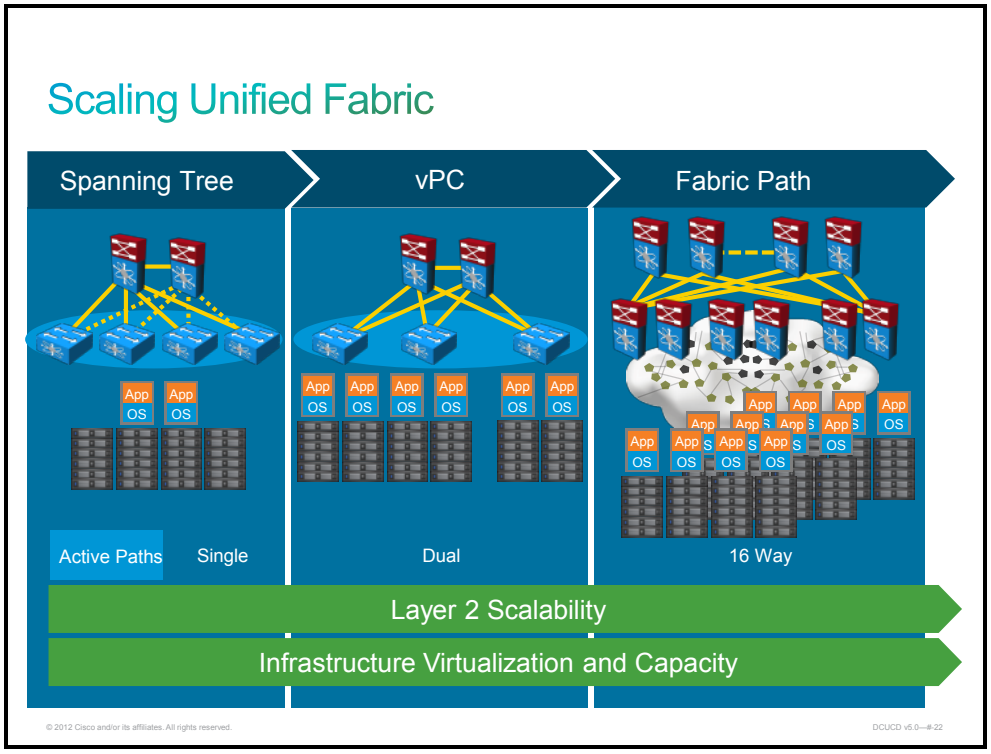
A virtual port channel (vPC) allows links that are physically connected to two different Cisco Nexus 5000 or 7000 Series devices to appear as a single port channel to a third device. The third device can be a Cisco Nexus 2000 Series FEX or a switch, server, or any other networking device.

A vPC can provide Layer 2 multipathing, which allows you to create redundancy by increasing bandwidth, enabling multiple parallel paths between nodes, and load balancing traffic where alternative paths exist.

https://t.me/learningnets

A vPC provides the following benefits:

- Allows a single device to use a port channel across two upstream devices
- Eliminates STP blocked ports
- Provides a loop-free topology
- Uses all available uplink bandwidth
- Provides fast convergence if either the link or a device fails
- Provides link-level resiliency
- Helps ensure high availability



Cisco FabricPath brings the stability and performance of Layer 3 routing to Layer 2 switched networks to build a highly resilient and scalable Layer 2 fabric. It is a foundation for building massively scalable and flexible data centers.

Cisco FabricPath addresses the challenges in current network design where data centers still include some form of Layer 2 switching. The challenges exist due to requirements set by certain solutions that expect Layer 2 connectivity, and also because of the administrative overhead and the lack of flexibility that IP addressing introduces. Setting up a server in a data center requires planning and implies the coordination of several independent teams: network team, server team, application team, storage team, and so on.

In a routed network, moving the location of a host requires changing its address, and because some applications identify servers by their IP addresses, changing the location of a server is basically equivalent to starting the server installation process all over again.

https://t.me/learningnets

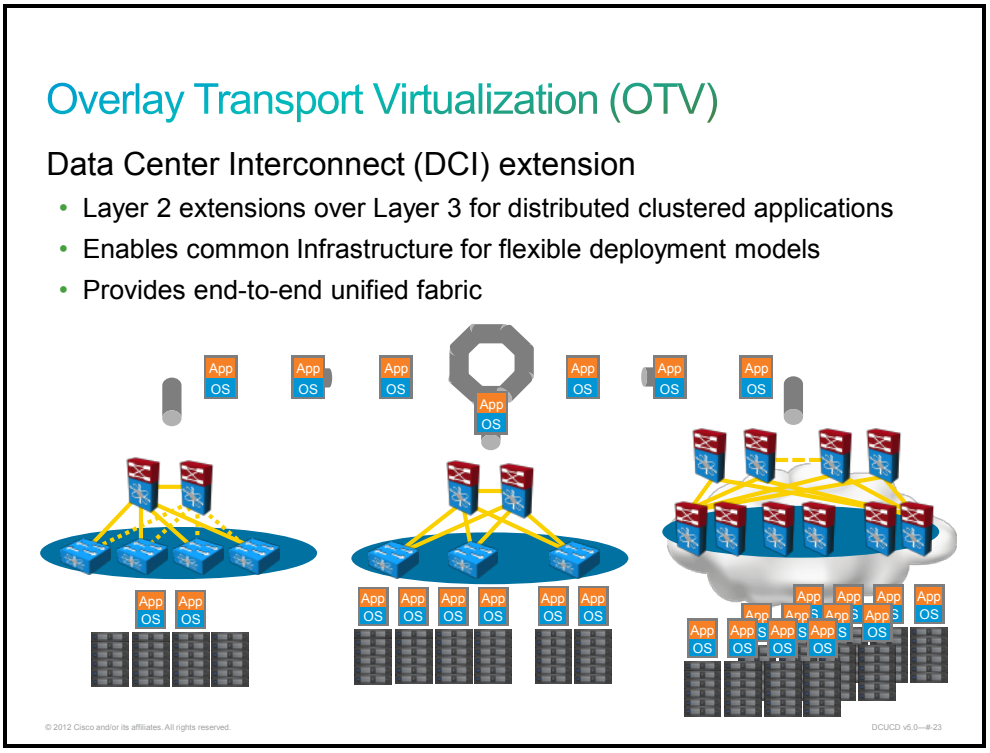
Layer 2 introduces flexibility by allowing the insertion or movement of a device in a transparent way from the perspective of the IP layer.

Virtualization technologies increase the density of managed virtual servers in the data center, making better use of the physical resources but also increasing the need for flexible Layer 2 networking.

Key functionalities of FabricPath are as follows:

- Eliminates spanning tree limitations
- Enables multipathing across all links with high cross-sectional bandwidth
- Provides high resiliency and faster network reconvergence
- Enables use of any VLAN anywhere in the fabric, and thus eliminates VLAN scoping

https://t.me/learningnets



Cisco Overlay Transport Virtualization (OTV) enables a Layer 2 extension (that is, Data Center Interconnect [DCI]), over Layer 3 connectivity to provide end-to-end unified fabric and a common infrastructure for flexible deployment models.

Cisco OTV has the following key characteristics:

- Enables Ethernet LAN extension over any network
  - Works over dark fiber, Multiprotocol Label Switching (MPLS), or IP
  - Can scale over multiple data centers
- Simplified configuration and operation
  - Seamless overlay—it does not require network redesign
  - Quick implementation with single-touch site configuration
- High resiliency
  - Isolates failure domains
  - Provides seamless multihoming
- Maximizes available bandwidth
  - Automated multipathing
  - Optimal multicast replication

Data Center Interconnect

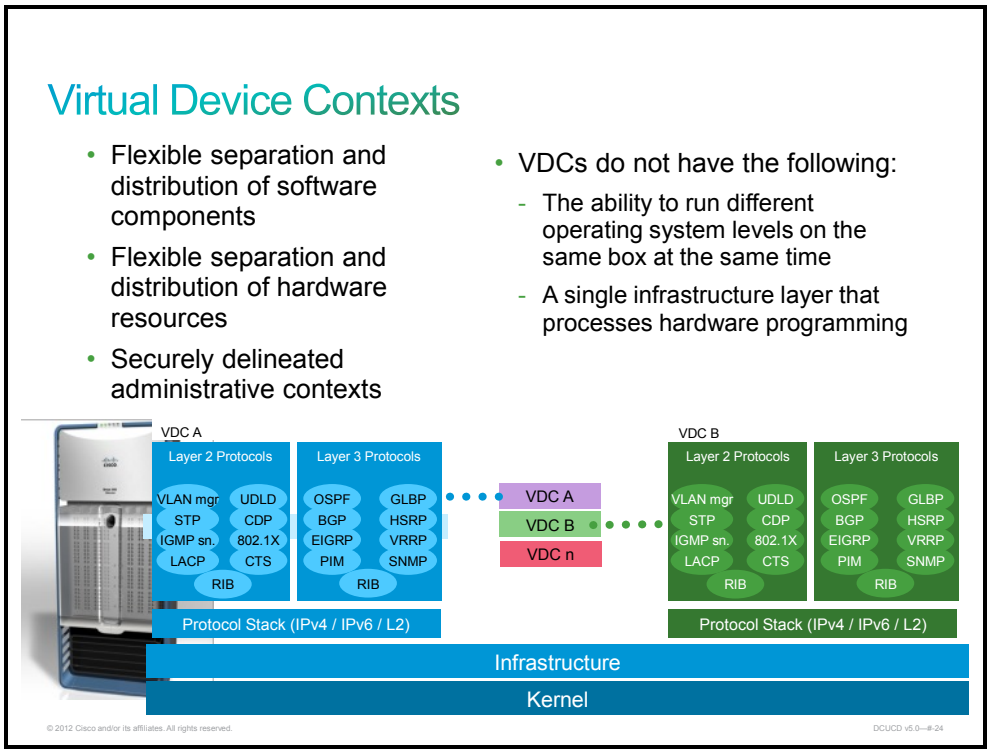
Data Center Interconnect (DCI) solutions are used to provide Data Center connectivity extensions over various connecting technologies.

DCI is most commonly required in the following situations:

- **Network and server virtualization:** Server virtualization is a baseline requirement used to prepare VMs for application mobility, with network virtualization as a baseline requirement to enable virtual network connectivity.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

- **LAN extensions:** LAN extensions are used to extend same VLANs across data centers to enable Layer 2 connectivity between VMs or clustered hosts.
- **Storage extensions:** Storage extensions are used to provide applications with access to local storage, and remote storage with correct storage attributes.
- **Path optimization:** Path optimization is used to route users to the data center where the application resides, while keeping symmetrical routing for IP services (for example, firewall and load balancer).

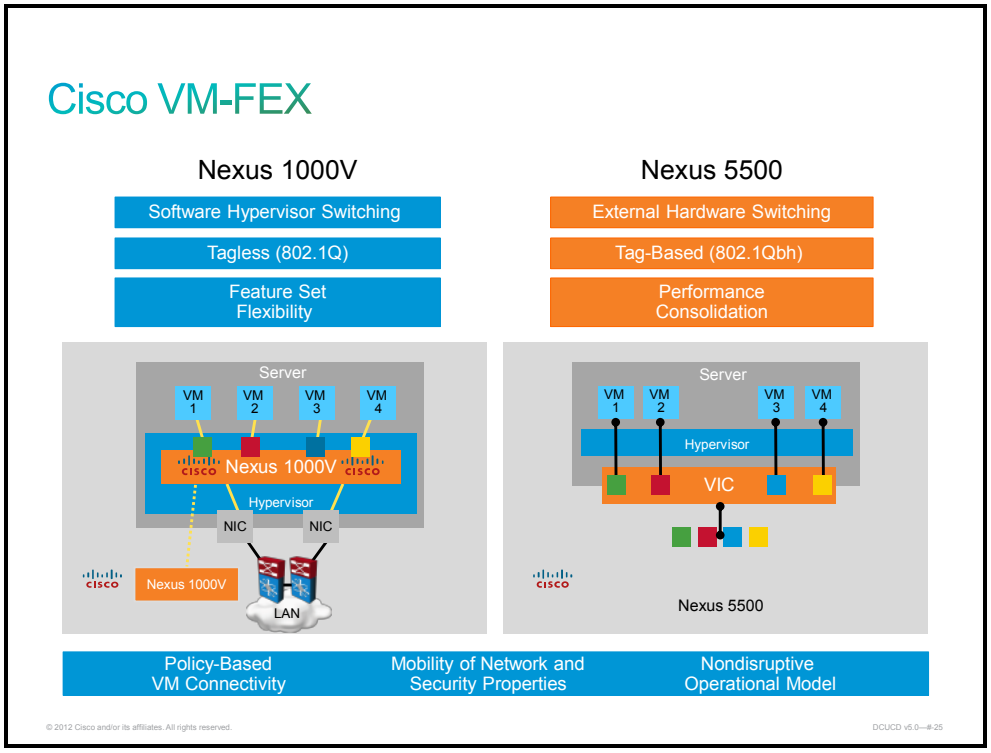


A VDC is used to virtualize the Cisco Nexus 7000 Switch, presenting the physical switch as multiple logical devices. Each VDC contains its own unique and independent VLANs and virtual routing and forwarding (VRF), and each VDC is assigned its own physical ports.

VDCs provide the following benefits:

- They can secure the network partition between different users on the same physical switch.
- They can provide departments with the ability to administer and maintain their own configurations.
- They can be dedicated for testing purposes without impacting production systems.
- They can consolidate the switch platforms of multiple departments on to a single physical platform.
- They can be used by network administrators and operators for training purposes.

https://t.me/learningnets



Cisco VM-FEX encompasses a number of products and technologies that work together to improve server virtualization strategies:

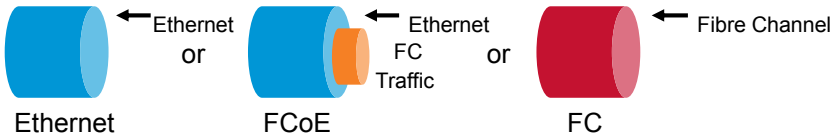
- **Cisco Nexus 1000V Distributed Virtual Switch:** This is a software-based switch that was developed in collaboration with VMware. The switch integrates directly with the VMware ESXi hypervisor. Because the switch can combine the network and server resources, the network and security policies automatically follow a VM that is being migrated with VMware VMotion.
- **NIV:** This VM networking protocol was jointly developed by Cisco and VMware, allowing the Cisco VM-FEX functions to be performed in hardware.
- **Cisco N-Port Virtualizer (NPV):** This function is currently available on the Cisco MDS 9000 family of multilayer switches and the Cisco Nexus 5000 and 5500 Series Switches. The Cisco NPV allows storage services to follow a VM as the VM moves.

Cisco VM-FEX provides visibility down to the VM level, simplifying management, troubleshooting, and regulatory compliance.

https://t.me/learningnets

Unified Ports

- One port for all types of server I/O
  - Any port can be configured as
    - 1/10 Gigabit Ethernet
    - DCB (lossless Ethernet)
    - FCoE on 10 Gigabit Ethernet (dedicated or converged link)
    - 8/4/2/1 Gb/s native Fibre Channel
- Flexibility of use
  - One standard chassis for all data center I/O needs



© 2012 Cisco and/or its affiliates. All rights reserved.

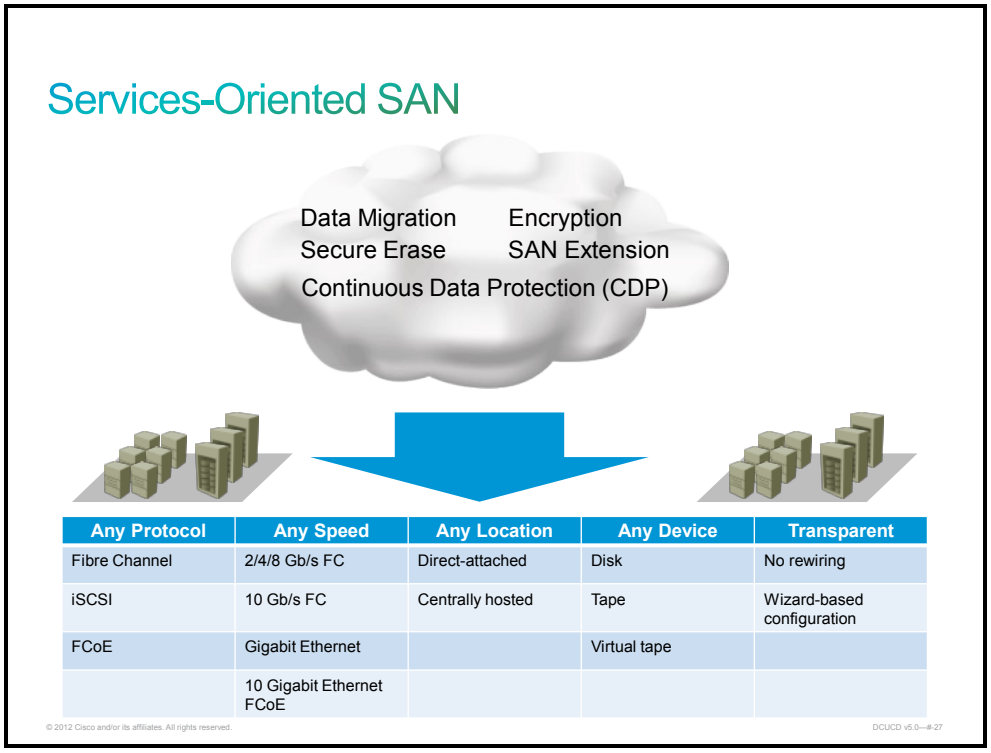
Unified ports are ports that can be configured as Ethernet or Fiber Channel, and are supported on Cisco Nexus 5500UP Series Switches and Cisco UCS 6200 Series Fabric Interconnects.

Unified ports support all existing port types: 1/10 Gigabit Ethernet, FCoE, and 1/2/4/8 Gb/s Fibre Channel interfaces. Cisco Nexus 5500UP Series ports can be configured as 1/10 Gigabit Ethernet, DCB (lossless Ethernet), FCoE on 10 Gigabit Ethernet (dedicated or converged link), or 8/4/2/1 Gb/s native Fibre Channel ports.

The benefits to unified ports are as follows:

- Deploys a switch (for example, Cisco Nexus 5500UP) as a data center switch capable of all important I/O
- Mixes Fibre Channel SAN to host as well as switch and target with FCoE SAN
- Implements with native Fibre Channel and enables smooth migration to FCoE in the future

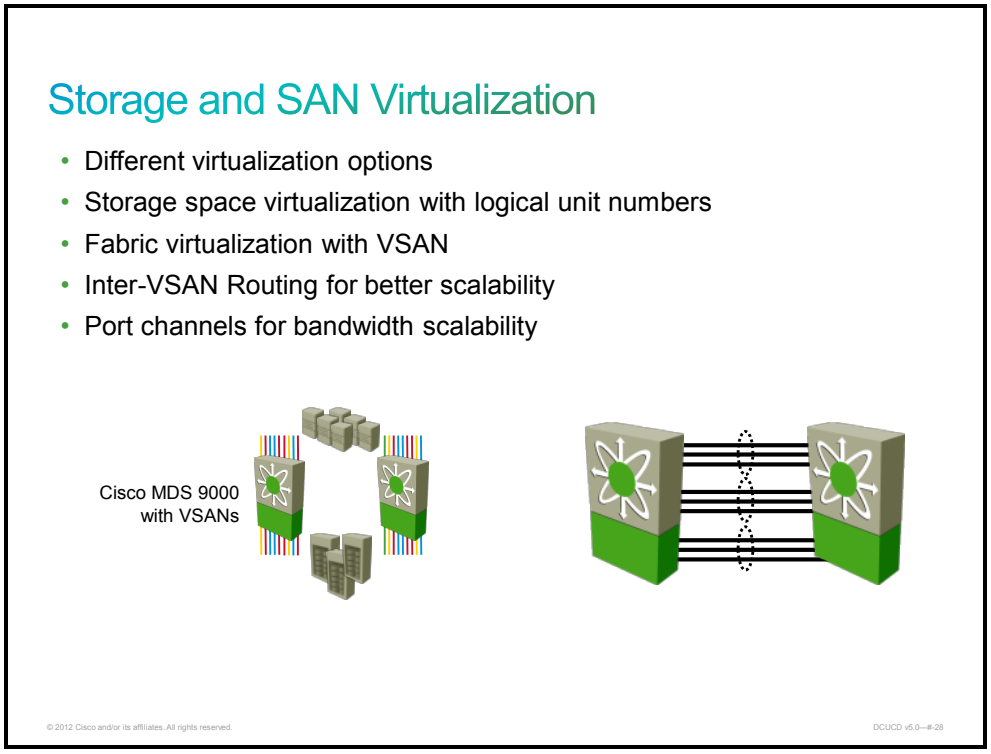
https://t.me/learningnets



Services-oriented SAN fabrics can transparently extend any of the following SAN services to any device within the fabric:

- Data Mobility Manager (DMM), which provides online data migration
- LinkSec encryption, which encrypts Fibre Channel traffic
- Secure Erase, which permanently erases data
- SAN extension features such as Write Acceleration and Tape Read Acceleration, which reduce overall latency
- Continuous data protection (CDP), which is enabled by the SANTap protocol

https://t.me/learningnets



SAN Islands

A SAN island refers to a physically isolated switch or group of switches that is used to connect hosts to storage devices. Today, SAN designers build separate fabrics, otherwise known as SAN islands, for various reasons.

Reasons for building SAN islands may include the desire to isolate different applications into their own fabrics or to raise availability by minimizing the impact of fabric-wide disruptive events. In addition, separate SAN islands also offer a higher degree of security because each physical infrastructure contains its own separate set of Fabric Services and management access.

VSAN Scalability

VSAN functionality is a feature that was developed by Cisco to leverage the advantages of isolated SAN fabrics with capabilities that address the limitations of isolated SAN islands.

VSANs provide a method for allocating ports within a physical fabric to create virtual fabrics. Independent physical SAN islands are virtualized onto a common SAN infrastructure.

An analogy is that VSANs on Fibre Channel switches are like VDCs on Cisco Nexus 7000 Series Switches. VSANs can virtualize the physical switch into many virtual switches.

Using VSANs, SAN designers can raise the efficiency of a SAN fabric and alleviate the need to build multiple physically isolated fabrics to meet organizational or application needs. Instead, fewer and less-costly redundant fabrics can be built, each housing multiple applications, and the VSANs can still provide island-like isolation.

VSANs provide not only hardware-based isolation but also a complete replicated set of Fibre Channel services for each VSAN. Therefore, when a VSAN is created, a completely separate set of Fabric Services, configuration management capability, and policies are created within the new VSAN.

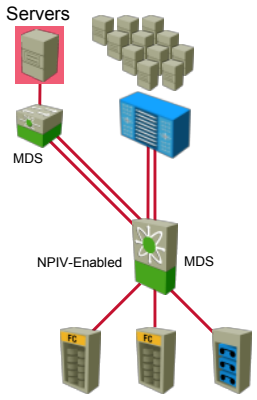
Each separate virtual fabric is isolated from other virtual fabrics by using a hardware-based, frame-tagging mechanism on VSAN member ports and Enhanced Inter-Switch Links (EISLs). The EISL link type includes added tagging information for each frame within the fabric. The

https://t.me/learningnets

EISL link is supported on links that interconnect any Cisco MDS 9000 Family Switch product. Membership to a VSAN is based on the physical port, and no physical port may belong to more than one VSAN. Therefore, a node is connected to a physical port and becomes a member of that VSAN port.

Cisco NPV-Enabled Environment

- NPV edge switches
  - Need to be enabled on switch
  - Changing to and from NPV mode is disruptive
    - The switch reboots.
    - The configuration is not saved.
  - Supports only F, SD, and NP modes
  - Local switching is not supported, all switching is done at the core,
- NPV core switches
  - Must enable the N-Port ID Virtualization (NPIV) feature
  - Supports up to 100 NPV edge switches
- NPV-enabled switches are standards-based and interoperable with other third-party switches in the SAN.



© 2012 Cisco and/or its affiliates. All rights reserved.

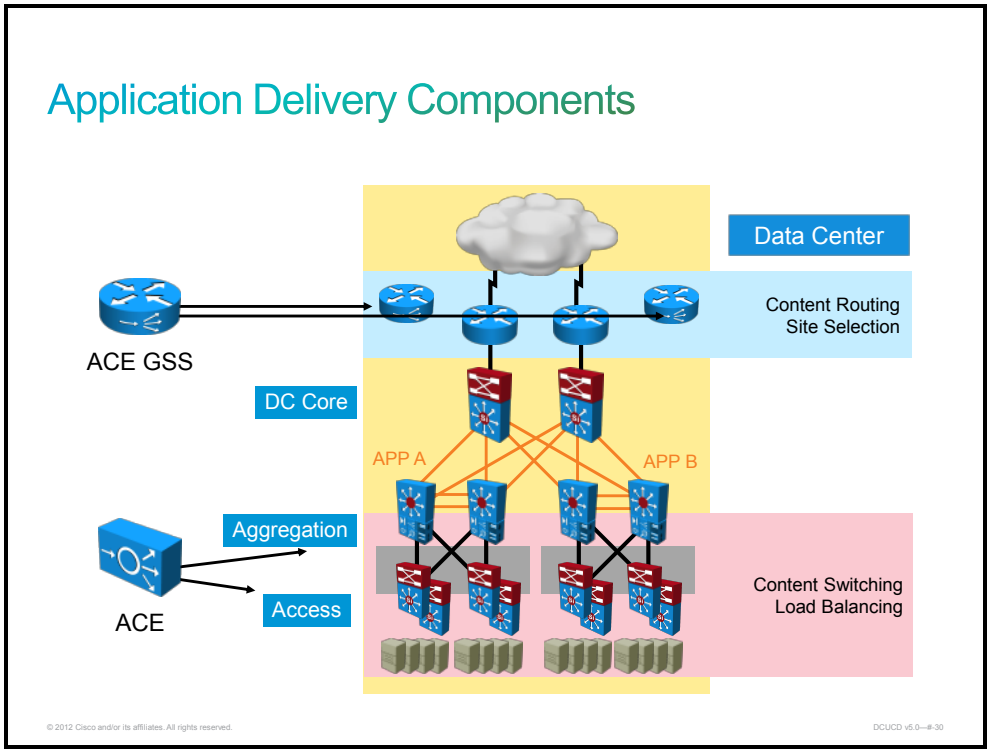
The Fibre Channel standards as defined by the ANSI T11 committee allow for up to 239 Fibre Channel domains per fabric or VSAN. However, the Optical Services Module (OSM) has only qualified up to 40 domains per fabric or VSAN.

Each Fibre Channel switch is identified by a single domain ID, so effectively there can be no more than 40 switches that are connected together.

Blade switches and top-of-rack access layer switches will also consume a domain ID, which will limit the number of domains that can be deployed in data centers.

The Cisco NPV addresses the increase in the number of domain IDs that are needed to deploy many ports by making a fabric or module switch appear as a host to the core Fibre Channel switch, and as a Fibre Channel switch to the servers in the fabric or blade switch. Cisco NPV aggregates multiple locally connected N Ports into one or more external N-Port links, which share the domain ID of the NPV core switch among multiple NPV switches. Cisco NPV also allows multiple devices to attach to the same port on the NPV core switch, and it reduces the need for more ports on the core.

https://t.me/learningnets



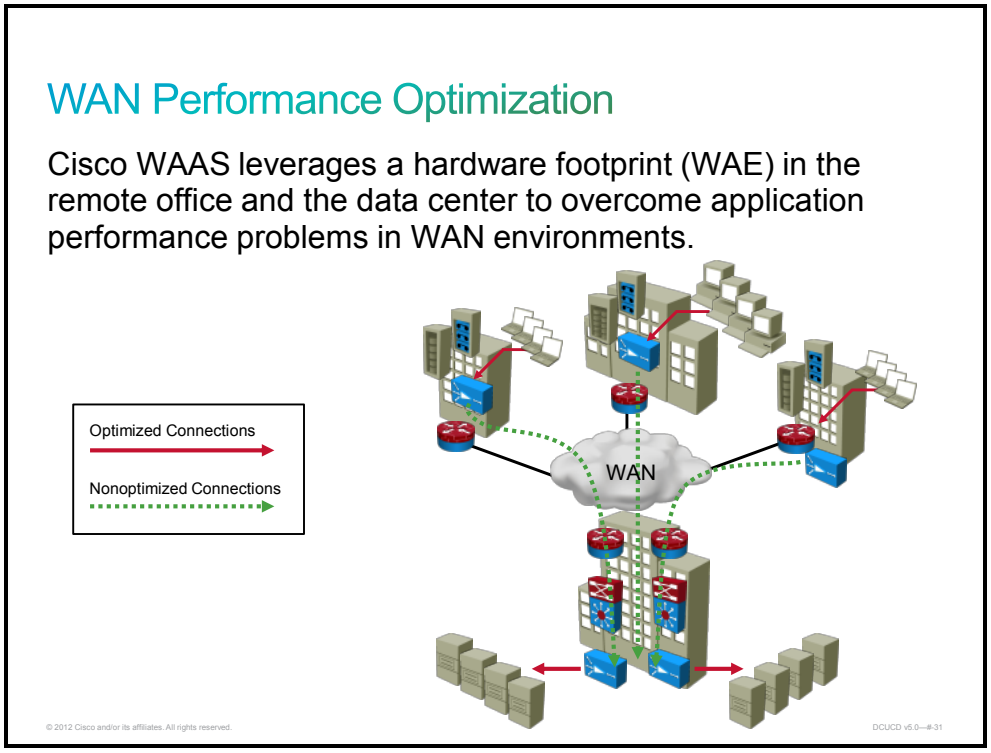
Server Load-Balancing

Load balancing is a computer networking methodology to distribute workload across multiple computers or a computer cluster, network links, central processing units, disk drives, or other resources, to achieve optimal resource utilization, maximize throughput, minimize response time, and avoid overload. Using multiple components with load balancing, instead of a single component, may increase reliability through redundancy. The load-balancing service is usually provided by dedicated software or hardware, such as a multilayer switch or a Domain Name System (DNS) server.

Server load balancing is the process of deciding to which server a load-balancing device should send a client request for service. For example, a client request may consist of an HTTP GET request for a web page or an FTP GET request to download a file. The job of the load balancer is to select the server that can successfully fulfill the client request and do so in the shortest amount of time without overloading either the server or the server farm as a whole.

The figure illustrates the application delivery components in the data center network. At the content routing layer, site selection is provided by global site selection (GSS). At the content switching layer, load balancing is provided by the Cisco Application Control Engine (ACE) Module or appliance. The ACE appliance is deployed in the data center access layer as well.

https://t.me/learningnets



WAN Challenges

The goal of the typical distributed enterprise is to consolidate as much of this infrastructure as possible into the data center without overloading the WAN, and without compromising the performance expectations of remote office users who are accustomed to working with local resources.

Latency

Latency is the most silent yet greatest detractor of application performance over the WAN. Latency is problematic because of the volume of message traffic that must be sent and received. Some messages are very small, yet even with substantial compression and flow optimizations, these messages must be exchanged between the client and the server to maintain protocol correctness, data integrity, and so on.

Bandwidth

Bandwidth utilization is another application performance killer. Transferring a file multiple times can consume significant WAN bandwidth. If a validated copy of a file or other object is stored locally in an application cache, it can be served to the user without using the WAN.

Cisco WAAS

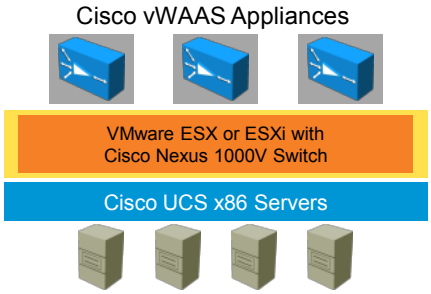
Cisco Wide Area Application Services (WAAS) is a solution that overcomes the challenges presented by the WAN. Cisco WAAS is a software package that runs on the Cisco Wide Area Application Engine (WAE) that transparently integrates with the network to optimize applications without client, server, or network feature changes.

A Cisco WAE is deployed in each remote office, regional office, and data center of the enterprise. With Cisco WAAS, flows to be optimized are transparently redirected to the Cisco WAE, which overcomes the restrictions presented by the WAN, including bandwidth disparity, packet loss, congestion, and latency. Cisco WAAS enables application flows to overcome restrictive WAN characteristics to enable the consolidation of distributed servers, save precious WAN bandwidth, and improve the performance of applications that are already centralized.

https://t.me/learningnets

Cisco Virtual WAAS

- Cisco vWAAS is cloud-ready WAN optimization.
- Virtual appliance that accelerates applications delivered from private and virtual private cloud infrastructures.
- Runs on the VMware ESXi hypervisor and Cisco UCS x86 servers.
- Cisco vWAAS can be deployed in two ways:
  - Transparently at the WAN network edge
  - Within the data center with application servers



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco virtual WAAS (vWAAS) is a cloud-ready WAN optimization solution. Cisco vWAAS is a virtual appliance that accelerates business applications delivered from private and virtual private cloud infrastructures, helping to ensure an optimal user experience. Cisco vWAAS runs on the VMware ESXi hypervisor and Cisco UCS x86 servers, providing an agile, elastic, and multitenant deployment.

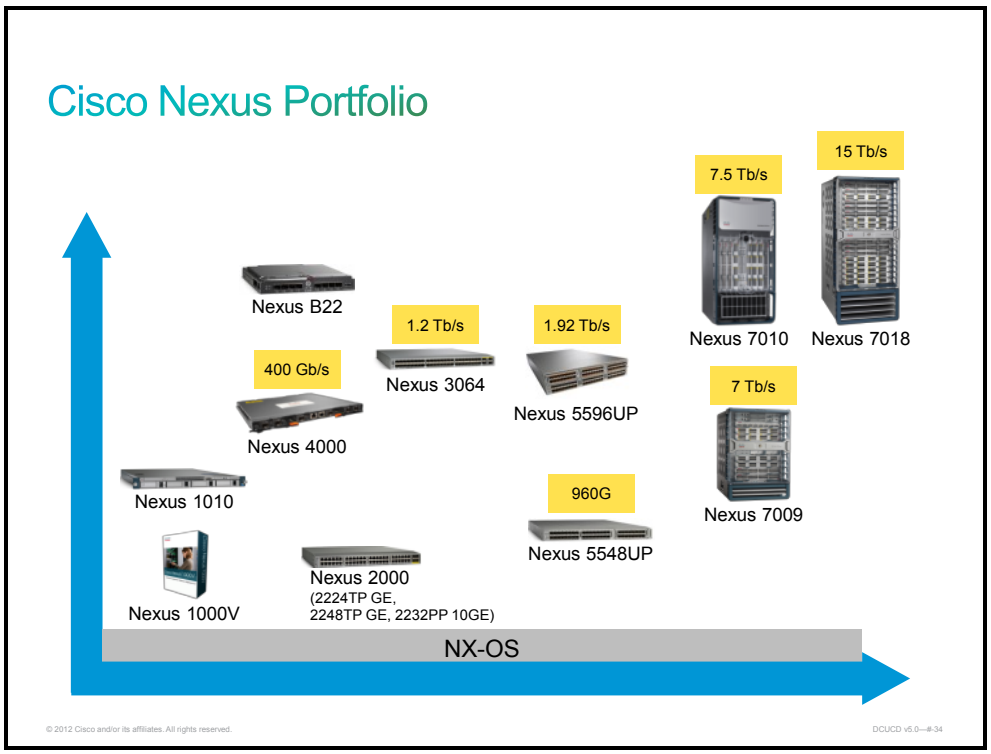
Cisco vWAAS can be deployed in two ways:

- Transparently at the WAN network edge, using out-of-path interception technology such as Web Cache Communication Protocol (WCCP), similar to deployment of a physical Cisco WAAS appliance
- Within the data center with application servers, using a virtual network services framework based on Cisco Nexus 1000V Series Switches to offer a cloud-optimized application service in response to instantiation of application server VMs

https://t.me/learningnets

Cisco Data Center Network Equipment

This topic describes the Cisco Data Center network equipment.



The Cisco Nexus product family comprises the following switches:

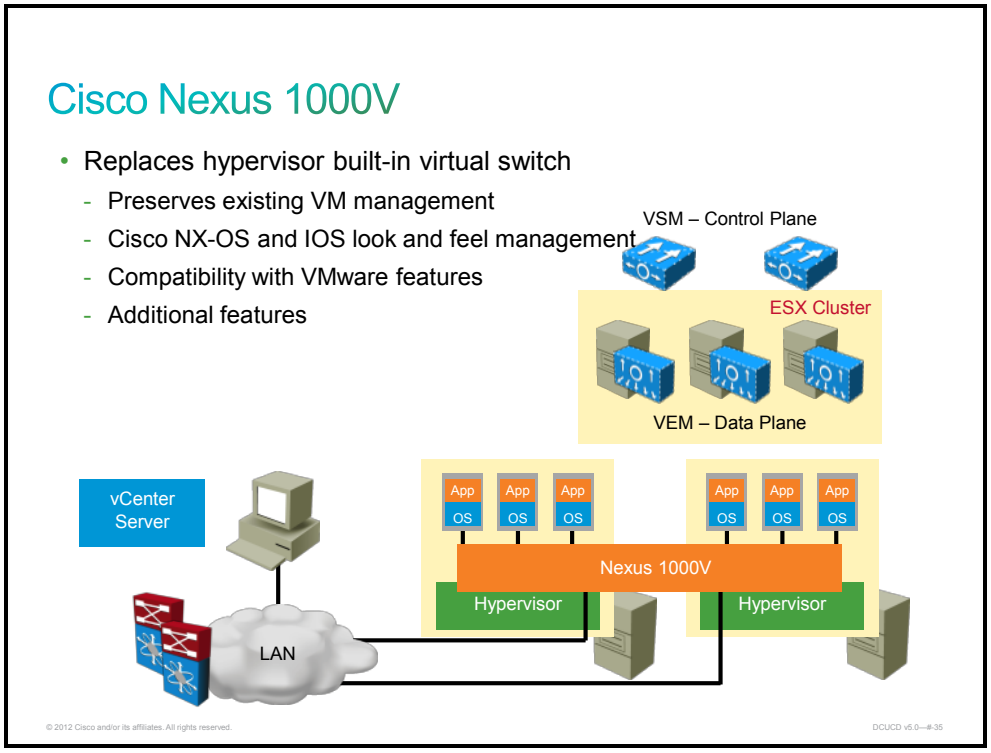
- **Cisco Nexus 1000V:** A virtual machine access switch that is an intelligent software switch implementation for VMware vSphere environments running the Cisco Nexus Operating System (NX-OS) Software. The Cisco Nexus 1000V operates inside the VMware ESX hypervisor, and supports the Cisco VN-Link server virtualization technology to provide the following:
  - Policy-based virtual machine connectivity
  - Mobile virtual machine security and network policy
  - Nondisruptive operational model for server virtualization and networking teams
- **Cisco Nexus 1010 Virtual Services Appliance:** The appliance is a member of the Cisco Nexus 1000V Series Switches and hosts the Cisco Nexus 1000V Virtual Supervisor Module (VSM). It also supports the Cisco Nexus 1000V Network Analysis Module (NAM) Virtual Service Blade and provides a comprehensive solution for virtual access switching. The Cisco Nexus 1010 provides dedicated hardware for the VSM, making access switch deployment easier for the network administrator.
- **Cisco Nexus 2000 Series Fabric Extender (FEX):** A category of data center products that are designed to simplify data center access architecture and operations. The Cisco Nexus 2000 Series uses the Cisco FEX-Link architecture to provide a highly scalable unified server-access platform across a range of 100-Mb/s Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet, unified fabric, copper and fiber connectivity, and rack and blade server environments. The Cisco Nexus 2000 Series Fabric Extenders act as remote line cards for the Cisco Nexus 5500 Series and the Cisco Nexus 7000 Series Switches.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

- **Cisco Nexus 3000 Series Switches:** The Cisco Nexus 3000 Series Switches extend the comprehensive, proven innovations of the Cisco Data Center Business Advantage architecture into the high-frequency trading (HFT) market. The Cisco Nexus 3064 Switch supports 48 fixed 1/10-Gb/s enhanced small form-factor pluggable plus (SFP+) ports and 4 fixed quad SFP+ (QSFP+) ports, which allow smooth transition from 10 Gigabit Ethernet to 40 Gigabit Ethernet. The Cisco Nexus 3064 Switch is well suited for financial colocation deployments, delivering features such as latency of less than a microsecond, line-rate Layer 2 and 3 unicast and multicast switching, and the support for 40 Gigabit Ethernet technologies.
- **Cisco Nexus 4000 Switch Module for IBM BladeCenter:** A blade switch solution for IBM BladeCenter-H and HT chassis. This switch provides the server I/O solution that is required for high-performance, scale-out, virtualized and nonvirtualized x86 computing architectures. It is a line-rate, extremely low-latency, nonblocking, Layer 2, 10-Gb/s blade switch that is fully compliant with the International Committee for Information Technology Standards (INCITS) FCoE and IEEE 802.1 DCB standards.
- **Cisco Nexus B22 Blade Fabric Extender:** The Nexus B22 Blade FEX behaves like extensions of a parent Cisco Nexus 5000 Series Switch, forming a distributed modular system. It enables simplified data center access architecture of integrated blade switches for third-party blade chassis.
- **Cisco Nexus 5500 Series Switches:** A family of line-rate, low-latency, lossless 10 Gigabit Ethernet and FCoE switches for data center applications. The Cisco Nexus 5000 Series Switches are designed for data centers transitioning to 10 Gigabit Ethernet, as well as data centers ready to deploy a unified fabric that can manage LAN, SAN, and server clusters. This capability provides networking over a single link, with dual links used for redundancy.
- **Cisco Nexus 7000 Series Switch:** A modular data center-class switch that is designed for highly scalable 10 Gigabit Ethernet networks with a fabric architecture that scales beyond 15 Tb/s. The switch is designed to deliver continuous system operation and virtualized services. The Cisco Nexus 7000 Series Switches incorporate significant enhancements in design, power, airflow, cooling, and cabling. The 10-slot chassis has front-to-back airflow, making it a good solution for hot-aisle and cold-aisle deployments. The 18-slot chassis uses side-to-side airflow to delivery high density in a compact form factor.

https://t.me/learningnets



The Cisco Nexus 1000V provides Layer 2 switching functions in a virtualized server environment, and replaces virtual switches within the ESX servers. This allows users to configure and monitor the virtual switch using the Cisco NX-OS CLI. The Cisco Nexus 1000V provides visibility in to the networking components of the ESX servers and access to the virtual switches within the network.

The vCenter Server defines the data center that the Cisco Nexus 1000V will manage, with each server being represented as a line card and managed as if it were a line card in a physical Cisco switch.

There are two components that are part of the Cisco Nexus 1000V implementation:

- **Virtual Supervisor Module (VSM):** This is the control software of the Cisco Nexus 1000V distributed virtual switch, and runs on either a VM or as an appliance. It is based on the Cisco NX-OS Software.
- **Virtual Ethernet Module (VEM):** This is the part that actually switches the data traffic and runs on a VMware ESX 4.0 host. The VSM can control several VEMs, with the VEMs forming a switch domain that should be in the same virtual data center that is defined by VMware vCenter.

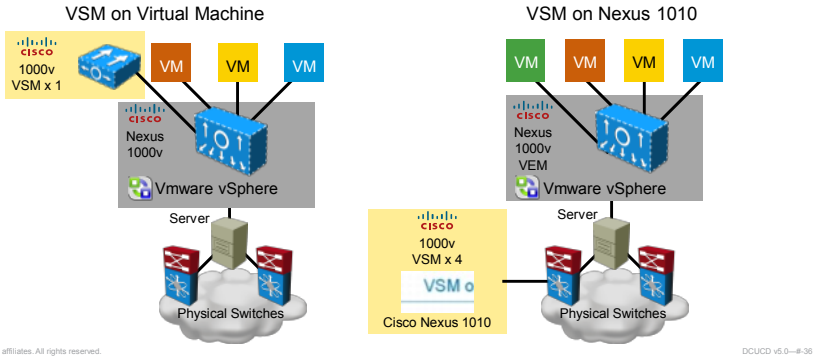
The Cisco Nexus 1000V is effectively a virtual chassis. It is modular, and ports can either be physical or virtual. The servers are modules on the switch, with each physical network interface virtualization (NIV) port on a module being a physical Ethernet port. Modules 1 and 2 are reserved for the VSM, with the first server or host automatically being assigned to the next available module number. The ports to which the virtual network interface card (vNIC) interfaces connect are virtual ports on the Cisco Nexus 1000V, where they are assigned a global number.

https://t.me/learningnets

Cisco Nexus 1010V Appliance

- Hardware platform for Nexus 1000V VSM
- Provides VSM independence of existing production hosts
- Cisco Nexus 1010V comes bundled with licenses
- Platform for additional services: Cisco Virtual NAM, VSG





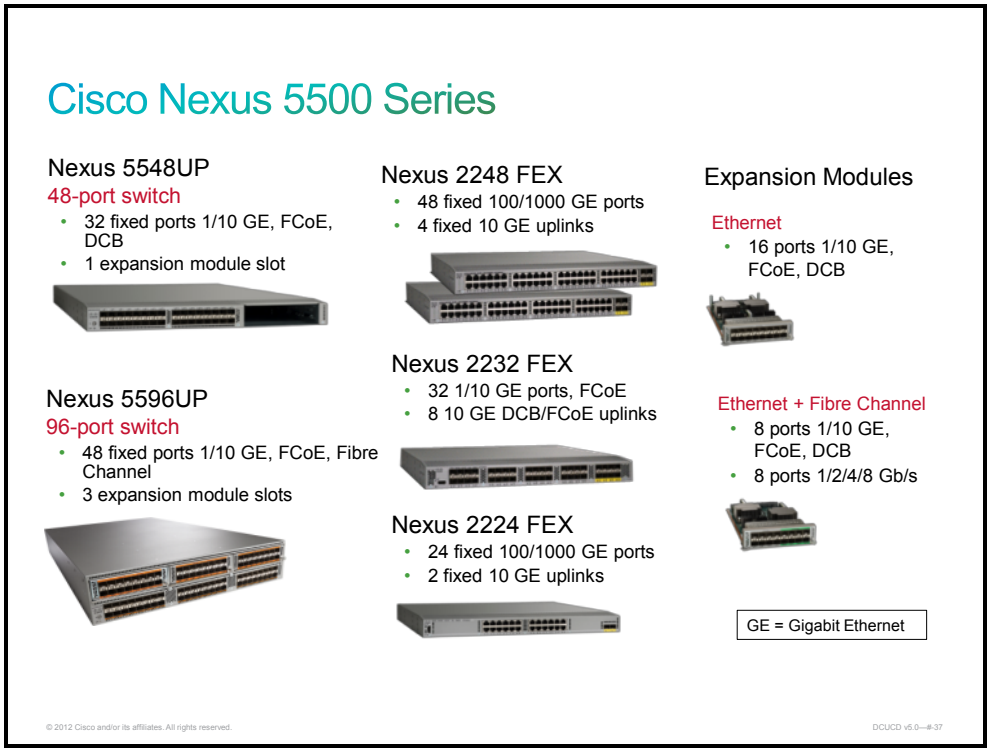
© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 1010V is the hardware platform for the VSM. It is designed for those customers who wish to provide independence for the VSM so that it does not share the production infrastructure.

As an additional benefit, the Nexus 1010V comes bundled with VEM licenses.

The Nexus 1010V serves also as a hardware platform for various additional services, including Cisco virtual NAM, Cisco Virtual Services Gateway, and so on.

https://t.me/learningnets



The Cisco Nexus 5500 Series Switches use a cut-through architecture that supports line-rate 10 Gigabit Ethernet on all ports, maintaining a consistent low latency independent of packet size, and service-enabled. The switches support a set of network technologies that are known collectively as IEEE DCB, which increases reliability, efficiency, and scalability of Ethernet networks. These features allow the switches to support multiple traffic classes over a lossless Ethernet fabric, thus enabling consolidation of LAN, SAN, and cluster environments. The ability to connect FCoE to native Fibre Channel protects existing storage system investments, which dramatically simplifies in-rack cabling.

The Cisco Nexus 5500 Series Switches integrate with multifunction adapters, called converged network adapters (CNAs), to provide Unified Fabric convergence. The adapters combine the functions of Ethernet NICs and Fibre Channel host bus adapters (HBAs). This functionality makes the transition to a single, unified network fabric transparent and consistent with existing practices, management software, and operating system drivers. The switch family is compatible with integrated transceivers and twinax cabling solutions that deliver cost-effective connectivity for 10 Gigabit Ethernet to servers at the rack level. This compatibility eliminates the need for expensive optical transceivers.

Cisco Nexus 5548UP Switch

The Cisco Nexus 5548UP is the first of the Cisco Nexus 5500 platform. It is a 1-RU 10 Gigabit Ethernet and FCoE switch offering up to 960-Gb/s throughput and up to 48 ports. The switch has 32 1/10-Gb/s FCOE/Fibre Channel ports and 1 expansion slot.

Cisco Nexus 5596UP Switch

The Cisco Nexus 5596UP Switch has 48 fixed ports capable of supporting 1/10-Gb/s FCOE/Fibre Channel and three additional slots. The three additional slots will accommodate any of the expansion modules for the Cisco 5500 Series Switches, taking the maximum capacity for the switch to 96 ports. These ports are unified ports that provide flexibility with the connectivity requirements, and the switch offers 1.92-Tb/s throughput.

https://t.me/learningnets

Both of the Cisco Nexus 5500 Series Switches support Cisco FabricPath, and with the Layer 3 routing module, both Layer 2 and Layer 3 support is provided. The Cisco Nexus 5500 Series supports the same Cisco Nexus 2200 Series Fabric Extenders.

Expansion Modules for the Cisco 5500 Series Switches

The Cisco Nexus 5500 Series Switches support the following expansion modules:

- Ethernet module that provides 16 1/10 Gigabit Ethernet and FCoE ports using SFP+ interfaces
- Fibre Channel plus Ethernet module that provides eight 1/10 Gigabit Ethernet and FCoE ports using the SFP+ interface, and eight ports of 8/4/2/1-Gb/s native Fibre Channel connectivity using the SFP interface
- A Layer 3 daughter card for routing functionality

The modules for the Cisco Nexus 5500 Series Switches are not backward-compatible with the Cisco Nexus 5000 Series Switches.

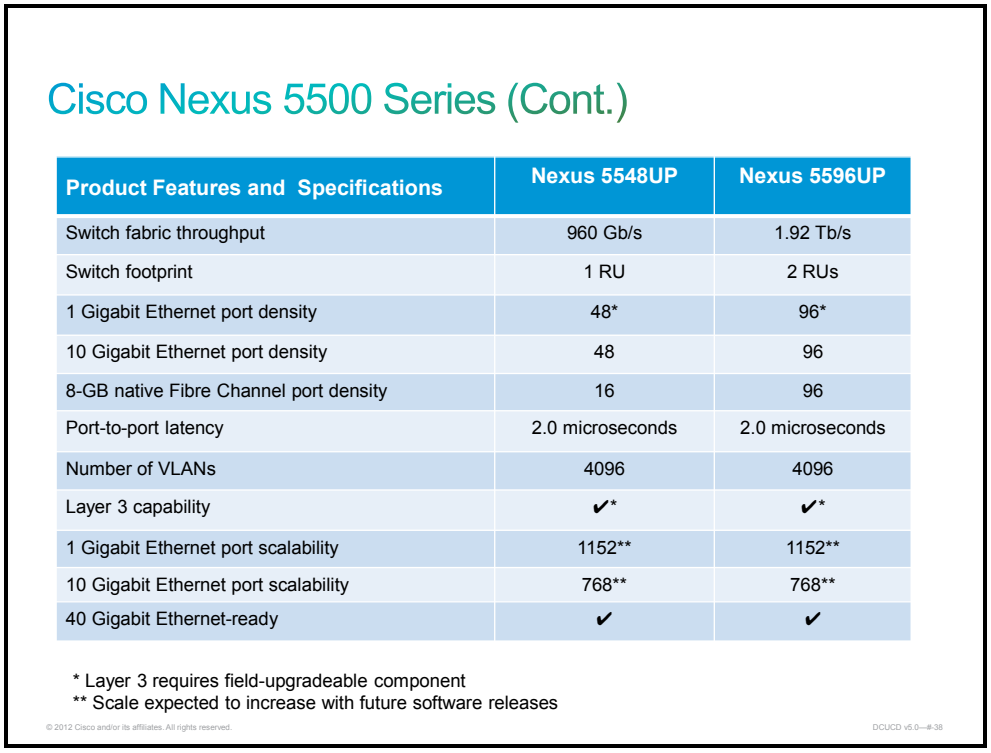
Cisco Nexus 2000 Series Fabric Extenders

The Cisco Nexus 2000 Series offers front-to-back cooling, compatibility with data center hot-aisle and cold-aisle designs, placement of all switch ports at the rear of the unit in close proximity to server ports, and accessibility of all user-serviceable components from the front panel. The Cisco Nexus 2000 Series has redundant hot-swappable power supplies and a hot-swappable fan tray with redundant fans, and is a 1-RU form factor.

The Cisco Nexus 2000 Series has two types of ports: ports for end-host attachment and uplink ports. The Cisco Nexus 2000 Series is an external line module for the Cisco Nexus 5000 and 5500 Series Switches, and for the Cisco Nexus 7000 Series.

- **Cisco Nexus 2224TP GE:** 24 100/1000BASE-T ports and 2 10 Gigabit Ethernet uplinks (SFP+)
- **Cisco Nexus 2248TP GE:** 48 100/1000BASE-T ports and 4 10 Gigabit Ethernet uplinks (SFP+)—this model is supported as an external line module for the Cisco Nexus 7000 Series using Cisco NX-OS 5.1(2) software.
- **Cisco Nexus 2232PP 10GE:** 32 1/10 Gigabit Ethernet/Fibre Channel over Ethernet (FCoE) ports (SFP+) and 8 10 Gigabit Ethernet/FCoE uplinks (SFP+)

https://t.me/learningnets



The table in the figure describes the differences between the Cisco Nexus 5548UP and 5596UP Series Switches. The port counts are based on 24 Cisco Nexus 2000 Fabric Extenders per Cisco Nexus 5500 Series Switch.

Cisco Nexus 5500 Platform Features

The Cisco Nexus 5500 Series is the second generation of access switches for 10 Gigabit Ethernet connectivity. The Cisco Nexus 5500 platform provides a rich feature set that makes it well suited for top-of-rack (ToR), middle-of-row (MoR), or end-of-row (EoR) access layer applications. It protects investments in data center racks with standards-based 1 and 10 Gigabit Ethernet and FCoE features, and virtual machine awareness features that allow IT departments to consolidate networks. The combination of high port density, lossless Ethernet, wire-speed performance, and extremely low latency makes the switch family well suited to meet the growing demand for 10 Gigabit Ethernet. The family can support unified fabric in enterprise and service provider data centers, protecting the investments of enterprises. The switch family has sufficient port density to support single and multiple racks that are fully populated with blade- and rack-mount servers.

- **High density and high availability:** The Cisco Nexus 5548P provides 48 1/10-Gb/s ports in 1 RU, and the upcoming Cisco Nexus 5596UP Switch provides a density of 96 1/10-Gb/s ports in 2 RUs. The Cisco Nexus 5500 Series is designed with redundant and hot-swappable power and fan modules that can be accessed from the front panel, where status lights offer an at-a-glance view of switch operation. To support efficient data center hot- and cold-aisle designs, front-to-back cooling is used for consistency with server designs.
- **Nonblocking line-rate performance:** All the 10 Gigabit Ethernet ports on the Cisco Nexus 5500 platform can manage packet flows at wire speed. The absence of resource sharing helps ensure the best performance of each port regardless of the traffic patterns on other ports. The Cisco Nexus 5548P can have 48 Ethernet ports at 10 Gb/s, sending packets simultaneously without any effect on performance, offering true 960-Gb/s bidirectional bandwidth. The upcoming Cisco Nexus 5596UP can have 96 Ethernet ports at 10 Gb/s, offering true 1.92-Tb/s bidirectional bandwidth.

https://t.me/learningnets

- **Low latency:** The cut-through switching technology that is used in the ASICs of the Cisco Nexus 5500 Series enables the product to offer a low latency of 2 microseconds, which remains constant regardless of the size of the packet being switched. This latency was measured on fully configured interfaces, with access control lists (ACLs), quality of service (QoS), and all other data path features turned on. The low latency on the Cisco Nexus 5500 Series together with a dedicated buffer per port and the congestion management features make the Cisco Nexus 5500 platform an excellent choice for latency-sensitive environments.
- **Single-stage fabric:** The crossbar fabric on the Cisco Nexus 5500 Series is implemented as a single-stage fabric, thus eliminating any bottleneck within the switches. Single-stage fabric means that a single crossbar fabric scheduler has complete visibility into the entire system and can therefore make optimal scheduling decisions without building congestion within the switch. With a single-stage fabric, the congestion becomes exclusively a function of your network design—the switch does not contribute to it.
- **Congestion management:** Keeping latency low is not the only critical element for a high-performance network solution. Servers tend to generate traffic in bursts, and when too many bursts occur at the same time, a short period of congestion occurs. Depending on how the burst of congestion is smoothed out, the overall network performance can be affected. The Cisco Nexus 5500 Series platform offers a complete range of congestion management features to reduce congestion. These features address congestion at different stages and offer granular control over the performance of the network.
  - **Virtual output queues:** The Cisco Nexus 5500 platform implements virtual output queues (VOQs) on all ingress interfaces, so that a congested egress port does not affect traffic that is directed to other egress ports. Every 802.1p CoS uses a separate VOQ in the Cisco Nexus 5500 platform architecture, resulting in a total of 8 VOQs per egress on each ingress interface, or a total of 384 VOQs per ingress interface on the Cisco Nexus 5548P, and a total of 768 VOQs per ingress interface on the Cisco Nexus 5596UP. The extensive use of VOQs in the system helps ensure high throughput on a per-egress, per-CoS basis. Congestion on one egress port in one CoS does not affect traffic that is destined for other CoS or other egress interfaces. This ability avoids head-of-line (HOL) blocking, which would otherwise cause congestion to spread.
  - **Separate egress queues for unicast and multicast:** Traditionally, switches support eight egress queues per output port, each servicing one 802.1p CoS. The Cisco Nexus 5500 platform increases the number of egress queues by supporting eight egress queues for unicast and eight egress queues for multicast. This support allows separation of unicast and multicast, which are contending for system resources within the same CoS, and provides more fairness between unicast and multicast. Through configuration, the user can control the amount of egress port bandwidth for each of the 16 egress queues.
  - **Lossless Ethernet with priority flow control (PFC):** By default, Ethernet is designed to drop packets when a switching node cannot sustain the pace of the incoming traffic. Packet drops make Ethernet very flexible in managing random traffic patterns that are injected into the network. However, they effectively make Ethernet unreliable and push the burden of flow control and congestion management up to a higher level in the network stack.

https://t.me/learningnets

PFC offers point-to-point flow control of Ethernet traffic that is based on 802.1p CoS. With a flow-control mechanism in place, congestion does not result in drops, transforming Ethernet into a reliable medium. The CoS granularity then allows some CoS to gain a no-drop, reliable behavior while allowing other classes to retain traditional best-effort Ethernet behavior. The no-drop benefits are significant for any protocol that assumes reliability at the media level, such as FCoE.

- **Explicit congestion notification (ECN) marking:** ECN is an extension to TCP/IP, defined in RFC 3168. ECN allows end-to-end notification of network congestion without dropping packets. Traditionally, TCP detects network congestion by observing dropped packets. When congestion is detected, the TCP sender takes action by controlling the flow of traffic. However, dropped packets can sometimes lead to long TCP timeouts and consequent loss of throughput. The Cisco Nexus 5500 platform can set a mark in the IP header so that, instead of dropping a packet, it sends a signal about impending congestion. The receiver of the packet echoes the congestion indicator to the sender, which must respond as though congestion had been indicated by packet drops.

- **FCoE:** FCoE is a standards-based encapsulation of Fibre Channel frames into Ethernet frames. By implementing FCoE, the Cisco Nexus 5500 platform enables storage I/O consolidation in addition to Ethernet.

- **NIV architecture:** The introduction of blade servers and server virtualization has increased the number of access layer switches that need to be managed. In both cases, an embedded switch or softswitch requires separate management. NIV enables a central switch to create an association with the intermediate switch, whereby the intermediate switch will become the data path to the central forwarding and policy enforcement under the central switch control. This scheme enables both a single point of management and a uniform set of features and capabilities across all access layer switches.

  One critical implementation of NIV in the Cisco Nexus 5000 and 5500 Series is the Cisco Nexus 2000 Series Fabric Extenders and their deployment in data centers. A Cisco Nexus 2000 Series Fabric Extender behaves as a virtualized remote I/O module, enabling the Cisco Nexus 5500 platform to operate as a virtual modular chassis.

- **IEEE 1588 Precision Time Protocol (PTP):** In financial environments, particularly high-frequency trading environments, transactions occur in less than a millisecond. For accurate application performance monitoring and measurement, the systems supporting electronic trading applications must be synchronized with extremely high accuracy (to less than a microsecond). IEEE 1588 is designed for local systems requiring very high accuracy beyond that attainable using Network Time Protocol (NTP). The Cisco Nexus 5500 platform supports IEEE 1588 boundary clock synchronization. In other words, the Cisco Nexus 5500 platform will run PTP and synchronize to an attached master clock, and the boundary clock will then act as a master clock for all attached slaves. The Cisco Nexus 5500 platform also supports packet time stamping by including the IEEE 1588 time stamp in the encapsulated remote switched port analyzer (ERSPAN) header.

- **Cisco FabricPath and TRILL:** Existing Layer 2 networks that are based on STP have a number of challenges to overcome. These challenges include suboptimal path selection, underutilized network bandwidth, control-plane scalability, and slow convergence. Although enhancements to STP and features such as Cisco vPC technology help mitigate some of these limitations, these Layer 2 networks lack fundamentals that limit their scalability.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco FabricPath and TRansparent Interconnection of Lots of Links (TRILL) are two emerging solutions for creating scalable and highly available Layer 2 networks. Cisco Nexus 5500 Series hardware is capable of switching packets that are based on Cisco FabricPath headers or TRILL headers. This capability enables customers to deploy scalable Layer 2 networks with native Layer 2 multipathing.

- **Layer 3:** The design of the access layer varies depending on whether Layer 2 or Layer 3 is used at the access layer. The access layer in the data center is typically built at Layer 2. Building at Layer 2 allows better sharing of service devices across multiple servers and allows the use of Layer 2 clustering, which requires the servers to be adjacent to Layer 2. In some designs, such as two-tier designs, the access layer may be Layer 3, although this may not imply that every port on these switches is a Layer 3 port. The Cisco Nexus 5500 Series platform can operate in Layer 3 mode with the addition of a routing module.
- **Hardware-level I/O consolidation:** The Cisco Nexus 5500 Series platform ASICs can transparently forward Ethernet, Fibre Channel, FCoE, Cisco FabricPath, and TRILL, providing true I/O consolidation at the hardware level. The solution that is adopted by the Cisco Nexus 5500 platform reduces the costs of consolidation through a high level of integration in the ASICs. The result is a full-featured Ethernet switch and a full-featured Fibre Channel switch that is combined into one product.

https://t.me/learningnets

Cisco Nexus 7000 Series

- 15+ Tb/s system
- DCB and FCoE support
- Continuous operations
- Device virtualization
- Modular operating system
- Cisco Trust Sec







	Nexus 7009	Nexus 7010	Nexus 7018
Slots	7 I/O + 2 sup	8 I/O + 2 sup	16 I/O + 2 sup
Height	14 RUs	21 RUs	25 RUs
BW / Slot Fab 1	N/A	230 Gb/s per slot	230 Gb/s per slot
BW / Slot Fab 2	550 Gb/s per slot	550 Gb/s per slot	550 Gb/s per slot

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 7000 Series Switches offer a modular data center-class product that is designed for highly scalable 10 Gigabit Ethernet networks with a fabric architecture that scales beyond 15 Tb/s. The Cisco Nexus 7000 Series provides integrated resilience that is combined with features optimized specifically for the data center for availability, reliability, scalability, and ease of management.

The Cisco Nexus 7000 Series Switch runs the Cisco NX-OS Software to deliver a rich set of features with nonstop operation.

- Front-to-back airflow with 10 front-accessed vertical module slots and an integrated cable management system facilitates installation, operation, and cooling in both new and existing facilities.
- 18 front-accessed module slots with side-to-side airflow in a compact horizontal form factor with purpose-built integrated cable management, easing operation and reducing complexity.
- The system is designed for reliability and maximum availability. All interface and supervisor modules are accessible from the front. Redundant power supplies, fan trays, and fabric modules are accessible from the rear to ensure that cabling is not disrupted during maintenance.
- The system uses dual dedicated supervisor modules and fully distributed fabric architecture. There are five rear-mounted fabric modules, which, combined with the chassis midplane, deliver up to 230 Gb/s per slot for 4.1-Tb/s of forwarding capacity in the 10-slot form factor, and 7.8-Tb/s in the 18-slot form factor using the Cisco Fabric Module 1. Migrating to the Cisco Fabric Module 2 increases the bandwidth per slot to 550 Gb/s. This increases the forwarding capacity on the 10-slot form factor to 9.9 Tb/s and on the 18-slot form factor to 18.7 Tb/s.
- The midplane design supports flexible technology upgrades as your needs change and provides ongoing investment protection.

https://t.me/learningnets

Cisco Nexus 7000 Series 9-Slot Chassis

- The Cisco Nexus 7000 Series 9-slot chassis with up to seven I/O module slots supports up to 224 10 Gigabit Ethernet or 336 Gigabit Ethernet ports.
- Airflow is from side-to-side.
- The integrated cable management system is designed to support the cabling requirements of a fully configured system to either or both sides of the switch, allowing maximum flexibility. All system components can easily be removed with the cabling in place, providing ease of maintenance tasks with minimal disruption.
- A series of LEDs at the top of the chassis provides a clear summary of the status of the major system components. The LEDs report the status of the power supply, fan, fabric, supervisor, and I/O module.
- The purpose-built optional front module door provides protection from accidental interference with both the cabling and modules that are installed in the system. The transparent front door allows easy observation of cabling and module indicators and status lights without the need to open the doors. The door supports a dual-opening capability for flexible operation and cable installation. The door can be completely removed for both initial cabling and day-to-day management of the system.
- Independent variable-speed system and fabric fans provide efficient cooling capacity to the entire system. Fan tray redundancy features help ensure reliability of the system and support for hot swapping of fan trays.
- The crossbar fabric modules are located in the front of the chassis, with support for two supervisors.

Cisco Nexus 7000 Series 10-Slot Chassis

- The Cisco Nexus 7000 Series 10-slot chassis with up to eight I/O module slots supports up to 256 10 Gigabit Ethernet or 384 Gigabit Ethernet ports, meeting the demands of large deployments.
- Front-to-back airflow helps ensure that use of the Cisco Nexus 7000 Series 10-slot chassis addresses the requirement for hot-aisle and cold-aisle deployments without additional complexity.
- The system uses dual system and fabric fan trays for cooling. Each fan tray is redundant and composed of independent variable-speed fans that automatically adjust to the ambient temperature. This adjustment helps reduce power consumption in well-managed facilities while providing optimum operation of the switch. The system design increases cooling efficiency and provides redundancy capabilities, allowing hot swapping without affecting the system. If either a single fan or a complete fan tray fails, the system continues to operate without a significant degradation in cooling capacity.
- The integrated cable management system is designed for fully configured systems. The system allows cabling either to a single side or to both sides for maximum flexibility without obstructing any important components. This flexibility eases maintenance even when the system is fully cabled.
- The system supports an optional air filter to help ensure clean airflow through the system. The addition of the air filter satisfies Network Equipment Building Standards (NEBS) requirements.
- A series of LEDs at the top of the chassis provides a clear summary of the status of the major system components. The LEDs alert operators to the need to conduct further

https://t.me/learningnets

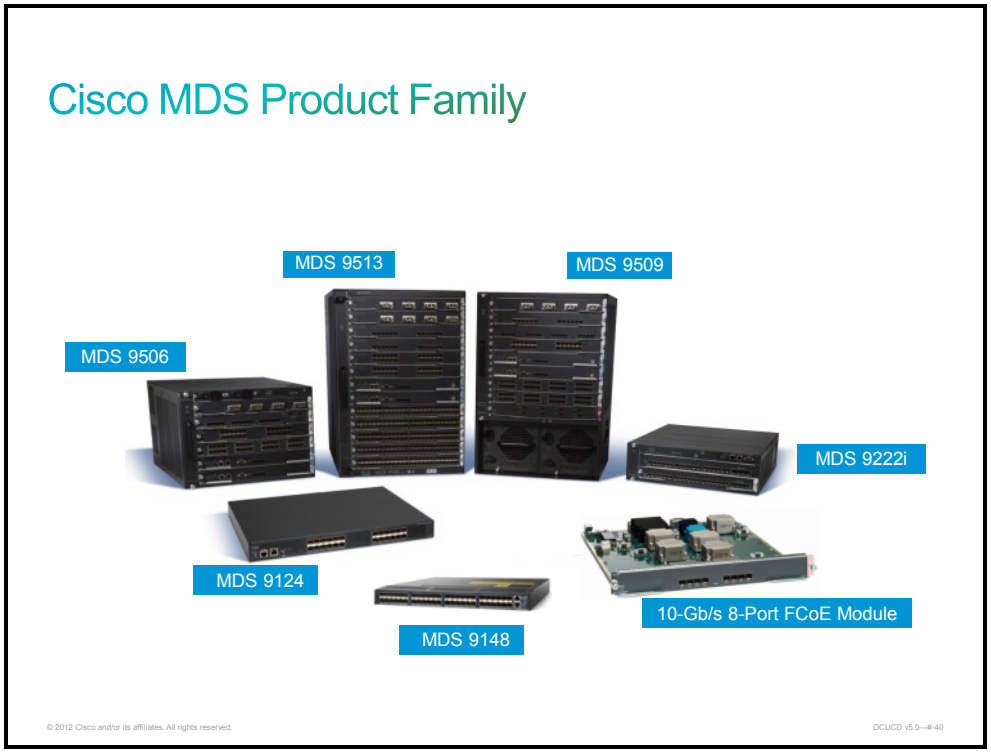
investigation. These LEDs report the power supply, fan, fabric, supervisor, and I/O module status.

- The cable management cover and optional front module doors provide protection from accidental interference with both the cabling and modules that are installed in the system. The transparent front door allows observation of cabling and module indicator and status lights.

Cisco Nexus 7000 Series 18-Slot Chassis

- The Cisco Nexus 7000 Series 18-slot chassis with up to 16 I/O module slots supports up to 512 10 Gigabit Ethernet or 768 Gigabit Ethernet ports, meeting the demands of the largest deployments.
- Side-to-side airflow increases the system density within a 25-RU footprint, optimizing the use of rack space. The optimized density provides more than 16 RUs of free space in a standard 42-RU rack for cable management and patching systems.
- The integrated cable management system is designed to support the cabling requirements of a fully configured system to either or both sides of the switch, allowing maximum flexibility. All system components can easily be removed with the cabling in place, providing ease of maintenance tasks with minimal disruption.
- A series of LEDs at the top of the chassis provides a clear summary of the status of the major system components. The LEDs alert operators to the need to conduct further investigation. These LEDs report the power supply, fan, fabric, supervisor, and I/O module status.
- The purpose-built optional front module door provides protection from accidental interference with both the cabling and modules that are installed in the system. The transparent front door allows easy observation of cabling and module indicators and status lights without any need to open the doors. The door supports a dual-opening capability for flexible operation and cable installation while fitted. The door can be completely removed for both initial cabling and day-to-day management of the system.
- Independent variable-speed system and fabric fans provide efficient cooling capacity to the entire system. Fan tray redundancy features help ensure reliability of the system and support for hot swapping of fan trays.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Cisco MDS SAN Switches

Cisco Multilayer Director Switches (MDS) and directors and the line card modules provide connectivity in a SAN. Since their introduction in 2002, the Cisco MDS switches and directors have embodied many innovative features that help improve performance and help overcome some of the limitations present in many SANs today. One of the benefits of the Cisco MDS products is that the chassis support several generations of line card modules without modification. As Fibre Channel speeds have increased from 2 to 4 Gb/s and are now 8 Gb/s, new line card modules have been introduced to support those faster data rates .These line card modules can be installed in existing chassis without having to replace them with new modules.

Multilayer switches are switching platforms with multiple layers of intelligent features, such as the following:

- Ultra-high availability
- Scalable architecture
- Comprehensive security features
- Ease of management
- Advanced diagnostics and troubleshooting capabilities
- Transparent integration of multiple technologies
- Multiprotocol support

The Cisco MDS 9000 products offer industry-leading investment protection and offer a scalable architecture with highly available hardware and software. Based on the Cisco MDS 9000 Family operating system and a comprehensive management platform in Cisco Fabric Manager, the Cisco MDS 9000 Family offers various application line card modules and a scalable architecture from an entry-level fabric switch to director-class systems.

https://t.me/learningnets

This architecture is forward and backward compatible with the Generation 1 and Generation 2 line cards, including the 12-port, 24-port, and 48-port Fibre Channel line cards that provide 1, 2, or 4 Gb/s and the 4-port, 10-Gb/s Fibre Channel line card.

The DS-X9304-18K9 line card has 18 1-, 2-, or 4-Gb/s Fibre Channel ports and 4 Gigabit Ethernet ports, and it supports hardware compression and encryption (IP Security [IPsec]).This line card natively supports the Cisco MDS Storage Media Encryption (SME) solution, which encrypts data that is at rest on heterogeneous tape drives and virtual tape libraries (VTLs) in a SAN environment that is using secure IEEE standard Advanced Encryption Standard (AES) 256-bit algorithms.

The Cisco MDS 9222i Multiservice Modular Switch uses the 18/4 architecture of the DS-X9304-18K9 line card and includes native support for Cisco SME along with all the features of the Cisco MDS 9216i Multilayer Fabric Switch.

The Cisco MDS 9148 Switch is an 8-Gb/s Fibre Channel switch providing 48 2-, 4-, or 8-Gb/s Fibre Channel ports. The base license supports 16-, 32-, or 48-port models, but can be expanded to use the 8-port license.

The Cisco MDS DS-X9708-K9 module has eight 10 Gigabit Ethernet multihop-capable FCoE ports. It enables extension of FCoE beyond the access layer into the core of the data center with a full line-rate FCoE module for the Cisco MDS 9500 Series Multilayer Directors.

© 2012 Cisco Systems, Inc

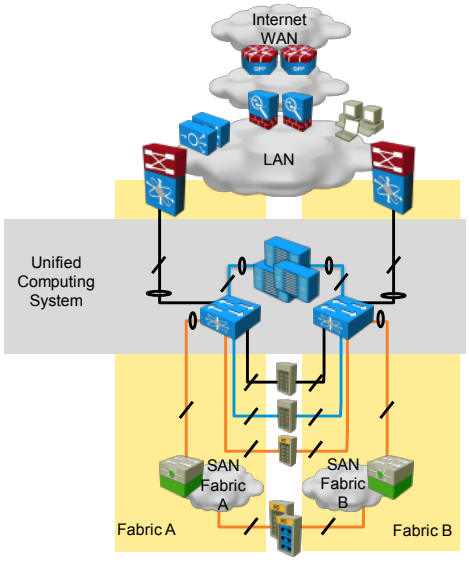
https://t.me/learningnets

Cisco Data Center Compute Architecture

This topic describes the compute component of the Cisco Data Center architectural framework.

Cisco Unified Computing System

- Complete server infrastructure
  - Servers, LAN, SAN, management
- Completely stateless servers
  - No hardware-dependent identities with service profiles
- Ultimate flexible connectivity
  - Automatic NIC failover
  - Multiple logical adapters
  - Administrative traffic engineering
  - Class of service
- Embedded management
  - No dedicated server required
  - High availability
  - Single-server infrastructure management point
- Close integration with ecopartner solutions



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco UCS is a data center platform that represents a pool of compute resources that is connected to existing LAN and SAN core infrastructures. The system is designed to perform the following activities:

- Improve responsiveness to new and changing business demands
- Ease and accelerate design and deployment of new applications and services
- Provide a simple, reliable, and secure infrastructure

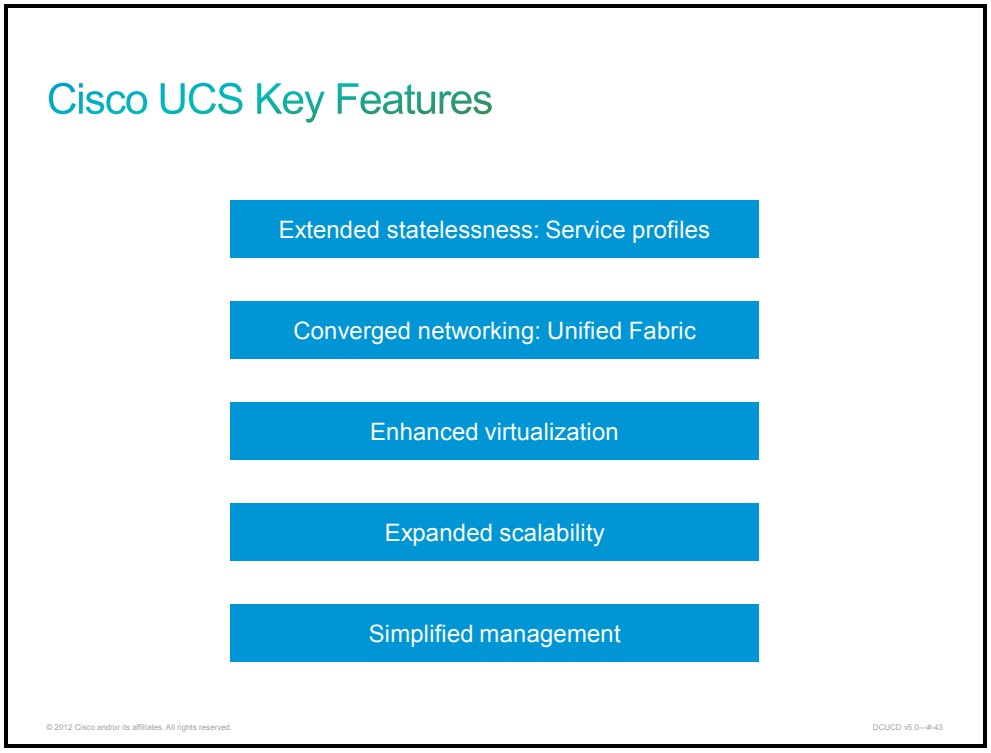
From the perspective of server deployment, the Cisco UCS represents a cable-once, dynamic environment that enables the rapid provisioning of new services. The unified fabric is an integral part of the Cisco UCS, so fewer cables are required to connect the system components, and fewer adapters need to be installed in the servers.

The network part of the system is realized with the fabric extender concept, which results in fewer switch devices.

Fewer system components result in lower power consumption, which makes the Cisco UCS solution greener. You will achieve a better power consumption ratio per computing resource.

The Cisco UCS offers great scalability because a single use of the system can consist of up to 40 chassis, with each chassis hosting up to 8 server blades. This scalability means that the administrator has a single management and configuration point for up to 320 server blades in the future.

https://t.me/learningnets



Extended Statelessness: Service Profiles

Using service profiles, the Cisco UCS is able to abstract the items that make a server unique. This ability allows the system to see the blades as being agnostic. As a result, moving, repurposing, upgrading, and making servers highly available is easy in the Cisco UCS.

Converged Networking: Unified Fabric

A unified fabric consolidates the different network types that are used for various communications on to a single 10 Gigabit Ethernet network. The Cisco UCS uses a single link for all communications (data and storage), with the fabric provided through the system-controlling device. Therefore, no matter which blade that you move a server to, the fabric and communications remain the same.

Enhanced Virtualization

As part of the statelessness, the Cisco UCS is designed to provide visibility and control to the networking adapters within the system. This visibility and control are achieved with the software running on the Cisco UCS and the implementation of the virtual interface card adapters, which, in addition to allowing the creation of virtualized adapters, also increase performance by alleviating the management overhead that is normally managed by the hypervisor.

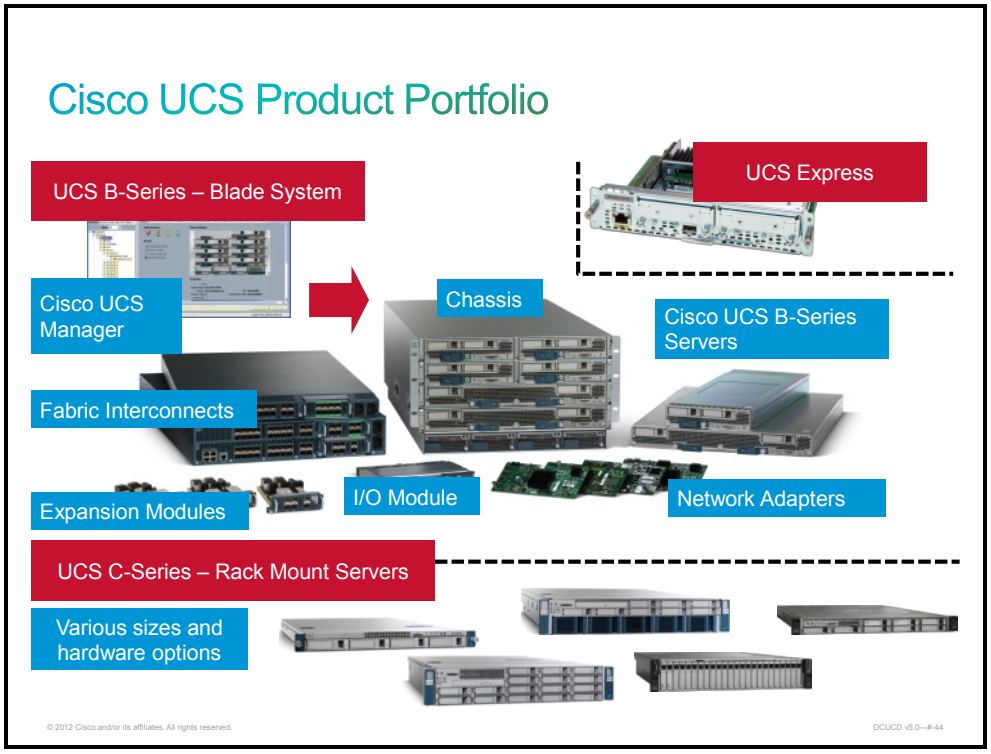
Expanded Scalability

The larger memory footprint of the Cisco UCS B250M1 2-Socket Extended Memory Blade Server offers a number of advantages to applications that require a larger memory space. One of those advantages is the ability to provide the large memory footprint, using standard-sized and lower-cost DIMMs.

Simplified Management

Cisco UCS Manager is embedded device-management software that manages the system from end to end as a single logical entity through either an intuitive GUI, a CLI, or an XML API.

https://t.me/learningnets



Cisco UCS B-Series Servers

The Cisco UCS B-Series server is composed of the following:

- Two Cisco UCS 6100/6200 Series Fabric Interconnect Switches in a cluster deployment:
  - These provide a single point of management by running the Cisco UCS Manager application.
  - They use a single physical medium for LAN and SAN connectivity (consolidating the I/O with FCoE).
  - They provide simultaneous traffic switching in a high-availability configuration.
- Up to 40 Cisco UCS 5108 Server Chassis units per system if Cisco UCS 6140 Fabric Interconnect Switches are used:
  - Two Cisco UCS 2104/2208 XP I/O Modules (or Fabric Extenders) per chassis.
  - Each chassis can host up to 8 server blades (altogether up to 320 server blades).
    - The server is equipped with one or two network adapters.

Cisco UCS C-Series Servers

The Cisco UCS C-Series rack-mount servers come in four different versions:

- **UCS C200 M2 High-Density Rack-Mount Server:** This high-density server has balanced compute performance and I/O flexibility.
- **UCS C210 M2 General-Purpose Rack-Mount Server:** This general-purpose server for workloads requires economical, high-capacity, internal storage.
- **UCS C250 M2 Extended-Memory Rack-Mount Server:** This high-performance, memory-intensive server is meant for virtualized and large-data-set workloads.
- **UCS C460 M1 High-Performance Rack-Mount Server:** This computing-intensive and memory-intensive server is best for enterprise-critical workloads.

https://t.me/learningnets

The M2 Series servers are the next generation of C200, C201, and C250 Series rack-mount servers. The difference between them and the first-generation M1 servers is in the processor type that is supported.

The M2 Series servers support Intel Xeon 5500 and 5600 Series processors, and the first-generation (M1) servers supported Intel Xeon 5500 only.

The difference between Intel Xeon 5500 and 5600 processors is in speed and number of cores. The latter comes in 4- and 6-core setups, whereas the Intel Xeon 5500 is available only in a 4-core setup.

Cisco UCS Express

The Cisco UCS Express is composed of multiple components:

- **Cisco Integrated Services Router Generation 2 (ISR G2):**This router is a single-device network integration that can house all devices such as servers and networking appliances. It uses a multigigabit fabric (MGF) backplane switch to connect compute, switching, and other networking devices together without taxing the router CPU.
- **Cisco Services-Ready Engine (SRE) server module:** This module has been designed to function both as an x86 blade server and as a hardware platform for networking appliances. It is a multipurpose x86 blade.
- **Dedicated blade management:** The SRE blades are managed by Cisco Integrated Management Controller Express (Cisco IMC Express), which provides a configuration, monitoring, and troubleshooting interface. The interface provides consistent management for the Cisco UCS Family, and it has the same design as the Cisco IMC for the Cisco UCS B-Series and C-Series servers.

The Cisco UCS Express can be used as a server virtualization platform by employing VMware vSphere Hypervisor (ESXi) version 4.1 or as a platform for edge services by hosting Microsoft Windows 2000 Server (certified by the Microsoft Windows Hardware Quality Labs and by the Microsoft Server Virtualization Validation Program).

The server modules can be used in a chassis manner with one-, two-, and four-blade-slot options, depending on the ISR G2 model. Models 2911 and 2921 each have one slot, 2951 and 3925 each have two slots, and 3945 has four slots.

© 2012 Cisco Systems, Inc

https://t.me/learningnets

Cisco UCS Second Generation Network Equipment



Product Features and Specs	6120XP	6140XP	6248UP	6296UP
Switch fabric throughput	520 Gb/s	1.04 Tb/s	960 Gb/s	1.92 Tb/s
Switch footprint	1 RU	2 RUs	1 RU	2 RUs
1 Gigabit Ethernet port density	8	16	48	96
10 Gigabit Ethernet port density	26	52	48	96
8 Gb native Fibre Channel port density	6	12	48	96
Port-to-port latency	3.2 µs	3.2 µs	2.0 µs	2.0 µs
Number of VLANs	1024	1024	4096*	4096*

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco UCS 6200 Fabric Interconnect Series

Cisco UCS 6200 Series uses a cut-through architecture, supporting deterministic, low-latency, line-rate 10 Gigabit Ethernet on all ports, switching capacity of 2 terabits (Tb), and 320-Gb/s bandwidth per chassis.

The Cisco UCS 6200 Series is built to consolidate LAN and SAN traffic onto a single unified fabric, saving the capital expenditures (CapEx) and operating expenses (OpEx) associated with multiple parallel networks, different types of adapter cards, switching infrastructure, and cabling within racks.

The unified ports support allows either base or expansion module ports in the interconnect to support direct connections from Cisco UCS to existing native Fibre Channel SANs.

The capability to connect FCoE to native Fibre Channel protects existing storage system investments while dramatically simplifying in-rack cabling.

Cisco UCS 6248UP

The Cisco UCS 6248UP 48-Port Fabric Interconnect is a one-rack-unit (1-RU) 10 Gigabit Ethernet, FCoE and Fiber Channel switch offering up to 960-Gb/s throughput and up to 48 ports. The switch has 32 1/10-Gb/s fixed Ethernet, FCoE, and Fibre Channel ports and one expansion slot.

Cisco UCS 6296UP

The Cisco UCS 6296UP 96-Port Fabric Interconnect is a 2-RU 10 Gigabit Ethernet, FCoE and native Fibre Channel switch offering up to 1920-Gb/s throughput and up to 96 ports. The switch has 48 1/10-Gb/s fixed Ethernet, FCoE and Fiber Channel ports and three expansion slots.

https://t.me/learningnets

Unified Port 16-Port Expansion Module

The Cisco UCS 6200 Series supports an expansion module that can be used to increase the number of 10 Gigabit Ethernet, FCoE, and Fibre Channel ports. This unified port module provides up to 16 ports that can be configured for 10 Gigabit Ethernet, FCoE, and 1/2/4/8-Gb/s native Fibre Channel using the SFP or SFP+3 interface for transparent connectivity with existing Fibre Channel networks.

Cisco UCS 2200 IOM Series

The Cisco UCS 2200 Series extends the I/O fabric between the Cisco UCS 6100 and 6200 Series Fabric Interconnects and the Cisco UCS 5100 Series Blade Server Chassis, enabling a lossless and deterministic FCoE fabric to connect all blades and chassis together. The fabric extender is similar to a distributed line card, so it does not perform any switching and is managed as an extension of the fabric interconnects. This approach removes switching from the chassis, reducing overall infrastructure complexity and enabling Cisco UCS to scale to many chassis without multiplying the number of switches needed, reducing TCO and allowing all chassis to be managed as a single, high-availability management domain.

Cisco UCS 2208XP Fabric Extender

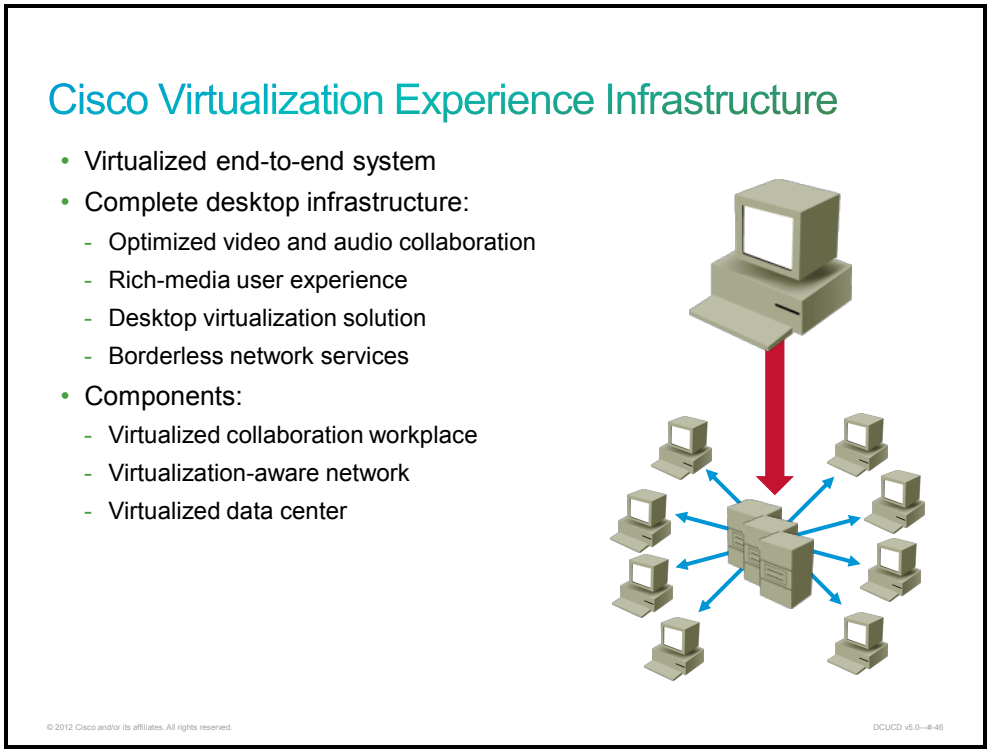
The Cisco UCS 2208XP Fabric Extender has eight 10 Gigabit Ethernet, FCoE-capable, SFP+ ports that connect the blade chassis to the fabric interconnect. Each Cisco UCS 2208XP has 32 10 Gigabit Ethernet ports connected through the midplane to each half-width slot in the chassis. Typically configured in pairs for redundancy, two fabric extenders provide up to 160 Gb/s of I/O to the chassis.

Cisco UCS 2204XP Fabric Extender

The Cisco UCS 2204XP Fabric Extender has four 10 Gigabit Ethernet, FCoE-capable, SFP+ ports that connect the blade chassis to the fabric interconnect. Each Cisco UCS 2204XP has 16 10 Gigabit Ethernet ports connected through the midplane to each half-width slot in the chassis. Typically configured in pairs for redundancy, two fabric extenders provide up to 80 Gb/s of I/O to the chassis.

© 2012 Cisco Systems, Inc

https://t.me/learningnets



Desktop virtualization separates a PC desktop environment from a physical machine by using a client-server model of computing. The model stores the resulting virtualized desktop on a remote central server instead of on the local storage of a remote client. Then, when a user works from a remote desktop client, all of the programs, applications, processes, and data that are used are kept and run centrally. This scenario allows a user to access a desktop on any capable device, such as a traditional PC, notebook computer, smart phone, or thin client. In simple terms, VMs run on the server (for each client) and clients can connect to their virtual computers by using remote desktop software.

Desktop Evolution Drivers

The desktop landscape is evolving due to the following drivers:

- **Disaster recovery and business continuity:** The continuous availability of desktops that is enabled by making high availability and disaster recovery solutions more cost-effective, simpler, and more reliable
- **Desktops as a secure service:** The elimination of the need for moves, adds, or changes, which can allow third parties to access corporate applications in a secure, controlled way
- **Desktop replacement:** The replacement of the "thick-client" PC with centralized, hosted virtual desktops (HVDs) for better control and efficient management
- **User experience:** A "thick-client," PC-like experience is provided with a choice of endpoints, and the system is personalized for the end-user environment, while allowing access from anywhere

https://t.me/learningnets

Cisco VXI Architecture

Cisco VXI is an architectural approach with rich end-to-end services, which delivers the best virtual desktop experience while lowering operational costs.

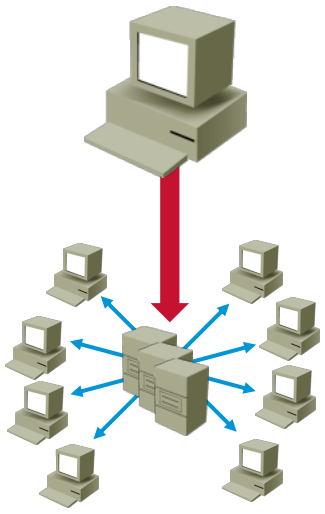
It is a service-optimized desktop virtualization platform that is meant to deliver any application to any device in any workplace environment. The Cisco VXI solution has components from both Cisco and third-party technology partners.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Cisco Virtualization Experience Infrastructure

- Virtualized end-to-end system
- Complete desktop infrastructure:
  - Optimized video and audio collaboration
  - Rich-media user experience
  - Desktop virtualization solution
  - Borderless network services
- Components:
  - Virtualized collaboration workplace
  - Virtualization-aware network
  - Virtualized data center



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco VXI-enabled data center architecture includes these mandatory components:

- Compute (Cisco)
- Hypervisor (technology partner)
- Virtual desktop infrastructure software (technology partner)

The Cisco VXI architecture encompasses the network and data center. The solutions that are already used and in place for a virtualized collaboration workplace are as follows:

- **Cisco Cius:** The Cius platform with a virtualization client is an enterprise tablet for a business that is based on the Google Android operating system. It supports applications for desktop virtualization from both Citrix XenDesktop and VMware View.
- **Cisco integrated zero client:** This device is part of the Cisco Unified 8900 and 9900 IP Phones that have integrated video. The client can connect from an accessory port, Power over Ethernet (PoE), and so it has become an elegant method of deployment. It contains both a phone and zero client.
- **Cisco zero client:** This client is a tower free-standing device with PoE. It can be used with any of the other IP phones.

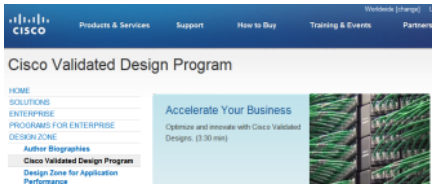
https://t.me/learningnets

Cisco Validated Designs

This topic describes Cisco Validated Designs.

Cisco Validated Designs Overview

- Set of different solution design guides (blueprints)
  - Integrated solutions consolidating Cisco UCS, Nexus, ACE, WAAS, and ecopartner products (Microsoft, EMC, NetApp, VMware, and so on)
- Tested, verified, and documented designs that have defined the following:
  - Customer requirements
  - Solution architecture
  - Hardware and software components (BOM)
  - Infrastructure topology
  - Explanation of components used
- More information available at Cisco.com:
  - Design Zone for Data Center
  - Cisco Validated Design Program



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco Validated Designs consist of systems and solutions that are designed, tested, and documented to facilitate and improve customer deployments. These designs incorporate a wide range of technologies and products into a portfolio of solutions that have been developed to address the business needs of our customers. Cisco Validated Designs are organized by solution areas.

Cisco UCS-based validated designs are blueprints that incorporate not only Cisco UCS but also other Cisco Data Center products and technologies along with applications of various ecopartners (Microsoft, EMC, NetApp, VMware, and so on).

The individual blueprint covers the following aspects:

- Solution requirements, from an application standpoint
- Overall solution architecture with all the components that fit together
- Required hardware and software BOM
- Topology layout
- Description of the components used and their functionalities

https://t.me/learningnets

Microsoft Applications and Solutions

- Validated design guides that focus on the Microsoft products and deployment on Cisco UCS and other Cisco Data Center products and technologies
  - Virtualized Microsoft applications
  - Microsoft Hyper-V and Exchange 2010
  - Microsoft Exchange 2010 on VMware vSphere and NetApp storage
  - Microsoft Sharepoint on Hyper-V



© 2012 Cisco and/or its affiliates. All rights reserved.

A set of the validated designs is focused around the Microsoft solutions and applications, with all the specifics.

Microsoft Hyper-V and Exchange 2010

This design guide presents an end-to-end solution architecture that demonstrates how enterprises can virtualize their Exchange 2010 environment on Cisco Unified Computing System. This is accomplished by using Microsoft Hyper-V virtualization and NetApp storage with application optimization services provided by Cisco Application Control Engine and Cisco Wide Area Application Services.

As enterprises face the critical problem of growing their data center when they reach maximum capacity for space, cabling, power, cooling, and management, they should consider consolidating and virtualizing their application servers into their data center. At the top of the list of servers to consolidate and virtualize are those mission-critical enterprise applications that typically require deployment of a large number of physical servers. Microsoft Exchange 2010 is one such application, as it requires four different server roles for a complete system, and often requires at least double that number of servers to provide minimal redundancy in the server roles. A virtualization infrastructure design that supports an enterprise-level Exchange 2010 deployment also provides the necessary services to address manageability of solution components, disaster recovery, high availability, rapid provisioning, and application optimization and delivery across the WAN.

Microsoft Exchange 2010 on VMware vSphere and NetApp Storage

This design and deployment guide demonstrates how enterprises can apply best practices for VSphere 4.0, Microsoft Exchange 2010, and NetApp Fiber-Attached Storage arrays to run virtualized and native installations of Exchange 2010 servers on the Cisco UCS. VSphere and Exchange 2010 are deployed in a Cisco Data Center Business Advantage architecture that provides redundancy, high availability, network virtualization, network monitoring, and application services. The Cisco Nexus 1000V virtual switch is deployed as part of the VSphere infrastructure and integrated with the Cisco Network Analysis Module appliance to provide

https://t.me/learningnets

visibility and monitoring into the virtual server farm. This solution was validated within a multisite data center scenario with a realistically-sized Exchange deployment using Microsoft Exchange 2010 Load Generator tool to simulate realistic user loads. The goal of the validation was to verify that the Cisco UCS, NetApp storage, and network link sizing was sufficient to accommodate the Load Generator user workloads. Cisco Global Site Selector (GSS) provides site failover in a multisite Exchange environment by communicating securely and optimally with the Cisco ACE load balancer to determine application server health. User connections from branch office and remote sites are optimized across the WAN with Cisco Wide Area Application Services (WAAS).

Microsoft SharePoint on Hyper-V

This document describes architecture for the deployment of the Microsoft Office SharePoint 2007 and Microsoft SQL Server 2008 using the Cisco UCS with Microsoft Hyper-V virtualization. The document provides guidance to engineers interested in deploying Cisco Data Center 3.0 architectures, including network, compute, and application services. The design options and implementation details provided are intended to be a reference for an enterprise data center.

The document is intended for enterprises interested in an end-to-end solution for deploying Microsoft Office SharePoint 2007 and Microsoft SQL Server 2008 in a Microsoft Hyper-V virtualized environment using Cisco UCS, Cisco ACE, and Cisco WAAS technologies.

© 2012 Cisco Systems, Inc

https://t.me/learningnets

VMware Solutions

- Validated design guides that focus on the VMware solutions
- Server virtualization with VMware vSphere
  - Designs incorporating also other vendor applications (Microsoft Exchange, Citrix XenDesktop)
- Desktop virtualization with VMware View
  - Complete VMware VDI solution with EMC or NetApp storage



© 2012 Cisco and/or its affiliates. All rights reserved.

A set of the validated designs is focused around the VMware server and desktop solutions, with all the specifics.

Server Virtualization with VMware vSphere

VMware vSphere is the server virtualization solution that can be used as a basis for many applications. These validated designs describe the deployment of VMware vSphere with Cisco UCS and other Cisco Data Center products, with applications such as Microsoft Exchange 2010, Citrix XenDesktop, and so on.

Desktop Virtualization with VMware View

These validated designs focus on the VDI solutions, namely, using VMware View with necessary server virtualization and hardware, such as VMware vSphere, Cisco UCS, and so on.

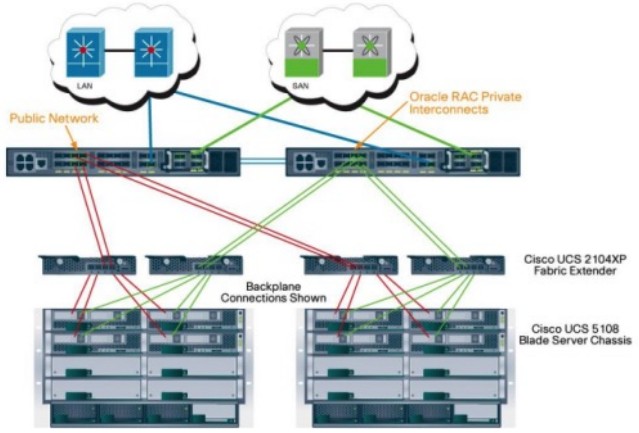
A VMware View 4.5 on Cisco UCS and EMC Unified Storage design guide reports the results of a study evaluating the scalability of VMware View 4.5 on Cisco UCS B-Series blade servers connected to EMC Celerra storage array. Best practice design recommendations and sizing guidelines for large-scale customer deployments are also provided.

A VMware View 4.5 on Cisco UCS and NetApp Storage design guide reports the results of a study evaluating the scalability of VMware View 4.5 on Cisco UCS B-Series blade servers connected to NetApp Storage array. Best practice design recommendations and sizing guidelines for large-scale customer deployments are also provided.

https://t.me/learningnets

Oracle Solutions

- Validated design guides that focus on the Oracle solution
  - Cisco UCS with EMC Clariion validated solution
  - Cisco UCS with NetApp storage validated solution



© 2012 Cisco and/or its affiliates. All rights reserved.

Oracle 11gR2 Real Application Clusters on the Cisco UCS with EMC CLARiiON Storage

This design guide describes how the Cisco UCS can be used with EMC CLARiiON storage systems to implement an Oracle Real Application Clusters (RAC) solution that is an Oracle Certified Configuration. The Cisco Unified Computing System provides the compute, network, and storage access components of the cluster, deployed as a single cohesive system. The result is an implementation that addresses many of the challenges that database administrators and their IT departments face, including needs for a simplified deployment and operation model, high performance for Oracle RAC software, and lower TCO. The document introduces the Cisco UCS and provides instructions for implementing it, and concludes with an analysis of the cluster performance and reliability characteristics.

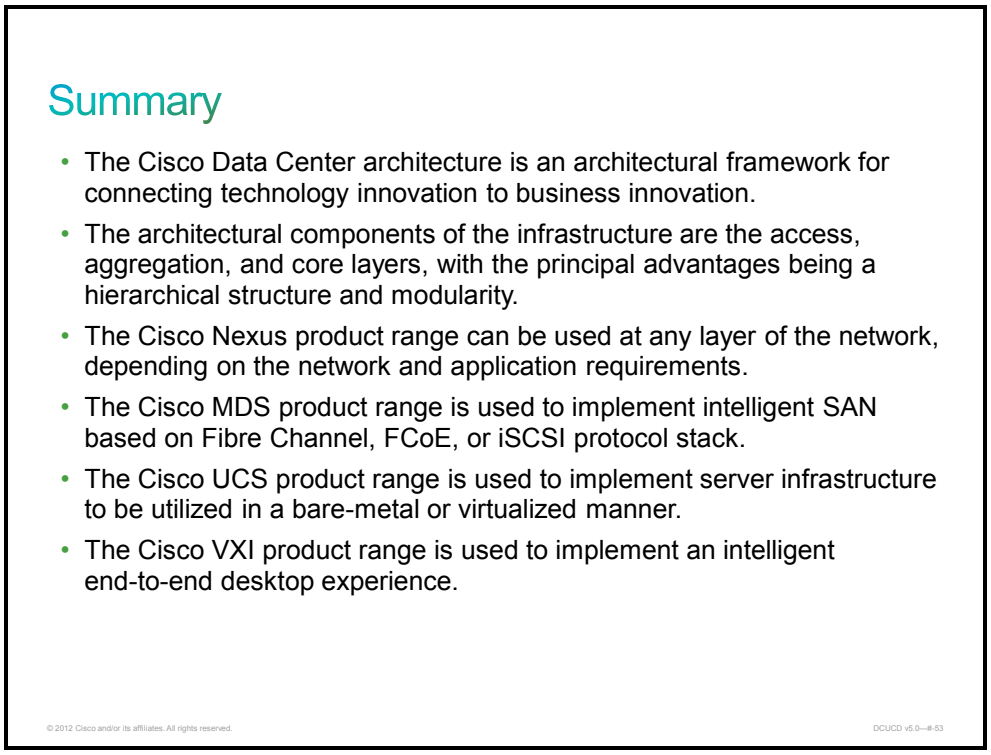
Cisco UCS and NetApp Solution for Oracle Real Application Clusters

This Cisco validated design describes how the Cisco UCS can be used with NetApp FAS unified storage systems to implement a decision-support system (DSS) or online transaction processing (OLTP) database utilizing an Oracle RAC system. The Cisco UCS provides the compute, network, and storage access components of the cluster, deployed as a single cohesive system. The result is an implementation that addresses many of the challenges that database administrators and their IT departments face, including requirements for a simplified deployment and operation model, high performance for Oracle RAC software, and lower TCO. This guide introduces the Cisco UCS and NetApp architecture and provides implementation instructions. It concludes with an analysis of the cluster' performance, reliability characteristics, and data management capabilities.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- http://www.cisco.com/en/US/netsol/ns743/networking_solutions_program_home.html
- http://www.cisco.com/en/US/netsol/ns741/networking_solutions_program_home.html
- http://www.cisco.com/en/US/netsol/ns340/ns394/ns224/index.html
- http://www.cisco.com/en/US/products/hw/switches/index.html
- http://www.cisco.com/en/US/products/hw/ps4159/index.html
- http://www.cisco.com/en/US/products/ps10265/index.html
- http://www.cisco.com/en/US/products/hw/contnetw/index.html

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Summary

This topic summarizes the primary points that were discussed in this module.

Module Summary

- The data center solution is an architectural framework for dynamic networked organizations, which should allow organizations to create services faster, improve profitability, and reduce the risk of implementing new business models.
- The data center solution architecture is influenced by the application characteristics.
- The Cisco Data Center architecture includes technologies for switching, application networking, security, storage, operating system, management, and compute aspects of the data center.
- The architectural components of the infrastructure are the access, aggregation, and core layers, with the principal advantages being a hierarchical structure and modularity.
- The Cisco UCS solution design process consists of assessment, plan, and verification phases.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) Which two items define how much equipment can be deployed in a data center facility? (Choose two.) (Source: Identifying Data Center Solutions)

A) virtualization support
B) available power
C) business demands
D) organizational structure
E) space

Q2) Which data center component is a network load balancer? (Source: Identifying Data Center Solutions)

A) network
A) compute
B) security
C) application services

Q3) Which two trends are predominant in data center solutions? (Choose two.) (Source: Identifying Data Center Solutions)

A) virtualization
B) self-service
C) chargeback and showback
D) consolidation
E) fixed cabling

Q4) Which component of server virtualization abstracts physical hardware from virtual machines? (Source: Identifying Data Center Applications)

A) virtual desktop
B) unified fabric
C) hypervisor
D) Hadoop

Q5) In which cloud computing service category is the Cisco VMDC? (Source: Identifying Data Center Applications)

A) virtual private cloud
B) ITaaS
C) SaaS
D) PaaS

Q6) Which three options are benefits of cloud computing? (Choose three.) (Source: Identifying Cloud Computing)

A) on-demand availability
B) standard provisioning
C) efficient utilization
D) pay per use
E) customer control

https://t.me/learningnets

Q7) Which three features make the Cisco UCS a natural selection for VDI deployments? (Choose three.) (Source: Identifying Cisco Data Center Architecture and Components)

A) server recovery
B) support for large memory
C) service profiles
D) fast server repurpose
E) virtualized adapters

Q8) Which feature enables virtual machine-level network visibility? (Source: Identifying Cisco Data Center Architecture and Components)

A) unified fabric
B) VM-FEX
C) OTV
D) FabricPath

© 2012 Cisco Systems, Inc

https://t.me/learningnets

Module Self-Check Answer Key

Q1) B, E

Q2) D

Q3) A, D

Q4) C

Q5) B

Q6) A, C, D

Q7) B, C, E

Q8) B

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module 2

Assess Data Center Computing Requirements

Overview

To design a sound data center solution with architecture that meets user requirements, the design process needs to follow proper steps. Any design needs to begin with an audit and analysis of requirements for either a new or existing environment. For an existing environment, the benefit is that by using the analysis tools, more quality input information can be gathered for the design.

This module recognizes the design process and steps, identifies the relevant network, storage, and compute historical performance characteristics, and discusses the relevant characteristics and functions of the reconnaissance and analysis tools for the Cisco Unified Computing System (UCS) solution.

Module Objectives

Upon completing this module, you will be able to design a solution by applying the recommended practice exercises and assess the requirements and performance characteristics of the data center computing solutions. This ability includes being able to meet these objectives:

- Define the tasks and phases of the design process for the Cisco UCS solution
- Assess the requirements and performance characteristics for the given data center computing solutions
- Use the reconnaissance and analysis tools to examine performance characteristics of the given computing solution

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Lesson 1

Defining a Cisco Unified Computing System Solution Design

Overview

This lesson defines the tasks and phases of the design process for the Cisco Unified Computing System (UCS) solution.

Objectives

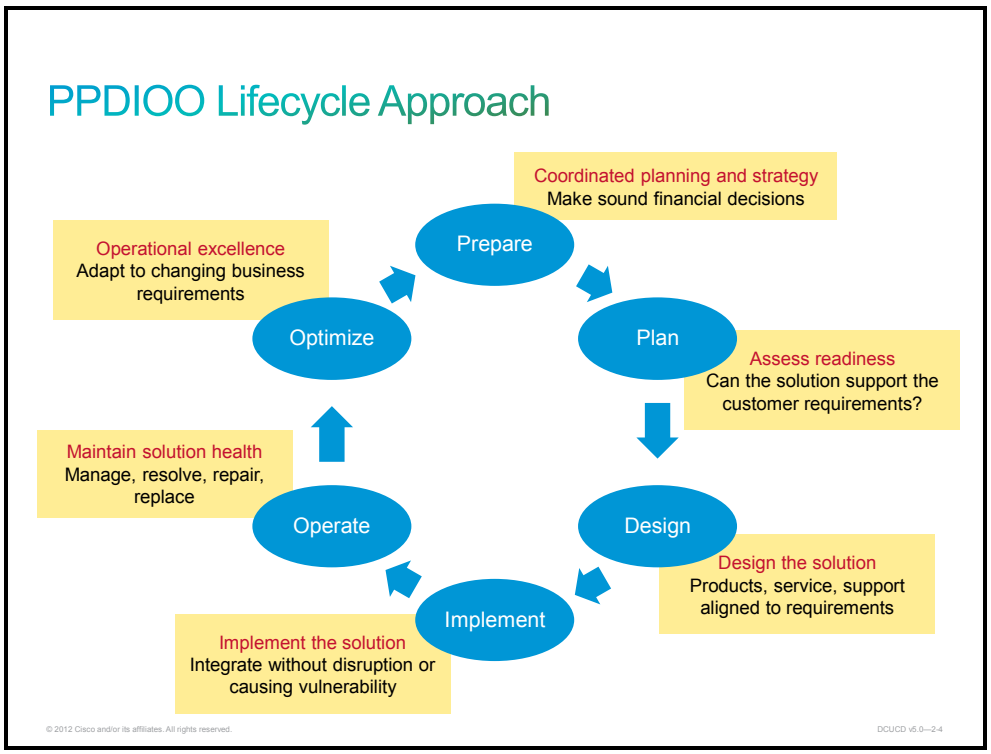
Upon completing this lesson, you will be able to define the tasks and phases of the design process for the Cisco UCS solution. You will be able to meet the following objectives:

- Describe the design process for the Cisco UCS solution
- Evaluate the design process phases for the Cisco UCS solution
- Assess the deliverables of the Cisco UCS solution

https://t.me/learningnets

Design Process

This topic describes the design process for the Cisco UCS solution.



Cisco has formalized the lifecycle of a solution into six phases: Prepare, Plan, Design, Implement, Operate, and Optimize (PPDIOO). For the design of the Cisco Unified Computing solution, the first three phases are used.

The PPDIOO solution lifecycle approach reflects the lifecycle phases of a standard solution. The PPDIOO phases are as follows:

- **Prepare:** The prepare phase involves establishing the organizational requirements, developing a solution strategy, and proposing a high-level conceptual architecture that identifies technologies that can best support the architecture. The prepare phase can establish a financial justification for a solution strategy by assessing the business case for the proposed architecture.
- **Plan:** The plan phase involves identifying initial solution requirements based on goals, facilities, user needs, and so on. The plan phase involves characterizing sites and assessing any existing environment and performing a gap analysis to determine whether the existing system infrastructure, sites, and operational environment are able to support the proposed system. A project plan is useful to help manage the tasks, responsibilities, critical milestones, and resources required to implement changes to the solution. The project plan should align with the scope, cost, and resource parameters established in the original business requirements.
- **Design:** The initial requirements that were derived in the planning phase direct the activities of the solution design specialists. The solution design specification is a comprehensive detailed design that meets current business and technical requirements and incorporates specifications to support availability, reliability, security, scalability, and performance. The design specification is the basis for the implementation activities.

https://t.me/learningnets

- **Implement:** After the design has been approved, implementation (and verification) begins. The solution is built or additional components are incorporated according to the design specifications, with the goal of integrating devices without disrupting the existing environment or creating points of vulnerability.
- **Operate:** Operation is the final test of the appropriateness of the design. The operational phase involves maintaining solution health through day-to-day operations, including maintaining high availability and reducing expenses. The fault detection, correction, and performance monitoring that occur in daily operations provide initial data for the optimization phase.
- **Optimize:** The optimization phase involves proactive management of the solution. The goal of proactive management is to identify and resolve issues before they affect the organization. Reactive fault detection and correction (troubleshooting) are needed when proactive management cannot predict and mitigate failures. In the PPDIOO process, the optimization phase may prompt a network redesign if too many solution problems and errors arise, if performance does not meet expectations, or if new applications are identified to support organizational and technical requirements.

Note	Although design is listed as one of the six PPDIOO phases, some design elements may be present in all the other phases.

https://t.me/learningnets

Benefits of PPDIOO Lifecycle Approach

- Lower the total cost of ownership:
  - Reduce operating expenses
  - Improve efficiency
  - Accelerate successful implementation
- Increase solution availability:
  - Assess ability to accommodate the proposed system
  - Produce a sound design
  - Stage and test; validate operation
  - Proactively monitor and address availability, security
- Improve business agility:
  - Establish business requirements and technology strategies
  - Ready your sites to support the system
  - Continually enhance performance
- Speed access to applications:
  - Access and improve operational preparedness
  - Improve service, delivery efficiency, and effectiveness
  - Improve availability, reliability, and stability of the solution

© 2012 Cisco and/or its affiliates. All rights reserved.

The solution lifecycle approach provides four main benefits:

- Lowering the total cost of solution ownership
- Increasing solution availability
- Improving business agility
- Speeding access to applications and services

The total cost of solution ownership is lowered by these actions:

- Identifying and validating technology requirements
- Planning for infrastructure changes and resource requirements
- Developing a sound solution design aligned with technical requirements and business goals
- Accelerating successful implementation
- Improving the efficiency of your solution and of the staff supporting it
- Reducing operating expenses by improving the efficiency of operation processes and tools

Solution availability is increased by these actions:

- Assessing the security state of the solution and its ability to support the proposed design
- Specifying the correct set of hardware and software releases and keeping them operational and current
- Producing a sound operations design and validating the solution operation
- Staging and testing the proposed system before deployment
- Improving staff skills
- Proactively monitoring the system and assessing availability trends and alerts
- Proactively identifying security breaches and defining remediation plans

https://t.me/learningnets

Business agility is improved by these actions:

- Establishing business requirements and technology strategies
- Readying sites to support the system that you want to implement
- Integrating technical requirements and business goals into a detailed design and demonstrating that the solution is functioning as specified
- Expertly installing, configuring, and integrating system components
- Continually enhancing performance

Access to applications and services is accelerated by these actions:

- Assessing and improving operational preparedness to support current and planned solution technologies and services
- Improving service delivery efficiency and effectiveness by increasing availability, resource capacity, and performance
- Improving the availability, reliability, and stability of the solution and the applications running on that solution
- Managing and resolving problems affecting your system and keeping software applications current

https://t.me/learningnets

Design Methodology Under PPDIOO

Three steps in the design methodology:

1. Identify the customer requirements
2. Characterize the existing environment
3. Design the topology and solution architecture

© 2012 Cisco and/or its affiliates. All rights reserved.

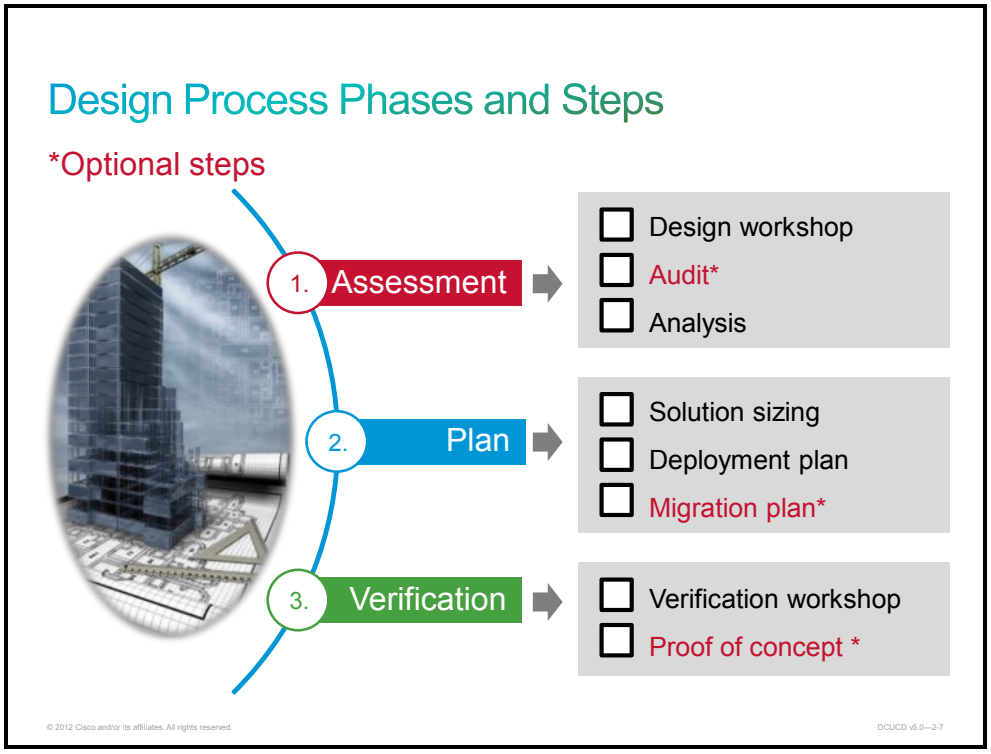
The design methodology under PPDIOO consists of three basic steps:

Step 1 **Identify customer requirements:** In this step, key decision makers identify the initial requirements. Based on these requirements, a high-level conceptual architecture is proposed. This step is typically done within the PPDIOO prepare phase.

Step 2 **Characterize the existing network and sites:** The plan phase involves characterizing sites and assessing any existing networks and performing a gap analysis to determine whether the existing system infrastructure, sites, and operational environment can support the proposed system. Characterization of the existing environment includes the existing environment audit and analysis. During the audit, the existing environment is thoroughly checked for integrity and quality. During the analysis, environment behavior (traffic, congestion, and so on) is analyzed. This investigation is typically done within the PPDIOO plan phase.

Step 3 **Design the network topology and solutions:** In this step, you develop the detailed design. Decisions on solution infrastructure, intelligent services, and solutions are made. You may also build a pilot or prototype solution to verify the design. You also write a detailed design document.

https://t.me/learningnets



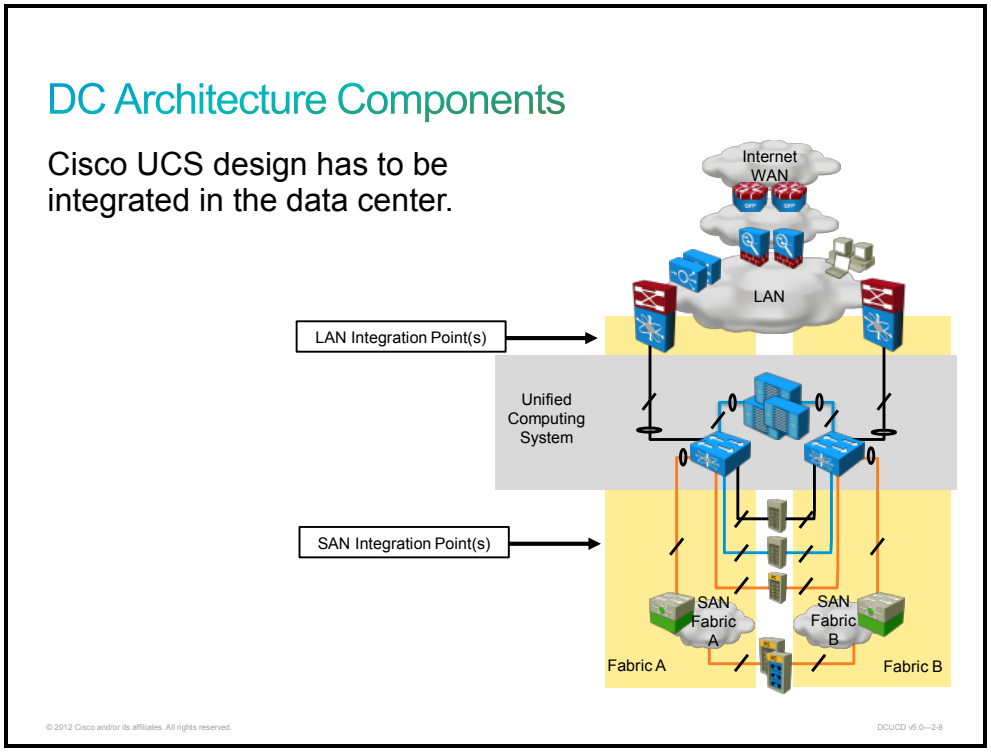
To design a solution that meets customer needs, the organizational goals, organizational constraints, technical goals, and technical constraints must first be identified. In general, the design process can be divided into three major phases:

- **Assessment phase:** This phase is vital for the project to be successful and to meet the customer needs and expectations. In this phase, all information that is relevant for the design has to be collected.
- **Plan phase:** In this phase, the solution architect creates the solution architecture by using the assessment phase results as input data.
- **Verification phase:** To ensure that the designed solution architecture meets customer expectations, the solution should be verified and confirmed by the customer.

Each phase of the design process has steps that need to be taken in order to complete that phase. Some of the steps are mandatory and some are optional. The decision about which steps are necessary is governed by the customer requirements and the type of project (for example, new deployment versus migration).

The checklist in the figure can aid the effort to track the design process progress as well as the completed and open actions.

https://t.me/learningnets



The design of the Cisco UCS solution needs to take into account other data center and IT infrastructure components and segments as the design needs to integrate and coexist with them.

From the design perspective, the Cisco UCS needs to integrate with existing or new LAN (data network) and SAN (storage network and devices) infrastructure and incorporate the following aspects:

- Number and type of LAN uplinks (that is, fiber optics versus copper, terminating on a single upstream switch versus multiple switches, and so on)
- VLAN numbering scheme and policies
- High availability on Layer 2 as well as on Layer 3 (default gateway location and connectivity)
- Number and type of SAN uplinks (that is, length, short-range versus long-range, port channel capability, and so on)
- Required and available throughput
- Virtual storage area network (VSAN) numbering scheme and policies
- Net present value (NPV) and N-Port ID Virtualization (NPIV) availability and support
- Network segment types and special requirements, such as disjointed Layer 2 domains
- Type and attachment options for storage devices (that is, directly attached storage versus Fibre Channel SAN-based versus network-attached storage)

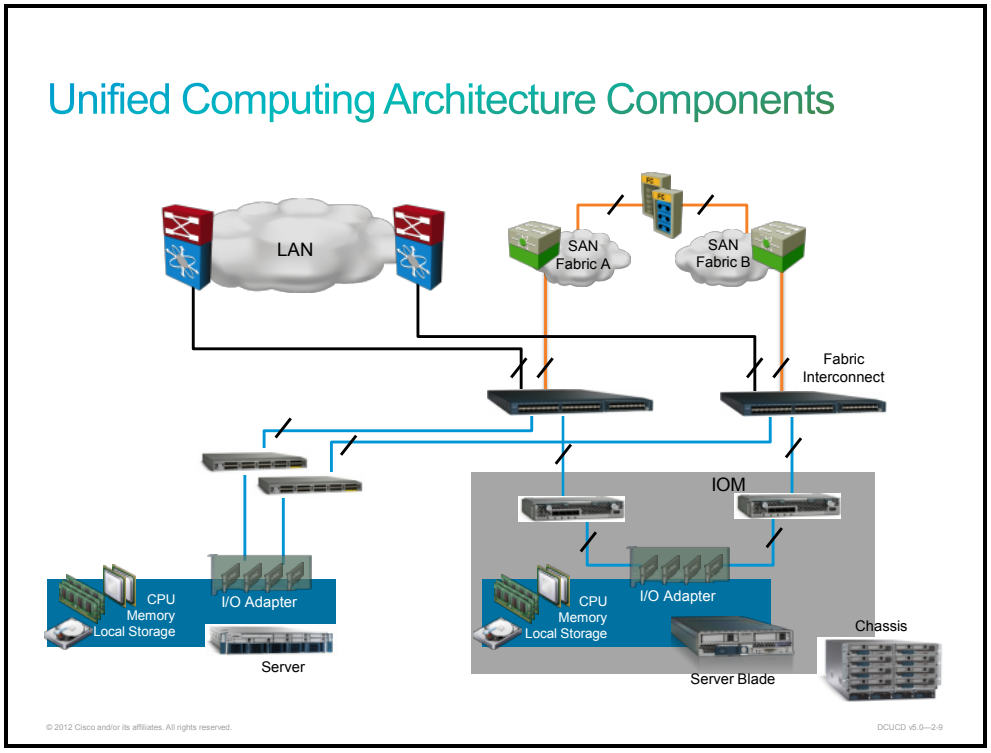
https://t.me/learningnets

The integration of Cisco UCS with LAN and SAN means that the Cisco UCS designer needs to also assess existing or new LAN and SAN and, if necessary, either:

- Adjust the Cisco UCS design with regard to equipment components as well as logical design (that is, VLANs, numbering and addressing, high-availability design aspects like port channels versus dynamic or administrative pinning, implementing traffic engineering to send demilitarized zone (DMZ)-destined traffic to proper segment, and so on).
- List which LAN and/or SAN components need to be added in order to adequately integrate Cisco UCS with both

From the overall data center fabric perspective, the fabric design (fabric being LAN and SAN) should be implemented in a dual fabric fashion, meaning there are two distinct paths for server connectivity in order to achieve proper redundancy and high availability. Typically Fabric A and Fabric B are the two fabrics, where on the LAN side they can coexist on the same core equipment, but in the SAN they are typically implemented in at least two separate core devices.

https://t.me/learningnets



The UCS solution architecture includes several components. The figure emphasizes those that need to be designed.

The two distinct types of Cisco UCS equipment are B-Series blade system and C-Series rack-mount servers, which can be incorporated into a B-Series system.

From the Cisco UCS server architecture perspective, the following components need to be selected and sized:

- CPU
- Memory
- Network adapter for LAN, SAN, or Cisco Unified Fabric
- Server (that is, blade or rack-mount)
- Local storage (that is, disk drive, SSD, or flash/USB)
- Local storage controller (that is, Redundant Array of Independent Disks [RAID])

From the B-Series system, additional components need to be selected:

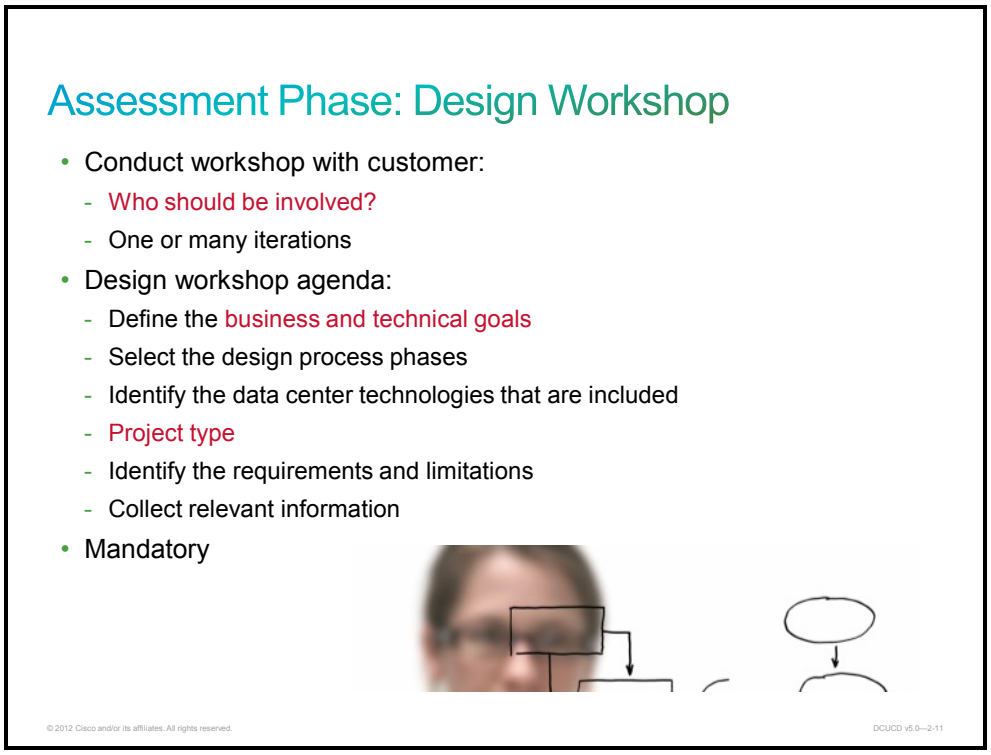
- Fabric Interconnects
- I/O module
- Chassis
- Cabling

This scheme can be used as a checklist when in the process of sizing the Cisco UCS solution.

https://t.me/learningnets

Design Process Phases

This topic evaluates the design process phases for the Cisco UCS solution.



The first action of the design process and the first step of the assessment phase is the design workshop. The workshop has to be conducted with proper customer IT personnel and can take several iterations in order to collect the relevant and valid information. In the design workshop, a draft high-level architecture may already be defined.

The high-level agenda of the design workshop should include these tasks:

- **Define the business goals:** This step is important for several reasons. First, you should ensure that the project follows customer business goals and ensure that the project is successful. With the list of goals defined, the solution architects can then learn and write down what the customer wants to achieve with the project and what the customer expects from the project.
- **Define the technical goals:** This step ensures that the project also follows customer technical goals and expectations and thus also ensures that the project is successful. With this information, the solution architect will know the technical requirements of the project.
- **Identify the data center technologies:** This task is used to clarify which data center technologies are covered by the project and is also the basis for how the experts determine what is needed for the solution design.
- **Define the project type:** There are two main types of projects— new deployments or the migration of existing solutions.

https://t.me/learningnets

- **Identify the requirements and limitations:** The requirements and limitations are the details that significantly govern the equipment selection, the connectivity that is used, the integration level, and the equipment configuration details. For migration projects, this step is the first part of identifying relevant requirements and limitations. The second part is the audit of the existing environment with the proper reconnaissance and analysis tools.

The workshop can be conducted in person or it can be accomplished virtually by using Cisco WebEx or a Cisco TelePresence solution.

The design workshop is a mandatory step of the assessment phase because without it, there is no relevant information upon which the design can be created.



It is very important to gather all of the relevant people in the design workshop to cover all of the aspects of the solution. (The design workshop can be a multiday event.)

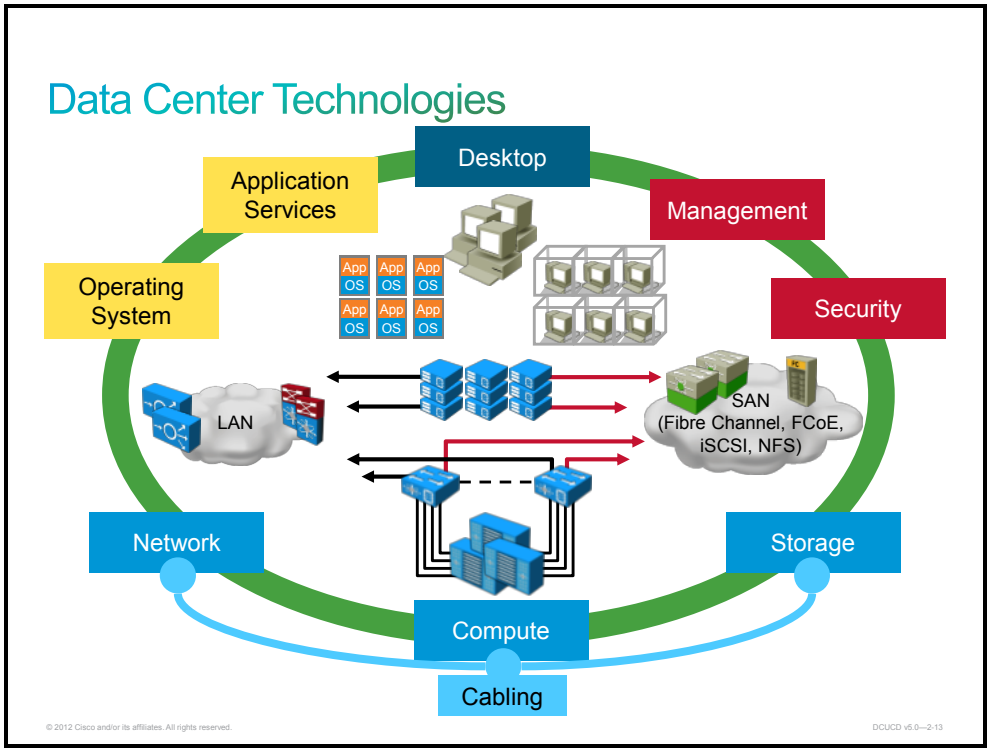
The Cisco UCS solution is effectively part of the data center and, as such, the system must comply with all data center policies and demands. The following customer personnel must attend the workshop (or should at least provide information that is requested by the solution architect):

- **Facility administrators:** They are in charge of the physical facility and have the relevant information about environmental conditions like available power, cooling capacity, available space and floor loading, cabling, physical security, and so on. This information is important for the physical deployment design and can also influence the equipment selection.
- **Network administrators:** They ensure that the network properly connects all the bits and pieces of the data center and thus also the equipment of the future Cisco UCS solution. It is vital to receive all the information about the network: throughput, port and connector types, Layer 2 and Layer 3 topologies, high-availability mechanisms, addressing, and so on. The network administrators may report certain requirements for the solution.

https://t.me/learningnets

- **Storage administrators:** They deal with the relevant information, which encompasses storage capacity (available and used), storage design and redundancy mechanisms (logical unit numbers [LUNs], RAID groups, service processor ports, and failover), storage access speed, type (Fibre Channel, Internet Small Computer Systems Interface [iSCSI], Network File System [NFS]), replication policy and access security, and so on).
- **Server and application administrators:** They know the details of the server requirements, operating systems, and application dependencies and interrelations. The solution architect learns which operating systems and versions are or will be used, what the requirements of the operating systems are from the connectivity perspective (one network interface card [NIC], two NICs, NIC teaming, and so on), and which applications will be deployed on what operating systems and what the application requirements will be (connectivity, high availability, traffic throughput, typical memory and CPU utilization, and so on).
- **Security administrators:** The solution limitations can also be known from the customer security requirements (for example, the need to use separate physical VMware vSphere hosts for a DMZ and private segments). The security policy also defines the control of equipment administrative access as well as the allowed and restricted services (for example, Telnet versus Secure Shell [SSH]), and so on.

https://t.me/learningnets



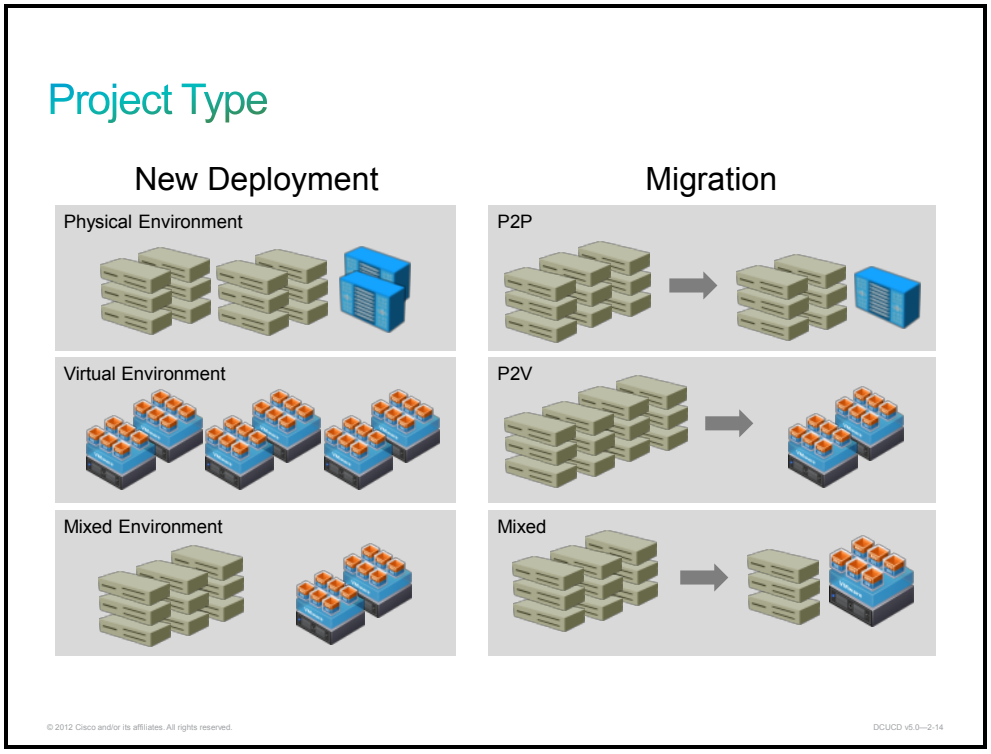
The Cisco Data Center architectural framework encompasses technologies that address the network, storage, compute, operating system, application services, management, and security aspects of the solution.

Because each project is different, in the design workshop, the solution architect should define the aspects and technologies that should be part of the design with the customer.

The technologies should be paired with relevant equipment when discussing needs with the customer so as to more closely match the requirements and to also bring out features that the customer is either unaware of or does not think are relevant.

Thus, the attendees in the design workshop should discuss the Cisco Nexus switching portfolio from Nexus 7000 to Nexus 1000V, the UCS B-Series and C-Series products, Cisco MDS SAN switches, application service equipment like Cisco Application Control Engine (ACE) and Cisco Wide Area Application Services (WAAS), hypervisor integration with VMware vSphere and other vendor virtualization systems, management applications like Cisco UCS Manager and Cisco Data Center Network Manager (DCNM), and so on.

https://t.me/learningnets



The project type determines not only whether some of the design process steps can or should be taken, but also the shape of the solution design itself.

There are two major project types:

- **New deployment:** Where there is no existing environment that could be inspected to gather the requirements, the design workshop is conducted with the customer in order to gather all relevant information. It is very important that the design workshop is very thorough in order to collect all of the relevant information and to understand customer expectations.
- **Migration of existing environment:** Where the customer has an existing solution, which reached the end of its lifecycle and needs to be replaced with a new solution, a thorough audit in the assessment phase should be conducted to gather relevant performance characteristics and to verify the customer requirements. The audit result can also be used to help the customer to come up with proper requirements (from the aspect of growth, resource utilization, current inventory, and so on).

Both a new deployment and a migration project can fit different environments. Environment choices include a purely physical server deployment, a pure virtual server deployment, or a mixed physical and virtual server deployment.

With migration, the difference is that it can also include a transformation from a physical into a virtual or mixed environment, or even a migration from one virtual environment to another.

https://t.me/learningnets

Assessment Phase: Audit

- Conducted for migration projects
- Audit aspects:
  - Resource utilization levels
  - High-availability requirements
  - Security policies
  - Network, storage, servers, and applications
  - Dependencies and interrelations
  - Limitations
- Select and use the reconnaissance and analysis tools
- Optional

© 2012 Cisco and/or its affiliates. All rights reserved.

The audit step of the assessment should typically be taken for the migration projects. It is not necessary, but it is strongly advised in order to audit the existing environment.

For the proper design, it is of the utmost importance to have the relevant information upon which the design is based:

- Memory and CPU resources
- Storage space
- Inventory details
- Historical growth report
- Security policies that are in place
- High-availability mechanisms
- Dependencies between the data center elements (that is, applications, operating system, server, storage, and so on)
- The limitations of the current infrastructure

From the description of the audit aspect, it is clear that some information should be collected over a longer time in order to be relevant. (For example, the levels of server memory and CPU utilization that are measured over the weekend are significantly lower than during weekdays.) Other details can be gathered by inspecting the equipment configuration (for example, administrative access, logging, Simple Network Management Protocol (SNMP) management, and so on).

Information can be collected with the various reconnaissance and analysis tools that are available from different vendors. If the project involves a migration to a VMware vSphere environment from physical servers, the VMware Capacity Planner will help with collecting the information about the servers, and it can even suggest the type of servers that are appropriate for the new design (regarding processor power, memory size, and so on).

https://t.me/learningnets

Assessment Phase: Analysis

- Baseline and optimize the requirements:
  - Computer—memory, CPU, storage
  - Network—throughput, redundancy
  - Storage—space, features, IOPS
  - Operating system and applications specifics
- Identify limitations and maximums
- Mandatory

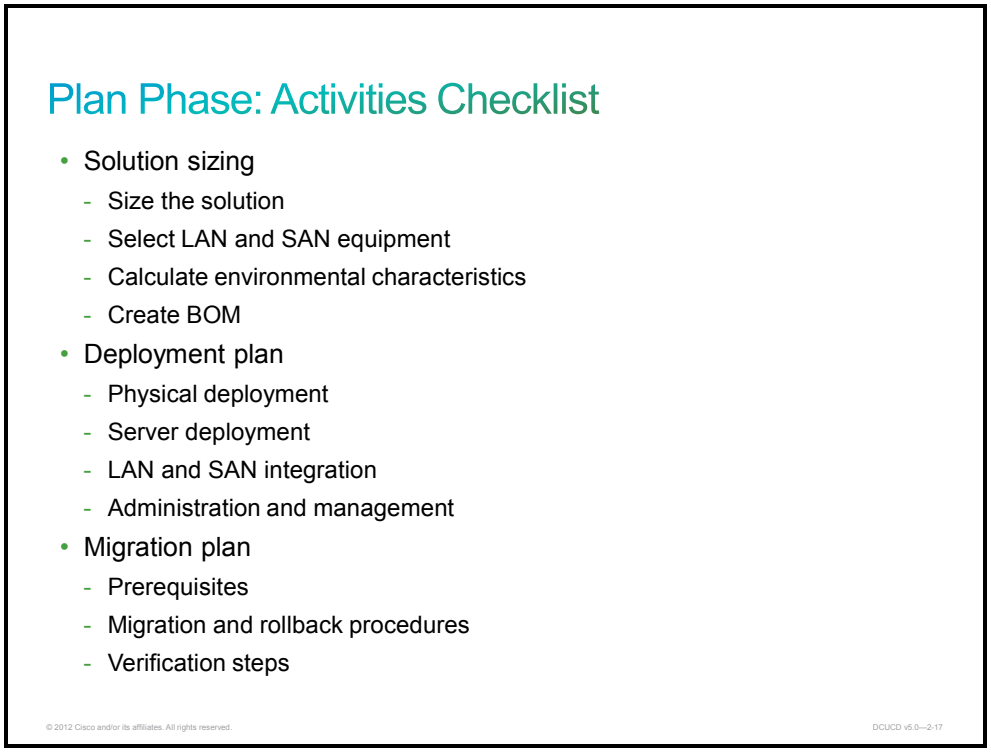
© 2012 Cisco and/or its affiliates. All rights reserved.

The analysis is the last part of the assessment phase. The solution architects must review all of the collected information and then select only the important details.

The architects must baseline and optimize the requirements. The requirements can then be directly translated into the proper equipment, software, and configurations.

The analysis is mandatory for creating a designed solution that will meet project goals and customer expectations.

https://t.me/learningnets



Once the assessment phase is completed and the solution architect has the analysis results, the design or plan phase can commence.

This phase (like the assessment phase) contains several steps and substeps. Some steps are mandatory and some are optional. There are three major steps:

- **Solution sizing:** In this step, the hardware and software that are used will be defined.
  - First, the Cisco UCS must be sized, and it must be decided whether B-Series or C-Series equipment should be used. This decision can be a complex process, and the architect must consider all the requirements and limitations of the operating systems and applications that will run in addition to the servers.
  - Second, LAN and SAN equipment that is required for connecting the system has to be selected. The equipment can be small form-factor pluggable (SFP) modules, a new module, or even Cisco Nexus and Multilayer Director Switch (MDS) switches and licenses.
  - Third, once the equipment is selected, the environmental requirements need to be determined by using the Cisco Power Calculator. You will need to calculate the power, cooling, and weight measurements for the Cisco UCS, Nexus, MDS, and other devices.
  - Lastly, the Bill of Materials (BOM), which is a detailed list of the equipment parts, needs to be created. The BOM includes not only the Cisco UCS or Nexus products but also all of the necessary patch cables, power inlets, and so on.

https://t.me/learningnets

- **Deployment plan:** This step can be divided into the following substeps:
  - The physical deployment plan, which details where and how the equipment will be placed into the racks for racking and stacking.
  - The server deployment plan, which details the server infrastructure configuration, such as the LAN and SAN access layer configuration; VLANs and VSANs; port connectivity; MAC, world wide name (WWN), and universally unique identifier (UUID) addressing; and management access, firmware versions, and high-availability settings. In the case of the Cisco UCS, all details are defined from a single management point in the Cisco UCS Manager.
  - The LAN and SAN integration plan, which details the physical connectivity and configuration of core data center devices (the Cisco Nexus and MDS switches, VLAN and VSAN configuration on the core side, and the high-availability settings).
  - The administration and management plan, detailing how the new solution will be managed and how it integrates into the existing management infrastructure (when present).
- **Migration plan:** Applicable for migration projects, this plan needs to detail when, how, and with what technologies the migration from an existing solution to a new deployment will be performed. A vital part of the migration plan is the series of verification steps that confirm or disprove the successfulness of migration. Equally important (although hopefully not used) are the rollback procedures that should be taken in case of failures or problems during migration.

Different deployments have different requirements and thus different designs. There are typical solutions to common requirements that are described in the Cisco validated designs (for example, Citrix XenDesktop with VMware and the Cisco UCS, an Oracle database and the Cisco UCS, and so on).

https://t.me/learningnets

Verification Phase

- Verification workshop with customer—mandatory
  - Confirm and sign off on the solution design
- Proof of concept—optional
  - Implement partial solution in a lab environment
  - Confirm and verify the designed solution



© 2012 Cisco and/or its affiliates. All rights reserved.

Once the design phase is completed, the solution must be verified and approved by the customer. This approval is typically achieved by conducting a verification workshop with the customer personnel who are responsible for the project. The customer also receives the complete information about the designed solution.

The second step of the verification phase can be the proof of concept, which is how the customer and architect can confirm that the proposed solution meets the expected goals. The proof of concept is typically a smaller set of the proposed solution that encompasses all the vital and necessary components to confirm the proper operation.

The solution architect must define the subset of the designed solution that needs to be tested and must conduct the necessary tests with expected results.

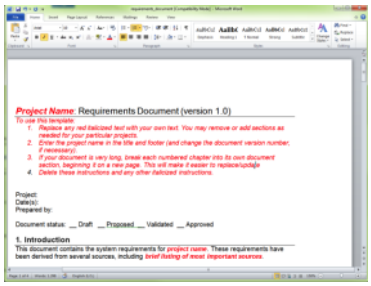
https://t.me/learningnets

Design Deliverables

This topic describes how to assess the deliverables of the Cisco UCS solution.

Deliverables: Customer Initial Input

- Customer requirements document:
  - Created by customer
- Sections:
  - Summary:
    - Description of the existing environment
    - Business and technical goals
    - Project scope
    - Expected outcome of the project
  - Customer requirements:
    - List of services and applications with goals
    - Solution requirements—performance, redundancy and high-availability, security, management, and so on



© 2012 Cisco and/or its affiliates. All rights reserved.

Every project should begin with a clear understanding of the requirement of the customer. Thus the Customer Requirements Document (CRD) should be used to detail the customer requirements for a project for which a solution will be proposed. It must be completed upon the request of the department or project leader from the customer team.

The following sections should be part of the CRD:

- **Existing environment:** Existing environment describes the customer and desired technology/solution that applies to this project.
- **Expected outcome:** Expected outcome provides an overview of the intentions, future direction, and summarizing of the services and applications that the customer intends to introduce. Define what is the strategic impact of this project to the customer (for example, is the requested technology/solution required in order to solve an important issue, make the customer more profitable, give the customer a competitive edge over competition, and so on).
- **Project scope:** Project scope defines the range of the project with regard to the design, (for example, which data center components are involved, which technologies should be covered, and so on).
- **List of services and applications with goals:** List of services and applications with goals lists the objectives and requirements for this service (that is, details about the type of services the customer plans to offer and introduce with the proposed solution). Apart from connectivity, security, and so on, this includes details about the applications and services planned to be deployed.

https://t.me/learningnets

- **Solution requirements:** Solution requirements encompass the following characteristics for the solution as a whole as well as for the individual parts:
  - Requirements considering system availability, behavior under failure scenarios, service restoration
  - All performance requirements
  - All security requirements
  - All critical features required in order to provide this service/application, including those that are not yet implemented
  - All aspects of solution management features like:
    - Service fulfillment (that is, service provisioning and configuration management)
    - Service assurance (that is, fault management and performance management)
    - Billing and accounting

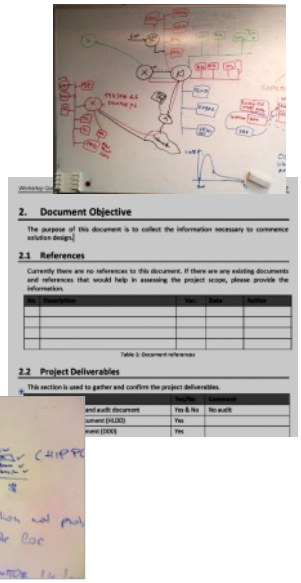
It is also advisable that the CRD holds the high-level timelines of the project so that the solution architect can plan accordingly.

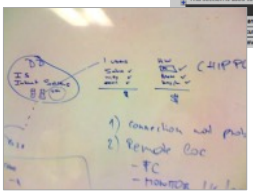
The CRD thus clearly defines what the customer wants from the solution, and it is also the basis and input information for the assessment phase.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Deliverables: Assessment Phase

- Design workshop:
  - Questionnaire
  - Meeting minutes
- Analysis document:
  - Relevant information from design workshop:
    - Business and technical goals
    - Existing environment (if applicable)
  - Audit results—existing environment characteristics
  - Analysis result with detailed description:
    - Design requirements
    - Design limitations





© 2012 Cisco and/or its affiliates. All rights reserved.

Each phase of the design process should result in documents that are necessary not only for tracking the efforts of the design team, but also for presenting the results and progress to the customer.

The supporting documentation for the assessment phase can include the following:

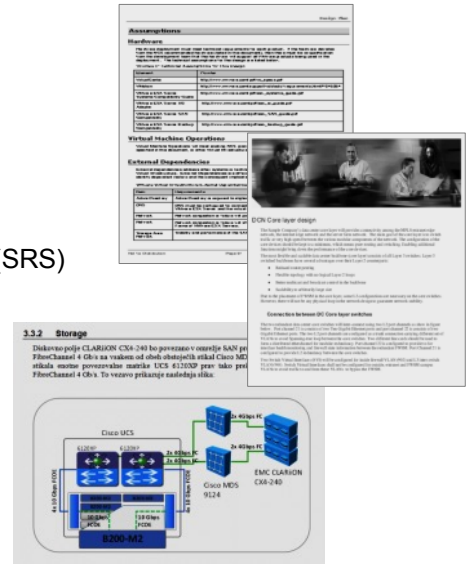
- **Questionnaire:** The questionnaire can be distributed to the customer personnel in order to prepare for the design workshop or to provide relevant information in written form when verbal communication is not possible.
- **Meeting minutes:** This document contains the relevant information from the design workshop.

The assessment phase should finally result in the analysis document, which must include all of the information that is gathered in the assessment phase.

https://t.me/learningnets

Deliverables: Plan Phase

- High-level design (HLD):
  - Conceptual design description
  - High-level equipment list
- Low-level design (LLD):
  - Detailed description of the design
  - Bill of Materials (BoM)
- Site requirements specification (SRS)
- Site survey form (SSF)
- Migration plan (MP)



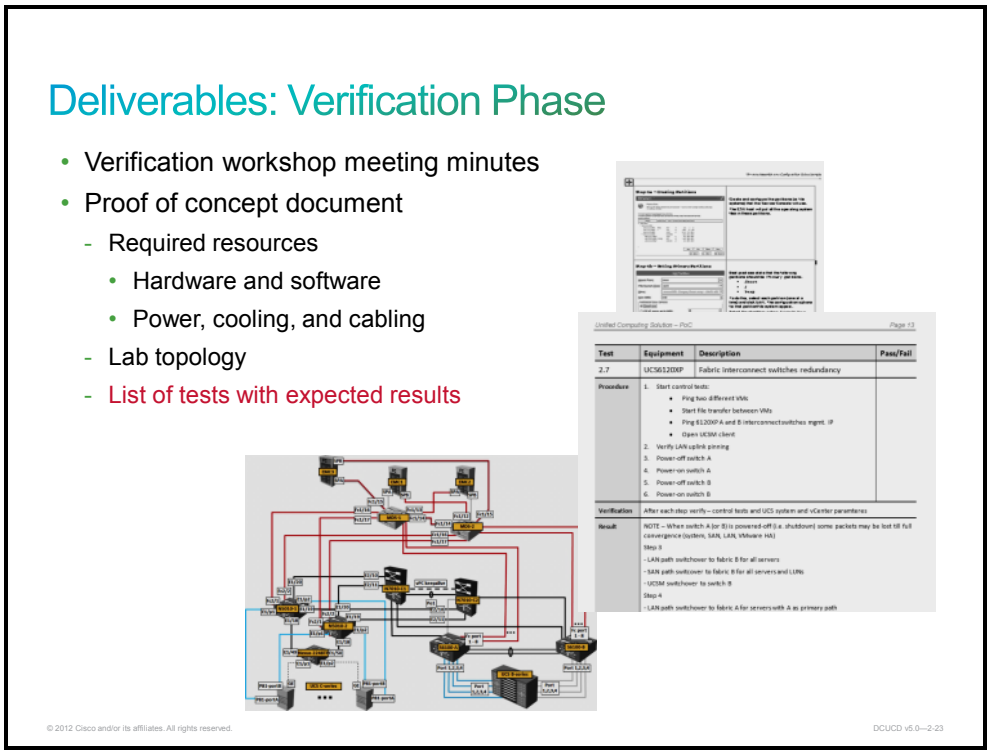
© 2012 Cisco and/or its affiliates. All rights reserved.

The design phase is the most document-intensive phase. It should result in the following documentation:

- **High-level design (HLD):** This document describes the conceptual design of the solution, such as the solution components, what equipment is used (not detailed information), how the high-availability mechanisms work, how the business continuance is achieved, and so on.
- **Low-level design (LLD) (also referred to as detailed design):** This document describes the design in details, such as comprehensive list of equipment, the plan of how the devices will be connected physically, the plan of how the devices will be deployed in the racks, as well as information about the relevant configurations, addressing, address pools and naming conventions, resource pools, management IP addressing, service profiles, VLANs, and VSANs.
- **Site requirements specification (SRS):** This document (or more than one document when the solution applies to more than one facility) will specify the equipment environmental characteristics, such as power, cooling capacity, weight, and cabling.
- **Site survey form:** This document (or more than one document when the solution applies to more than one facility) is used by the engineers or technicians to conduct the survey of a facility in order to determine the environmental specifications.

https://t.me/learningnets

- **Migration plan:** This document is necessary when the project is a migration. It must contain, at minimum, the following sections:
  - **Required resources:** Specifies the resources that are necessary to conduct the migration (for example, extra space on the storage, extra Ethernet ports to connect new equipment before the old one is decommissioned, or extra staff or even external specialists).
  - **Migration procedures:** Specifies the actions for conducting the migration (in the correct order) with verification tests and expected results.
  - **Rollback procedures:** Specifies the actions that are necessary to revert to a previous state if there are problems during the migration.



Because the first step of the verification phase is the verification workshop, the meeting minutes should be taken in order to track the workshop.

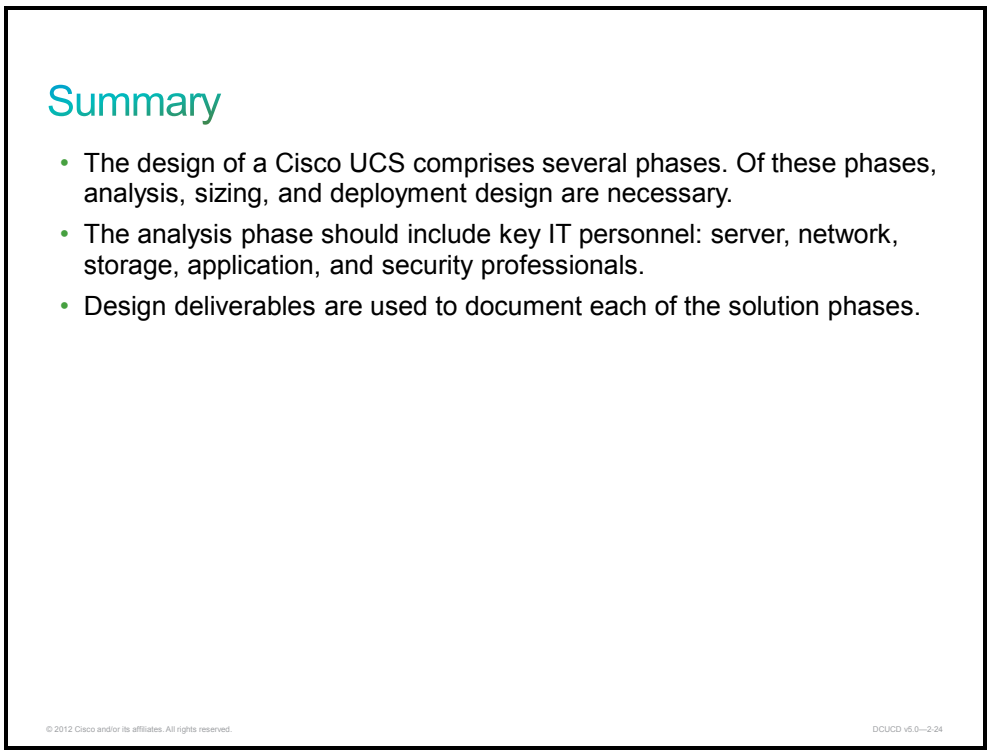
If the customer confirms that the solution design is approved, it needs to be signed with approval.

Secondly, when the proof of concept needs to be conducted, the proof of concept document should be produced. The document is a subset of the detailed design document for the equipment that will be used in the proof of concept. In addition, the document must specify what resources are required for conducting the proof of concept (not only the equipment but also the environmental requirements), and it should list the tests and the expected results with which the solution is verified.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- http://www.cisco.com/global/EMEA/IPNGN/ppdioo_method.html
- http://www.ciscopress.com/articles/article.asp?p=1608131&seqNum=3

https://t.me/learningnets

Lesson 2

Analyzing Computing Solutions Characteristics

Overview

This lesson assesses the performance characteristics and requirements for the computing solution.

Objectives

Upon completing this lesson, you will be able to assess the requirements and performance characteristics for the given data center computing solutions. You will be able to meet these objectives:

- Identify performance characteristics
- Assess server virtualization performance characteristics
- Assess desktop virtualization performance characteristics
- Assess small VMware vSphere deployment requirements
- Assess small Hyper-V deployment requirements
- Assess VMware VDI deployment requirements

https://t.me/learningnets

Performance Characteristics

This topic describes performance characteristics.

Network Performance Characteristics

Network equipment performance

- CPU load and memory usage
- Forwarding rate
- Equipment architecture

Link throughput and utilization

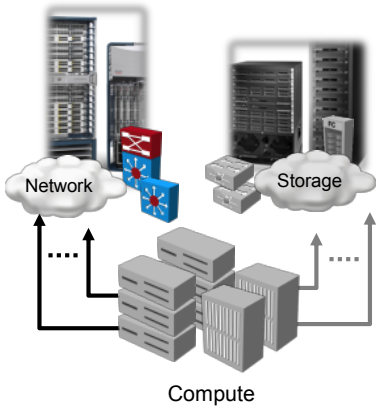
- Average vs. maximum

Performance affected by

- Insufficient forwarding capacity
- Unreliable connectivity

LAN considerations for Cisco UCS

- Nonblocking architecture
- Link quantity consideration



© 2012 Cisco and/or its affiliates. All rights reserved.

The network performance characteristics are defined by the network device performance and connection performance.

The device performance is defined by the following:

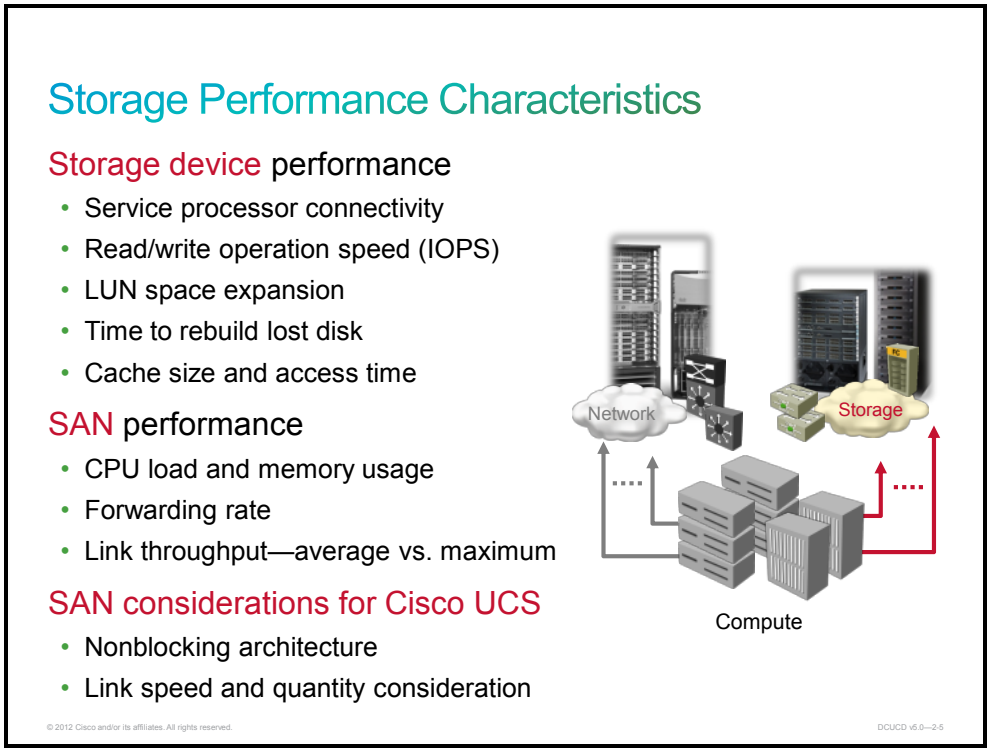
- CPU power and load
- Memory usage
- Number of packets that the device can manage per second
- Architecture of the equipment

The network performance may especially be affected when packet manipulation is done or when there is congestion on the network, such as the server pushing more traffic than the link is capable of handling.

The connection performance is defined by the throughput, or the bits per second.

The network performance can also be affected by an unstable environment, such as an environment that has flapping or erroneous links.

https://t.me/learningnets



Storage Performance Characteristics

Storage performance characteristics are defined by the SAN storage device performance.

Storage device performance is limited by the following:

- Number of interfaces and the speed of the interfaces that connects the device to the SAN
- Disk read/write operation speed
- Logical unit number (LUN) space expansion
- The load that is presented by the rebuild operation in case of a disk failure
- The size and speed of the controller cache

SAN Performance Characteristics

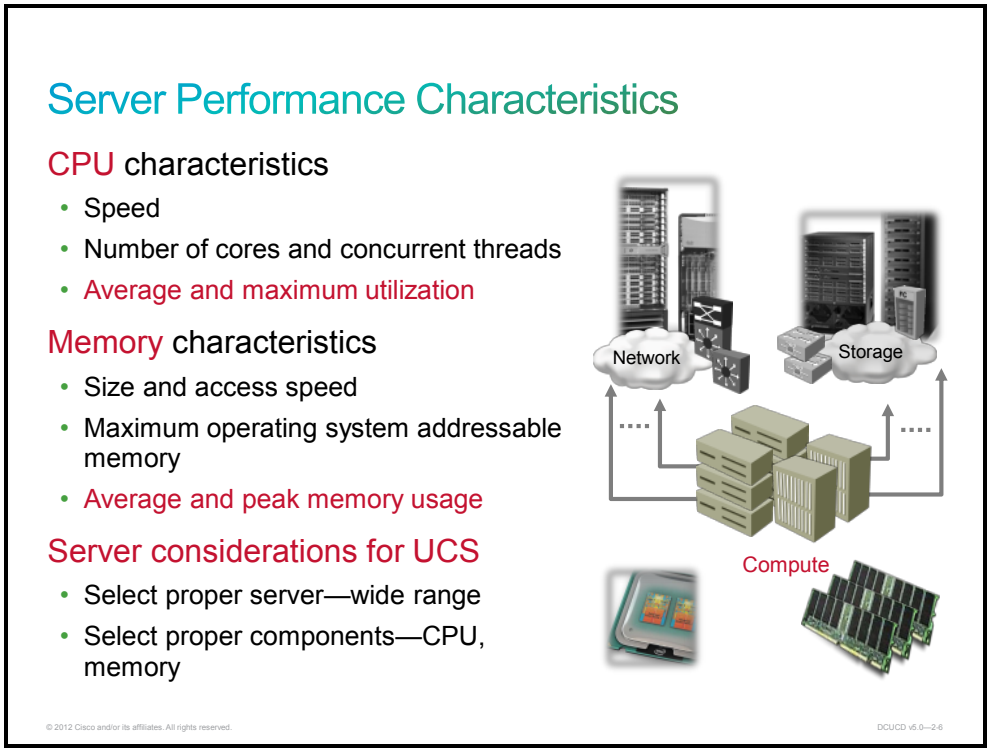
The SAN performance is affected by the following:

- CPU power and load
- Memory usage
- Number of packets the device can manage per second (IOPS)
- Amount of traffic that is sent via Fibre Channel link (or Fibre Channel over Ethernet [FCoE] or Fibre Channel over IP [FCIP], or any other type of link)

The connection performance is defined by the throughput, or the bits per second.

The SAN performance is strongly affected by an unstable environment, such as an environment that has flapping or erroneous links.

https://t.me/learningnets



The actual characteristics of the server performance depend on the following:

- Average and peak CPU load
- Average and peak memory usage
- Average and peak LAN and SAN throughput
- Required storage space and access speed

Server CPU Characteristics

CPU characteristics are defined by CPU speed, the number of cores, and concurrent threads.

Even if the CPU offers advanced functionality, the operating system might not be capable of fully using this functionality. Thus, performance is the least common factor between CPU characteristics and operating system capabilities.

Server Memory Characteristics

Memory characteristics are limited by the maximum amount of memory that can be installed and the memory access speed.

As with the CPU, the capability of the operating system might be the limiting factor for memory usage.

https://t.me/learningnets

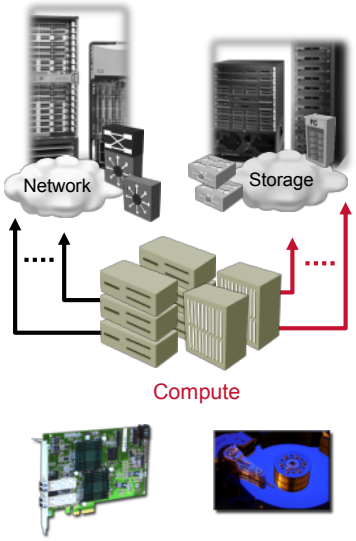
Server Performance Characteristics (Cont.)

LAN and SAN connectivity

- Separate adapters for SAN and LAN connectivity
- Multiple adapters for
  - Redundancy
  - Higher throughput
- Adapter speed (from 1 to 10 Gb/s)
- Average and peak throughput

Storage

- Disk subsystem performance
- Read/write operation speed
- Cache size (battery backup) and access time



© 2012 Cisco and/or its affiliates. All rights reserved.

Server I/O Characteristics

Server I/O characteristics are defined by the LAN and SAN connectivity.

Separate adapters can be used for LAN and SAN connectivity with different raw speeds.

The I/O subsystem throughput may be affected by the CPU performance in cases where the CPU must manipulate packets that are received or sent—for example, in an Internet Small Computer Systems Interface (iSCSI) deployment without the TCP/IP Offload Engine (TOE), or in an FCoE deployment with software-based FCoE functionality.

Multiple adapters are typically used to scale performance and to achieve redundancy.

Server Storage Characteristics

The server storage performance is affected by the storage subsystem.

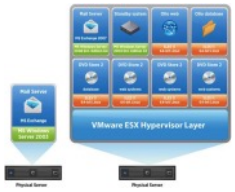
If the volume uses a local disk, the disk performance defines the server storage performance.

More often, the server storage is decoupled from the server hardware to a storage array. The performance of that array, combined with the SAN performance, defines the server storage performance characteristics.

https://t.me/learningnets

Server Performance Benchmark Tools

- SPEC = Standard Performance Evaluation Corporation
  - SPEC CPU2006, CFP2006, and CINT2006: Performance measurements that compare compute-intensive workloads on different computer systems
  - SPECjbb2005: Performance measurement for typical Java business applications
  - SPECpower_ssj2008: Evaluates the power and performance characteristics of volume server class and multinode class computers
- VMware VMmark
  - Measure of virtualization platform performance
  - Platform-level workloads plus traditional application-level workloads
    - Dynamic VM-relocation (VMotion) and dynamic datastore-relocation (storage VMotion)



© 2012 Cisco and/or its affiliates. All rights reserved.

Standard Performance Benchmark (SPEC) Tools

The Standard Performance Evaluation Corporation (SPEC) is a nonprofit corporation formed to establish, maintain, and endorse a standardized set of relevant benchmarks that can be applied to the newest generation of high-performance computers. Here are some of the benchmarks relevant to Cisco Unified Computing System (UCS):

- **SPEC CPU2006**: Designed to provide performance measurements that can be used to compare compute-intensive workloads on different computer systems, SPEC CPU2006 contains two benchmark suites:
  - CINT2006 for measuring and comparing compute-intensive integer performance
  - CFP2006 for measuring and comparing compute-intensive floating point performance
- **SPECjbb2005**: A benchmark for evaluating the performance of servers running typical Java business applications, jbb2005 represents an order-processing application for a wholesale supplier. The benchmark can be used to evaluate the performance of hardware and software aspects of Java Virtual Machine (JVM) servers.
- **SPECpower_ssj2008**: The first industry-standard SPEC benchmark that evaluates the power and performance characteristics of volume server class computers. The initial benchmark addresses the performance of server-side Java, and additional workloads are planned.

https://t.me/learningnets

VMware VMmark

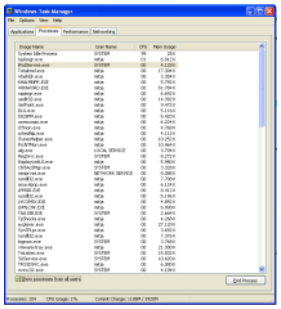
The VMmark 2.0 benchmark uses a tiled design that incorporates six real-world workloads to determine a virtualization score. Then it factors in VMware VMotion, Storage VMotion, and virtual machine provisioning times to determine an infrastructure score. The combination of these scores is the total benchmark score. Because Cisco UCS is a single cohesive system, it delivers both virtualization and infrastructure performance. Because the system virtualizes the hardware itself, it offers greater flexibility, running any workload on any server, much as cloud computing environments support virtual machine images.

VMware VMmark incorporates six benchmarks, including email, web, database, and file server workloads, into a tile. A tile represents a diverse, virtualized workload, and vendors increase the number of tiles running on a system during testing until a peak level of performance is observed. This procedure produces a VMware VMmark score and the number of tiles for the benchmark run.

Application Performance Characteristics

- Performance depends on
  - Server + storage + network performance
  - Application code used
  - Application protocol
    - Chatty protocols incur additional processing time, thus limiting application responsiveness
- Single host versus distributed applications
- Memory vs. CPU-bound applications

Application	Limit
In-memory database	Memory
File conversion (e.g. video)	CPU
File compression	CPU
Virus scanning	CPU
Graphic manipulation	Memory & CPU



© 2012 Cisco and/or its affiliates. All rights reserved.

Application performance characteristics are affected by all the server components as well as by the application performance itself. Applications that are written in a nonoptimal way can result in a slow response, even if the underlying server performance is sufficient.

The type of application also defines its characteristics—those requiring a great deal of memory have different design considerations than those requiring high CPU power.

https://t.me/learningnets

Application Requirements Example

Type of Application	Memory	Network	Storage
High-end database applications	High memory	Multiple NICs	SAN
Low/midrange database applications	Low to midsize memory	Dual NICs	NAS or SAN
Web hosting	Low memory	Dual NICs	Local disk/NAS
General computing	Low memory	Single NIC	Local disk/NAS
CRM/ERP front-end application servers	Midsize memory	Dual NICs	Local disk/NAS/SAN
Microsoft Exchange	Midsize memory	Dual NICs	Local disk/NAS/SAN
Market data applications/ algorithmic trading	High memory	Multiple NICs	SAN
Linux Grid Based Environments	Midsize memory	Dual NICs	Local disk/NAS

© 2012 Cisco and/or its affiliates. All rights reserved.

The following list presents typical memory, storage, and network use cases for certain applications:

- High-end database applications like Oracle RAC, MS SQL HA Cluster, and Sybase HA Cluster require a large amount of memory, multiple network interface cards (NICs), SAN connectivity, and high availability.
- Low-to-midrange database applications like Linux or Microsoft Server-based MySQL and Postgress databases require low-to-midsize memory and dual NICs. They can use network-attached storage (NAS) or SAN for storage.
- Web-hosting applications like Linux Apache or Microsoft Internet Information Services (IIS) require a relatively small memory amount, local disk or NAS storage, and dual NICs.
- General computing applications like Microsoft SharePoint, file servers, print servers, etc., require a small amount of memory, local disk or NAS storage, and a single NIC.
- Customer relationship management (CRM) and enterprise resource planning (ERP) front-end application servers like SAP or PeopleSoft typically require a midsized amount of memory, dual NICs, and local disk or NAS or SAN storage.
- Microsoft Exchange (depending on the use case) may require a midsize amount of memory, dual NICs, and local disk or NAS or SAN storage.
- A high-guest-count virtual desktop infrastructure (VDI), like VMware with 128 to 300 guests per server, typically requires a large memory, multiple NICs, SAN or NAS storage, and high availability.
- Market data applications and algorithmic trading require a large memory, with multiple NIC connectivity, SAN or NAS storage, and high availability.
- Linux grid-based environments like DataSynapse require a midsize amount of memory, dual NICs, and local disk or NAS storage.

https://t.me/learningnets

Migration Requirements

- Migrated server infrastructure should provide the same functionality with equal or better performance
- Consider individual migrated components

Resource	Requirement
CPU	Guaranteed CPU resources in virtual environment that are equal to or greater than existing CPU power
Memory	Equal to or greater than existing memory for virtual machine
Disk	Create virtual disk on the central storage that is equal to or greater than existing disk capacity
Network	Ensure equal or greater bandwidth to virtual machine (on server and network equipment)

- Audit the existing environment to get peak and average resource utilization, and provide guarantees for the identified peak periods (do not extend the P2V ratio beyond capacity).

© 2012 Cisco and/or its affiliates. All rights reserved.

When you are dimensioning a new virtualized data center, it is normal to assign multiple virtual servers per one high-performance server (that is, several CPUs or cores—or both—and large amounts of memory). Care should be taken not to assign too many, which could result in performance degradation after migration. Longer monitoring (for example, over several weeks) should be performed and the data analyzed to determine the average and peak resource utilization of each physical server. The peak utilization should be used to dimension the new data center and the physical to virtual (P2V) migration ratio.

The collection interval should be long enough to collect the relevant peak hours and periods when resources are under more stress. The collection interval can last for two months, three months, or even longer than three months. The recommended practice is to keep the collection running for at least one month, which on average provides the desired results.

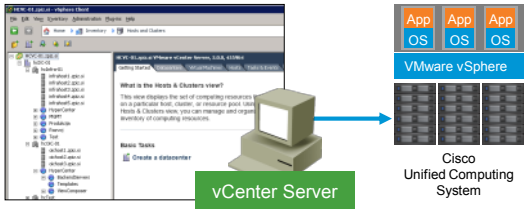
https://t.me/learningnets

Assess Server Virtualization Characteristics

This topic describes how to assess server virtualization performance characteristics.

VMware vSphere Characteristics

- Components
- Scalability aspects
  - Resource maximums
    - CPU, memory, network, storage
    - Per component – ESXi host, cluster, VM
  - Licensing – per CPU and memory
- Virtualization ratio
  - no. of VMs per UCS server

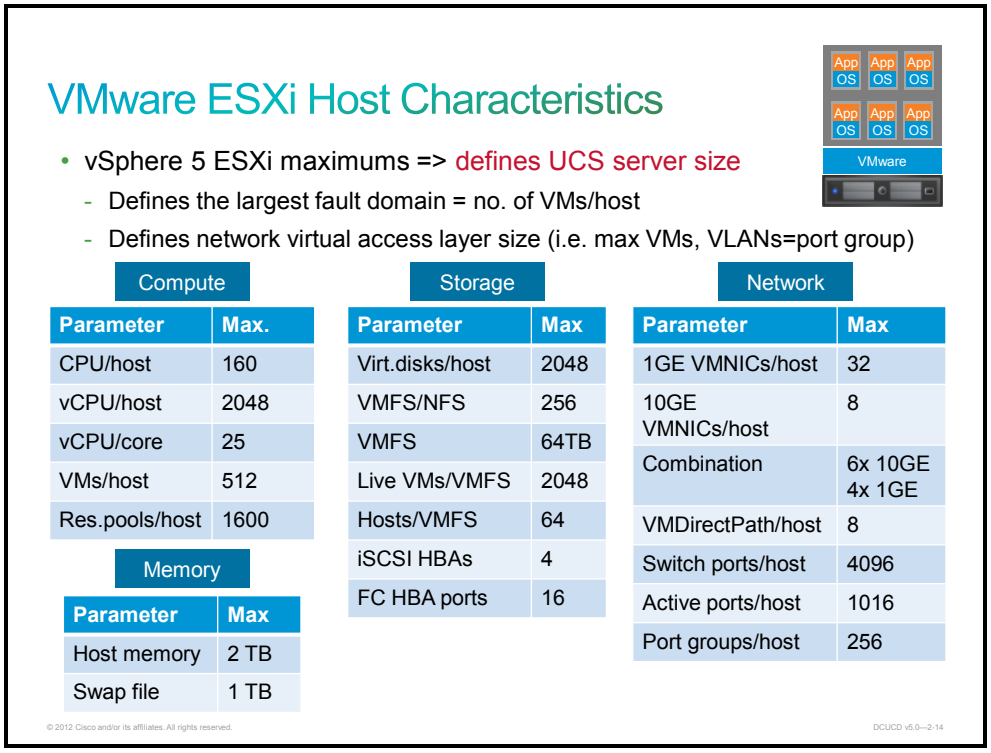


© 2012 Cisco and/or its affiliates. All rights reserved.

The VMware vSphere solution has many components, which have scalability limitations that need to be observed in the design:

- ESX and ESXi hypervisor
- Dynamic Resource Scheduler (DRS)
- Virtual Machine File System (VMFS) native ESX cluster
- Distributed switch that can be native to VMware or Cisco Nexus 1000V
- VMotion, Storage VMotion, high availability, fault tolerance, and data recovery availability tools
- Virtual Symmetric Multiprocessing (SMP), enabling virtual machines (VMs) to use multiple physical processors—that is, upon creation, the administrator can assign multiple virtual CPUs to a VM

https://t.me/learningnets



The ESXi host and VM characteristics govern the VMware solution scalability.

The ESXi host scalability characteristics define the performance and consolidation rates available per single physical server, whereas the VM scalability characteristics define the amount of workload a VM can manage.

Note	ESX version 5 supports real-time (or “hot”) additions of resources like memory, CPU, network connectivity, and storage. With this feature, it is no longer necessary to shut down the VMs in order to make configuration changes.

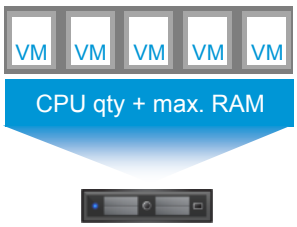
The following characteristics apply to the current VMware ESXi version 5:

- The ESXi host scalability is limited to:
  - 2 TB of memory
  - 2048 virtual CPUs per host
  - 512 VMs
- The VM scalability is limited to:
  - 32 virtual CPUs with Virtual SMP
  - 1000 GB of memory

https://t.me/learningnets

Licensing Characteristics

- Governs memory, CPU amount per host => sizing
  - Licensed per
    - physical CPU (socket)
    - maximum entitled RAM – above additional fee applies
- License imposed maximums
  - More smaller vs. fewer larger UCS servers
  - Fault domain size is important for recovery time



Entitlement/socket	Standard	Enterprise	Enterprise plus
vRAM	32 GB	64 GB	96 GB
vCPU	8	8	32

© 2012 Cisco and/or its affiliates. All rights reserved.

In addition to the host maximums, the licensing imposes a second level of maximums, and the total amount of memory entitled per socket differs based on the license. Depending on your maximum memory needs, you may need a different license that allows you to use more memory on the hosts.

When sizing hosts—UCS servers—the following should be weighted:

- Amount of memory per host should be considered from licensing perspective also
- Fault domain size should also be taken into consideration. Fewer large UCS servers create large fault domains (when one server fails, many VMs need to be restarted), and more "smaller" UCS servers create smaller fault domains (because fewer VMs would be run on a single server).

https://t.me/learningnets

Cluster Characteristics

- vSphere cluster maximums
  - Define resource pool size = number of UCS servers in a cluster
- vCenter Server maximums
  - Define the total virtual infrastructure size = total number of UCS servers managed by a single vCenter Server instance

Cluster

Parameter	Max
Hosts/cluster	32
VMs/cluster	3000
Concurrent HA failovers	32
Resource pools/cluster	1600

Network

Parameter	Max
VDS ports/vCenter	30000
Port groups/vCenter	5000
Hosts/VDS	350
VDS/vCenter	32



© 2012 Cisco and/or its affiliates. All rights reserved.

While the individual hosts have maximums, so too does the entire environment. Two aspects influence this:

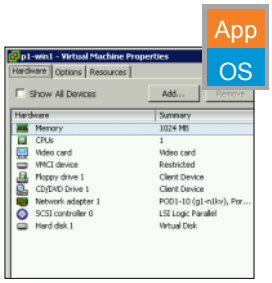
- **Cluster maximums:** This governs how UCS servers can be part of a vSphere cluster
- **vCenter Server maximums:** This governs how many UCS servers in total can be managed by a single vCenter Server instance

Because the UCS server hardware configuration is very flexible (that is, going from 4 GB to 1 TB of memory), the size of the virtual domain (that is, the amount of VMs that need to fit into an environment) also governs how the UCS servers will be sized.

https://t.me/learningnets

VM Characteristics

- Per VM maximums per vSphere platform
  - Lower limits may apply
    - Per guest operating system used
    - Per application installed in guest operating system
- Overcommitment
  - higher VM/host ratio
  - Even more VMs per UCS server



Compute, Memory, Network

Parameter	Max.
vCPU	32
Memory	1 TB
Swap file	1 TB
vNICs	10
Throughput	>36 Gb/s

Storage

Parameter	Max
SCSI adapters	4
Virtual disks	60
Virtual disk size	2–512 TB
IDE devices	4
IOPS	1,000,000

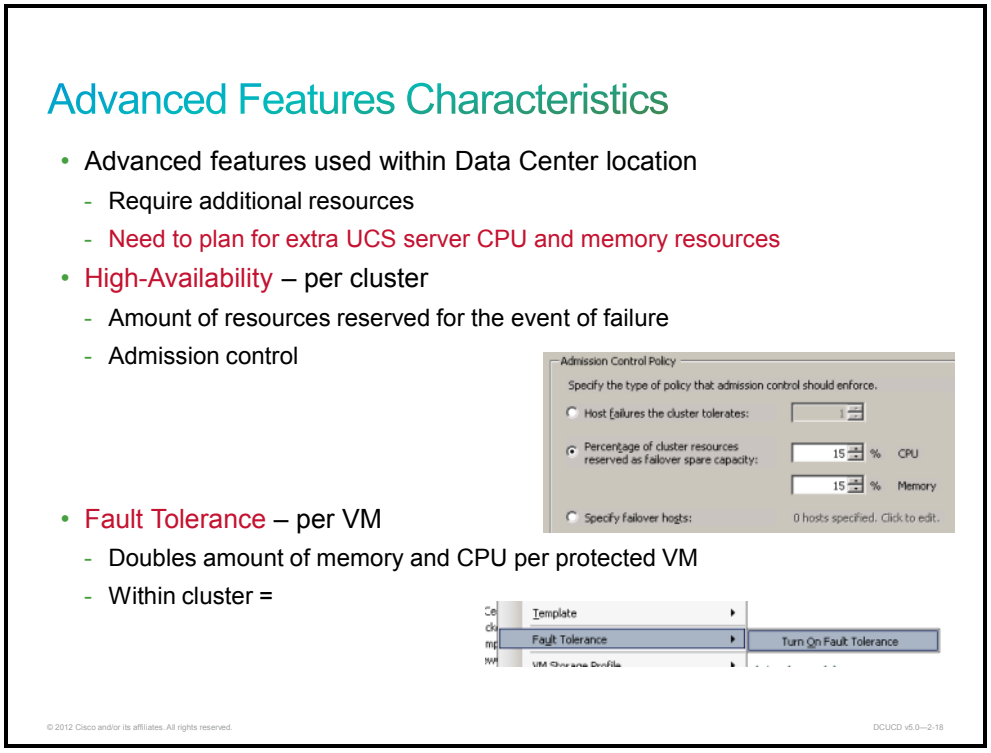
© 2012 Cisco and/or its affiliates. All rights reserved.

The table describes VMware ESX 4 and VMware ESXi 5 per VM maximums.

VMware ESXi 4 Versus VMware ESX 5 Per VM Maximums

Parameter	VMware ESX 4	VMware ESXi 5
Size of SCSI disk	2 TB	32 TB
Number of vCPUs	8	32
Memory size	255 GB	1000 GB

https://t.me/learningnets



The VMware solution incorporates two different mechanisms to achieve VM high availability:

- VMware high availability
- VMware fault tolerance

VMware HA

VMware HA enables automatic failover of VMs upon ESX host failure. The high-availability automatic failover restarts the VM that was running on a failed ESX host on another ESX host that is part of the high-availability cluster.

Because the VM is restarted and thus the operating system has to boot, high availability does not provide automatic service or application failover in the sense of maintaining client sessions. Upon failure, a short period of downtime occurs. The exact amount of downtime depends on the time needed to boot the VM or VMs. Because the failover is achieved with VM restart, it is possible that some data may be lost because of ESX host failure.

VMware HA requires the following:

- Dedicated network segment with assigned IP subnet
- ESX or ESXi hosts configured in a high-availability cluster
- Operational DNS server
- ESX or ESXi hosts must have access to the same shared storage
- ESX or ESXi hosts must have identical networking configuration (either by configuring standard vSwitch the same way on all ESX hosts, or with distributed vSwitch)

When designing VMware HA, it is important to observe whether the remaining ESX hosts in a high-availability cluster will be overcommitted upon member failure.

https://t.me/learningnets

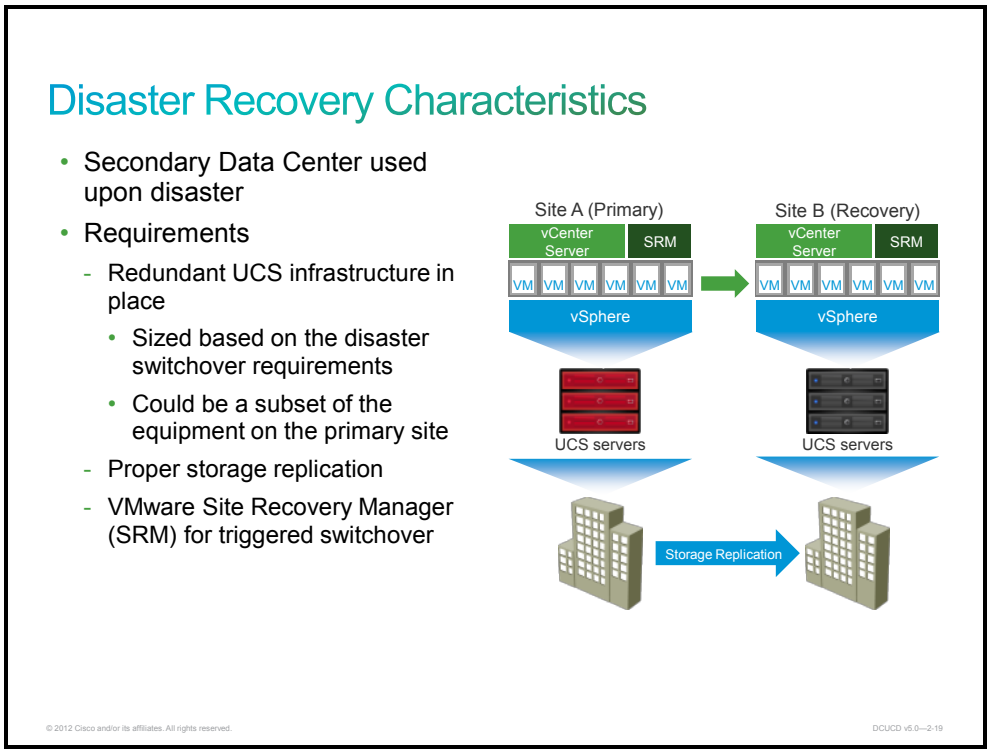
VMware FT

VMware FT advances the high-availability functionality by enabling true zero downtime switchover time. This is achieved by running primary and secondary VMs, where the secondary is an exact copy of the primary. The VMs run in lockstep using VMware vLockstep so that the secondary VM is in the same state as the primary one. The difference between the two VMs is that the primary one controls the network connectivity.

Upon failure, the switchover to the secondary VM preserves the live client session because the VM is not restarted. FT is enabled per VM.

To be able to use FT, the following requirements, among others, have to be met:

- DRS for the VM is disabled.
- Thin disk provisioning is converted to zeroized thick disk provisioning.
- Hosts used for FT must be in the same cluster.



Disaster Recovery (DR) adds to the total amount of resources required. With VMware this typically means a redundant set of UCS infrastructure is placed on the secondary site.

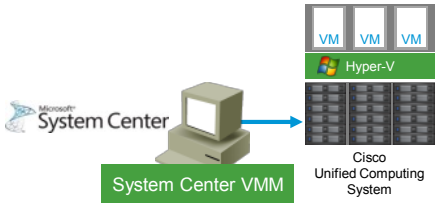
The size of redundant UCS infrastructure (the resources) depends on whether a complete primary site is switched over in case of a disaster—or if it is only a subset.

For the VMware vSphere environment, DR policy usually means that VMware Site Recovery Manager (SRM) is used. The tool provides a single point of DR policy configuration and execution upon primary site failure. From the DR site Cisco UCS infrastructure perspective, the disaster means that the selected VMs will be restarted on the DR site. Sizing the DR site Cisco UCS infrastructure means that the architect needs to understand and know which VMs are vital and need to be restarted. The Cisco UCS infrastructure at the DR site would be smaller than the one on the primary site because only a subset of VMs needs to be restarted.

https://t.me/learningnets

Microsoft Hyper-V R2 Characteristics

- Components
- Scalability aspects
  - Resource maximums
    - CPU, memory, network, storage
    - Per component – Hyper-V host, cluster, VM
  - Licensing – per host or CPU
- Virtualization ratio
  - Number of VMs per UCS server



© 2012 Cisco and/or its affiliates. All rights reserved.

Hyper-V is built on the architecture of Windows Server 2008 Hyper-V and enables integration with new technologies.

Hyper-V can be used in the following scenarios:

- Increased server consolidation
- Dynamic data center
- Virtualized centralized desktop

Hyper-V is part of Windows Server 2008 Hypervisor-based architecture and is available as a new virtualization role that can be configured as a full role with local user interface or as a Server Core role.

Hyper-V as a role in Windows Server 2008 provides you with the tools and services to create a virtualized server computing environment. This type of environment is useful because you can create and manage VMs, which allows you to run multiple operating systems on one physical computer and isolate the operating systems from each other. This topic introduces Hyper-V by providing instructions for installing this role and configuring a VM.

Hyper-V has specific requirements. Hyper-V requires an x64-based processor, hardware-assisted virtualization, and hardware data execution prevention (DEP). Hyper-V is available in x64-based versions of Windows Server 2008—specifically, the x64-based versions of Windows Server 2008 Standard, Windows Server 2008 Enterprise, and Windows Server 2008 Datacenter.

https://t.me/learningnets

Supported Guest Operating Systems

The operating systems that Microsoft supports as guest sessions under Windows 2008 Hyper-V are:

- Windows Server 2008 x86 and x64
- Windows Server 2003 SP2 or higher x86 and x64
- Windows 2000 Server with SP4 and Windows 2000 Advanced Server with SP4
- Windows Vista x86 and x64
- Windows XP SP2 or later x86 and x64
- SUSE Linux Enterprise Server 10 SP1 or later x86 and x64

Failover Cluster

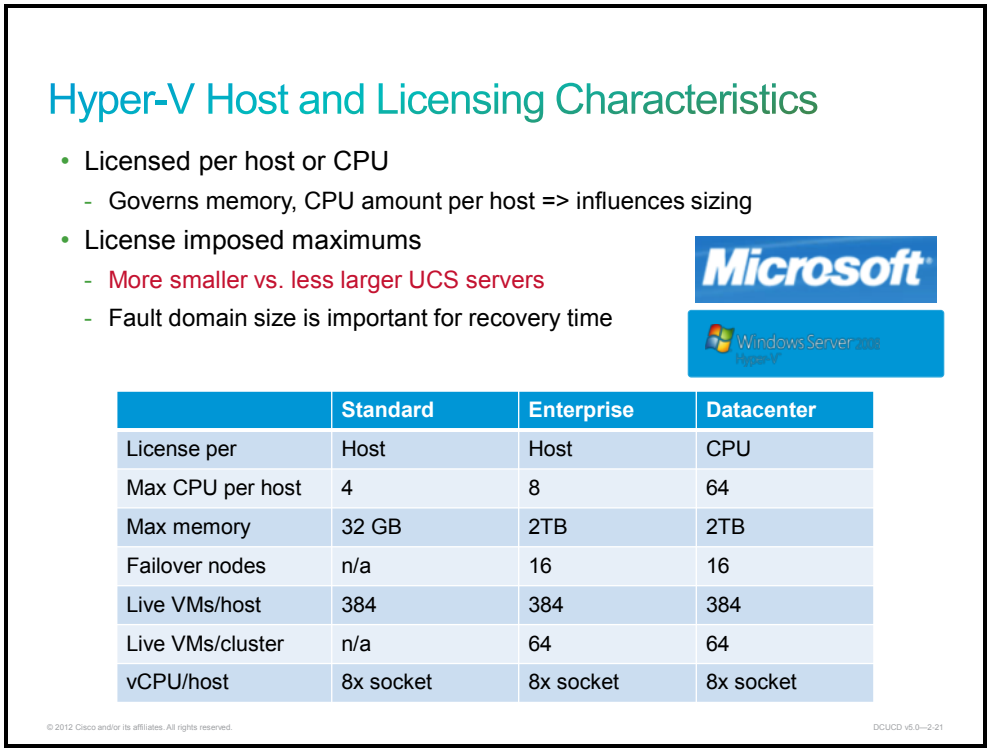
Failover clusters typically protect against hardware failure. Overall system failures (system unavailability) are not usually the result of server failures but are more commonly caused by power outages, network stoppages, security issues, or misconfiguration. A redundant server generally will not protect against an unplanned outage such as lightning striking a power substation, a backhoe cutting a data link, an administrator inadvertently deleting a machine or service account, or the misapplication of a zoning update in a Fibre Channel fabric.

A failover cluster is a group of similar computers (referred to as nodes) working in a coordinated fashion to increase the availability of specific services or applications. One way of increasing the availability of specific services or applications is by protecting against the loss of a single physical server from an unanticipated hardware failure. Another possible method is through proactive maintenance.

Cluster Shared Volumes

Cluster Shared Volume (CSV) is a feature of failover clustering available in Windows Server 2008 R2 for use with the Hyper-V role. A CSV is a standard cluster disk containing an NTFS volume that is made accessible for read and write operations by all nodes within the cluster. This gives the VM complete mobility throughout the cluster, as any node can be an owner, and changing owners is fairly simple.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



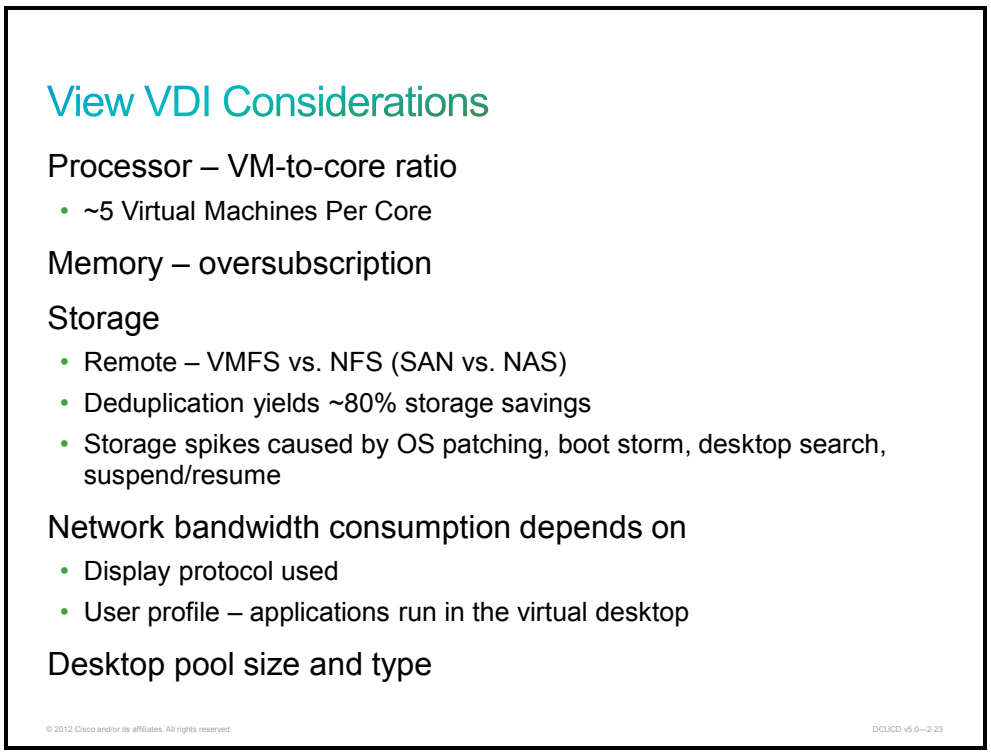
Hyper-V delivers high levels of availability for production workloads via flexible and dynamic management while reducing overall costs through efficient server consolidation via:

- Better flexibility:
  - Live migration
  - Cluster shared volumes
  - Hot swappable (real-time addition and removal of) storage
  - Processor compatibility mode for live migration
- Improved performance:
  - Improved memory management
  - TCP offload support
  - Virtual Machine Queue (VMQ) support
  - Improved networking
- Greater scalability:
  - Support for 64 logical processors
  - Enhance green IT with core parking

https://t.me/learningnets

Assess Desktop Virtualization Performance Characteristics

This topic describes how to assess desktop virtualization performance characteristics.



The VMware View solution is an end-to-end solution encompassing data center components as well as remote locations and WAN. The VMware View solution integrates the data center, DMZ, security servers providing mobile or remote access for View clients, as well as the branch with WAN optimization to a data center type of connection.

Considerations

The sizing considerations for the desktop servers apply to the CPU, memory, storage, and network aspects.

For the processor, about five virtual machines per core (for example, a physical machine with 8 cores can host 40 virtual machines) is pretty common on average.

On the memory side, users typically use a gigabyte per VM. With VMware ESXi, an advantage is a memory oversubscription.

VDI Desktop Type

The VMware View architecture enables usage of different types of VDI desktops, which also influences the compute, network, and storage aspects of the data center solution:

- Individual Desktop
  - Static 1:1 relationship between user and virtual desktop
  - Can be assigned with individual settings and resource allocations

https://t.me/learningnets

- Automated Desktop Pools
  - Persistent Desktop Pool
    - Automatically created and assigned to users by View Connection Server
    - Once assigned, user is always connected to the same desktop
    - Desktop changes persist after logoff
  - Nonpersistent Desktop Pool
    - Automatically created and assigned to users by View Connection Server
    - Assignment is on a session-by-session basis
    - Desktops are returned to the pool for reallocation after logoff
    - Desktops persist after disconnects
- Manual Desktop Pools
- Microsoft Terminal Services Desktop Pool

Individual Desktops

There is a static 1:1 relationship between a user and a specific virtual desktop. Individual desktops are typically configured for a particular user. This can include specific applications, data access and resource—such as RAM—allocations. Individual desktops allow a high degree of customization for the user.

Persistent Pools

The persistent pool contains multiple pooled virtual desktops, which are initially cloned from the same template. When a group of users is entitled to a persistent pool, each user is entitled to access any of the virtual desktops within the pool. The VDM Connection Server will allocate users to a virtual desktop upon the initial request. This allocation is then retained for subsequent connections. When the user connects to the persistent pool on subsequent occasions, the VDM Connection Server will connect the user to the same virtual desktop that was initially allocated.

Persistent pools provide a simple automated mechanism for initial cloning and deployment of the virtual desktop and allow the user to customize the desktop. The initial creation effort is less than for Individual Desktops because only a single template and entitlement is required to provision a virtual desktop for every user in a large group.

Nonpersistent Pools

The nonpersistent pool also contains multiple hosted virtual desktops, which are initially identical and cloned from the same template. The VDM Connection Server will allocate entitled users to a virtual desktop from the nonpersistent pool. This allocation is not retained. When the user logs off, the desktop and the virtual desktop are placed back into the nonpersistent pool for reallocation to other entitled users. When the user connects to the nonpersistent pool on subsequent occasions, the VDM Connection Server will connect the user to any available virtual desktop in the nonpersistent pool.

Nonpersistent pools provide the most efficient many-2-many configuration. Simple automated mechanisms for cloning and deploying the virtual desktops reduce desktop provisioning efforts, and the virtual desktops can be reused by different users. Nonpersistent pools present a good solution for hoteling, shift workers, or environments where a more dynamic desktop environment is desired.

https://t.me/learningnets

Desktop Provisioning Type

VDI desktops can also be provisioned in different manners—full clones versus linked clones.

Full Clones

Full clone is a copy of a VM (at a given point) with a separate identity. It can be powered on, suspended, snapshot, reconfigured, and is independent of the VM it was cloned from.

Link Clones

Link clones reduce storage usage by 50 to 90 percent over full clones. They redirect folders to a separate optional user disk and enable rapid provisioning of desktops versus full cloning.

VDI Operations

The following operations can be applied to VDI desktops:

- **Refresh:** Clean desktop, pristine image
- **Recompose:** Migrate existing desktops from one version to the other
- **Rebalance:** Relocate desktops to enable efficient usage of the storage available (add more storage or retire existing array)

© 2012 Cisco Systems, Inc

https://t.me/learningnets

View VDI Recommendations

Processor

- Use fastest Intel Xeon CPU processors (power consumption)

Memory

- Use memory overcommit even with large memory (1.25 – 1.5)
- Watch for swapping and ballooning

Storage

- Plan for spikes with storage access
- Minimize storage footprint: thin provisioning, ThinApp, deduplication

Network

- Plan for traffic spikes (e.g. virus scans)
- Plan staggering storm activities or schedule it off-hours

© 2012 Cisco and/or its affiliates. All rights reserved.

There are some general recommendations that can be applied when building a VDI solution, but the final figures must be derived from the workload.

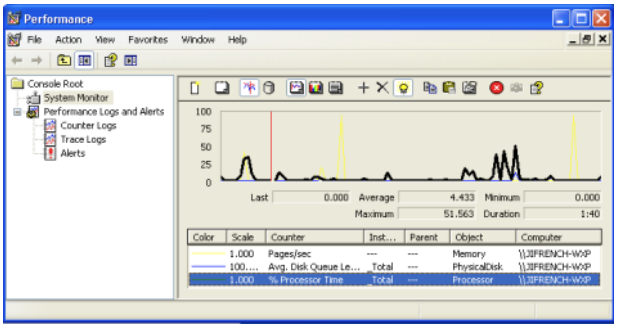
It must also be noted that the VDI deployments can be optimized from the workload perspective if the following is taken into account:

- Remove unnecessary services, device drivers, add-ons
  - nLite (Windows XP), vLite (Windows Vista) to optimize the OS build
  - Disable graphical screen savers
- When data is accessed over WAN – the protocol becomes key (RDP/PCoIP/ICS/EOP)
  - Assess your network while deploying over WAN
  - Consider using WAAS solutions
- Consider ThinApp as a means of virtualizing an application that is running on the guest operating system.

https://t.me/learningnets

View VDI Server CPU Sizing Example

- Processor Time average 5% on 2-GHz core
- Requires 100 MHz per desktop
- 100 desktops require 10 GHz processing
- Add overhead for virtualization, display protocol, and buffer for spike
- 100 desktops achieved on ~4 cores to achieve 12 GHz



© 2012 Cisco and/or its affiliates. All rights reserved.

The desktop server CPU sizing is based on using a standard VMware View workload. You can also size based on knowing how much that you are using on the PC today.

View VDI Server Memory Sizing Example

- Virtual platform memory optimization
  - VMware ESXi Transparent Page Sharing – shares master copy of memory pages among virtual machines
- Windows XP desktop memory requirements
  - Per desktop
    - Average = 254 MB
    - Minimum (applying memory optimizations) = 175 MB
  - 100 desktops require
    - minimum = 17.5 GB min
    - Maximum = ~ 25 GB peak
  - Recommend provisioning 512 MB per desktop

Application/OS	Optimized Memory Use
Windows XP	175 MB
Microsoft Word	15 MB
Microsoft Excel	15 MB
Microsoft PowerPoint	10 MB
Microsoft Outlook	10 MB
Total	225 MB

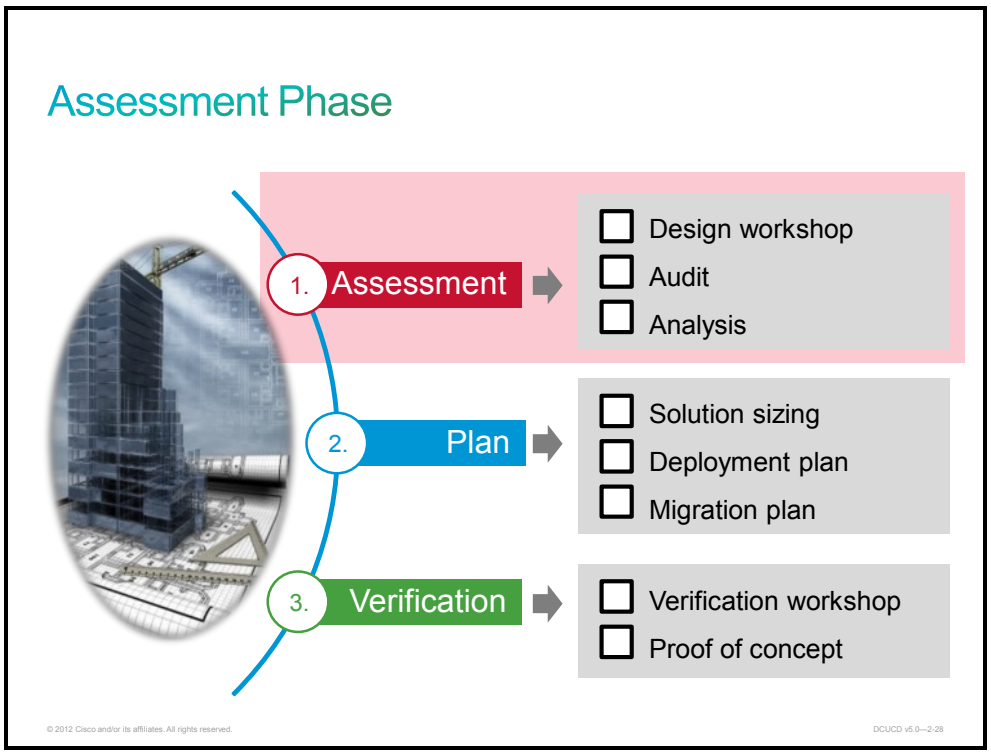
© 2012 Cisco and/or its affiliates. All rights reserved.

The memory desktop server sizing can be accomplished by inspecting the performance monitor of physical desktops. Or, you can use guidelines from VMware on how many resources are required by common applications like Word, Office, Outlook, and Office.

https://t.me/learningnets

Assess Small vSphere Deployment Requirements

This topic recognizes and assesses a given small VMware vSphere deployment requirements.



As with any other solution, a small VMware vSphere deployment has specific requirements. This topic describes an example of building such a solution from the beginning, which includes gathering the requirements along with the assessment. The requirements include the design workshop and an audit of the existing environment.

https://t.me/learningnets

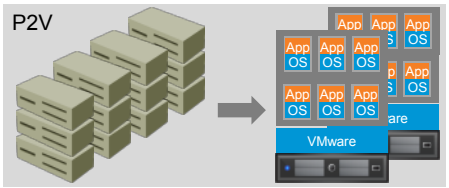
Design Workshop Input

Migration of existing environment

- Currently bare metal Windows 2003 environment
- Need to scale whole environment: more users, more data
- Migrate to the latest version of applications and OS

Requirements

- Flexibility
  - Ability to easily add memory
  - Ability to easily increase disk size
- High availability for services running
  - To protect against server failure for selected applications
  - To protect against application failure for selected applications
- Separate network for management and data
- No external storage



© 2012 Cisco and/or its affiliates. All rights reserved.

After running the design workshop, the initial requirements have been gathered.

These requirements include information about the current environment (that is, the current physical desktop environment based on Windows 2003) and the requirements that the customer wants to achieve with the migration:

- A new environment should scale from the user as well as data perspective.
- The version of operating system and applications should be brought up to the latest versions, which translates into higher resource requirements compared to what is currently available.
- Flexibility has been identified as a key requirement where the customer wants to be able to easily scale the servers by adding additional memory or enlarging the disk size.
- High availability is defined from two aspects:
  - Some applications need protection against server failure (against physical equipment failure).
  - Some more vital applications also require application-level protection.
- The management and data traffic should be separated to isolate the management of the solution from user traffic.
- The customer does not want to use external storage such as Fibre Channel, iSCSI, or NFS storage devices. The customer wishes to use the disk drives internal to the servers.

https://t.me/learningnets

Solution Requirements Summary

Application	Current Resources	Required	High Availability
2x application server (purpose-built application)	Windows 2003 8 GB; 1 CPU 100-GB disk	Windows 2008 R2 16 GB memory (ability to scale if required) 2 CPU; 150 GB disk	Yes
1x MS SQL 2005 server	Windows 2003 8GB; 1 CPU 100-GB disk	MS SQL 2008; Windows 2008 R2 24GB memory (ability to scale if required) 2 CPU; 200-GB disk	Yes – application level
2x Windows 2003 TS (60 users)	Windows 2003 4GB; 1 CPU 50-GB disk	Windows 2008 R2 16GB memory (ability to scale if required) 1 CPU; 80-GB disk	Yes
AD/DNS/DHCP server	Windows 2003 4 GB; 1 CPU 40-GB disk	Windows 2008 R2 4 GB memory 1 CPU; 50-GB disk	Yes – app. level i.e. two instances
Email server	Windows 2003 8 GB; 1 CPU 200 GB	Windows 2008 R2 16 GB memory 1 CPU; 200-GB disk	Yes
VSA	n/a	VM appliance 1GB memory; 2x 4GB disk	Yes
vCenter	n/a	VM appliance 4GB memory; 100 GB + ISO/software disk (variable)	Manual High Availability

© 2012 Cisco and/or its affiliates. All rights reserved.

The analysis of input gathered results in an overview of existing versus desired state of the server infrastructure.

The applications were identified per type (infrastructure versus user), number of instances, and type of high-availability, and have specified the current resource, operating system, application version, as well as the amount of resources, operating system version, and application version after migration.

https://t.me/learningnets

High-Level Design

Server virtualization with VMware vSphere

- Essential Plus = max 192GB memory, high availability
- VMware Storage Appliance(VSA) for shared storage
- Embedded database (up to 5 hosts, 50 VMs)

Application	Instance	High Availability
Application server	2	VMware HA
MS SQL 2008	2	High availability with failover cluster instance
TS	2	VMware HA
AD	2	No high availability, just two instances
Email server	1	VMware HA
VSA	1	High availability
vCenter	1	Manual high availability (operational upon vCenter failure)

Resource	Normal	Upon Failure
Memory	137 GB	109 GB
CPU	14	11
Net disk space	1,168 TB Approx. 400 GB per host, target 500 GB	918 GB

© 2012 Cisco and/or its affiliates. All rights reserved.

After discussion with the customer and based on the requirements the high-level design of the new solution defines that it will be virtualized – a VMware vSphere will be used to implement server virtualization with a selected set of features (i.e. VMware High-Availability and maximum amount of memory per host).

The characteristics of a new virtual environment are calculated from the desired state of new infrastructure with the limitations the customer has given i.e. no external storage. This demand along with virtualization platform requires the use of embedded shared datastore, because the VMs that will run the applications still need to failover to remaining ESXi host in case of an individual host failure.

https://t.me/learningnets

Storage Design with VSA

Characteristics

- Enables high availability, VMotion without external storage
- NFS based shared storage
- Up to 3 hosts
  - RAID 10 within host
  - Network mirror between servers
  - 4,6, or 8 disks per host
  - 2 GHz vCPU reservation

Solution

- 3 shared volumes each 500+GB with replica
- 1 local VMFS 100 GB for vCenter
- Minimum total = 6.2+ TB



Volume	Type	Size
Volume1	Primary-VSA	500
Volume2	Primary-VSA	500
Volume3	Primary-VSA	500
Volume1-R	Replica-VSA	500
Volume2-R	Replica-VSA	500
Volume3-R	Replica-VSA	500
Volume0	Primary-VMFS	100
VolumeX	Other	Remaining space

© 2012 Cisco and/or its affiliates. All rights reserved.

In VMware vSphere 5.0, VMware is releasing a new software storage appliance to the market called the vSphere Storage Appliance (VSA). This appliance provides an alternative shared storage solution for small-to-medium business (SMB) customers who might not be in a position to purchase a SAN or NAS array for their virtual infrastructure. Without shared storage configured in a vSphere environment, customers have not been able to exploit the unique features available in vSphere 5.0, such as vSphere High Availability (vSphere HA), vSphere VMotion, and vSphere Distributed Resource Scheduler.

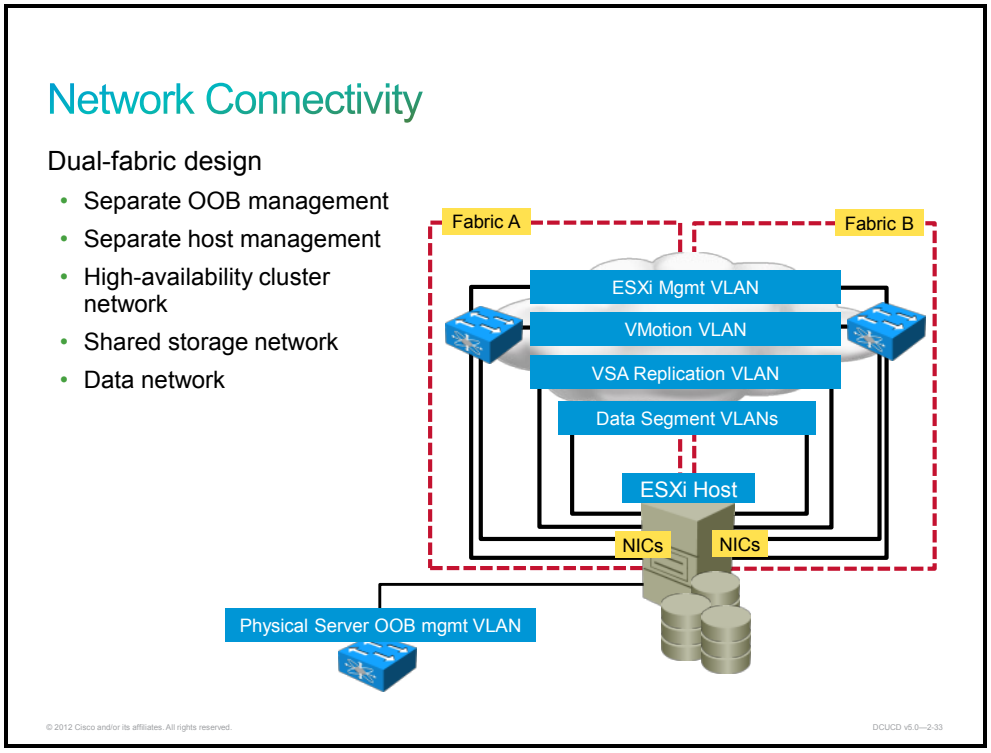
Architectural Overview

VSA can be deployed in a two-node or three-node configuration. Collectively, the two or three nodes in the VSA implementation are known as a VSA storage cluster. Each VMware ESXi server has a VSA instance deployed to it as a virtual machine. The VSA instance will then use the available space on the local disk(s) of the VMware ESXi servers to present one mirrored NFS volume per VMware ESXi. The NFS volume is presented to all VMware ESXi servers in the datacenter.

Each NFS datastore is a mirror, the source residing on one VSA (and thus, one VMware ESXi), and the target residing on a different VSA (and thus, a different VMware ESXi). Therefore, should one VSA (or one VMware ESXi) suffer a failure, the NFS datastore can still be presented, albeit from its mirror copy. This means that a failure in the cluster is transparent to any virtual machines running on that datastore.

The two-node VSA configuration uses a special VSA cluster service, which runs on the VMware vCenter Server. This behaves as a cluster member and is used to make sure that there is still a majority of members in the cluster, should one VMware ESXi server VSA member fail. In the figure above, the VSA datastores in the oval are NFS file systems presented as shared storage to the VMware ESXi servers in the datacenter.

https://t.me/learningnets



The network connectivity is also an important part of the assessment phase because it governs the adapter selection and connectivity type.

Each ESXi host in the design will have a separate out-of-band management i.e. Cisco UCS remote KVM via management Ethernet interface will be used.

Other segments will use dual-fabric design approach:

- ESXi host management VLAN segment will be connected via two virtual network interface cards (vNICs) in an active-standby manner.
- VMotion VLAN segment will be connected via two ESXi vNICs in a teaming manner.
- VSA Replication VLAN segment will be connected via two ESXi vNICs in a teaming manner.
- Data segment VLAN will also be connected via two ESXi vNICs in a teaming manner.

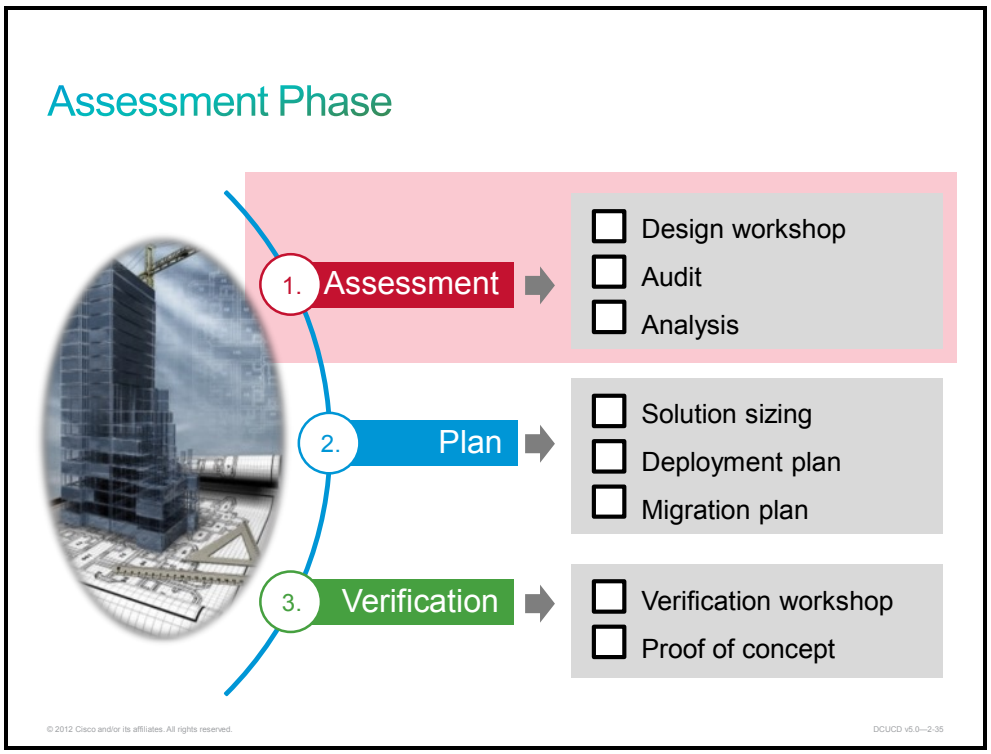
Altogether, nine vNICs will be used:

- Four vNICs connected to fabric A
- Four vNICs connected to fabric B
- One management vNIC connected

https://t.me/learningnets

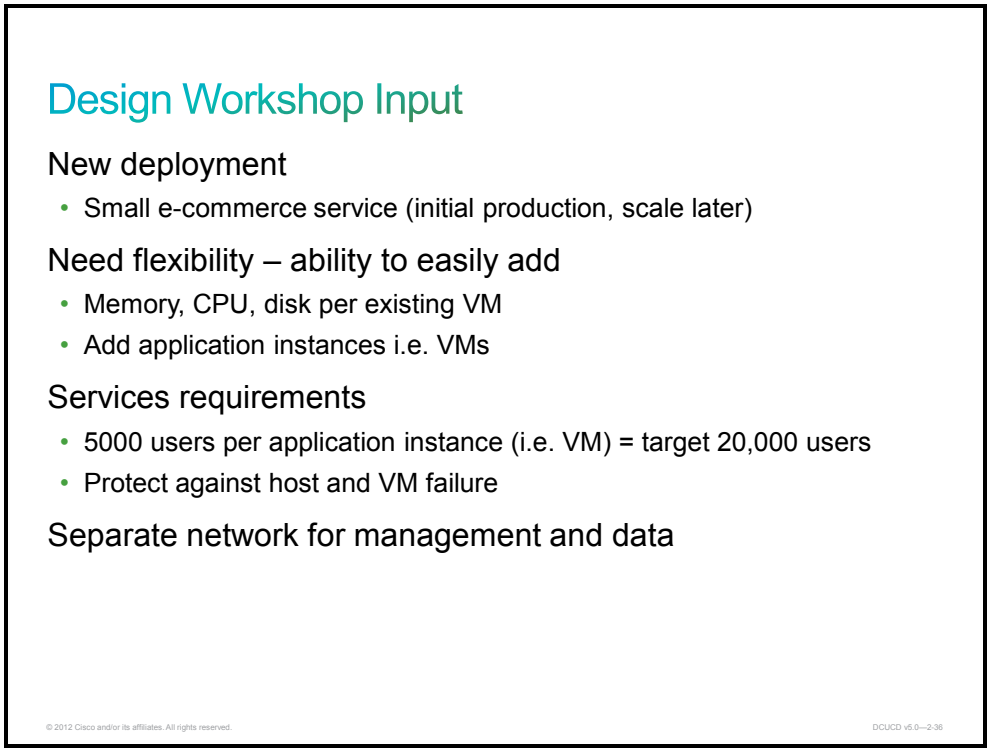
Assess Small Hyper-V Deployment Requirements

This topic recognizes and assesses a given small Hyper-V deployment requirements.



This second example is about a small Microsoft Hyper-V deployment that also has specific requirements. This topic describes an example of building such a solution from the beginning and gathering the requirements in with the assessment, which includes the design workshop and audit of the existing environment.

https://t.me/learningnets



The solution needs to meet the requirements of a customer building a new solution, a small e-commerce service where initially a smaller production needs to be created and later has to be allowed to be scaled.

The customer has identified the requirements from a services and flexibility perspective and has stated a wish to build the environment on the Microsoft Hyper-V virtualization platform.

Hyper-V Overview

Microsoft Hyper-V is offered as a server role that is packaged into the Windows Server 2008 R2 installation or as a standalone server. In either case, it is a hypervisor-based virtualization technology for x64 versions of Windows Server 2008. The hypervisor is a processor-specific virtualization platform that allows multiple isolated operating systems to share a single hardware platform. In order to run Hyper-V virtualization, the following system requirements must be met:

- An x86-64-capable processor running an x64 version of Windows Server 2008 Standard, Windows Server 2008 Enterprise, or Windows Server 2008 Datacenter.
- Hardware-assisted virtualization, which is available in processors that include a virtualization option—specifically Intel VT, which is an extension of the x86 architecture. The processor extension allows a virtual machine (VM) hypervisor to run an unmodified operating system without incurring significant performance penalties within operating system emulation.
- A CPU that is compatible with the no-execute (NX) bit must be available, and the hardware DEP bit must be enabled in the BIOS. For the Cisco UCS system, these are offered and enabled by default.
- At least 2 GB of memory must be available, and more memory may be needed based on the virtual operating system and application requirements. The standalone Hyper-V Server does not require an existing installation of Windows Server 2008. Minimum requirements are 1 GB of memory and 2 GB of disk space.

https://t.me/learningnets

Hyper-V isolates operating systems that are running on the virtual machines from each other through partitioning or logical isolation by the hypervisor. Each hypervisor instance has at least one parent partition that runs Windows Server 2008. The parent partition houses the virtualization stack, which has direct access to hardware devices such as network interface cards (NICs) and is responsible for creating the child partitions that host the guest operating systems. The parent partition creates these child partitions using the hypercall API, an application programming interface that is exposed by Hyper-V.

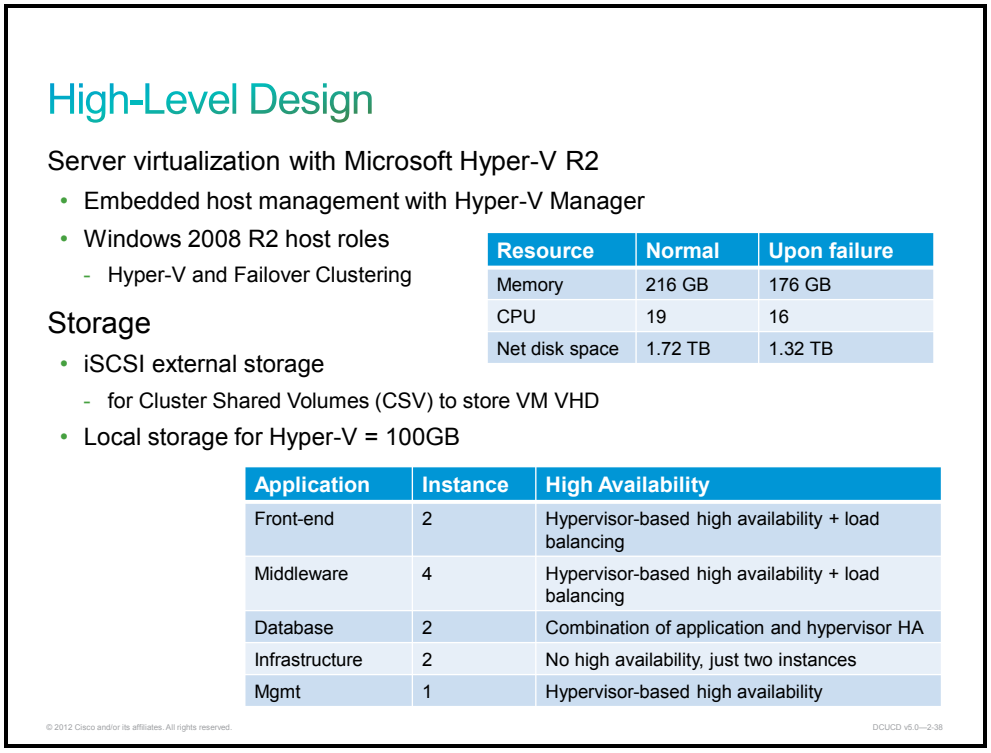
Solution Requirements Summary

Application	Required	Instances
Front-end (presentation)	Windows 2008 R2 16GB memory (ability to scale if required) 2 CPU; 60-GB disk	2
Middleware (application)	Windows 2008 R2 24GB memory (ability to scale if required) 2 CPU; 150-GB disk	4
Database backend	MS SQL 20008; Windows 2008 R2 32GB memory (ability to scale if required) 2 CPU; 300-GB disk	1+1
Infrastructure services (AD/DNS/DHCP)	Windows 2008 R2 8 GB memory 1 CPU; 100-GB disk	1+1
Mgmt	Windows 2008 R2 8GB memory; 1 CPU; 100 GB + 100 GB	1

© 2012 Cisco and/or its affiliates. All rights reserved.

Once again, the analysis of input gathered results in an overview of the existing versus the desired state of the server infrastructure. The applications are listed by their function along with resource requirements, operating system, and some with application versions. The important information also includes the number of instances of an individual application because that governs the amount of resources required.

https://t.me/learningnets



After discussion with the customer, and based on the requirements, the decision to virtualize the new environment has been taken.

The characteristics of a new virtual environment are calculated from the desired state of the new infrastructure with the limitations the customer has given.

To implement live migration and failover clustering to adhere to high-availability requirements, iSCSI external storage will be used. This means that CSV will be implemented to host the VM images.

Virtual Partitions

A virtualized partition does not have access to the physical processor, nor does it manage its real interrupts. Instead, it has a virtual view of the processor and runs in a guest virtual address space, which, depending on the configuration of the hypervisor, might not necessarily be the entire virtual address space. A hypervisor could choose to expose only a subset of the processors to each partition. The hypervisor intercepts the interrupts to the processor and redirects them to the respective partition using a logical synthetic interrupt controller (SynIC). Hyper-V can hardware accelerate the address translation between various guest virtual address spaces by using an I/O memory management unit (IOMMU), which operates independently of the memory management hardware that is used by the CPU.

Child partitions do not have direct access to hardware resources, but instead have a virtual view of the resources in terms of virtual devices. Any request to the virtual devices is redirected via the VMBus to the devices in the parent partition, which manage the requests. The VMBus is a logical channel that enables interpartition communication. The response is also redirected via the VMBus. If the devices in the parent partition are also virtual devices, the response is redirected further until it reaches the parent partition, where it gains access to the physical devices.

Parent partitions run a virtualization service provider (VSP), which connects to the VMBus and processes device access requests from child partitions. Child partition virtual devices internally run a virtualization service client (VSC), which redirects the request to VSPs in the parent partition via the VMBus. This entire process is transparent to the guest operating system.

https://t.me/learningnets

Cluster Shared Volumes

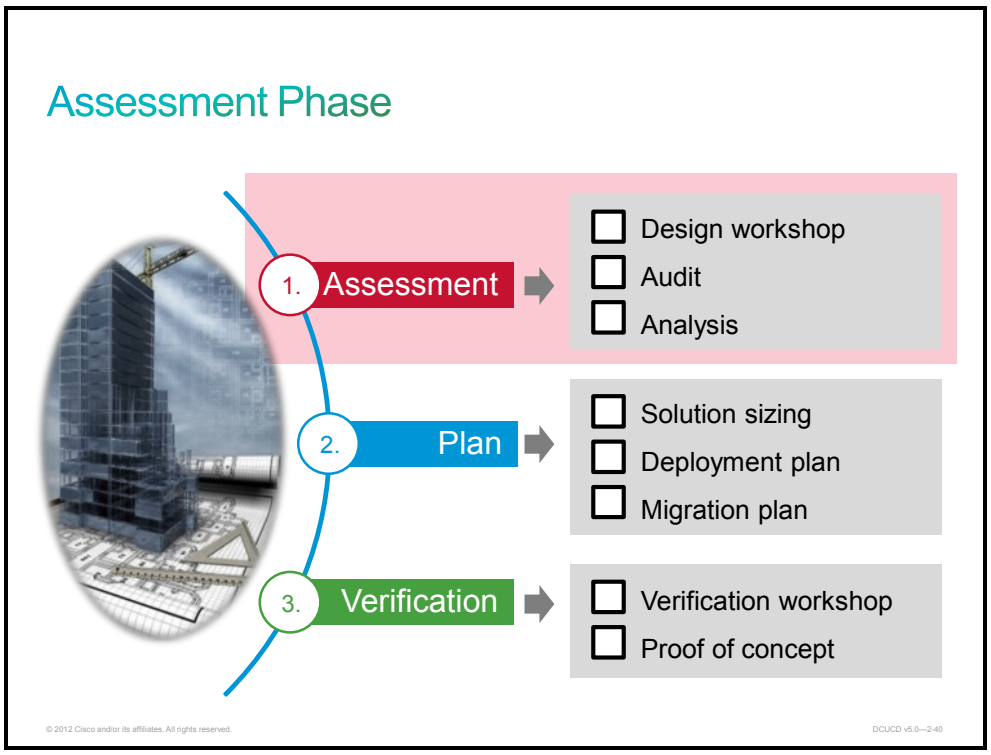
With Windows Server 2008 R2, Hyper-V uses CSV storage to support live migration of Hyper-V virtual machines from one Hyper-V server to another. CSV enables multiple Windows servers to access SAN storage using a single consistent namespace for all volumes on all hosts.

Multiple hosts can access the same logical unit number (LUN) on SAN storage so that virtual hard disk (VHD) files can be migrated from one hosting Hyper-V server to another. CSV is available as part of the failover clustering feature of Windows Server 2008 R2.

https://t.me/learningnets

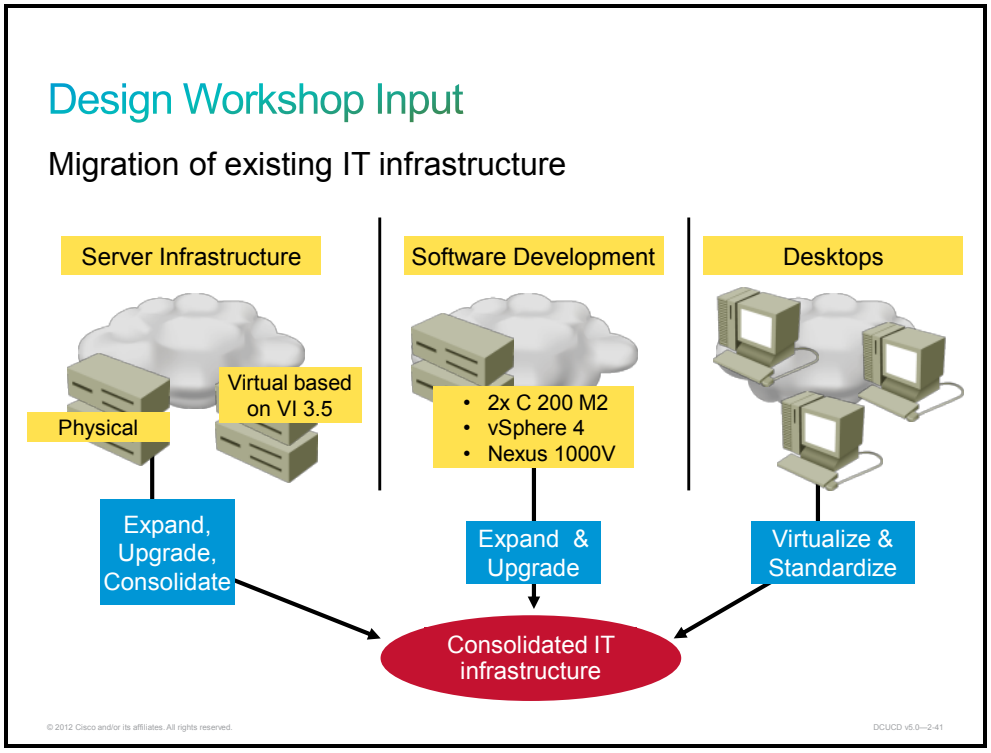
Assess VMware View VDI Deployment Requirements

This topic recognizes and assesses a given VMware VDI deployment requirements.



The third example concerns a larger environment where VDI will be added to the existing server virtualization to ease the user desktop and application management efforts. This topic describes an example of gathering the information and requirements in the design workshop and audit of the existing environment.

https://t.me/learningnets



Every VDI solution is constructed with building blocks that typically have the following characteristics:

- Desktop servers
  - Windows virtual desktops
  - Server blades hosting ESXi hosts
  - ESX cluster per blade server chassis
  - Resource pools per ESX cluster
- Infrastructure servers
  - Server blades hosting ESXi hosts
  - Virtual center/composer
  - View manager
  - Microsoft Windows Active Directory
- Storage
  - Storage device typically attached via Fibre Channel uplinks with properly sized LUNs and LUN masking

Design Workshop Output

The customer in this example has an existing environment that consists of the following segments:

- **Server infrastructure:** The existing infrastructure is based on the VMware Virtual Infrastructure 3.5 and some physical servers. The customer wants to upgrade and expand the physical as well as virtual platforms.

https://t.me/learningnets

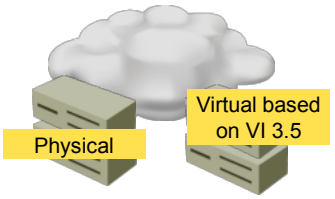
- **Software development infrastructure:** The software development department uses its own virtual infrastructure, which is based on the vSphere 4 with Cisco Nexus 1000V. The customer wants to expand, upgrade, and consolidate this infrastructure with the server infrastructure .

- **Physical desktops:** The existing physical desktop infrastructure is nearing its end of life, and the customer would need to replace many desktop computers. The customer has identified the VDI solution as a way to upgrade this part of the infrastructure. As with the software development, the goal is to consolidate on the same server infrastructure.

Server Infrastructure Input

Analysis

- Mixed environment of various servers
- 2 applications that cannot be virtualized
- Existing storage infrastructure
  - FC SAN based on MDS 9500
  - Drive arrays will be upgraded – capacity & performance wise



Requirements

- Upgrade virtual infrastructure and virtualize as much as possible
- Audit results

Resource	Value
Memory	600+ GB
CPU	50+

Application Type	HA Type	Resources
Mission-critical	FT	Memory =200+ GB CPU = 10
Regular	HA, 25% resources reserved	Memory = 400+ GB CPU = 40+

© 2012 Cisco and/or its affiliates. All rights reserved.

The existing server infrastructure has the following characteristics:

- It is a mixed environment of different type of servers—some quite old, some nearing the end of their life—and bought from various vendors and requiring a lot of management effort.
- There are two applications that cannot be virtualized, for these even with the new infrastructure the bare-metal OS installation will be used.
- There is an existing storage infrastructure – it is Fibre Channel-based and the customer already planned to upgrade it to fit the requirements of a larger virtual infrastructure.

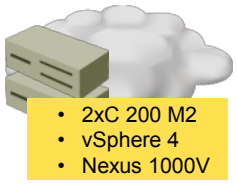
The goals of the migration are to upgrade the complete virtual infrastructure– the requirements are summarized in the tabular form.

https://t.me/learningnets

Software Development Infrastructure Input

Analysis

- Used by software development team (20 users)
- Windows 2003 VMs with development tool installed
  - VMs cloned from single image
  - Nexus 1000V with PVLANs to prevent name conflicts



Requirements

- Enable them to scale – more developers, larger VMs = double resources
- Simple HA for development VMs; 25% compute resources required
- Audit results

Resource	Value
Memory	80+ GB; 1:1.66 oversubscription ratio
CPU	20+; 1:4

Component	Value
CPU	Intel Xeon 5500, 4 core
Memory	48 (4-GB DIMMs)
Adapter	2GE LOM used
Disks	2x500G SATA in RAID1

© 2012 Cisco and/or its affiliates. All rights reserved.

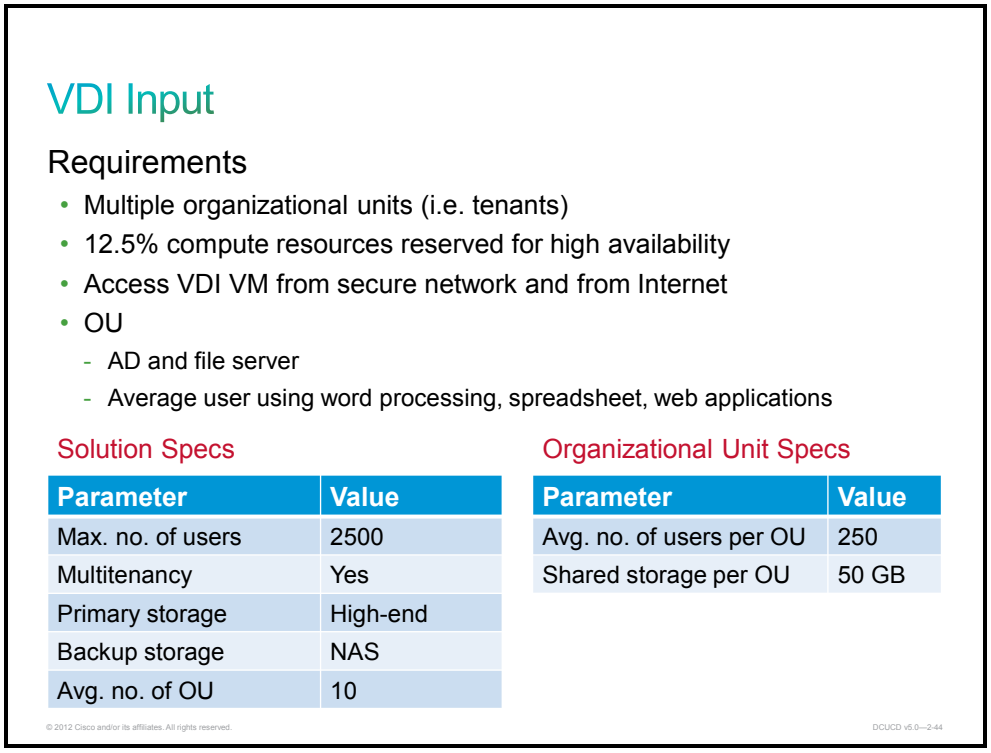
The existing software development environment has the following characteristics:

- It is used by 20 members of the development team.
- VMs are using Windows 2003 operating system.
- A member of the development team prepares the image for the development VM, which is then cloned so other members can use it on its own. This results in many VMs with the same settings—effectively, the hostname—which presents a challenge.
- To address the overlapping characteristics challenge, the environment is using Cisco Nexus 1000V with PVLAN functionality.

The requirements for this infrastructure are:

- The customer wants to scale the environment beyond 20 users; plus the amount of resources for individual VM have to be scaled.
- A simple high availability is sufficient for development VMs. It has been identified that 25 percent of resources should be enough for failover.
- The results of the design workshop and audit are presented in the figures.

https://t.me/learningnets



When we are talking about scalability, there are lots of factors to be considered. The most important is the workload definition and what kind of platform is needed to build that kind of a workload.

Secondly, a reasonable response time is required. Less than two seconds or three seconds is a decent response time.

From a scalability perspective, it is also important not to use 100 percent of the resources otherwise there will be a lot of thrashing or ballooning (in the case of VMware ESXi).

Finally, regarding storage, the amount needed is determined by the IOPS.

As with the previous segments, for VDI the requirements have been identified:

- There will be a multitenant environment because the customer wants to separate the VDI desktops of the organizational units.
- 12.5 percent is sufficient for high availability in case of host failure.
- There are some infrastructure services, and thus servers will be used by individual organizational units.
- The numbers (size, resource requirements) are given in the table in the figure.

https://t.me/learningnets

VDI Input (Cont.)

OU Virtual Server Specs (AD, File Server)

Parameter	Value
OS	Windows 2008 R2
VM memory	1 GB
Avg. disk space	32 GB
IOPS	10

User VDI VM specs

Parameter	Value
OS	Windows 7
VM memory	1.5 GB
Disk space	15 GB
IOPS	5

Oversubscription rates

Parameter	Ratio
VDI memory	1.5
VDI CPU	7
VS memory	1.4
VS CPU	5

© 2012 Cisco and/or its affiliates. All rights reserved.

The information about the desired VDI environment specifies:

- Requirements for the virtual servers, which will host infrastructure services per individual organizational unit
- Requirements for VDI VMs listing operating system version, assigned memory, disk space, and anticipated required IOPS—this defines the user profile
- Oversubscription rates, which are used to calculate physical resource requirements for the VDI as well as the virtual server VMs

https://t.me/learningnets

Solution Requirements Summary

Total

- Memory = 3848.5+ GB
- CPU = 488 +

Parameter	Memory	CPU	HA level	Total memory	Total CPU
Server VMs with FT	200+ GB	10+	100%	400+ GB	20+
Server VMs with HA	400+ GB	40+ CPU	25%	500+ GB	50+
Devel.VM with HA	2* 80 (oversub. 1:1.66)	2* 20 (oversub. 1:4)	25%	120 GB	12,5
VDI VMs	2500* 1.5GB (oversub 1:1.5)	2500*1 (oversub 1:7)	12,5%	2812, 5+GB	401+
VDI VSs	10OU * 2 srv * 1 GB (oversub 1:1.4)	10OU*2srv*1 (oversub 1:5)	Max. 12,5%	16+ GB	4.5+

© 2012 Cisco and/or its affiliates. All rights reserved.

The table summarizes the requirements of the entire environment calculated from the requirements about the individual segment.

https://t.me/learningnets

High-Level Design

Mixed environment

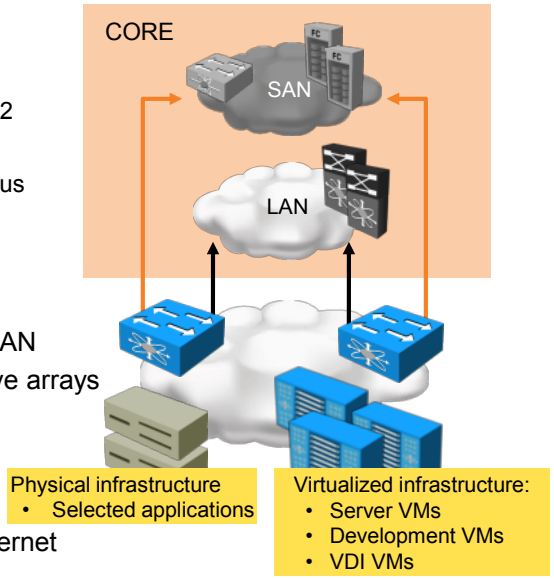
- Physical for selected apps
  - Repurpose existing C 200 M2
- Virtual infrastructure
  - VMware vSphere 5 – Ent. Plus
  - VMware View 5
  - Cisco Nexus 1000V

Storage

- Integrate with existing FC SAN
- Utilize upgraded central drive arrays
  - NFS for file sharing

Network

- Integrate with existing
- Upgraded to 10 Gigabit Ethernet



© 2012 Cisco and/or its affiliates. All rights reserved.

From the requirements, and in the discussion with the customer, a high-level design has been identified as presented in the figure scheme.

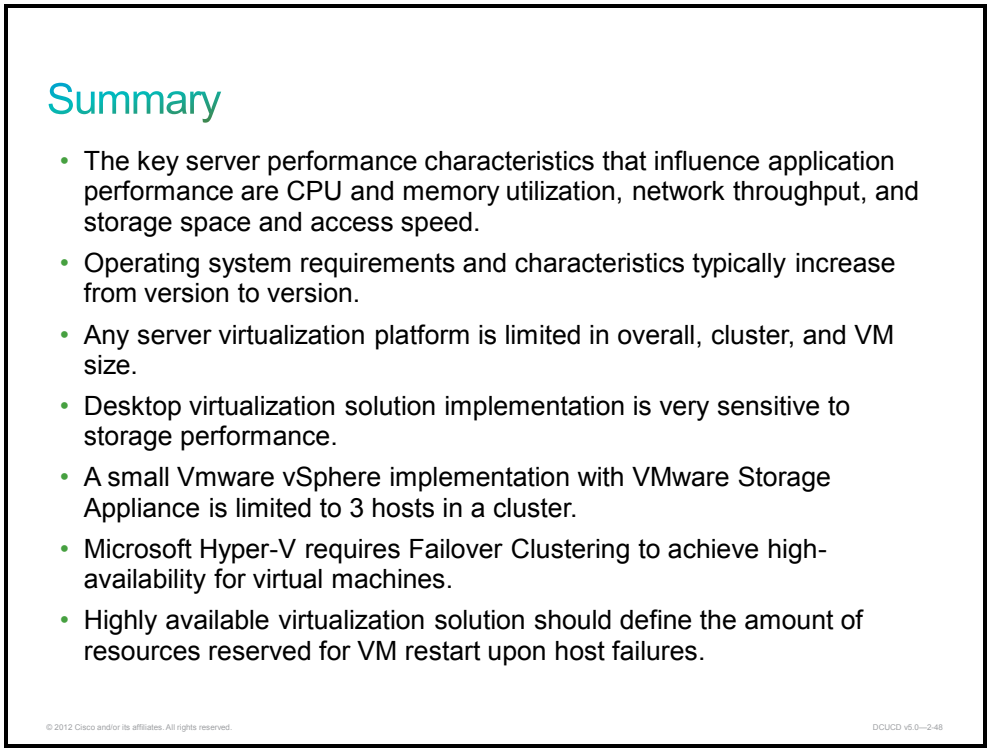
A new solution will be a mixed environment because it has been determined that the existing UCS C200 M2 servers can be repurposed for applications that cannot be virtualized.

Through the discussion with customer IT staff, it has been determined what kind of virtual platforms will be used to build a new solution.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- http://www.wmarow.com/strcalc/
- http://www.atlantiscomputing.com/products/reference-architecture/vmware-view
- http://myvirtualcloud.net/?page_id=1562
- http://myvirtualcloud.net/?page_id=2303
- http://myvirtualcloud.net/?page_id=1076

https://t.me/learningnets

Lesson 3

Employing Data Center Analysis Tools

Overview

Existing environments need to be examined not only to determine the input requirements for a new solution, but also to provide the performance baseline that can be used to verify if the new design has met expected goals and is performing better than the existing one. The task of gathering the historical performance information manually is near to impossible, and so the analysis tools are used for the purpose.

This lesson identifies the reconnaissance and analysis tools, their characteristics and required functionalities, and describes the use of such tools for examining the performance characteristics of the given computing solution.

Objectives

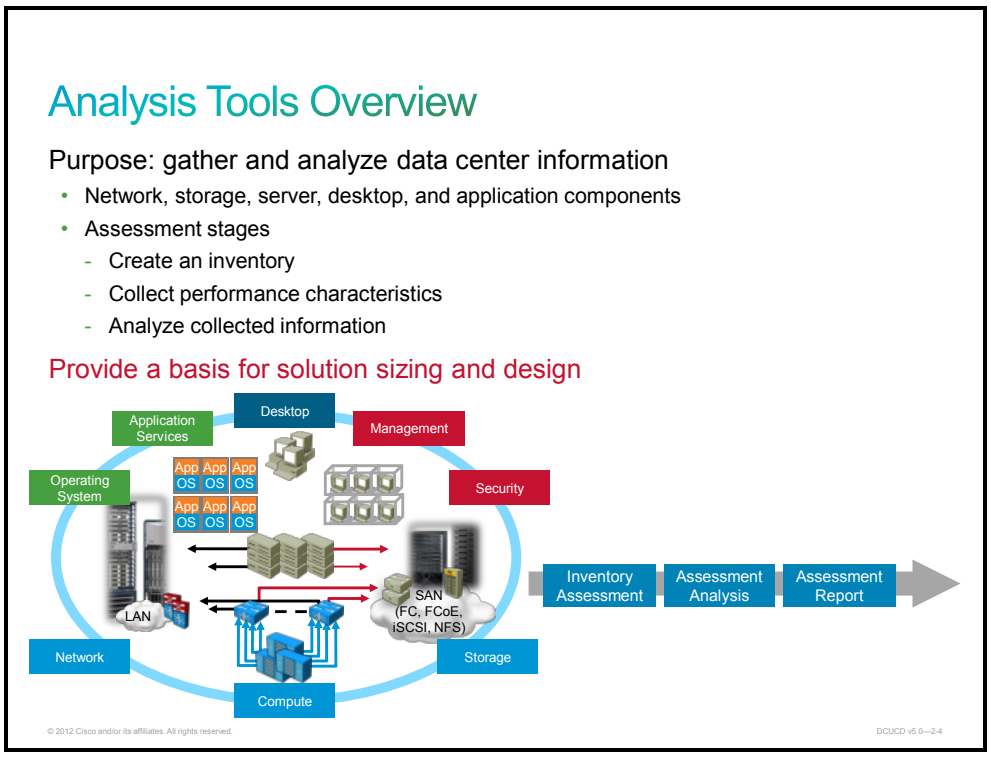
Upon completing this lesson, you will be able to use the reconnaissance and analysis tools to examine performance characteristics of the given computing solution. You will be able to meet these objectives:

- Evaluate reconnaissance and analysis tools
- Discuss general steps of running an analysis with the selected tool
- Perform existing computing solution analysis with VMware Capacity Planner
- Perform VMware vSphere analysis with VMware CapacityIQ
- Perform existing computing solution analysis with Microsoft Assessment and Planning toolkit
- Evaluate the Cisco UCS TCO/ROI Advisor tool

https://t.me/learningnets

Reconnaissance and Analysis Tools

This topic describes reconnaissance and analysis tools.



The reconnaissance and analysis tools can gather information about data center resource utilization for network, storage, server, and desktop components. The primary purpose of using these tools is to get a solid basis for sizing and designing a data center solution that is composed of all the components mentioned previously.

Because the data center consists of so many components, not every tool can assess and analyze all the data center aspects. More often, dedicated tools are used to assess individual components.

The analysis and reconnaissance tools that are used to assess and analyze the compute component of the data center (the compute component includes the server, desktop, and application aspects and characteristics) typically also provide information about the network and storage characteristics as they pertain to the compute component.

A tool can be used to list the inventory of the existing environment or to analyze the current utilization and workload. Usually, tools are used to prepare the ground for planning a new solution that can be either completely physical, a mixed physical and virtual environment, or even a completely virtualized environment. In any of these cases, the analysis outcome is the basis for the server consolidation, such as reducing the number of physical servers by using more powerful processors, more processors, and more memory per individual server. Combined with computing virtualization, the consolidation ratio can be even higher.

When planning for virtualization, the server historical performance is very important because it governs the physical server dimensions as well as how the physical server resources are divided between the virtual machines (VMs).

https://t.me/learningnets

Purpose

The analysis and reconnaissance tools are used most often for the following purposes:

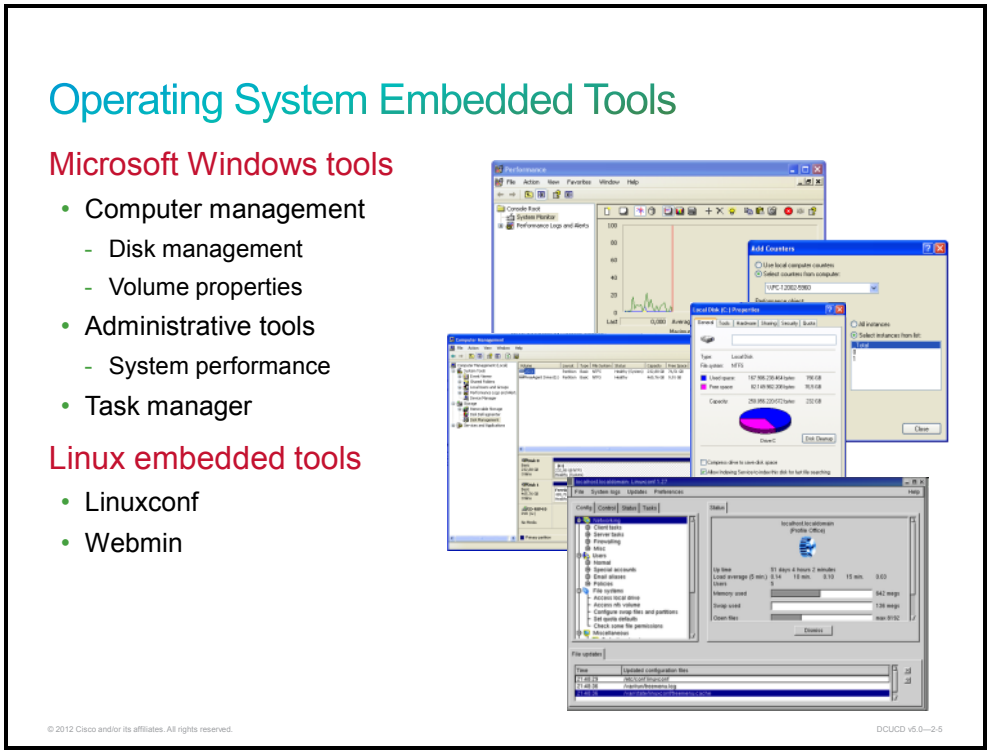
- Assess the current workload capacity of the IT infrastructure through comprehensive discovery and inventory of IT assets
- Measure system workloads and capacity utilization across various elements of the IT infrastructure, including by function, location, and environment
- Plan for capacity optimization through detailed utilization analysis, benchmarking, trending, and identification of capacity optimization alternatives
- Identify resources and establish a plan for virtualization, hardware purchase, or resource deployment
- Decide on the optimal solution by evaluating various alternatives through scenario modeling and "what-if" analyses
- Determine which alternative best meets the predefined criteria
- Monitor resource utilization through anomaly detection and alerts that are based on benchmarked thresholds
- Help to generate recommendations to ensure ongoing capacity optimization

Analysis Phases

The assessment and analysis can be divided into three distinct phases or stages:

- **Phase 1:** Creating the existing environment inventory
- **Phase 2:** Collecting performance characteristics and metrics of the existing environment that are based on the inventory information
- **Phase 3:** Utilizing data that is gathered to run current utilization analysis, "what-if" analyses, and various scenarios to create potential solutions with different consolidation ratios

https://t.me/learningnets



The operating systems typically have built-in tools that can be used for basic analysis and data collection.

Windows-Embedded Tools

The Microsoft Windows operating system offers a selection of embedded tools that can be used to gather the historical performance characteristics for a single server.

The figure introduces the following Microsoft Windows embedded tools that can be used to perform server analysis:

- Computer management
- Administrative tools
- Task manager

The Linux operating system is available with many applications and utilities, so the embedded tools that can be used to gather historical performance characteristics vary per Linux distribution.

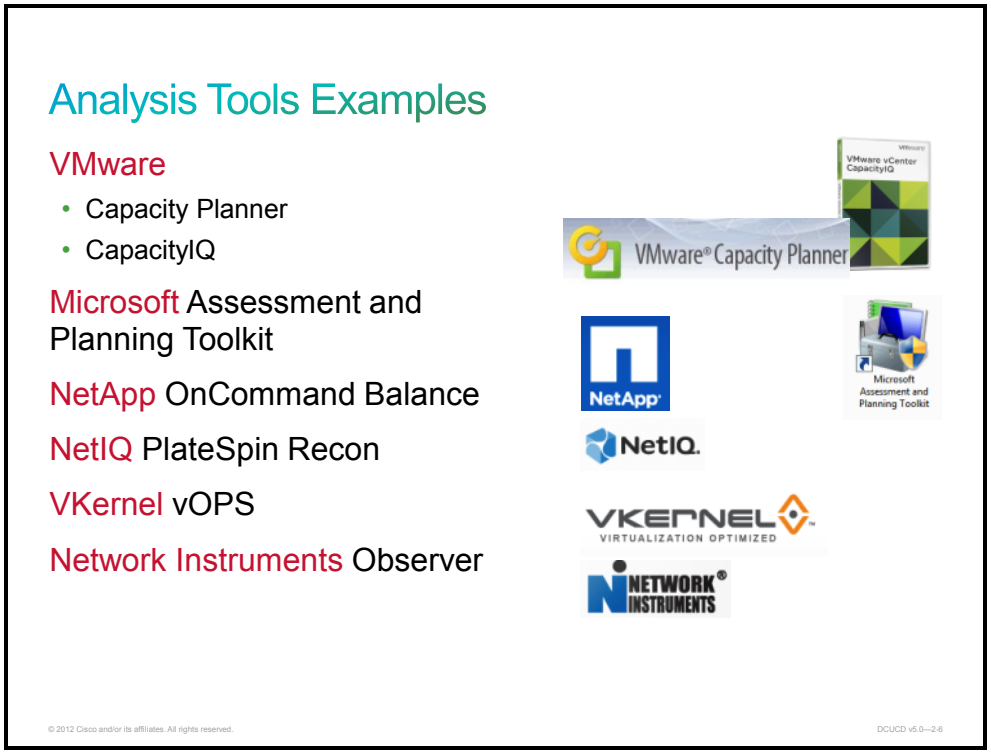
Linux-Embedded Tools

Linuxconf comes with Mandrake Linux and Red Hat Linux, but Linuxconf is also available for most modern Linux distributions. Multiple interfaces for Linuxconf are available: GUI, web, command line, and curses.

Webmin is a web-based modular application. It offers a set of core modules that manage the usual system administration functionality, and there are also third-party modules available for administering various packages and services. Webmin is available in a number of formats that are specific to different distributions.

Note	Any user can install Linuxconf, but Webmin must be installed by the root user. After installation, you can access this tool from any user account as long as you know the root password.

https://t.me/learningnets



VMware Capacity Planner and CapacityIQ

There are two tools available from VMware. The Capacity Planner is a previrtualization tool that is used mainly to analyze the pre-VMware virtualization environments. The second tool is CapacityIQ, which is used to analyze the VMware-virtualized environment.

More information about both tools is provided in the following topics.

Microsoft Assessment and Planning Toolkit

The Microsoft Assessment and Planning (MAP) Toolkit is similar to VMware Capacity Planner and CapacityIQ. More information is provided in the following topic.

NetApp OnCommand Balance

OnCommand Balance is another agentless management software with advanced analytics to help fix problems, optimize utilization, and improve performance in the virtualized data center. It helps companies to ensure that their data center server and storage virtual and physical systems deliver the best application performance possible.

The tools can help reduce operations and infrastructure costs by using servers and storage more efficiently and reduce the time and resources that are spent on management through automation. BalancePoint supports a wide range of applications, servers, and storage systems. It can assist with the following tasks:

- Manage performance across applications, servers, and storage with cross-domain visualization and performance-based service alerting. Hidden contention, hotspots, and bottlenecks can be found using this tool. In addition, when problems are found, the program provides direct automatic troubleshooting analysis.
- Optimize application performance and resource utilization. You can manage IT as a business with performance indicators that specify the optimal balance between application requirements and resource capabilities.

https://t.me/learningnets

NetIQ PlateSpin Recon

PlateSpin Recon offers reconnaissance tools that can be used for tasks similar to those performed by VMware Capacity Planner and OnCommand Balance.

These tools can be used for workload profiling. PlateSpin Recon tracks the actual CPU, disk, memory, and network utilization over time on both physical and virtual hosts. Every server workload and virtual host has utilization peaks and valleys, and PlateSpin Recon can build consolidation scenarios that are based on adjusting these peaks and valleys.

VKernel vOPS

VKernel offers performance and capacity management solutions that can be used for analysis:

- **Capacity Analyzer:** This tool can be used to avoid Hyper-V and VMware performance problems. The architect can perform capacity planning and capacity management to meet current and future application demands.
- **Optimization Pack:** This tool can be used to maintain VMware performance while reclaiming overallocated CPU, memory, and storage. It can also delete abandoned VMs, snapshots, and templates.
- **Inventory tool:** This tool can be used for organizing, filtering, detailing, and reporting on all VMs in an environment. It is maintained with thorough VM change records.

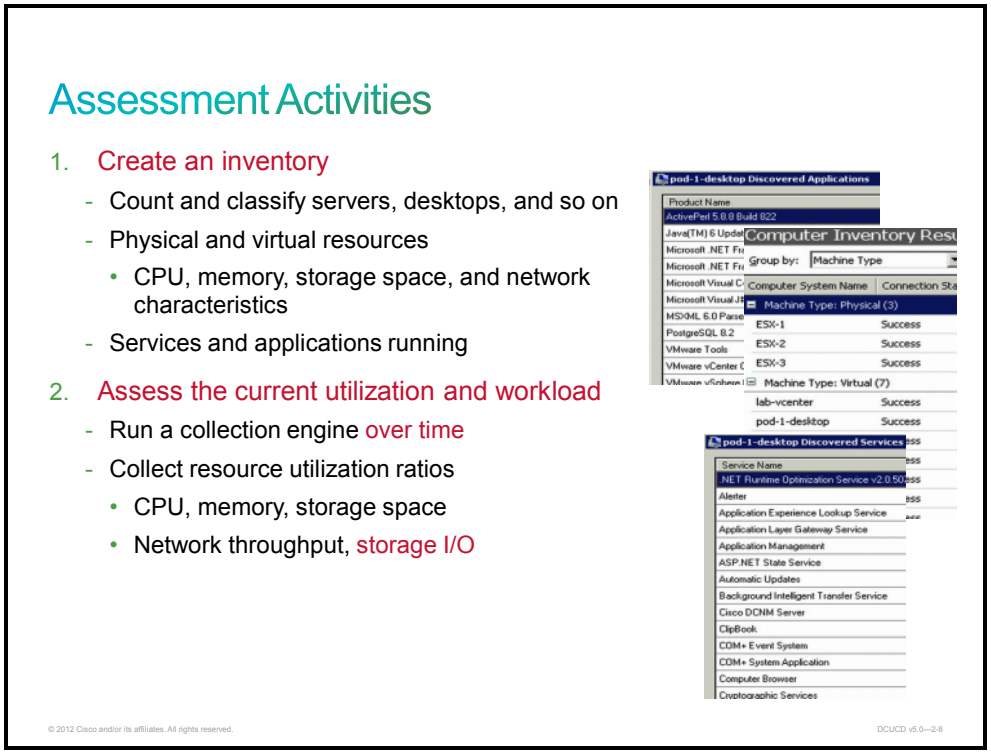
Network Instruments Observer

The Observer tool exists in three packages: Standard, Expert, and Suite. The tool helps architects to better know, optimize, and repair complex networks. This work requires complete visibility into the network and the devices that are residing on it.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Use Reconnaissance and Analysis Tools

This topic discusses general steps of running an analysis with the selected tool.



The analysis activities can be divided into assessment and postassessment activities. The assessment activities are the first two stages of the analysis.

Inventory Collection

In this stage, the operator uses the tool to build the inventory of the existing environment. One of the aspects is to count the number of servers and desktops in the environment. The other is to record the following physical and virtual characteristics of the individual machines (whether they are servers or desktops):

- **CPU characteristics:** Vendor and type, number of CPU sockets and cores per CPU, speed, cache size
- **Memory characteristics:** Vendor and type of DIMMs, number of DIMMs and total memory size, memory speed
- **Network characteristics:** Network interface card (NIC) type and quantity, speed, host bus adapter (HBA) type and quantity, speed
- **Storage space characteristics:** Type and size of local and remote storage, individual volume size, and number of storage spaces
- **Applications and services characteristics:** Application or service vendor, type, and version

https://t.me/learningnets

Utilization and Workload Collection

After the inventory database is built, it can be used to query the systems and gather the performance metrics (such as the counters) for individual resources that are used. The most important metrics are the following:

- **CPU utilization levels:** These should include not only the average levels but also the minimum and peak levels, with the busiest hours being most important.
- **Memory utilization levels (percentage):** This level is similar to the CPU utilization levels.
- **Storage space utilization level:** This level shows how storage space requirements have grown (or have been reduced, which happens less often) over time.
- **Network utilization:** This information shows the amount of traffic that is received and sent by the individual machine.
- **Storage I/Os:** This setting shows the amount of I/O operations per second (IOPS) that are used by the machine. If the required IOPS is over the IOPS available to the machine, it can reduce the application or service speed running on that machine.

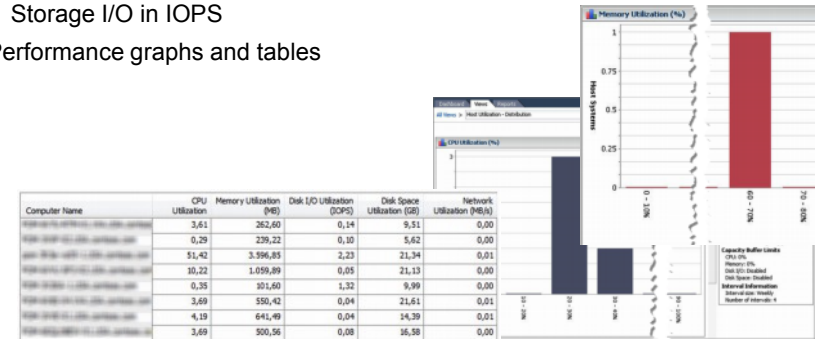
An important aspect of this phase is the collection interval. It should be long enough to collect during the relevant peak hours as well as periods when resources are under more stress. The collection interval can last for two months, three months, or even longer.

The recommended practice is to keep the collection running for at least one month, which on average produces the desired results. However, the optimal interval really depends on the customer business process (for example, periodic data manipulation, such as monthly payrolls, and so on) of the person who is conducting the analysis with the tool. The person using the tool should understand the processes, activities, and applications and the ways that they are used.

© 2012 Cisco Systems, Inc
https://t.me/learningnets

Postassessment Activities

3. Review the assessment reports
   - Minimum, peak, and average load and utilization ratios as applicable to individual resource type
     - CPU, memory percentage per total capacity
     - Storage space in gigabytes
     - Network utilization in megabits per second
     - Storage I/O in IOPS
   - Performance graphs and tables



Computer Name	CPU Utilization	Memory Utilization (MB)	Disk I/O Utilization (IOPS)	Disk Space Utilization (GB)	Network Utilization (MB/s)
[redacted]	3,61	262,60	0,14	9,51	0,00
[redacted]	0,29	239,22	0,10	5,62	0,00
[redacted]	51,42	3.596,85	2,23	21,34	0,01
[redacted]	10,22	1.059,89	0,05	21,13	0,00
[redacted]	0,35	101,60	1,32	9,99	0,00
[redacted]	3,69	550,42	0,04	21,61	0,01
[redacted]	4,19	641,49	0,04	14,39	0,01
[redacted]	3,69	500,56	0,08	16,58	0,00

© 2012 Cisco and/or its affiliates. All rights reserved.

The postassessment activities are the third stage of the analysis. These activities can be divided into two distinct groups: assessment reports and solution planning.

Assessment Reports

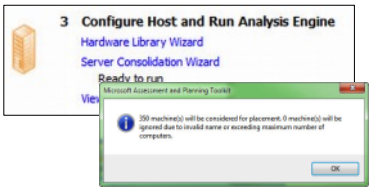
The assessment reports are used by the operator to review what has been collected by the tool. Operator interaction with the reports can be as simple as reviewing plain inventory data or more complex, such as running different reports to group and present the utilization and workload data. This data describes individual resources like CPU, memory, storage space, network, and storage connectivity in either performance graphs or tables.

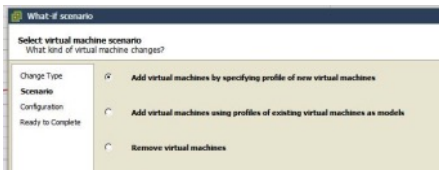
The data that is collected in the previous steps is usually saved in some form of database that can be queried by using prebuilt or custom-made queries to extract the information that is relevant to the person who is conducting the analysis.

https://t.me/learningnets

Postassessment Activities (Cont.)

4. Help with the solution planning
   - Provide benchmarking based on references
   - Plan capacity optimization
     - Server consolidation
     - Consolidation with virtualization
     - Planning per eco-partner specific solution
   - Perform "what-if" analysis
     - Add or remove resources and demonstrate changes to the utilization/workload





© 2012 Cisco and/or its affiliates. All rights reserved.

The second part of the postassessment activities presents the advanced functionality of the tool, which is running "what-if" scenarios and getting comparisons and details of consolidation scenarios. These activities can help the data center architect to plan and design the solution that best suits the customer needs.

The more advanced tools utilize large knowledge databases to perform the benchmarking of the environment and present a report that proposes the capacity and infrastructure optimization of the compute part of the data center. The person running the analysis usually has the following options:

- Server consolidation options from conservative to more aggressive for new environment characteristics
- Server consolidation with virtualization getting even more aggressive consolidation ratios along with the information about the virtualization environment requirements, such as licensing

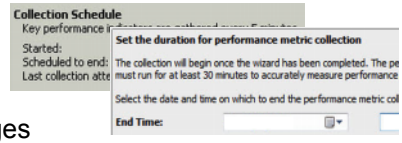
Another option of the tool is to conduct the "what-if" analysis to see how the environment can scale by adding or removing certain resources and see how that reflects on the utilization and workload. These predictions are based on the utilization and workload history that was collected in the previous phase.

https://t.me/learningnets

Preparing to Run Analysis

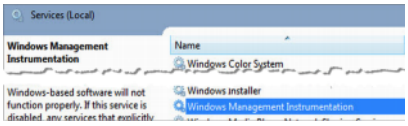
Define the analysis parameters.

- Size of the environment being analyzed (limitations)
- Collection duration and intervals
  - Should be run over a longer period to gather relevant data



Fulfill the collection requirements.

- User account with administrative privileges
  - Create dedicated time-limited administrative accounts
- Enable protocols (WMI, SSH) used for collection on systems to be analyzed
- Agent versus agentless collection



© 2012 Cisco and/or its affiliates. All rights reserved.

Before the analysis is conducted, some preparation must be done and certain requirements need to be met.

First, the following analysis parameters have to be defined:

- **Size of environment being analyzed:** This measurement does not have to be a precise number, but it has to give the person conducting the analysis an idea of how large the environment is in terms of number of servers and desktops. This number is important because analysis tools typically have collectors that can manage a certain maximum number of systems.
- **The collection duration and intervals:** As mentioned earlier, this evaluation is important from the analysis perspective of getting the relevant data. If the interval is too small, it might not collect all the peak utilization times and levels, which could in turn lead to an undersized solution if it is based on the outcome of such an analysis. The data center architect should understand the customer business process—and perhaps the applications that are used—in order to determine approximately how long the collection period should be so that the relevant utilization levels are collected. It is recommended that the collection period should be at least one month in order to gather the relevant data.

https://t.me/learningnets

Second, in order to be able to collect the data, the data center architect should know what means the analysis tools use to gather the data.

- The tool usually needs administrative access granted to the systems being surveyed. The recommended practice is to create time-limited, dedicated administration accounts for the time of the collection period.
- To collect the information, certain protocols are used that need to be enabled on the systems that are being analyzed if they are not already enabled by default. Typically, the Windows Management Instrumentation (WMI) is used to collect data about the Microsoft Windows-based machines, and Secure Shell (SSH) is used to run scripts to collect data on Linux and UNIX systems.
- If the tool uses some custom-based method of data collection, it may require an agent to be installed on the systems that are being surveyed. Typically, the tools are agentless, so there is no need for intervention and software installation on the surveyed systems.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Employ VMware Capacity Planner

This topic describes how to perform existing computing solution analysis with VMware Capacity Planner.

VMware Capacity Planner Overview

Hosted application service previrtualization tool used for server and desktop environments

Output—consolidation blueprint with VM configuration and recommendations



Answers the following questions:

- How many systems are present?
- What is the individual system performance and utilization profile?
- Which virtualization strategy should be used?
- What is the number of new systems that is required?
- What is the number of VMware infrastructure licenses that is required?

© 2012 Cisco and/or its affiliates. All rights reserved.

VMware Capacity Planner is a previrtualization tool that is used in physical environments to provide a server virtualization and consolidation-stacking plan. It is used to plan for capacity optimization and to design an optimal solution to achieve maximum performance. It assists in IT resource utilization and in the development of a virtualization plan for server containment and consolidation.

The tool collects comprehensive resource utilization data in heterogeneous environments and then compares it to industry standard reference data to provide analysis and decision-support modeling. Its detailed capacity analysis includes CPU, memory, network, and disk utilization across every server that is monitored.

The output of the analysis is the server consolidation blueprint with VMware in mind.

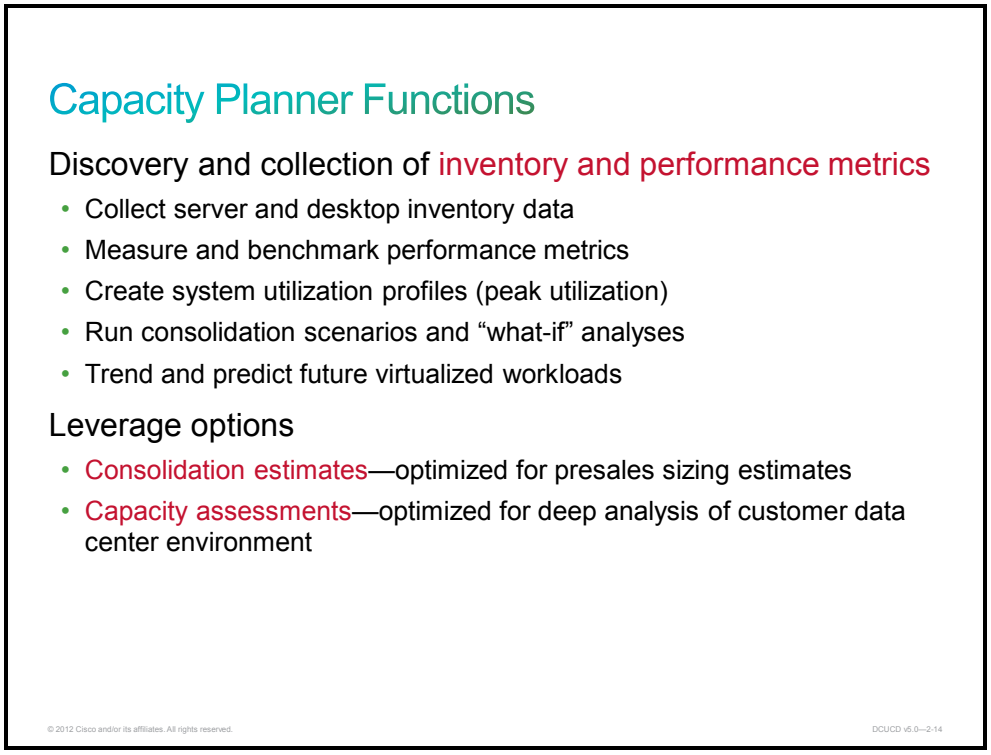
The architect using the tool can use the analysis to answer the following questions:

- How many systems do I have?
- What is my system performance and utilization profile?
- What is the best virtualization strategy for me?
- How many new systems and VMware infrastructure licenses should I purchase?

Web-Based Hosted Application Service

The Capacity Planner Dashboard component is delivered as a web-based application that delivers rich analysis, modeling, and decision-support capabilities that are based on the data that is collected from your data center. Service providers can use this interface to access prebuilt report templates and create custom reports depending on the type of assessment that is being delivered.

https://t.me/learningnets



VMware Capacity Planner includes components that deliver integrated capacity-planning functions, including data collection, data analysis, capacity monitoring, and decision-making capabilities.

Capacity Planner offers the following to the solution architect:

- Comprehensive collection:
  - Simple collector:
    - Supports up to 500 systems
    - Easy to scale by adding collectors as needed
    - Metrics reconciliation in the dashboard
  - Easy administration:
    - No maintenance
    - Remote Monitoring
  - Robust collection:
    - Support for many operating systems
    - Inventory
    - Performance
- Detailed analysis:
  - **Time-based:** Compatible systems, hourly patterns, peak hours
  - **All resources:** CPU, memory, disk, network
  - **Conversion factors:** CPU speed, virtualization overhead, virtualization benefits
  - **Comparisons:** Hardware, groups, and thresholds

https://t.me/learningnets

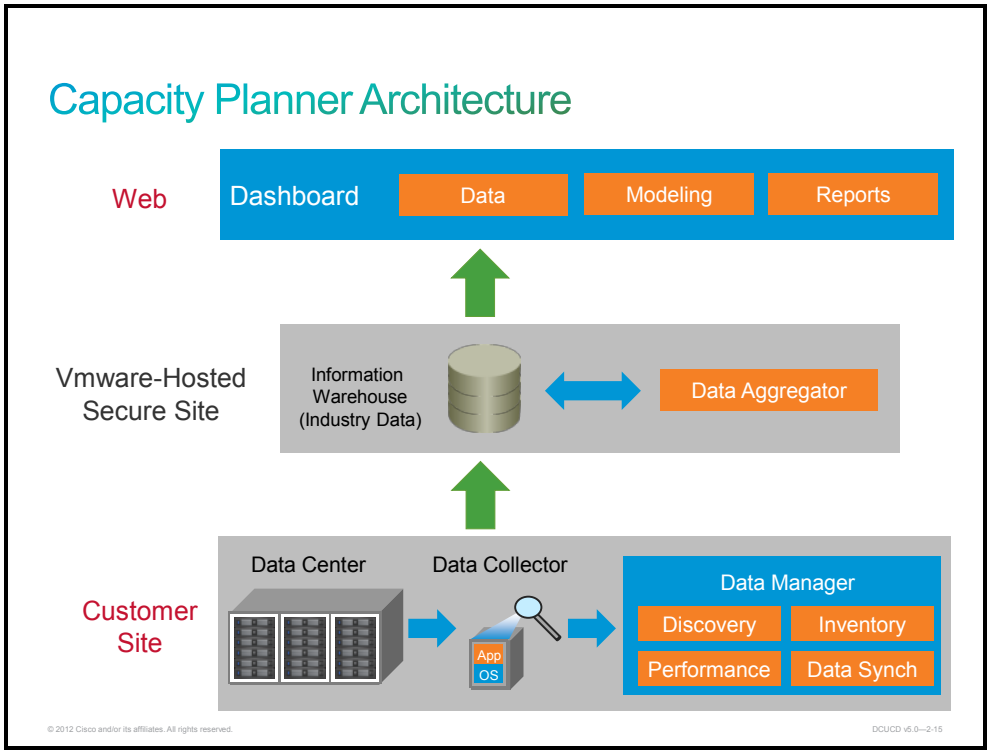
- Flexible reporting:
  - Provides environment summary and consolidation recommendation
  - **Canned reports:** Professional, easy to read, and quick to generate
  - **Custom reports:** Report builder with wizards

Leverage Options

The tool has two leverage options:

- **Consolidation estimates (CEs):** These are optimized for sales professionals to conduct presales sizing estimates of a customer data center environment. CEs provide guidance on what can be achieved via virtualization and consolidation assessments. Consolidation estimates show the potential state of the data center. The CE is a simplified, guided workflow of Capacity Planner and a defined process for sales professionals to conduct more sizing and lead-generation activity.
- **Capacity assessments (CAs):** These are optimized for consulting or services professionals who need to analyze a customer data center environment. CAs provide a detailed plan for how customers can achieve an optimized, virtualized data center. Capacity assessments show the implementation blueprint that can be used for the data center. CAs leverage the complete power and flexibility of Capacity Planner and the industry expertise of professional services to conduct more virtualization assessments and data center transformation.

https://t.me/learningnets



The tool is a web-based application that combines inventory and utilization data.

Data Collector

Installed locally at the client site, this component uses an agentless implementation to discover server or desktop systems, collect detailed hardware and software inventory data, and gather key performance metrics that are required for capacity-utilization analysis. The Data Collector can gather data from heterogeneous environments that are based on multiple platforms.

Data Manager

This component manages the data-collection process, providing an organized view of the collected information as well as administrative controls for the Data Collector. The Data Manager "anonymizes" and securely sends the collected data to the centralized Information Warehouse.

Information Warehouse

This component is a hosted data warehouse where the data collected from the client environment is sent to be scrubbed, aggregated, and prepared for analysis. The Information Warehouse also includes valuable industry benchmark data that can be leveraged for benchmarking, scenario modeling, and setting utilization thresholds.

Data Analyzer

This component serves as the core analytical engine that processes all the analyses that are required for intelligent capacity planning. It includes advanced algorithms that solve capacity-optimization problems and supports analysis capabilities such as aggregation, trending, and benchmarking. Scenario modeling and what-if analyses help to model and test various planning scenarios.

https://t.me/learningnets

Dashboard

This web-based, hosted application can deliver capacity analysis and planning capabilities to users through a browser interface. Users can remotely access a rich set of prebuilt analyses and "slice and dice" data, and they can create custom reports. Planning capabilities let you set objectives and constraints and also model and test scenarios to arrive at intelligent capacity decisions.

Capacity Planner Requirements

Eligibility

- VMware partners (license-free)—at least one VMware Capacity Planner-trained engineer
- Individuals can attend training and receive a flexible deployment license

Required resources

- Analysis engineer with basic system administration skills for Microsoft, Linux, or UNIX
- System running Data Manager



Parameter	Value
Operating system	Windows 2000/XP/2003/2008
Processor	At least 1.5 GHz CPU
Memory	At least 1 GB
Storage space	2 GB
Settings	English language Firewall deactivated

© 2012 Cisco and/or its affiliates. All rights reserved.

Any individual from a valid VMware partner may attend the two-day VMware Capacity Planner class. Upon completion of the course, your organization may purchase VMware Capacity Planner flexible deployment license bundles for your trained users to use at your customer sites.

Any partner organization is eligible to use VMware Capacity Planner under the following conditions:

- Participation in the VMware partner program (member in good standing)
- Completion of the two-day Capacity Planner training course (live or e-learning)

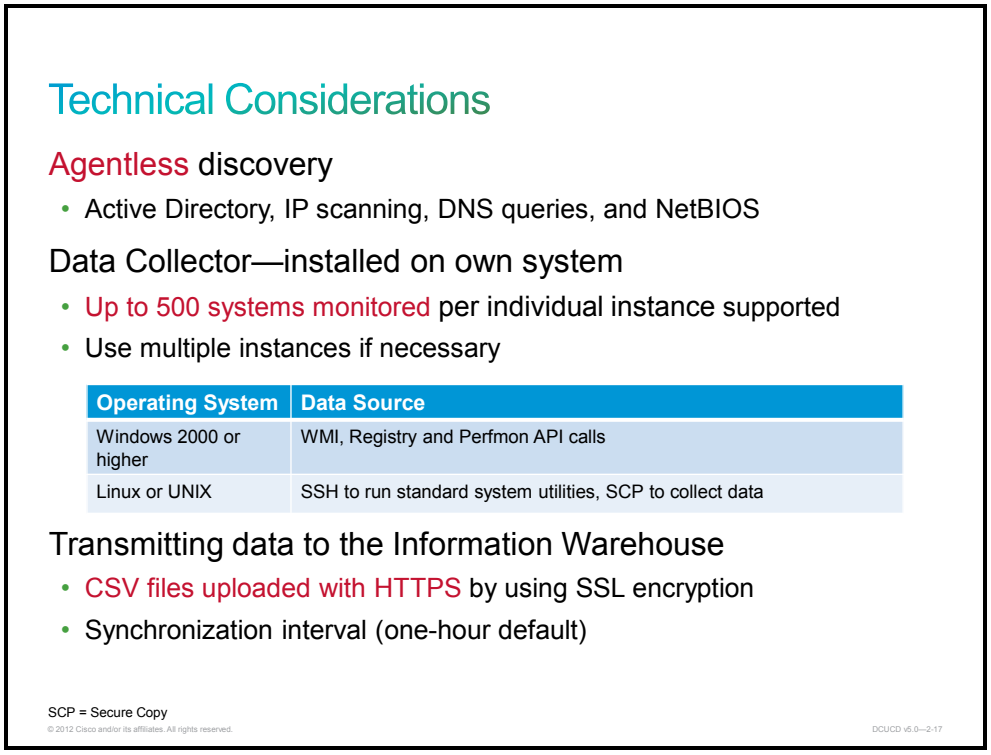
The license for the partner organization is complimentary.

To conduct the analysis, the Data Manager has to be installed on a dedicated system that fulfills the following minimum requirements:

- Microsoft Windows 2000, XP, 2003, or 2008
- At least 1.5 GHz CPU
- At least 1 GB of RAM
- 2 GB of free disk space
- ASCII language operating system (English) and firewall functionality disabled

This system can be a physical desktop, a server, or a virtual machine, as long as it is not used for anything else but running the Data Manager.

https://t.me/learningnets



Agent-Free Implementation

The VMware Capacity Planner Data Collector is installed onsite at the data center that is being assessed. This component collects detailed hardware and software metrics that are required for capacity utilization analysis across a broad range of platforms, but without the use of software agents.

The agentless discovery utilizes Microsoft Active Directory, IP scanning, Domain Name System (DNS) queries, and NetBIOS.

Data Collector

An individual data collector can survey up to approximately 500 systems. If a larger environment has to be analyzed, multiple collectors should be used. The data collectors should also be installed on separate Microsoft Windows-based systems.

As a source of the information and collection interface, the Data Collector uses (depending on the operating system type) the following data sources: WMI, registry and Perfmon API calls on Windows 2000 or higher, or remote SSH sessions that use UNIX and Linux utilities. The protocols that are mentioned have to be enabled on the systems that are surveyed and also allowed along the path between the Data Collector and the system. (This setting means that any firewalls have to be enabled with the proper configuration to allow the communication between the Collector and the system that is surveyed.)

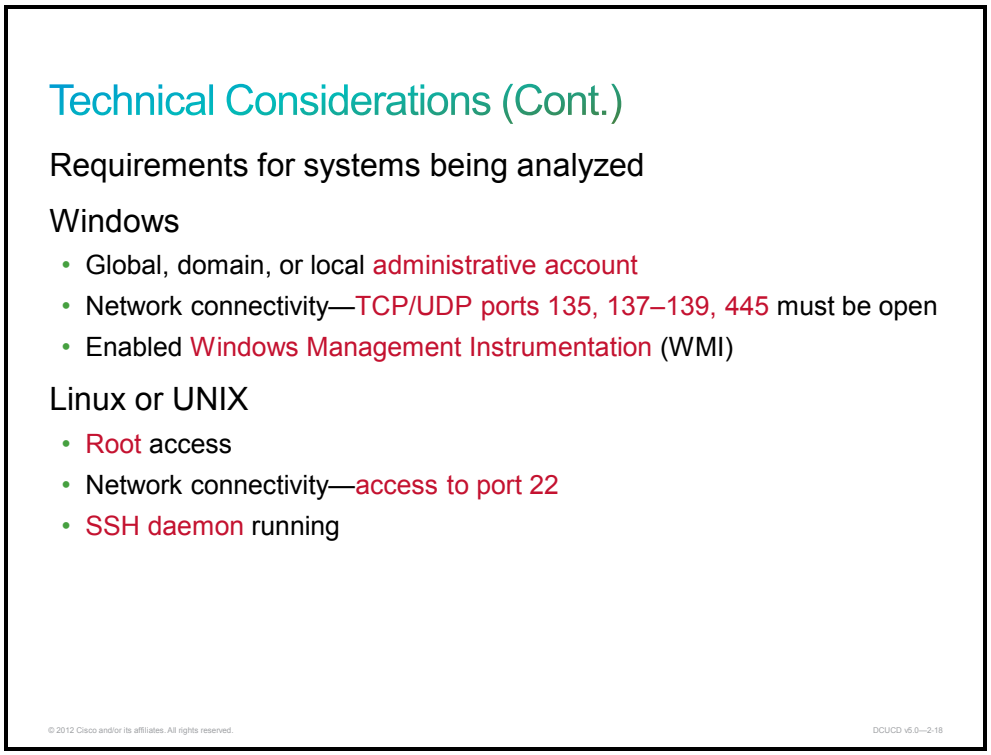
The Collector can receive information from the following systems: Windows 2000, XP, 2003, 2008, Vista, and 7; Red Hat Linux 8 and 9 and Enterprise Linux 3, 4, 5, and 6; SUSE Linux 8, 9, 10, SUSE Linux Enterprise 11 and Server 9; Hewlett-Packard UNIX (HP-UX) 10.xx, 11.0, 11.11, 11.22 (PA-RISC), and 11.23 (Itanium); and Sun Solaris 7, 8, 9, and 10 (Scalable Processor Architecture [SPARC]) and 9 and 10 (x86) operating environments, and AIX 5.1, 5.2, and 5.3.

https://t.me/learningnets

Information Transfer

When gathered, the data is uploaded to the Information Warehouse in comma-separated values (CSV) files in a secure manner by using HTTPS and Secure Sockets Layer (SSL) encryption. The Internet connectivity that is required should permit HTTPS.

When the analysis is conducted over a longer interval, the data is synchronized in intervals (by default once every hour).



The systems that are being analyzed must also meet certain requirements for the analysis and collection to be successful.

Windows Systems

On Windows target systems, the Collector uses WMI, the registry, and Perfmon to collect inventory and performance data. To collect this information, it must connect to the target systems using an account that has at least local administrative rights on the target systems. In many environments, the Collector uses a domain administrator account with rights on all or most of the target systems. This approach is the most convenient if the site security policies permit it.

The Collector must also be able to connect to all the Windows systems it is to analyze by using the TCP/UDP and ports 135, 137 to 139, and 445.

Linux and UNIX Systems

On UNIX and Linux systems, the Collector runs standard system utilities through an SSH connection, so every UNIX and Linux target system must have the SSH server daemon running and configured properly for a successful connection. Root permissions are required for each UNIX or Linux system. Not having root permissions can result in incomplete data collection while executing the scripts remotely because only user accounts with root privileges can run some of the utilities that the Collector uses.

On UNIX and Linux target systems, the Collector requires access to port 22 for an SSH connection.

https://t.me/learningnets

Analysis Security Considerations

Data collected

- Inventory—CPU, RAM, hard drive, network interfaces, chassis, software, and services
- Core performance metrics—memory, disk, network, processor, and application (more than 300 metrics)

Capacity Planner makes data anonymous

- CVS files are securely transmitted to the Information Warehouse
- Data sent includes domain and server names (can be masked)
  - Inventory—information on manufacturer, model, version, and status
  - Performance metrics—counter names and statistics related to those counters
- Data sent does not contain usernames, passwords, IP addresses, or shared information

Data is retained and available until archived after one year.

© 2012 Cisco and/or its affiliates. All rights reserved.

The Capacity Planner Data Collector systematically discovers domains and potential target systems within those domains, then inventories the target systems to provide data that is needed to assess capacity and utilization in the existing environment. It collects the inventory (hardware, software, and services) and over 300 core performance metrics (memory, disk, network, processor, and application).

After collecting the inventory and performance data, Capacity Planner makes the data anonymous, and then transmits the data over a secure connection to the Information Warehouse, where the Data Analyzer aggregates it. The Data Collector sends the data to the VMware data center in CSV files via an HTTPS connection that is using SSL encryption.

The inventory consists of data about CPU, RAM, hard drive, network interfaces, chassis, software, and services. Capacity Planner sends information about manufacturer, model, version, and status. The performance information includes counter names and statistics that are related to those counters, and the CSV files also contain domain names and server names. Optionally, the server and domain names can be masked before the data is transmitted.

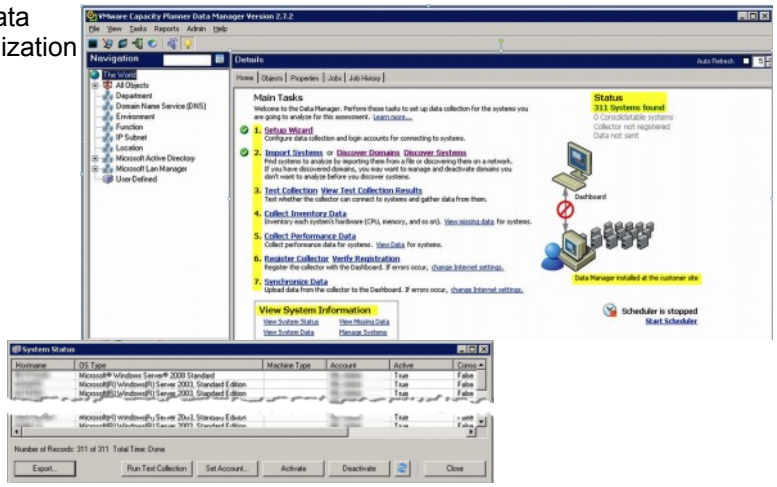
The CSV files that are sent from the Collector to the data center do not contain usernames, passwords, IP addresses, or shared information. The CSV files do, however, contain domain names and server names.

The data is retained and available until it is archived after one year.

https://t.me/learningnets

Conducting Analysis: Data Manager

- Run system discovery
- Run data collection
- Set up data synchronization



© 2012 Cisco and/or its affiliates. All rights reserved.

To use the Capacity Planner, the Capacity Planner Data Collector and its GUI—the Capacity Planner Data Manager—are installed on a computer at the customer site.

Agents do not need to be installed on any of the target systems. Capacity Planner can analyze target systems by running Windows, Linux, or UNIX operating systems.

First, the following discovery process has to be run:

- LANMan browser requests
- Lightweight Directory Access Protocol (LDAP) requests for Active Directory
- DNS queries for legacy
- IP scanning

The discovery task identifies the following:

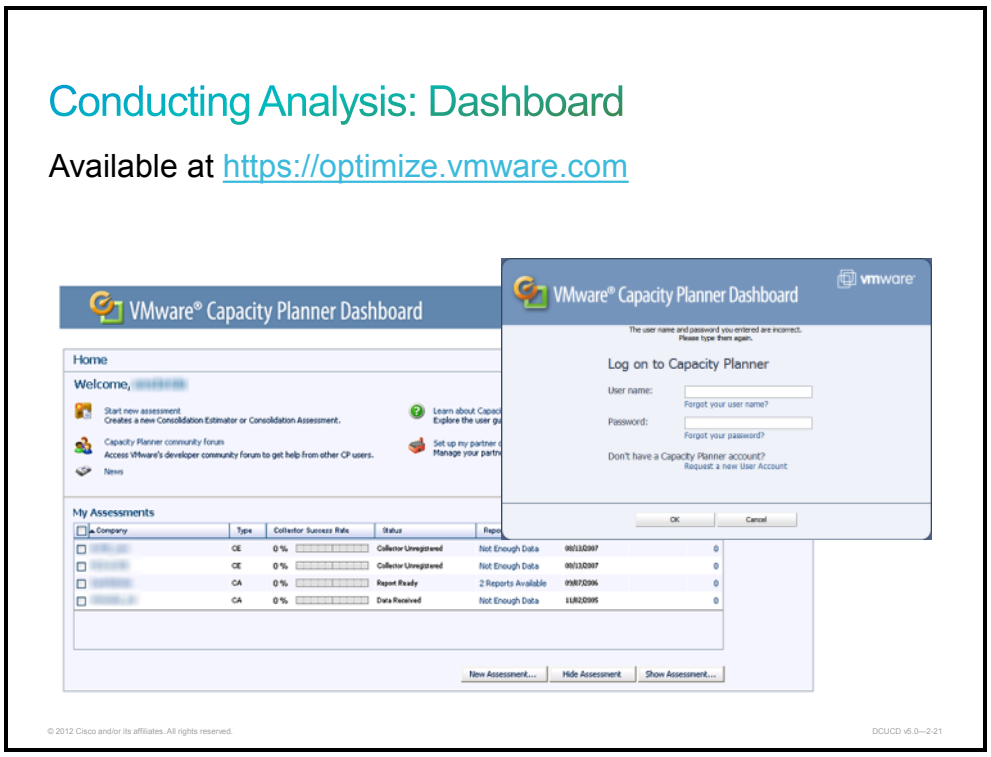
- Domains
- Systems
- Workgroups
- Active Directory nodes

The fact that the Data Collector discovers a system or node in the network does not mean that inventory or performance data must be collected from that system or node. Likewise, a node that is inventoried might not have performance data collected from it. The number of discovered nodes is often greater than the number of nodes that are inventoried or the number of nodes on which performance data is collected.

Second, performance information is collected by using one of two methods—one for Windows target systems and the other for Linux and UNIX target systems. Capacity Planner stores the performance information on the system that hosts the Data Collector.

https://t.me/learningnets

Third, the data synchronization to the Information Warehouse has to be configured. The first set of data is transmitted to the Information Warehouse in a manual process after the first round of data collection. Then, the Capacity Planner synchronizes data automatically every hour. A custom time interval or even a manual setting can be used for subsequent synchronizations.

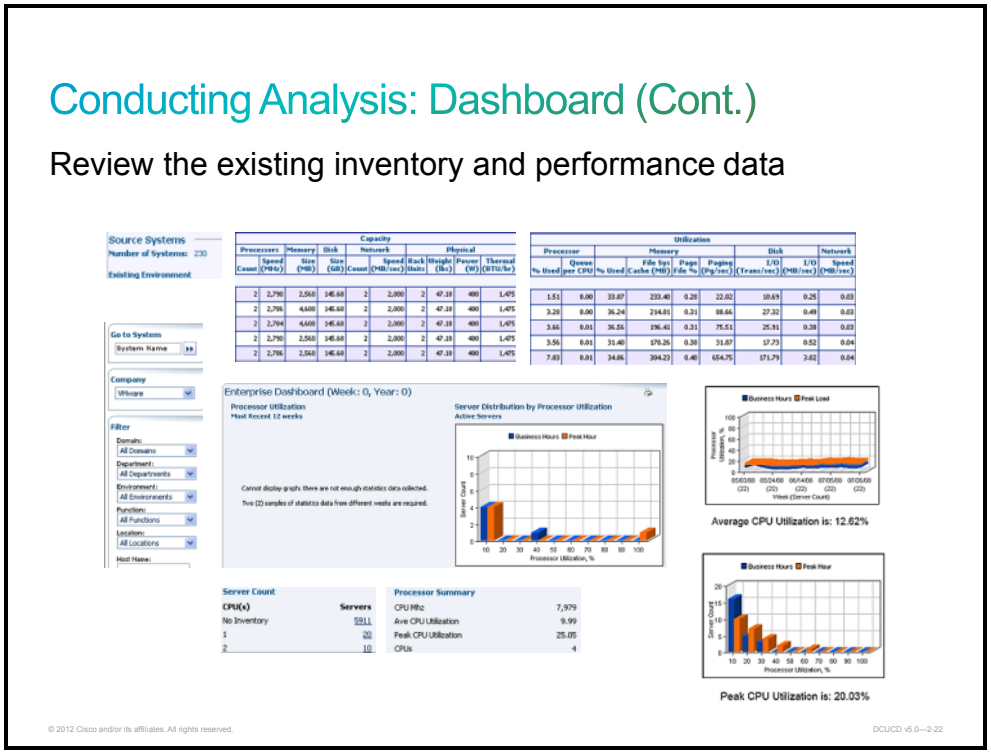


Once the assessment phases are complete, the architect can log in to https://optimize.vmware.com with a username and password that are obtained after the two-day training.

The Dashboard is a web-based hosted application that delivers capacity analysis and planning capabilities through a browser interface.

The architect can remotely access a set of prebuilt analyses and “slice and dice” data and can create custom reports. Planning capabilities allow setting objectives and constraints and also modeling and testing scenarios so that the architect can make intelligent capacity decisions. The monitoring capabilities of the Capacity Planner Dashboard enable proactive anomaly detection and alerts.

https://t.me/learningnets



The architect can examine the collected inventory information and the historical performance characteristics. This analysis includes the information about the total resource capacities and the utilization levels that were collected.

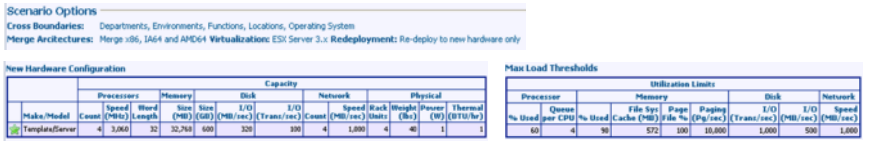
Reference Benchmarking

Analysis that is provided by the VMware Capacity Planner Dashboard is based on comparisons to reference data that is collected across the industry. This unique capability helps in guiding decisions around server consolidation and capacity optimization for your data center.

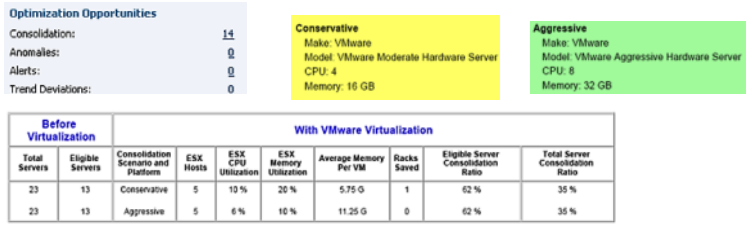
https://t.me/learningnets

Conducting Analysis: Scenarios

- Create and run a scenario



- Review the recommended server consolidations



© 2012 Cisco and/or its affiliates. All rights reserved.

Once the collected information is reviewed, the architect can see the recommended server consolidations. These recommendations can be conservative, where a single server has fewer CPU cores and memory, or they can be more aggressive, where more CPUs and more memory per server are used. For each option, the analysis shows the server consolidation ratio, the number of physical servers that are required, and the predicted average utilization levels of such hosts for the memory and CPU. The analysis also compares the previrtualization and postvirtualization server quantities.

The architect can also create and run different scenarios. For this option, these parameters for the scenario have to be entered:

- Processor architecture, virtualization type, deployment type (new versus reused hardware)
- Hardware configuration for new servers if they are used

https://t.me/learningnets

Employ VMware CapacityIQ

This topic describes how to perform VMware vSphere analysis with VMware CapacityIQ.

VMware vCenter CapacityIQ Overview

Analyze, forecast, and plan virtualized data center or desktop capacity

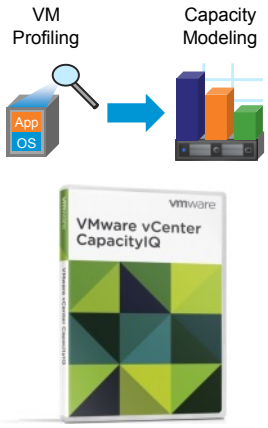
Postvirtualization tool

Explore existing capacity

- How many VMs can be deployed?
- What is the historical resource utilization?
- Can the existing capacity be used more efficiently?

Predict future needs

- When will the capacity limit be hit?
- What happens if capacity is added, removed, or reconfigured?



© 2012 Cisco and/or its affiliates. All rights reserved.

VMware vCenter provides a set of management vServices that greatly simplifies application and infrastructure management.

With CapacityIQ, continuous monitoring can be achieved. CapacityIQ continuously analyzes and plans capacity to ensure the optimal sizing of VMs, clusters, and entire data centers.

Here are the key features of VMware vCenter CapacityIQ:

- Performs “what-if” impact analyses to model the effects of capacity changes
- Identifies and reclaims unused capacity
- Forecasts the timing of capacity shortfalls and needs

Here are some benefits of VMware vCenter CapacityIQ:

- Delivers the right capacity at the right time
- Makes informed planning, purchasing, and provisioning decisions
- Enables capacity to be utilized most efficiently and cost-effectively

VMware vCenter CapacityIQ is a postvirtualization product that is used for the ongoing management of virtualized environments.

https://t.me/learningnets

CapacityIQ Use Case Example

How can VMware vCenter CapacityIQ be used to resolve the following questions?

To understand the current capacity usage

- How much capacity is being used right now?

To forecast future capacity needs

- How many more VMs can be added?

To predict the impact of capacity changes

- What happens to capacity if more VMs are added?

To maximize the utilization of existing capacity

- How much capacity can be reclaimed?

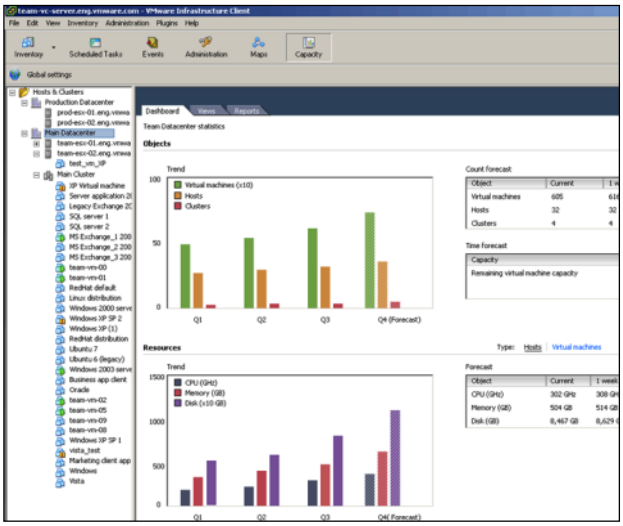
© 2012 Cisco and/or its affiliates. All rights reserved.

CapacityIQ can assist in answering the following questions concerning day-to-day operations:

- How much capacity is being used right now?
- How many more VMs can be added? (In other words, when will capacity run out?)
- What happens to capacity if more VMs are added?
- How much capacity can be reclaimed?

https://t.me/learningnets

CapacityIQ: Examining Current Resources

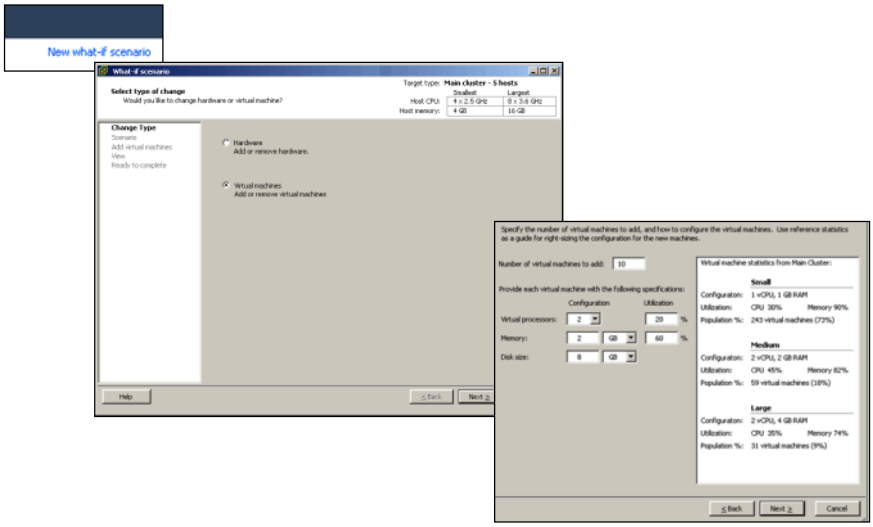


© 2012 Cisco and/or its affiliates. All rights reserved.

First, the Capacity Dashboard for the selected data center should be reviewed. This screen shows the data center-level view of the current and future state of capacity in terms of VMs/Hosts/Clusters and CPU/Memory/Disk.

https://t.me/learningnets

CapacityIQ: “What-If” Analysis



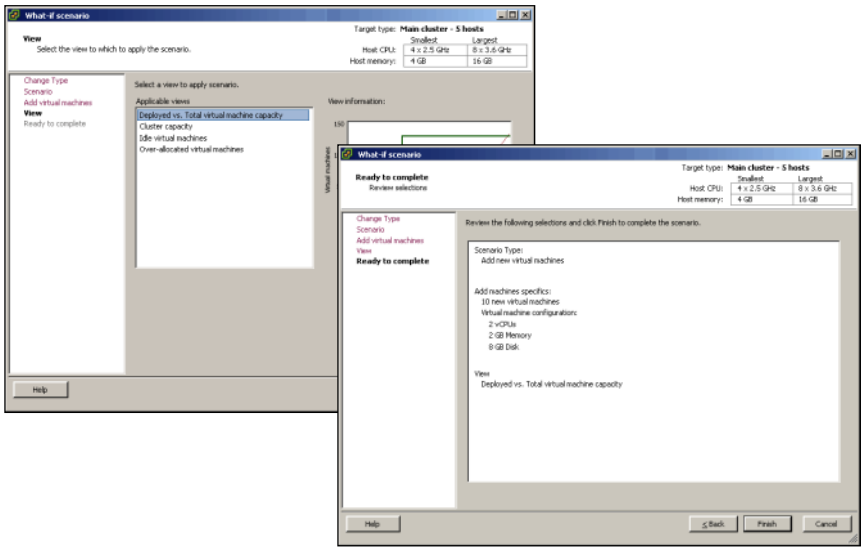
© 2012 Cisco and/or its affiliates. All rights reserved.

Second, a “what-if” analysis can be started. For example, what happens if 10 new VMs are added?

- Start a new “what-if” scenario and enter the parameters per the wizard.
- Choose the type of capacity change: hardware or virtual machine.
- Choose the type of VM capacity change in this scenario.
- Define the proposed VM configurations for this scenario.
- Review a summary of existing VM configurations as a reference for sizing VMs.

https://t.me/learningnets

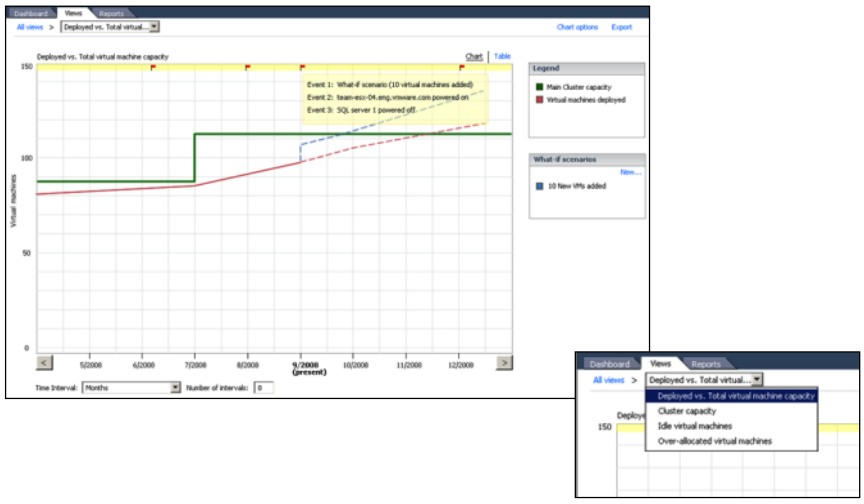
CapacityIQ: "What-If" Analysis (Cont.)



© 2012 Cisco and/or its affiliates. All rights reserved.

Choose how you would like to view the results, review your selections, and click **Finish** to complete the scenario.

CapacityIQ: "What-If" Analysis (Cont.)



© 2012 Cisco and/or its affiliates. All rights reserved.

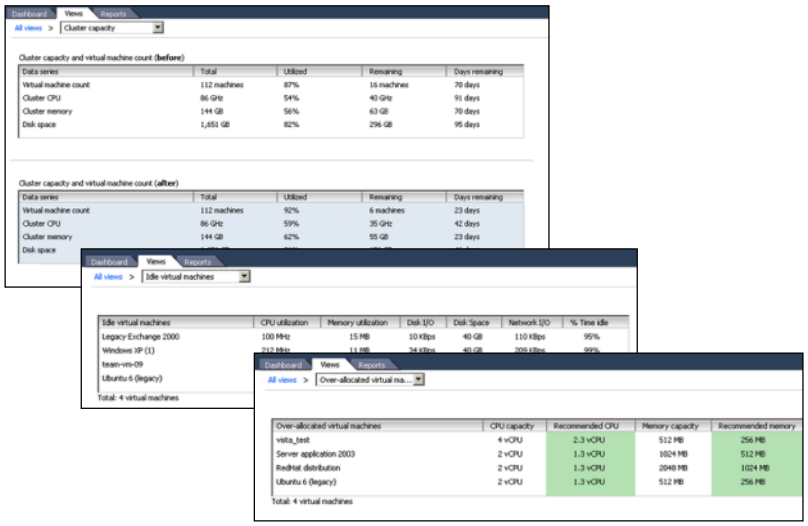
The "what-if" scenario result for the deployed versus total VM capacity (in the figure) shows that, at the current provisioning rate, capacity will run out in 60 days (this rate is represented by the red line).

If 10 new VMs are deployed today, capacity would run out in 23 days.

You can choose alternate views to examine additional information.

https://t.me/learningnets

CapacityIQ: “What-If” Analysis (Cont.)



© 2012 Cisco and/or its affiliates. All rights reserved.

All Views: Cluster Capacity

This option shows the cluster capacity before and after the “what-if” scenario of adding 10 new VMs.

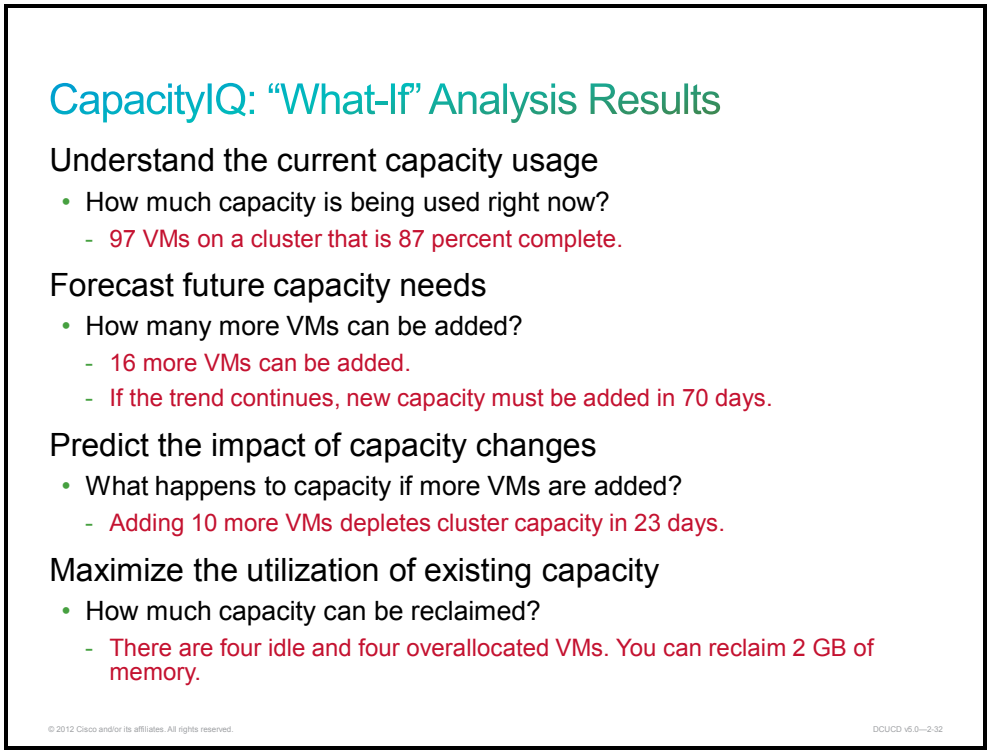
Idle Virtual Machines

This option shows that idle VMs are VMs that consistently have low resource utilization over a long period. The CapacityIQ has identified four idle VMs (out of 97 total VMs in this cluster). These idle VMs mean that there will be low resource utilization for more than 95 percent of the profiled lifetime of the system.

Overallocated Virtual Machines

Overallocated VMs are VMs that have been allocated more capacity than they actually demand. These are candidates for reconfiguration to actual capacity needs. This reconfiguration enables the reclaiming of resource capacity.

https://t.me/learningnets



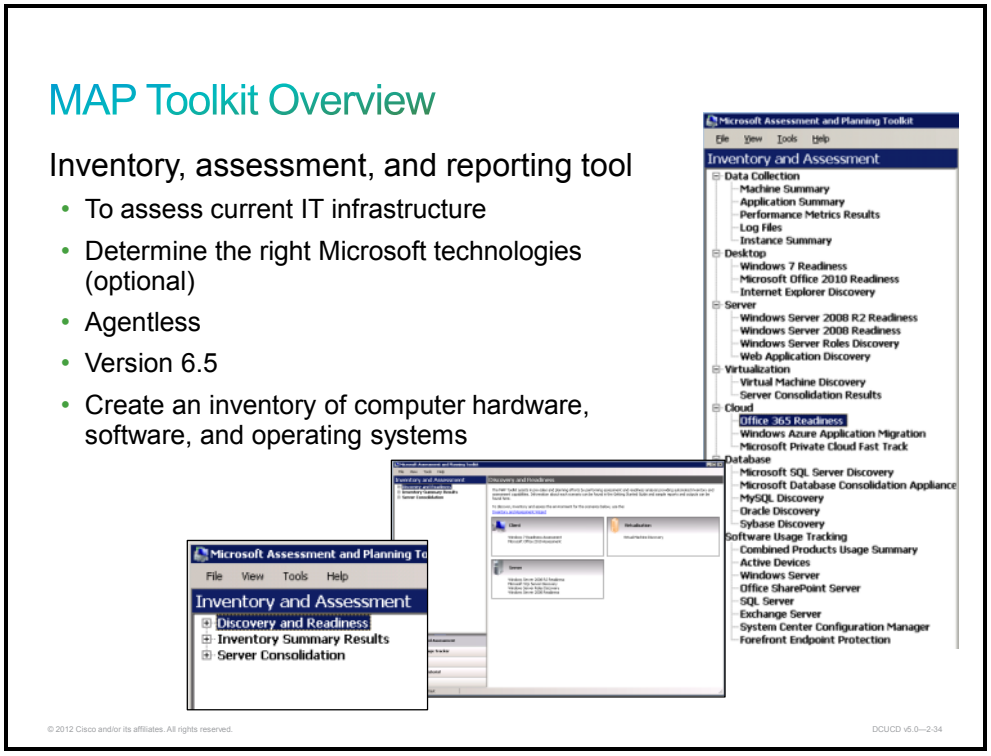
With the Dashboard and "what-if" analyses, you have answered the following questions:

- **How much capacity is being used right now?** This cluster currently has 97 VMs and is 87 percent complete.
- **How many more VMs can be added?** 16 more VMs can be added. More cluster capacity must be added within 70 days.
- **What happens to capacity if more VMs are added?** If 10 more VMs are added, cluster capacity will run out in 23 days.
- **How much capacity can be reclaimed?** There are four idle VMs and four overallocated VMs, so 2 GB of memory can be reclaimed.

https://t.me/learningnets

Employ MAP Toolkit

This topic describes how to perform existing computing solution analysis with the MAP Toolkit.



The MAP Toolkit is a powerful inventory, assessment, and reporting tool that makes it easier for architects to assess the current infrastructure for the customer and to determine the right Microsoft technologies to be used. This toolkit can inventory computer hardware, software, and operating systems in small or large IT environments without installing any agent software on the target computers.

The MAP Toolkit provides three key functions:

- **Secure and agentless inventory:** It gathers a network-wide inventory that scales from small businesses to large enterprises by collecting and organizing system resources and device information from a single networked computer. MAP uses WMI, the remote registry service, Active Directory domain services, and the computer browser service technologies to collect information without deploying agents.
- **Comprehensive data analysis:** MAP performs a detailed analysis of hardware and device compatibility for migration to Windows 7, Windows Server 2008 R2, Microsoft SQL Server 2008 R2, Office 365, and Microsoft Office 2010. The hardware assessment looks at the installed hardware and determines if migration is recommended or not.

  It also identifies heterogeneous IT environments consisting of Windows Server and Linux operating systems—including those running in a virtual environment—and can discover Linux-powered Linux, Apache, MySQL, PHP (LAMP) application stacks. The MAP VMware discovery feature identifies already-virtualized servers running under VMware that can be managed with the Microsoft System Center Virtual Machine Manager platform or that can be migrated to the Microsoft Hyper-V hypervisor.

https://t.me/learningnets

For server consolidation and virtualization through technologies such as Hyper-V, MAP helps to gather performance metrics and generate server consolidation recommendations that identify the candidates for server virtualization and suggest how the physical servers might be placed in a virtualized environment.

- **In-depth readiness reporting:** MAP generates reports that contain both summary and detailed assessment results for each migration scenario. The results are provided in Microsoft Excel workbooks and Microsoft Word documents.

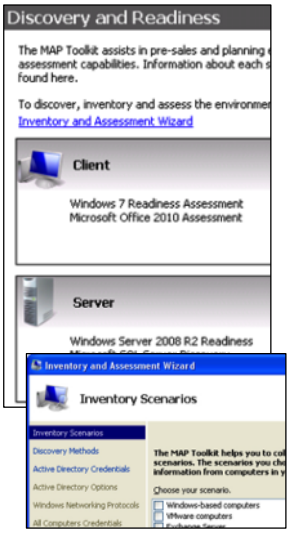
MAP Inventory Methods

Methods to gather inventory information

- WMI
- Remote Registry service
- VMware Web Service
- SSH

Requirements

- Accounts from local administrators group on inventoried computers



© 2012 Cisco and/or its affiliates. All rights reserved.

MAP can inventory the following technologies:

- Windows desktop operating systems, which include Windows 7, Vista, and XP Professional
- Office 2010 and previous versions
- Windows Server versions, including 2008 or 2008 R2, 2003 or 2003 R2, 2000 Professional, or 2000 Server
- VMware vSphere, vCenter, ESX/ESXi and VMware Server
- Linux distributions
- SQL Server 2008
- MySQL
- Oracle
- Sybase
- Hyper-V

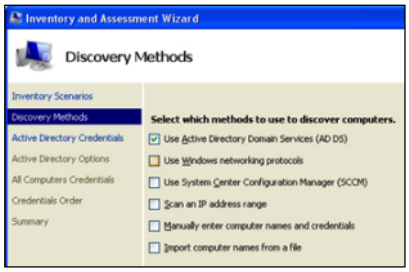
MAP uses WMI, Active Directory Services, Microsoft Systems Management Server (SMS) Provider, and other technologies to collect data in the environment without using agents. After the data is gathered, it can be parsed and evaluated for specific hardware and software needs and requirements. Ensure that you have administrative permissions for all computers and VMs that you want to assess.

https://t.me/learningnets

MAP Discovery Methods

Discover computers in environments with the following methods:

- Active Directory DS
- Windows networking protocols
- System Center Configuration Manager
- Import computer names from a file
- Scan an IP address range
- Manually enter computer names



© 2012 Cisco and/or its affiliates. All rights reserved.

Before the data collection can occur, the environment has to be surveyed and the computers in the environment have to be discovered.

Because the MAP Toolkit primarily uses WMI to collect hardware, device, and software information from the remote computers, you will need to configure your machines to inventory through WMI and also allow your firewall to allow remote access through WMI.

The MAP Toolkit can discover computers in your environment, or you can specify which computers to inventory using one of the following methods:

- **Active Directory DS:** Use this method if all computers and devices that you plan to inventory are in Active Directory DS.
- **Windows networking protocols:** Use this method if the computers in the network are not joined to an Active Directory DS domain.
- **Microsoft System Center Configuration Manager:** Use this method if you have System Center Configuration Manager in your environment and you need to discover computers that are managed by the System Center Configuration Manager servers.
- **Import computer names from a file:** Use this method if you have a list of up to 120,000 computer names that you want to inventory.
- **Scan an IP address range:** Use this method to target a specific set of computers in a branch office or specific subnets when you only want to inventory those computers. You can also use it to find devices and computers that cannot be found by browsing or through Active Directory DS.
- **Manually enter computer names:** Use this method if you want to inventory a few specific computers.

https://t.me/learningnets

MAP Advanced Functionality

Advanced analysis for the most widely used Microsoft technologies:

- Migration to Windows 7, Windows Server 2008 R2, Microsoft Office 2010 and readiness for Microsoft Cloud solutions (e.g. Office 356, Azure)
- Server virtualization with Hyper-V
- SQL Server consolidation appliance and migration to SQL Server 2008 R2

Software usage tracking for the following:

- Windows Server
- SharePoint Server
- System Center Configuration Manager
- Exchange Server
- SQL Server
- Forefront Endpoint Protection

Heterogeneous IT environment inventory



© 2012 Cisco and/or its affiliates. All rights reserved.

MAP Toolkit includes the following analysis for simplified IT infrastructure planning:

- Identification of currently installed Windows client operating systems, hardware, and recommendations for migration to Windows 7.
- Inventory and reporting of deployed web browsers, Microsoft ActiveX controls, and add-ons for migration to Windows 7-compatible versions of Internet Explorer.
- Identification of currently installed Microsoft Office software and recommendations for migration to Microsoft Office 2010 and Office 365.
- Identification of currently installed Windows Server operating systems, underlying hardware and devices, as well as recommendations for migration to Windows Server 2008 R2.
- Identification of currently installed Linux operating systems and underlying hardware for virtualization on Hyper-V or management by Microsoft System Center Operations Manager R2.
- Identification of virtual machines running on both Hyper-V and VMware, hosts, and details about hosts and guests.
- Assessment of Windows 2000 Server environments and inventory.
- Identification and analysis of web applications, IIS servers, and SQL Server databases for migration to the Windows Azure Platform.
- Detailed assessment and reporting of server utilization, as well as recommendations for server consolidation and virtual machine placement using Hyper-V.
- Discovery of SQL Server databases, instances, and selected characteristics.
- Heterogeneous database discovery of MySQL, Oracle, and Sybase instances.
- Identification of SQL Server host machines and SQL Server components.

https://t.me/learningnets

Installation Requirements

Hardware requirements (minimum)

- 1.6-GHz dual-core processor
- 1.5 GB of RAM
- 1 GB of available disk space
- Network adapter card
- Graphics adapter with 1024x768 or higher resolution

Software requirements

- Microsoft SQL Server 2008 Express Edition (download and installation during setup of MAP Toolkit)
- Microsoft Office Excel and Word 2007 SP2 or 2010
- .NET Framework 3.5SP1 (3.5.30729.01)
- Windows Installer 4.5

© 2012 Cisco and/or its affiliates. All rights reserved.

You can install the MAP Toolkit on a single computer that has access to the network on which you want to conduct an inventory and assessment. The MAP Toolkit Setup Wizard guides you through the installation.

The MAP Toolkit requires a nondefault installation of SQL Server 2008 R2 Express. If the computer is already running another instance of SQL Server 2008 R2 Express, the wizard must still install a new instance. This version is customized for the MAP Toolkit wizards and should not be modified. By default, access to this program is blocked from remote computers. Access to the program on the local computer is only enabled for users who have local administrator credentials.

The MAP Toolkit has the following minimum hardware installation requirements:

- 1.6 GHz processor (dual-core 1.5 GHz or faster recommended for Windows Vista, Windows 7, Windows Server 2008, or Windows Server 2008 R2)
- 1.5 GB of RAM (2 GB of RAM recommended for Windows Vista, Windows 7, or Windows Server 2008 R2)
- 1 GB of available disk space
- Network adapter card
- Graphics adapter that supports 1024x768 or higher resolution

The following are the software installation prerequisites for the MAP Toolkit:

- Any of the following operating systems:
  - Windows 7
  - Windows Vista Ultimate, Enterprise, or Business Editions
  - Windows XP Professional Edition
  - Windows XP Service Pack 3
  - Windows Server 2003 Service Pack 2

https://t.me/learningnets

- Windows Server 2008 R2
- Windows Server 2008

Note	MAP will run on either x86 or x64 versions of the operating system. Itanium processors are not supported. MAP will not install on Home or Home Server editions of Windows.

- .NET Framework 3.5 SP1 (3.5.30729.01)
- Windows Installer 4.5
- Microsoft Office Word 2007 SP2 or Word 2010
- Microsoft Office Excel 2007 SP2 or Excel 2010

Collecting Performance Data Overview

Step 1. Collect the inventory data

- Run the Inventory and Assessment Wizard

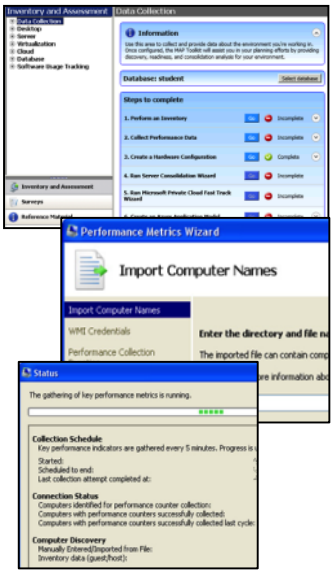
Step 2. Collect the performance data

- The collection interval is 5 minutes (not changeable)
- Specify the collection end time
- Run the Performance Metrics Wizard

Step 3. Review the collected data

- Create a report
- Review the data in an Excel spreadsheet

Step 4. Run server consolidation scenario(s).



© 2012 Cisco and/or its affiliates. All rights reserved.

Upon starting MAP Toolkit, the database, where data collected will be placed, needs to be created. Remember that with the MAP Toolkit MS SQL Server 2008 Express Edition was installed, and thus the database is a regular MS SQL Server 2008 database. Should you need to further manipulate collected data, the database can be accessed with common SQL Server tools and SQL language.

The MAP Toolkit can be used for analysis and reconnaissance in four steps:

- Running the inventory discovery
- Running the performance data collection
- Reviewing the collected data
- Running the server consolidation wizard

Because the MAP Toolkit is a Microsoft tool, the server consolidation wizard performs calculations and predictions for the Microsoft Windows 2008 Hyper-V or Microsoft Windows 2008 Hyper-V R2 environments.

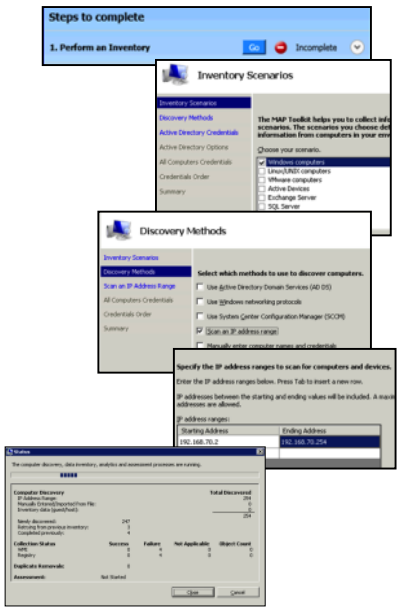
https://t.me/learningnets

Inventory Discovery

Specify how inventory should be collected

- Select Inventory Scenarios to instruct MAP what to collect (e.g. Windows, Unix)
- Select Discovery Method (e.g. AD, IP address scan, etc.)
- Enter data for discovery method (e.g. IP address range)
- Provide computer credentials

Run inventory discovery



© 2012 Cisco and/or its affiliates. All rights reserved.

As mentioned, the first step with MAP Toolkit is to build the inventory—more precisely, to run the inventory collection wizard.

This is achieved by telling the MAP what and how to collect:

- Select option(s) from the **Inventory Scenarios** to enable MAP to collect more detailed and relevant information for certain application types (e.g. SQL database, Exchange Server, and so on)
- Select **Discovery Method**, which means the MAP will seek for systems in the existing solution (such as Active Directory or IP address range scan)
- Provide the details for the selected discovery method (such as starting and ending IP address for the IP address range)
- Provide the credentials that MAP can use to hook into discovered systems in order to collect relevant inventory data.

Once the info is provided, the discovery can be started. Depending of the method selected and Data Center span will determine if the discovery takes more or less time to collect the information.

After the process is finished, the inventory data can be reviewed before proceeding to the performance metrics gathering.

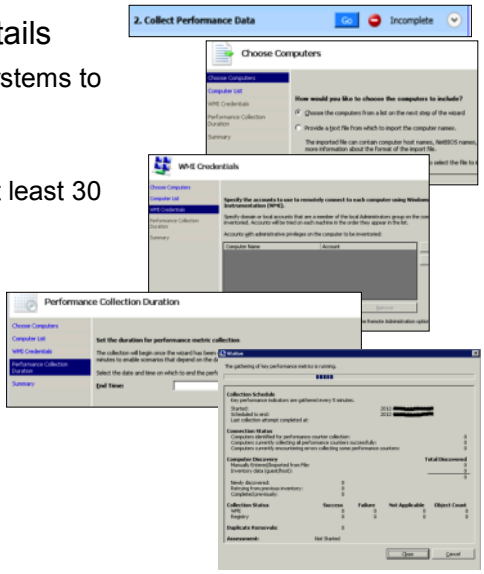
https://t.me/learningnets

Collecting Performance Data

Set performance collection details

- Set Computer List to tell which systems to monitor
- Provide credentials
- Set collection interval duration (at least 30 minutes)

Start collection process



© 2012 Cisco and/or its affiliates. All rights reserved.

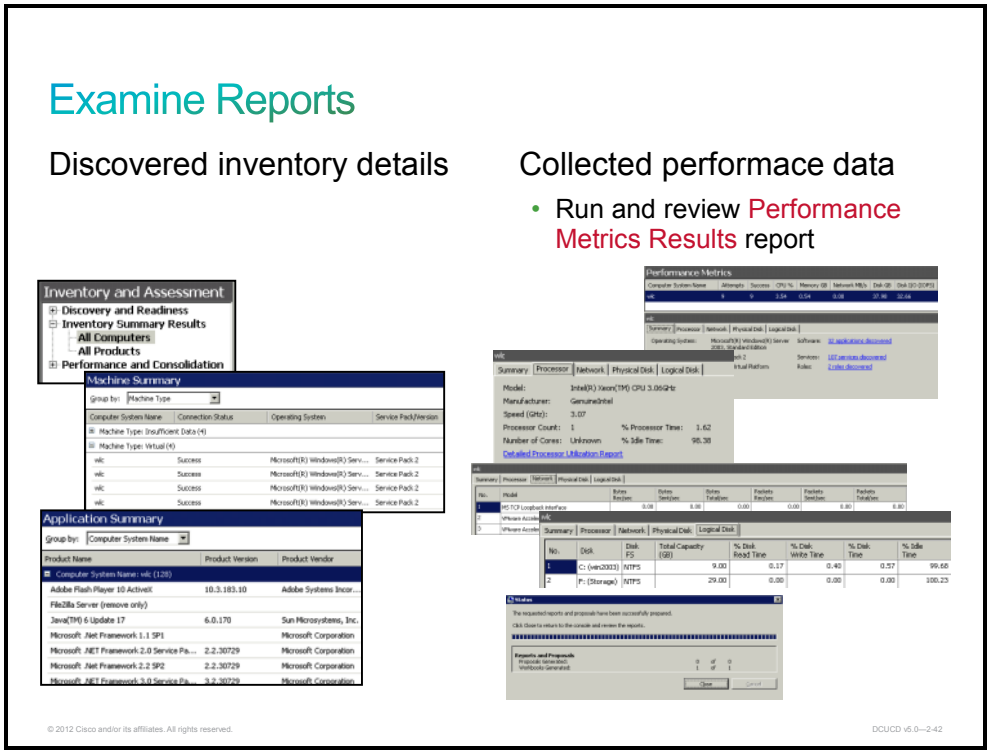
With the inventory data available, the performance data can be collected. This is done by running the performance collection data wizard.

For the performance data collection, the following needs to be provided:

- List of systems (computers) for which the collection is to be run—either all computers or only those that are relevant for the analysis can be selected from the list of computers
- Credentials for the MAP to hook into selected computers to collect performance data
- Collection interval duration—in order to make analysis relevant, the collection interval should be long enough

Remember that the wizard can be run multiple times in order to gather as much data as possible at different times.

https://t.me/learningnets

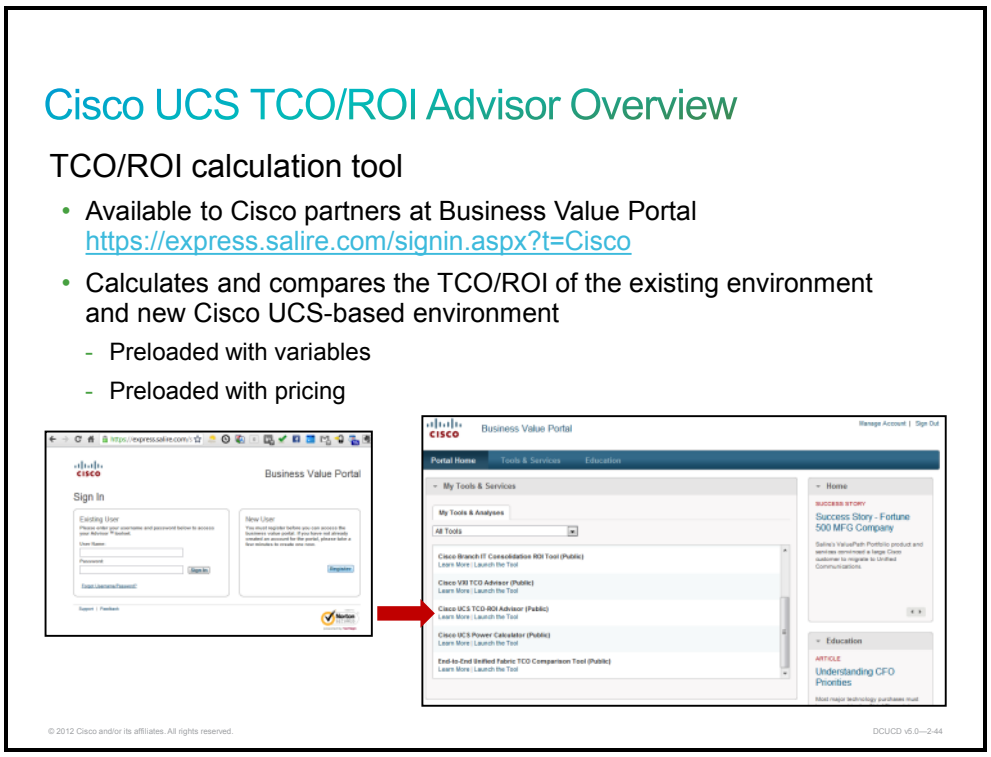


Once both processes are finished, the analysis report can be generated and exported in the Excel spreadsheet format or reviewed in the tool.

https://t.me/learningnets

Employ Cisco UCS TOC/ROI Advisor

This topic evaluates the Cisco Unified Computing System (UCS) TCO/ROI Advisor tool.



The Cisco UCS TCO/ROI Advisor tool compares the capital and operating costs of maintaining an existing computing infrastructure (server and network environment) with the cost of implementing a Cisco UCS solution (Cisco UCS and accompanying devices) over the same time period (typically five years).

The analysis measures the benefit of the Cisco UCS solution by calculating the difference in TCO between these two environments and the ROI in the proposed Cisco UCS solution. In other words, the analysis helps to discover how much data center capital and operating expenses can be reduced.

The tool delivers a comparison between existing rack-mount or blade servers and networks, and the new Cisco UCS environment with Cisco UCS B-Series blade servers or C-Series rack optimized servers and unified fabric networking.

The tool is available to Cisco partners at the Cisco Business Value Portal website. https://express.salire.com/signin.aspx?t=Cisco

Tool Purpose

The TCO/ROI tool is helpful in the following ways:

- Sophisticated calculation engine
- Preloaded with directional pricing
- Preloaded with directional data center environmental variables

https://t.me/learningnets

The tool is not designed for the following:

- A guide to UCS architecture and Data Center design. The solution design requires understanding and planning each and every component of the solution.
- An up-to-the-minute competitive pricing configurator. Accurate scenario modeling requires fresh pricing inputs by the user.
- It is not representative of the unique environment of the customer.

Terms of Use

The Cisco Unified Computing TCO/ROI Tool is based on information publicly available per the last update date as specified on the web page. Any tool output is intended to provide information to customers considering the purchase, license, lease, or other acquisition of Cisco products and services.

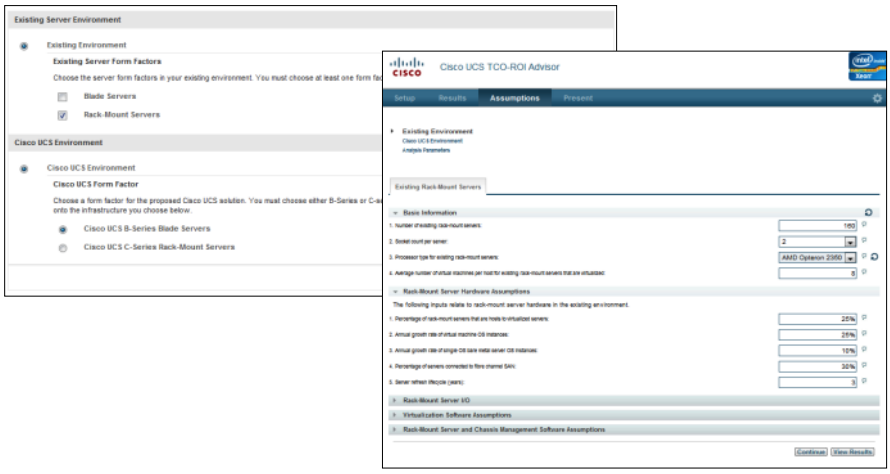
The tool is designed to provide general guidance only; it is not intended to provide business, purchasing, legal, accounting, tax or other professional advice. The Tool is not a substitute for the professional or business judgment of the customer.

© 2012 Cisco Systems, Inc
https://t.me/learningnets

Tool Wizard

Enter existing and new environment characteristics

- Preloaded variables can be changed under Assumptions menu



© 2012 Cisco and/or its affiliates. All rights reserved.

The tool estimates the financial effect of upgrading an existing server and associated network infrastructure (current state) to a Cisco UCS infrastructure (new state). The financial analysis model projects cost structures for the current and new state over the chosen period of the analysis (between three and 10 years).

Depending on the current-state environment, the tool allows a comparison between existing rack-optimized servers or blade servers and the associated network with a new state that is based on either of the following:

- Cisco UCS with B-Series blade servers and integrated unified fabric
- Cisco UCS with C-Series rack-mount servers and Cisco Nexus unified fabric

Current State

IT departments need to support new business applications and growth of existing workloads and data stores while maintaining service levels to end users and containing overall costs. Compute and storage requirements continue to grow at an unprecedented rate, while more than 70 percent of current IT budgets is spent simply to maintain and manage existing infrastructure.

The current state of infrastructure is amplifying these challenges. In most cases, data center environments are still assembled from individual compute, network, and storage network components. Administrators spend significant amounts of time manually accomplishing basic integration tasks rather than focusing on more strategic, proactive initiatives. One of the main initiatives that data center managers are undertaking to address this situation is consolidation of physical servers through server virtualization. While server virtualization can have immediate impact by enabling the improved utilization of server hardware, current infrastructures are generally not well suited to easy deployment, expansion, and migration of virtualized workloads to meet changing business needs.

https://t.me/learningnets

New State

The Cisco UCS solution is a next-generation data center platform that unites compute, network, storage access, and virtualization into a cohesive system designed to reduce TCO and increase business agility. The system integrates a low-latency, lossless 10 GB Ethernet unified network fabric with enterprise-class, x86-architecture servers. The system is an integrated, scalable, multichassis platform in which all resources participate in a unified management domain. IT departments have the choice of unified computing blade or rack form factors.

Primary Assumptions

These are the primary assumptions:

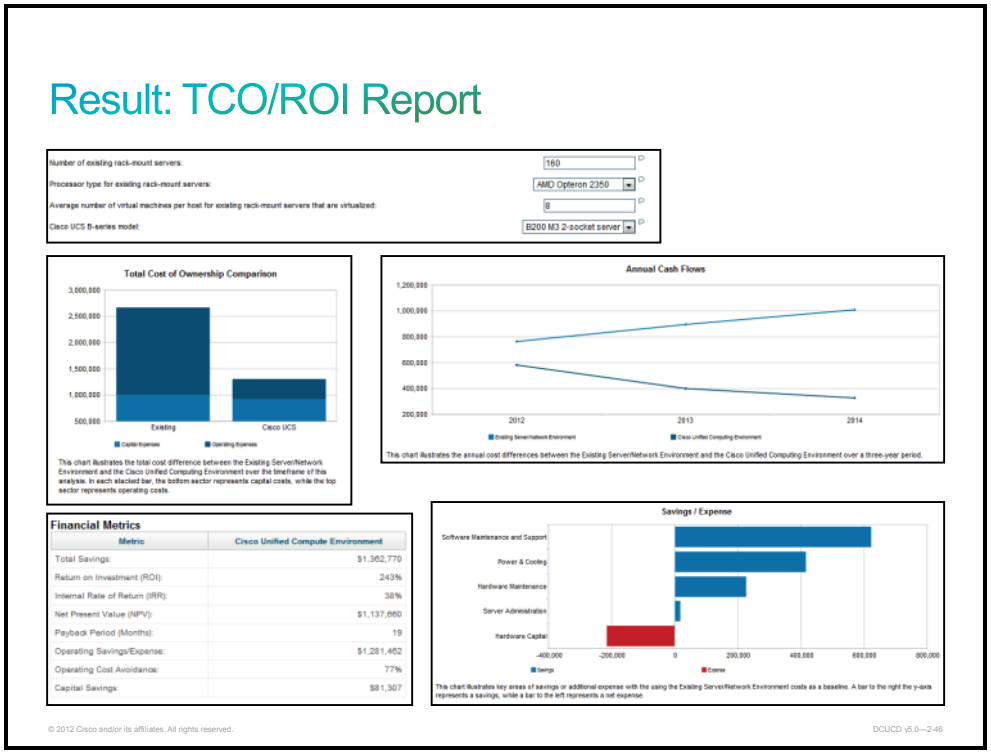
- The server consolidation ratio depends on the anticipated compute improvements that will be gained by moving from your current server CPU to the Cisco UCS solution based on the latest Intel Xeon processors. Your consolidation ratio may vary based on workload and may be higher in workloads that are memory-intensive.
- The model assumes that your current-state network infrastructure uses a 1 Gigabit Ethernet LAN and 1/2/4-Gb Fibre Channel SAN connections.
- The model assumes that you will move your server network to a Cisco 10-Gb unified fabric infrastructure for all Ethernet and Fibre Channel traffic. Review the financial results to determine the associated costs of your compute and I/O components to more clearly understand the cost impacts of each.
- The current-state network infrastructure defaults assume the use of Cisco Catalyst 6500 Series Switches and Cisco MDS 9500 Family end-of-row Fibre Channel switches.
- The default analysis assumes that your end-state compute environment will be a chassis-based Cisco UCS B-Series unified computing blade server environment. You can edit the assumptions if you want the model to use a Cisco UCS C-Series unified computing rack-server environment as the end state.
- The financial analysis is a cash flow-based model and does not take into account metrics such as depreciation, taxes, and amortization.

Analysis Inputs

When using the tool, the analysis is provided immediately upon choosing the type of existing environment—that is, either rack-optimized or blade server. To tailor the analysis to better reflect the customer environment, the following analysis inputs can be changed from the preloaded values:

- Existing server architecture type:
  - Rack-optimized server infrastructure
  - Blade server infrastructure
- Analysis parameters

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Once all of the parameters and characteristics are tailored to the specific customer, the analysis can be updated and the results reviewed. Note that the parameters can be changed at any time by editing them and updating the analysis.

The results of the analysis can be examined over the web or downloaded. Both sets of results consist of several sections.

TCO Comparison

The Total Cost of Ownership Comparison chart in the figure illustrates the difference in TCO between the current server and network infrastructure, and a Cisco UCS solution that leverages Intel Xeon processors. Hovering the mouse over the graph in the final report displays details about the costs that are related to each solution. The Financial Metrics table details financial metrics such as total savings, ROI, and others.

Savings / Expense

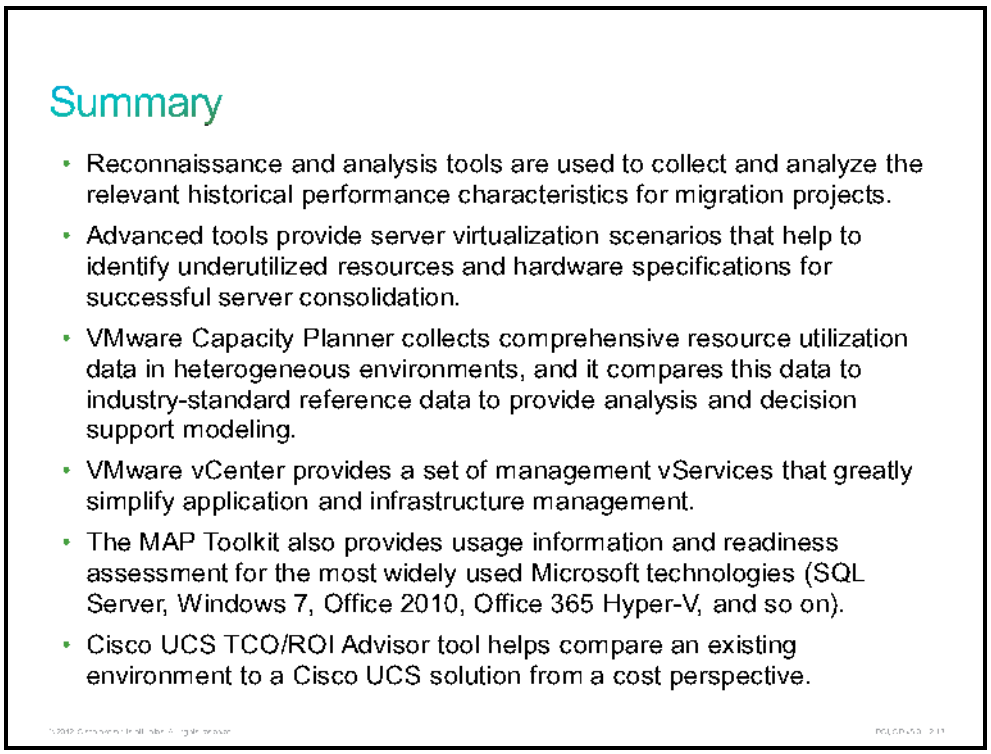
This section examines the cost structures of a unified computing environment using the existing infrastructure costs as a baseline.

The graph is illustrated with areas of savings that are positive or negative and organized per category.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.



References

For additional information, refer to these resources:

- https://express.salire.com
- http://www.microsoft.com/en-us/download/details.aspx?id=7826
- http://www.vmware.com/products/capacity-planner/overview.html
- https://optimize.vmware.com/index.cfm
- http://www.vmware.com/products/vcenter-capacityiq/overview.html
- https://www.netiq.com/products/recon/
- http://www.netapp.com/us/products/management-software/oncommand-balance/?euaccept=true
- http://www.netinst.com/products/observer/index.php

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Module Summary

This topic summarizes the primary points that were discussed in this module.

Module Summary

- The Cisco UCS solution design will meet the goals if thorough assessment is performed.
- The key server performance characteristics that influence application performance are CPU and memory utilization, network throughput, and storage space and IOPS.
- Reconnaissance and analysis tools build inventory of the existing environment and collect resource utilization information.
- VMware Capacity Planner and the Microsoft Assessment and Planning Toolkit are agentless premigration tools that are used to assess the existing server infrastructure to aid in a physical-to-virtual conversion.
- The Cisco UCS TCO/ROI Advisor tool can help in assessing the cost aspect of the solution.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Self-Check

Use the questions here to review what you have learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) Which deliverable serves as the customer-signed document with input and requirements? (Source: Defining a Cisco Unified Computing System Solution Design)

A) HLD
B) SRS
C) CRD
D) LLD

Q2) Which option affects actual server performance? (Source: Analyzing Computing Solutions Characteristics)

E) application chattiness
F) operating system management
G) CMOS
H) BIOS

Q3) Which tool can be used to evaluate requirements for converting physical Windows 2003 servers into virtual machines? (Source: Employing Data Center Analysis Tools)

A) VMware CapacityIQ
B) Microsoft MAP
C) Webmin
D) CLI

Q4) What needs to be carefully defined prior to running resource utilization information collection in order to gather relevant data? (Source: Employing Data Center Analysis Tools)

A) collection protocol
B) database
C) user accounts
D) collection duration

Q5) Which information needs to be collected prior to running utilization and workload assessment? (Source: Employing Data Center Analysis Tools)

A) inventory
B) operating system types
C) performance metrics
D) “what-if” analysis output

Q6) Which two options are required to conduct an analysis with the Capacity Planner? (Choose two.) (Source: Employing Data Center Analysis Tools)

A) server with built-in protocol analyzer
B) user account with administrative privilege
C) Internet connectivity
D) separate management network
E) SQL Server 2008 database

https://t.me/learningnets

Q7) What is required by the MAP Toolkit to examine Windows servers in the existing environment? (Source: Employing Data Center Analysis Tools)

A) connectivity to the hosted warehouse
B) Active Directory
C) Windows 2008 Server for MAP Toolkit
D) dedicated management NIC
E) WMI protocol that is enabled on systems that are surveyed

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Self-Check Answer Key

Q1) C

Q2) A

Q3) B

Q4) D

Q5) A

Q6) B, C

Q7) E

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module 3

Size Cisco Unified Computing Solutions

Overview

After the analysis of the requirements, the architect can select and size the Cisco Unified Computing System (UCS) hardware components. This process requires the architect to review the requirements and match them as closely as possible with the selected equipment. Once the equipment is selected, the architect needs to also size the physical deployment plan. Here, the architect needs to create a plan on how Cisco UCS equipment will be installed in the facility, rooms, and racks while having the necessary power, cooling, space, loading, and cabling available.

This module presents the UCS C-Series and B-Series hardware information, general steps in sizing Cisco UCS C-Series and Cisco B-Series solution, and three sizing examples.

Module Objectives

Upon completing this module, you will be able to size and prepare a physical deployment plan for Cisco UCS C-Series and B-Series solution per requirements. This ability includes being able to meet these objectives:

- Consider the standalone Cisco UCS C-Series server solution for a given set of requirements
- Consider the Cisco UCS B-Series server solution for a given set of requirements
- Create a physical deployment plan by using the Cisco UCS Power Calculator tool

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Lesson 1

Sizing the Cisco UCS C-Series Server Solution

Overview

When sizing the Cisco Unified Computing System (UCS) C-Series solution, the architect needs to review the gathered solution requirements and then size the individual servers—select proper components like CPU, memory DIMMs, and select adequate quantities—and pick the appropriate quantity for the server selected.

This lesson identifies how to size a Cisco UCS C-Series solution for a given set of requirements.

Objectives

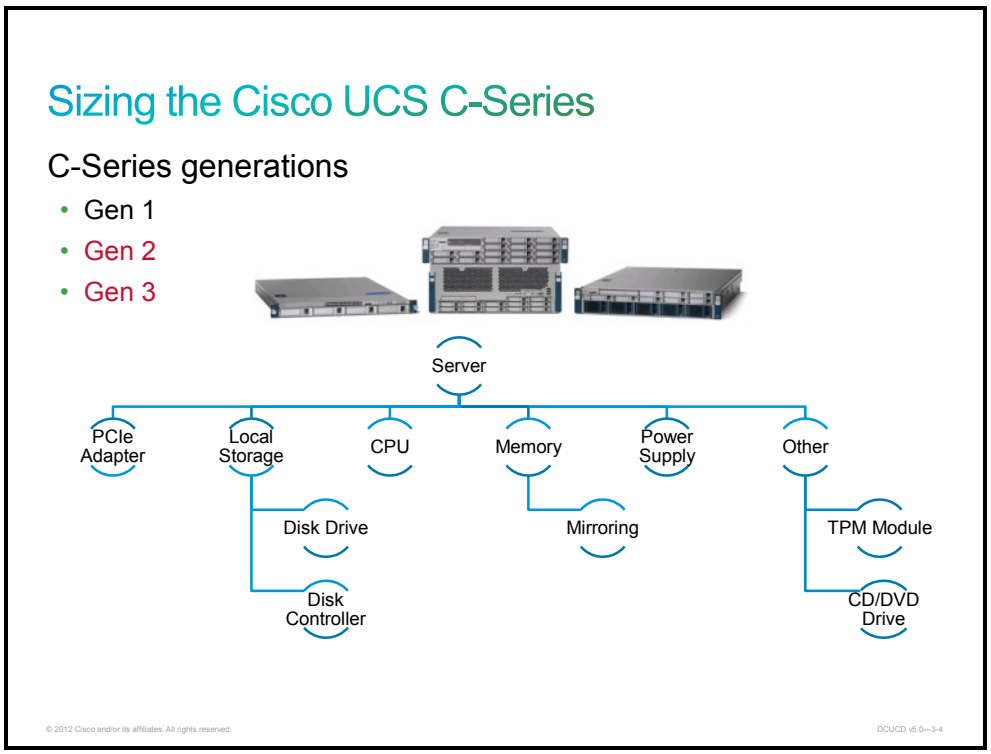
Upon completing this lesson, you will be able to consider the standalone Cisco UCS C-Series server solution for a given set of requirements. You will be able to meet these objectives:

- Recognize general steps for Cisco UCS C-Series server selection
- Identify the requirements of Cisco UCS C-Series integration with Cisco UCS Manager
- Select proper Cisco UCS C-Series server hardware based on the requirements for a given small VMware vSphere environment
- Select proper Cisco UCS C-Series server hardware based on the requirements for a given small Hyper-V vSphere environment

https://t.me/learningnets

Size the Cisco UCS C-Series Solution

This topic describes general steps for Cisco UCS C-Series server selection.



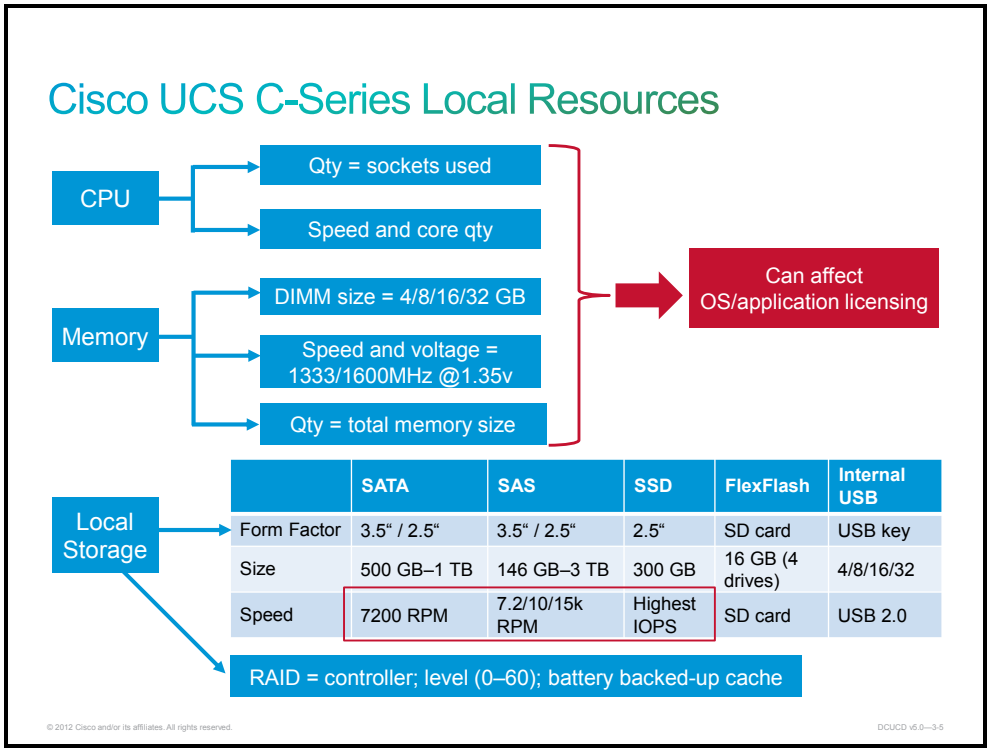
The sizing process for the Cisco UCS C-Series is used to determine the appropriate components that are needed to build the Cisco UCS per the requirements that were gathered in a design workshop.

The sizing process can be divided into two major categories:

- **Identify Cisco UCS C-Series server types:** With this step, the individual server type is identified and quantities are defined.
- **Size Cisco UCS C-Series servers:** With this step, the components of individual C-Series servers are selected.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



Each of the hardware components has its own specific parameters that need to be observed in order to closely match the selected hardware with the requirements. Improperly selected hardware can negatively affect application responsiveness as well as incur additional licensing costs.

The UCS C-Series server has hardware components that are local to the server, namely the CPU, memory, RAID controller and disk drive.

When selecting the CPU, the architect needs to observe the number of cores per CPU and select the proper quantity per server. The performance can also be affected by the cache size and bus speed, which governs the type of DIMMs that can be used from a speed and size perspective.

DIMMs came in various sizes, from 4 up to 32 GB, and supporting different speeds and voltage. Not all are supported by individual CPUs or servers. The architect needs to review the details about the server selected to use in order for the solution to properly populate the server with adequate number and accurately sized DIMMs.

The local storage configuration with C-Series is governed by the selected server, Redundant Array of Independent Disks (RAID) controller and disk drive type. It is vital to recognize that local storage performance is very dependent on the disk drive selected. For example, a 7200RPM SATA disk is much slower than the 15K RPM SAS disk.

https://t.me/learningnets

Cisco UCS C-Series M2 Server Comparison



	UCS C200 M2*	UCS C210 M2*	UCS C250 M2*
Processor	1/2 x Intel Xeon 5600	1/2 x Intel Xeon 5600	1/2 x Intel Xeon 5600
Form Factor	1 RU	2 RU	2 RU
Memory	12 DIMMs, max 192 GB	12 DIMMs, max 192 GB	48 DIMMs(Cisco ext.mem), max 384 GB
Disks	8x 2.5“ or 4x 3.5“ SAS/SATA, max 8 TB, hot-swap	16x 2.5“ SAS/SATA max 16 TB, hot-swap	8x 2.5“ SAS/SATA max 8TB, hot-swap
Built-in RAID	0, 1 (SATA only)	0,1 (5 SATA only)	None
Optional RAID	0,1,5,6,10 – 4 disks 0,1,5,6,10,50,60 – 8 disks	0,1,5,6,10,50,60 4,8,16 disks	0,1,5,6,10,50,60 up to 8 disks
LOM	2 GE 10/100 mgmt	2 GE 10/100 mgmt	4 GE 10 GE UF optional
I/O slots	2 half-length x8 PCIe 2.0: - 1 full-height x16 - 1 low-profile x8	5 full-height x8 PCIe 2.0: - 2 full-length x8 - 3 half-length, x16	5 half-length PCIe 2.0: - 3 low-profile x8 - 2 full-height x16

* 6 cores/socket; optional CD/DVD drive

© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares Cisco UCS C-series M2 maximum hardware configurations.

The M2 UCS C-Series servers are the second-generation rack-mount servers that can be equipped with up to two Intel Xeon 5500 or 5600 processors. They differ in the total amount of memory they can be equipped with, the number of local drives that can be installed, and the PCI Express (PCIe) slots.

https://t.me/learningnets

Cisco UCS C-Series M2 Server Comparison (Cont.)





	UCS C260 M2	UCS C460 M2*
Processor	1/2x Intel Xeon E7-2800 (up to 10 cores/socket)	2/4 x Intel Xeon E7-4800/8800 (up to 10 cores/socket)
Form Factor	2 RU	4 RU
Max mem.	64 DIMMs (Cisco ext.mem), max 1 TB	64 DIMMs, max 1TB
Disks	16x 2.5“ SAS/SATA2/SSD max 16TB, hot-swap	12x 2.5“ SAS/SATA2/SSD max 12TB, hot-swap
Built-in RAID	None	None
Optional RAID	0,1,5,6,10,50,60 Up to 16 SAS/SATA2 disks	0,1,5,6,10,50,60 Up to 12 SAS/SATA2 disks
LOM	2 GE 2 10GE SFP+ 2 10/100 mgmt	2 GE 2 10/100 mgmt
I/O slots	7 PCIe 2.0: - 2 full-height, half-length x16 - 4 low-profile, half-length x8 - 1 low-profile, half-length x4	10 full-height PCIe 2.0: - 2 Gen1 + 8 Gen2 - 4 half-length + 6 3/4-length

* optional CD/DVD drive

© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares Cisco UCS C-Series M2 maximum hardware configurations.

The M2 UCS C-Series servers are the second-generation rack-mount servers that can be equipped with up to two Intel Xeon E7-2800 or E7-4800/8800 processors and thus increase the total number of CPU cores per server as well as maximize the memory size to 1 TB per server. This makes the servers suitable for memory-intensive applications like in-memory databases or for large virtualization projects where a very high virtualization ratio needs to be achieved.

https://t.me/learningnets

Cisco UCS C-Series M3 Server Comparison





	UCS C220 M3*	UCS C240 M3*
Processor	1/2x Intel Xeon E5-2600 (up to 10 cores/socket)	1/2 x Intel Xeon E5-2600 (up to 10 cores/socket)
Form Factor	1 RU	2 RU
Max mem.	16 DIMMs, max 256 GB	24 DIMMs, max 384 GB
Disks	8x 2.5“ SAS/SATA/SSD max 8 TB, hot-swap	24x 2.5“ SAS/SATA2/SSD max 24 TB, hot-swap
Built-in RAID	None	None
Optional RAID	0,1,5,6,10,50,60	0,1,5,6,10,50,60
LOM	2 GE 1 10/100/1000 mgmt	4 GE 1 10/100/1000 mgmt
I/O slots	2 PCIe 3.0: - 1 half-height, half-length x8 - 1 full-height, 3/4-length x16	5 PCIe 3.0: - 2 full-height, 1/2 + 3/4 length x16 - 2 full-height, full + half length x8 - 1 half-height, half-length x8

* internal USB; 16GB Cisco FlexFlash (SD card)

© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares Cisco UCS C-Series M3 maximum hardware configurations.

The generation 3 M3 UCS C-Series servers are the latest addition to the rack-mount server family. They enable the architect to use the servers with Intel Xeon E5-2600 processors, making them price-optimized while still preserving the high core-per-server count—with processors that have 10 cores.

https://t.me/learningnets

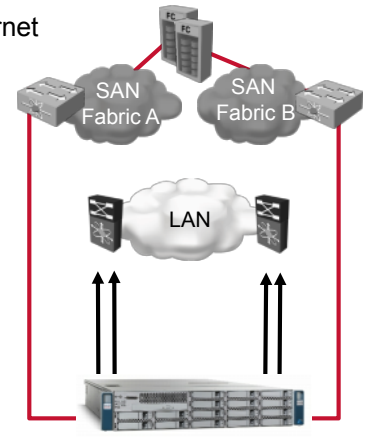
Sizing Cisco UCS C-Series Connectivity

Connectivity aspects

- OOB management for Cisco IMC
- LAN—1 Gigabit versus 10 Gigabit Ethernet
- SAN
  - Fibre Channel with HBA
  - iSCSI with TOE/iSCSI offload
  - FCoE with CNA or VIC

Design aspects

- Throughput and oversubscription
- Redundancy—multifabric design



© 2012 Cisco and/or its affiliates. All rights reserved.

The UCS C-Series sizing is primarily influenced by two factors—the required throughput and allowed oversubscription.

Both factors should be observed from different perspectives:

- LAN connectivity
- SAN connectivity

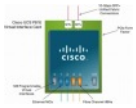
The Cisco UCS C-Series connectivity architecture follows a multifabric design. Such a design is used to achieve high availability (with failover on the Cisco UCS level using P81E VIC, or on the operating system level using other adapters), to achieve more throughput or to achieve a combination of redundancy and higher throughput.

When determining the number of logical adapters that need to be presented to the operating system or hypervisor (for example, VMware ESX/ESXi or Microsoft Hyper-V host), the Cisco UCS P81E VIC adapter can be selected for consolidated LAN and SAN connectivity when up to two host bus adapters (HBAs) and up to 16 network interface cards (NICs) are required. When connectivity consolidation is not required, the server can also be equipped with a multiport Ethernet adapter and a multiport Fibre Channel HBA.

The maximum number of physical adapters is governed by the number of PCIe slots on the server motherboard.

https://t.me/learningnets

Cisco UCS C-Series VIC and CNA Adapter Comparison





	P81E VIC	OneConnect	QLE8152	NetXtreme II 57712**
Vendor	Cisco	Emulex	QLogic	Broadcom
Total Interfaces	16+2(128 integrated)	4	4	4
Interface Type	Dynamic	Fixed	Fixed	SR-IOV,ToE
Ethernet NICs	16 (128)	2	2	2
FC HBAs	2 (128)	2	2	2
VM-FEX	HW/SW	SW	SW	n/a
Eth.NIC Teaming	Hardware, no driver needed	Software, bonding driver	Software, bonding driver	802.3ad
Physical Ports	2x 10 GB	2x 10 GB	2x 10 GB	2x 10Bb
Srv.Compatibillity	M1/M2/M3	M1/M2	M1/M2	M2/M3

© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares the features of virtual interface card (VIC) and converged network adapters (CNAs) that are available for C-Series servers. These adapters allow C-Series to be configured with Ethernet NICs as well as with Fibre Channel over Ethernet (FCoE) HBAs.

Cisco UCS C-Series Ethernet Adapter Comparison

Adapter	Vendor	Ports	Type	Servers
NetXtreme II 5709*	Broadcom	4	10/100/1000Base-T, PCIe x4	M1/M2/M3
NetXtreme II 5709	Broadcom	2	10/100/1000Base-T, PCIe x4	M1/M2
NetXtreme II 57711*	Broadcom	2	10 GB E SFP+, PCIe x8 Optical / Copper direct attach cable	M1/M2
GB ET dual/quad**	Intel	2/4	10/100/1000Base-T, PCIe x4	M1/M2/M3
GB EF dual-port**	Intel	2	1000Base-SX 62.5/50 MMF	
X520-SR1/SR2/LR1/DA1***	Intel	1/2/1/2	10GBase-SR/SR/LR/CU SFP+	M1/M2/M3

Broadcom

* TCP offload, iSCSI boot, 64 VLANs



Intel

** LinkSec, iSCSI/PXE boot, 4096 VLANs

*** LinkSec, iSCSI/PXE boot, 4096 VLANs, 802.3ad



© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares the features of Ethernet adapters that are available for C-Series servers.

https://t.me/learningnets

Cisco UCS C-Series HBA Adapter Comparison



Adapter	Vendor	Ports	Type	Features	Servers
LPe 11002	Emulex	2	4Gb FC 62.5/50 MMF	NPIV FC-SP	M1/M2/M3
LPe 12002	Emulex	2	8Gb FC 62.5/50 MMF	NPIV FC-SP	M1/M2/M3
QLE2462	QLogic	2	4Gb FC 62.5/50 MMF	NPIV VSAN	M1/M2/M3
QLE 2562	QLogic	2	8G FC 62.5/50 MMF	NPIV VSAN	M1/M2/M3

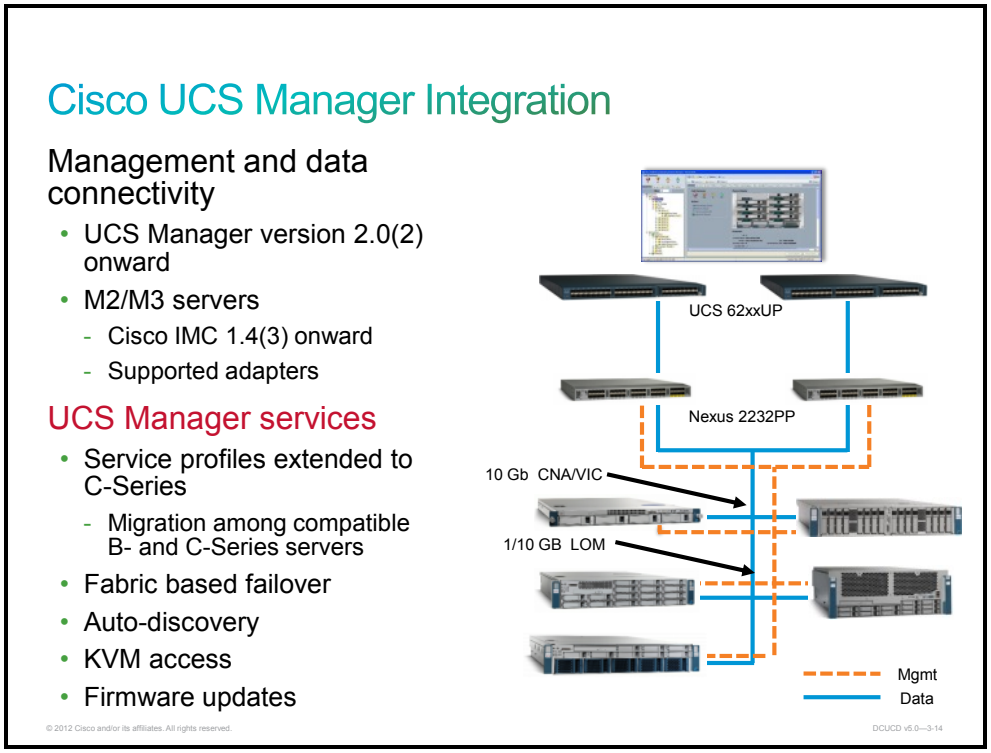
© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares the features of Fibre Channel HBA adapters that are available for C-Series servers.

https://t.me/learningnets

Cisco UCS C-Series Integration with UCS Manager

This topic describes the requirements of Cisco UCS C-Series integration with Cisco UCS Manager.



The C-Series servers can be integrated with the B-Series system—namely the Cisco UCS Manager—to utilize all of the benefits of consolidated management and common unified fabric.

The C-Series need to be connected to the fabric interconnects via Cisco Nexus 2232PP Fabric Extenders (FEX).

The following requirements apply to the integration of the C-Series with Cisco UCS Manager:

- Cisco UCS Manager has to be at software release 2.0(2xx) or later.
- C-Series Cisco Integrated Management Controller (IMC) has to be at 1.4(3c) or later firmware release.

Cisco Nexus 2232PP FEX has to be connected to Cisco UCS 6200 Fabric Interconnects using one of the following options:

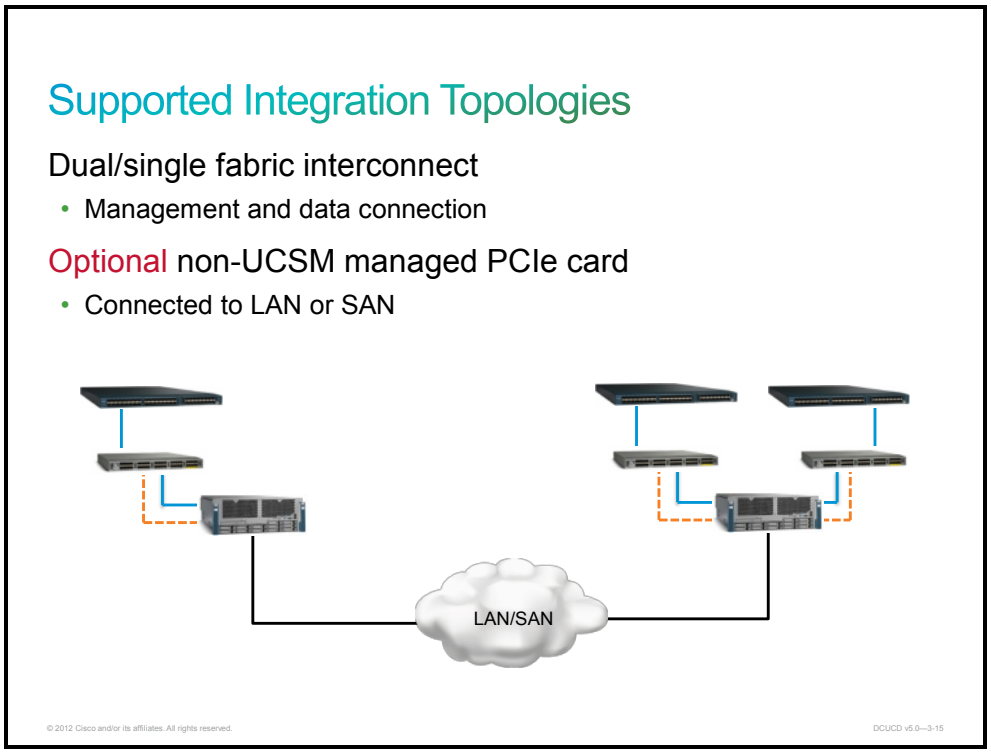
- **Hard-pinning mode:** In this mode, the server-facing ports of the FEX are pinned to the connected uplink ports as soon as the FEX is discovered. The Cisco UCS Manager software pins the server-facing ports to the uplink ports based on the number of uplink ports that are acknowledged. If a new uplink is connected later, or if an existing uplink is deleted, you must manually acknowledge the FEX again to make the changes take effect.
- **Port channel mode:** In this mode, all uplink ports are members of a single port channel that acts as the uplink to all server-facing ports. (There is no pinning.) If a port channel member goes down, traffic is automatically distributed to another member.

© 2012 Cisco Systems, Inc

https://t.me/learningnets

When planning the Cisco Nexus 2232FEX to fabric interconnect connectivity, the architect needs to observe that the available virtual interface (VI) namespace varies depending on where the uplinks are connected to the ports of the fabric interconnect:

- When port channel uplinks from the FEX are connected only within a set of eight ports managed by a single chip, Cisco UCS Manager maximizes the number of VIFs used in service profiles deployed on the servers.
- If uplink connections are distributed across ports managed by separate chips, the VIF count is decreased. For example, if you connect seven members of the port channel to ports 1–7, but connect the eighth member to port 9, this port channel can only support VIFs as though it had only one member.



The scheme details the supported C-Series integration topologies.

There are two general supported topologies:

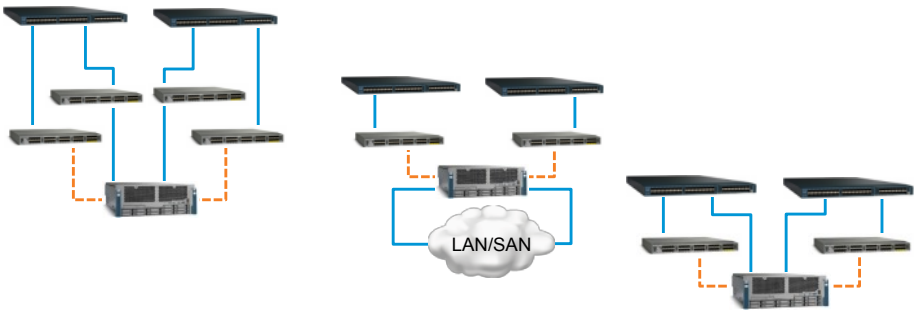
- Single fabric connection, which is not recommended because there is no connectivity redundancy
- Dual fabric connection, with redundant physical connectivity for Cisco UCS C-Series server management as well as data traffic.

If required, there can be an interface on the server that is not connected to the Cisco UCS Manager via Cisco Nexus 2232 FEX. Keep in mind that in such a case, the interface is not managed by the Cisco UCS Manager. Such deployments are rare but still possible.

https://t.me/learningnets

Nonsupported Integration Topologies

- Multi-FEX topology
- No data connection from C-Series
- Direct data connection from C-Series



© 2012 Cisco and/or its affiliates. All rights reserved.

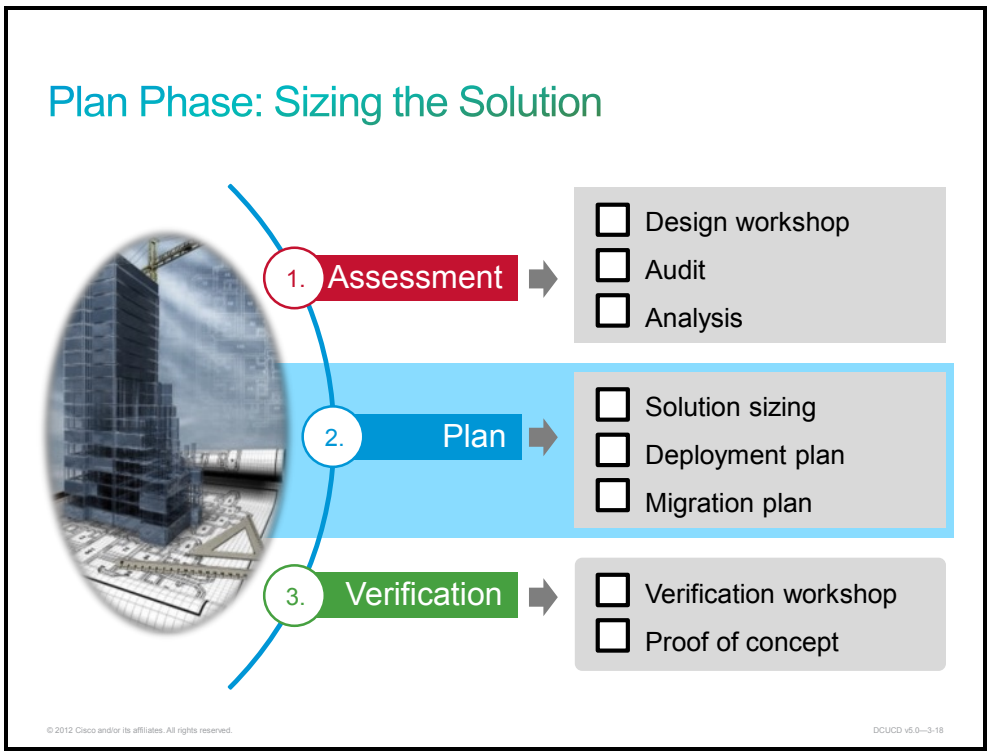
The scheme details the nonsupported C-Series integration topologies:

- Multi-FEX topology, where one set of Cisco Nexus 2232PP is used for management traffic and the second one for data traffic
- No-data connection from C-Series to Cisco UCS Manager, where only management is integrated.
- Direct connection of data interface to Fabric Interconnect s.

https://t.me/learningnets

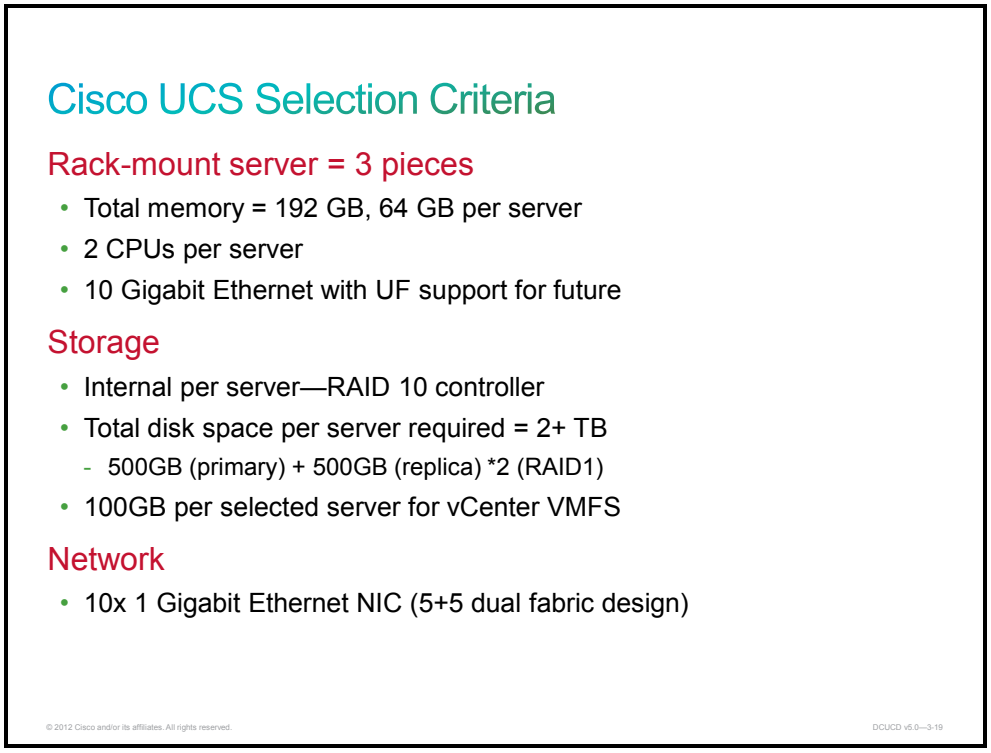
Size the Small VMware vSphere Solution—Plan

This topic describes how to select proper Cisco UCS C-Series server hardware based on the requirements for a given small VMware vSphere environment.



The second step in the design is the plan. Sizing the solution actually means selecting the proper hardware. This example continues with the small VMware vSphere solution design, but here the sizing is detailed.

https://t.me/learningnets



The high-level design specified the selected platform and characteristics that define the overall server characteristics:

- Total memory of 192 GB.
- Local RAID 10 setup for vendor-specific attribute (VSA) deployment for shared datastores.
- The VSA deployment limits the number of hosts in vSphere cluster to a maximum of three. In this example, three are used in order to deploy a solution with a smaller fault domain. (The failure of a host presents a loss of 33 percent of resources instead of 50 percent in the case of only two hosts being used.)
- Because three hosts are used, a single host will have 64 GB of memory and be equipped with two CPUs.
- The network connectivity will be based on 10 Gigabit Ethernet with FCoE support for future use.

https://t.me/learningnets

Sizing Cisco UCS Hardware: Example

Server selected: UCS 220 M3

- Price/performance + investment protection (scaleable HW)

Hardware components per server

Component	Part	Qty	Comment
Processor	2.30 GHz E5-2630/95W 6C	2	Performance & value criteria
DIMM	8GB DDR3-1600-MHz RDIMM/PC3-12800/dual rank/1.35v	8	Performance & value criteria
RAID controller	2008 Mezz Card for UCS C220 server	1	Performance & value criteria
Disk #1	300 GB 6 GB SAS 10K RPM	8	Performance & value criteria, for VSA
Disk #2	4 GB USB Drive	1	For ESXi
Network adapter	UCS P81E VIC	1	10 GE with UF support for future

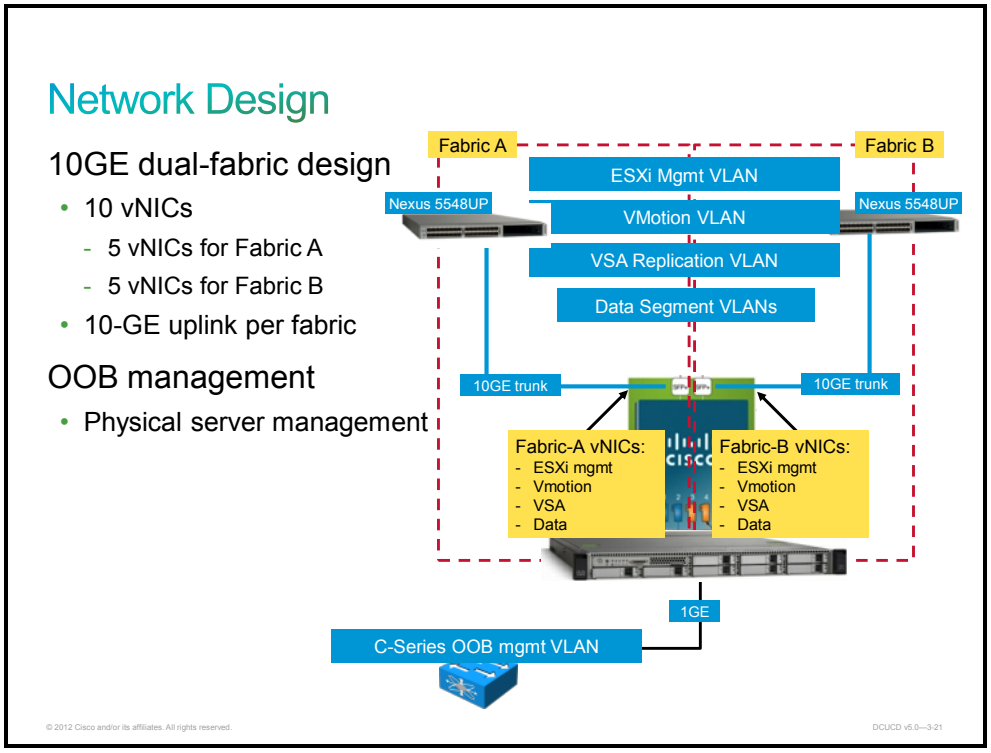
© 2012 Cisco and/or its affiliates. All rights reserved.

The high-level design is the basis for selecting the server hardware.

In the example, the C220 M3 series server was selected with the hardware components as specified in the table above.

The disk selection is based on the VSA deployment requirements, with the emphasis on I/) operations per second (IOPS) performances and price. This is the reason for selecting the 300 GB 10K RPM SAS disk drives.

https://t.me/learningnets



The network connectivity is planned according to the identified requirements and constraints as well as the network adapters selected.

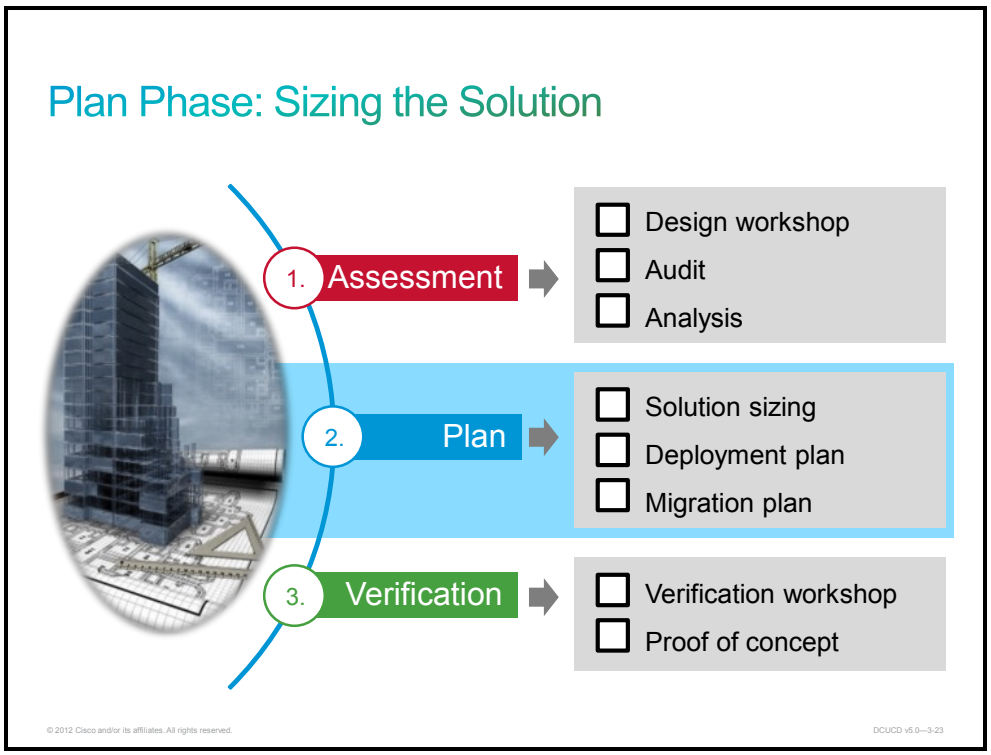
Cisco P81 adapter enables the creation of multiple NICs per fabric. In the example, four are required per fabric:

- ESXi management NIC
- VMotion segment NIC
- VSA replication NIC
- Data NIC

https://t.me/learningnets

Size the Small Hyper-V vSphere Solution—Plan

This topic describes how to select proper Cisco UCS C-Series server hardware based on the requirements for a given small Hyper-V vSphere environment.



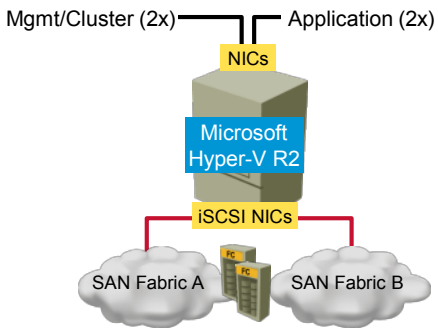
This example continues with the small Hyper-V solution design. The second step in the design is the plan. Sizing the solution actually means selecting the proper hardware.

https://t.me/learningnets

Cisco UCS Selection Criteria

Rack-mount server = 3 pieces

- Total memory = 288 GB, 96 GB per server
- 2 CPUs per server
  - 4 cores per CPU = TOTAL 24 cores
- 1GE NICs



Storage

- Internal per server with RAID 1
- Total local disk space per server required = 200 GB

© 2012 Cisco and/or its affiliates. All rights reserved.

Based on the requirements, a high-level design is proposed.

The network connectivity is also an important part of the assessment phase because it governs the adapter selection and connectivity type.

LAN Connectivity Requirements

Microsoft recommends assigning dedicated links to various traffic types that are critical to the operation of certain applications and network services.

Some of the traffic types involve communication to the Hyper-V server (the parent operating system), and some involve traffic up to the virtual guest operating systems themselves:

- Live migration and virtual machine (VM) management by Hyper-V Manager
- Application data
- Internet Small Computer Systems Interface (iSCSI) storage

https://t.me/learningnets

Sizing Cisco UCS Hardware: Example

Server selected: UCS 200 M2

- Price/performance + investment protection (scalable HW)

Hardware components per server

Component	Part	Qty	Comment
Processor	2.40 GHz Xeon E5620 80W CPU/12MB cache/DDR3 1066MHz	2	24 cores in total
DIMM	8 GB DDR3-1333-MHz RDIMM/PC3-10600/2R/1.35v	12	96 GB required, performance & value criteria
RAID controller	Embedded – RAID1	1	External storage for VMs
Disk #1	Gen 2 500-GB SATA 7.2K RPM 3.5 HDD	2	for Hyper-V, performance & value criteria
Network adapter	Broadcom 5709 Quad Port 10/100/1-GB NIC w/TOE iSCSI	1	For iSCSI TOE + data

© 2012 Cisco and/or its affiliates. All rights reserved.

The high-level design is the basis for selecting the server hardware. In this example, UCS C200 M2 rack-mount server with the hardware components as specified in the table is selected.

The emphasis is on the network adapter, where Broadcom 5709 Quad-Port Adapter also supports iSCSI TCP/IP Offload Engine (ToE) functionality, which in this solution brings additional performance benefits because the storage is an iSCSI base.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.

Summary

- Cisco UCS C-Series sizing process encompasses selection of server type and components (CPU, memory, PCIe adapter, RAID controller, disk drives, optional components).
- Cisco UCS C-Series can be integrated with Cisco UCS Manager via Cisco Nexus 2232PP FEX.
- Use of VSA influences Cisco UCS C-Series local disk and RAID sizing.
- Ethernet-only adapters can be used when SAN connectivity is not required.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Lesson 2

Sizing the Cisco UCS B-Series Server Solution

Overview

When sizing the Cisco Unified Computing System (UCS) B-Series, solution the architect sizes the individual servers—selecting proper components like CPU, memory DIMMs, and adequate quantities—picks appropriate quantity of server selected, and selects and sizes fabric interconnect switches, chassis, and I/O module (IOM).

This lesson discusses the Cisco UCS B-Series server solution for a given example.

Objectives

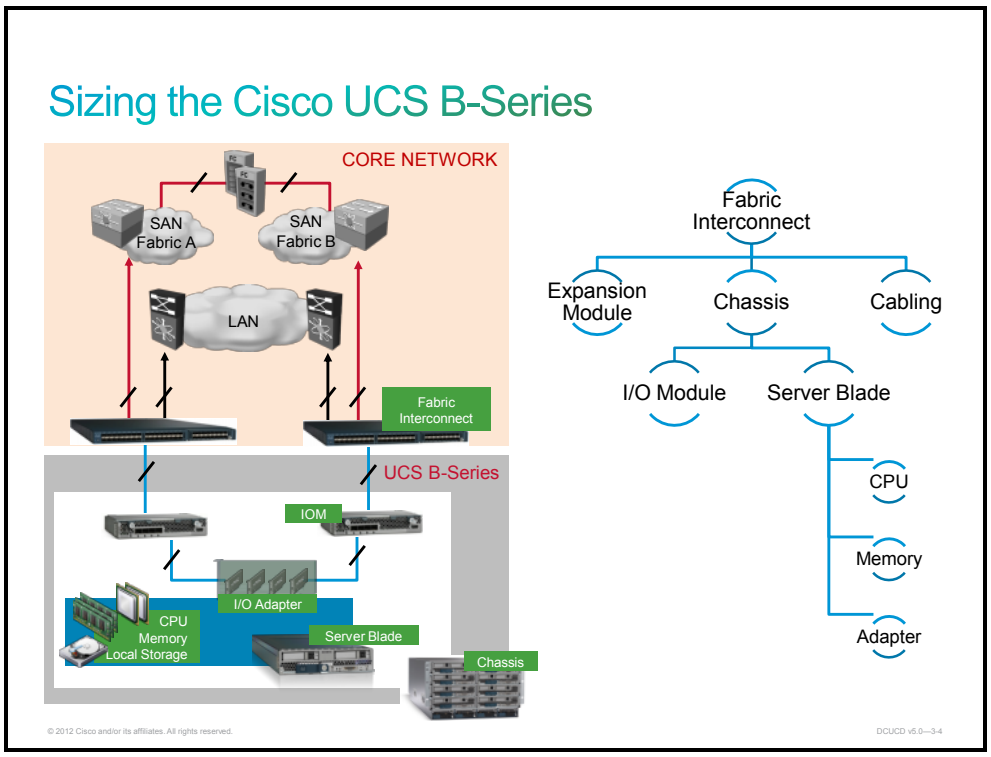
Upon completing this lesson, you will be able to consider the Cisco UCS B-series server solution for a given set of requirements. You will be able to meet these objectives:

- Recognize the general Cisco UCS B-Series server hardware sizing aspects
- Describe an example of gathering requirements for a given VMware View desktop virtualization solution

https://t.me/learningnets

Size the Cisco UCS B-Series Solution

This topic describes the general Cisco UCS B-Series server hardware sizing aspects.

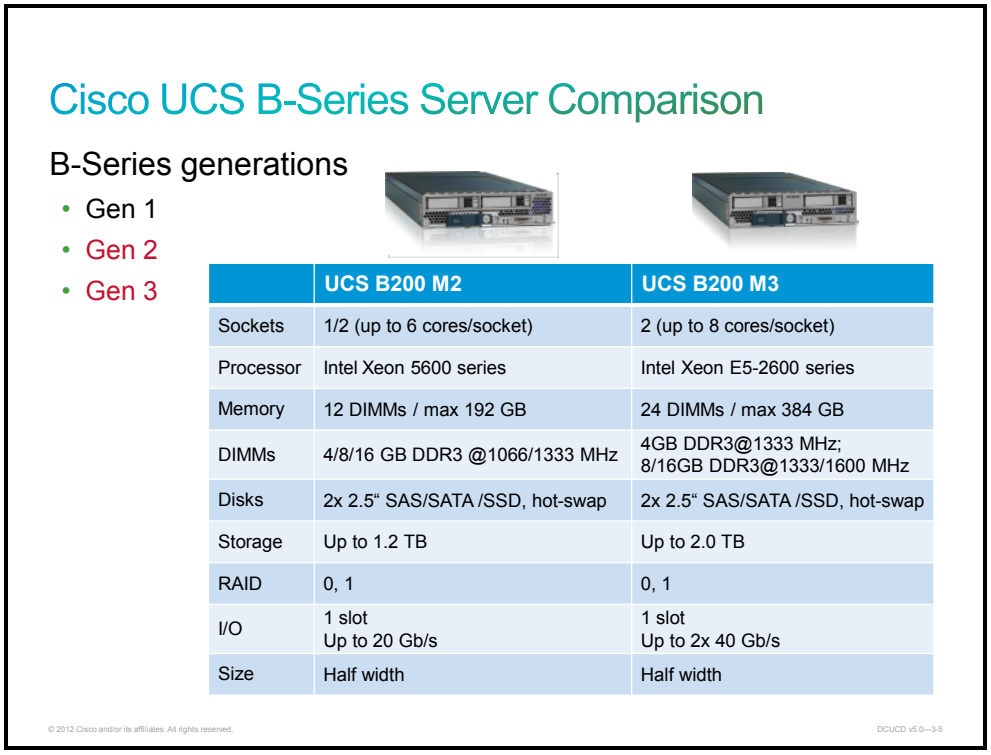


The sizing process for Cisco UCS deployment is used to determine the appropriate components that are needed to build the Cisco UCS per the requirements that were gathered in a design workshop.

The sizing process can be divided into three major categories:

- **Identify and size Cisco UCS server classes:** With this step, the individual server type is identified and sized per requirements.
- **Identify and define Cisco UCS chassis classes:** With this step, the server blades are put into the chassis—that is, the chassis classes are identified and properly sized—for the purpose of identifying the correct number of server uplinks.
- **Identify and size Cisco UCS fabric interconnect clusters:** With this step, the fabric interconnects, the expansion modules, and the number and type of chassis connected to the individual fabric interconnect clusters are identified.

https://t.me/learningnets



The figure compares Cisco UCS B200 M2 and M3 with maximum hardware configurations.

Cisco UCS B200 M2 Server Blade

The Cisco UCS B200 M2 server blade has the following features:

- Up to two Intel Xeon 5600 Series processors, which adjust server performance according to application needs.
- Up to 192 GB of double data rate 3 (DDR3) memory, which balances memory capacity and overall density.
- Two optional Small Form-Factor (SFF) Serial Attached SCSI (SAS) hard drives or 15-mm Serial AT Attachment (SATA) solid-state drive (SSD) drives, with an LSI Logic 1064e controller and integrated Redundant Array of Independent Disks (RAID).

Cisco UCS B200 M3 Server Blade

The Cisco UCS B200 M3 server blade has the following features:

- Up to two processors from the Intel Xeon E5-2600 product family
- Up to 384 GB of RAM with 24 DIMM slots
- Up to two SAS/SATA/SSD
- Cisco UCS virtual interface card (VIC) 1240 is a 4 x 10 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE)-capable modular LAN on motherboard (LOM) and, when combined with an optional I/O expander, allows up to 8 x 10GE blade bandwidth
- 1 mezzanine, PCI Express (PCIe) slot Gen 3

https://t.me/learningnets

Cisco UCS B-Series Server Comparison (Cont.)



	UCS B250 M2	UCS B230 M2	UCS B440 M2
Sockets	1/2 (up to 6 cores/socket)	2 (up to 10 cores/socket)	2/4 (up to 10 cores/socket)
Processor	Intel Xeon 5600 Series	Intel Xeon E7-2800 series	Intel Xeon E7-4800 series
Memory	48 DIMMs / max 384 GB	32 DIMMs / max 512 GB	32 DIMMs / max 1 TB
DIMMs	4/8 GB DDR3@1066/1333 MHz	4/8/16/32 GB DDR3 @1066/1333 MHz	4/8/16/32 GB DDR3 @1066/1333 MHz
Disks	2x 2.5“ SAS/SATA /SSD, hot swap	2x 2.5“ SSD, hot swap	4x 2.5“ SAS/SATA /SSD, hot swap
Storage	Up to 1.2 TB	Up to 200 GB (SSDs)	Up to 2.4 TB
RAID	0, 1	0, 1	0, 1, 5, 6
I/O	2 slots Up to 40 Gb/s	1 slot Up to 20Gb/s	2 slots Up to 40 Gb/s
Size	Full-width	Half-width	Full-width

© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares other Cisco UCS B-Series M2 with maximum hardware configurations.

Cisco UCS B250 M2 Server Blade

The Cisco UCS B250 M2 server blade has the following features:

- Up to two Intel Xeon 5600 Series processors, which adjust server performance according to application needs
- Up to 384 GB based on Samsung’s 40 nanometer class (DDR3) memory technology, for demanding virtualization and large dataset applications or for a more cost-effective memory footprint for less-demanding workloads
- Two optional Small Form-Factor (SFF) Serial Attached SCSI (SAS) hard drives available in 73 GB, 15,000 RPM, and 146 GB, 10,000 RPM versions with an LSI Logic 1064e controller and integrated RAID
- Two dual-port mezzanine cards for up to 40 Gb/s of I/O per blade. Mezzanine card options include either a Cisco UCS VIC M81KR Virtual Interface Card, a converged network adapter (Emulex or QLogic compatible), or a single 10 Gigabit Ethernet adapter.

Cisco UCS B230 M2 Server Blade

The Cisco UCS B230 M2 server blade has the following features:

- Supports Intel Xeon® processor E7-2800 product family
- Enhanced memory capacity and bandwidth to support your virtualized environment
- Advanced reliability features for mission-critical workloads
- It has one dual-port mezzanine card for up to 20 Gb/s I/O per blade. Options include a Cisco UCS M81KR Virtual Interface Card or converged network adapter (Emulex or QLogic compatible). Other features include:
- 32 DIMM slots

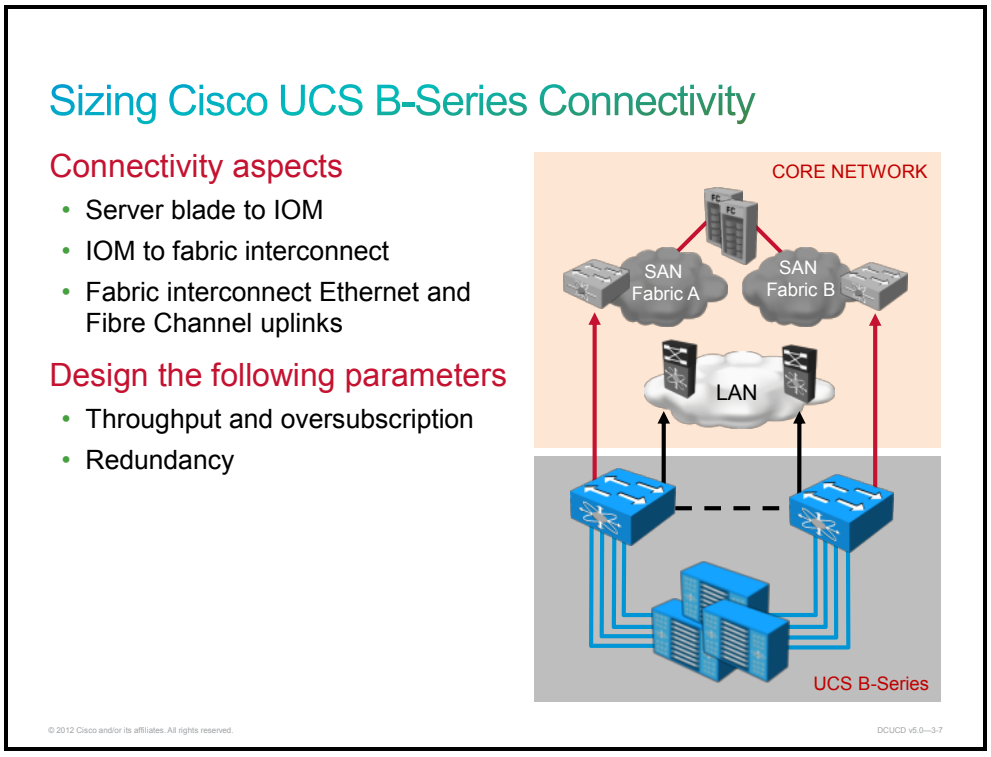
https://t.me/learningnets

- Up to 512 GB memory at 1066 MHz, based on Samsung 40 nanometer class (DDR3) technology
- Two optional front-accessible, hot-swappable SSDs and an LSI SAS2108 RAID Controller

Cisco UCS B2440 M2 Server Blade

The Cisco UCS B230 M2 server blade has the following features:

- Supports Intel Xeon processor E7-4800 product family
- 32 DIMM slots
- Up to 1 TB memory at 1066 MHz, based on Samsung's 40 nanometer class (DDR3) technology
- Four optional front-accessible, hot-swappable small form-factor pluggable (SFP) drives and an LSI SAS2108 RAID Controller
- Two dual-port mezzanine cards for up to 40 Gb/s I/O per blade.



The UCS B-Series sizing is influenced by two factors—the required throughput and the allowed oversubscription.

Both factors should be observed from different perspectives:

- The individual server blade
- Individual chassis and cumulative requirements of the server blades installed in the chassis
- Individual fabric interconnect and cumulative requirements of the chassis connected to the switch, and the upstream LAN and SAN connectivity requirements

https://t.me/learningnets

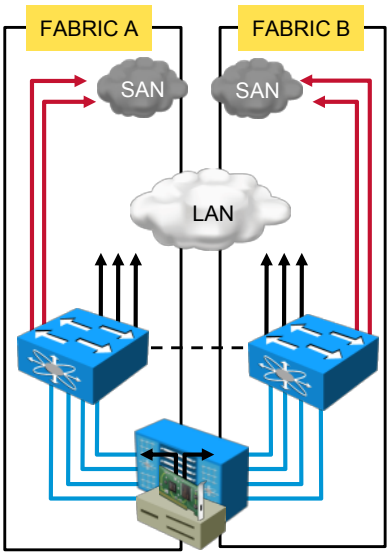
Redundant Connectivity

Dual fabric design

- Two fabric interconnects and IOMs

Determine number of

- Server downlinks
- LAN and/or SAN uplinks
- Links for directly attached devices



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco UCS B-Series connectivity architecture typically follows the dual fabric design. Such design can be used to either achieve high availability (with failover on the Cisco UCS level or on the operating system level) or to achieve more throughput by directing traffic from some servers to Fabric A and from other servers to Fabric B. It is also possible to direct traffic to both fabrics, combining redundancy and higher throughput.

Unified Fabric Design

The “physical” fabric will consist of two fabrics—Fabric A (left) and Fabric B (right).

The LAN fabrics are physically separated on the Cisco UCS fabric interconnect level but can be (and typically will be) on the same devices in the LAN core.

The SAN fabrics are physically separated on the Cisco UCS fabric interconnect level and typically remain separated on the devices in the SAN core.

https://t.me/learningnets

Fabric Interconnect Comparison

Generations

- Gen 1
- Gen 2



	UCS 6248UP	UCS 6296UP
Form Factor	1RU	2RU
Expansion Slot	1	3
Total/Fixed Ports	48/32	96/48
Additional ports	16 with expansion module	48 with 3 expansion modules
Throughput	960 Gb/s	1920 Gb/s
SFP+	10G SR/LR/CU (1,3,5 twinax) 10G FET MMF	10G SR/LR/CU (1,3,5 twinax) 10G FET MMF
SFP – all ports	1G SW/LW/UTP 4/8G FC SW/LW SFP	1G SW/LW/UTP 4/8G FC SW/LW SFP
Port licensing	12 + additional licenses	18 + additional licenses
VLANs	1024	1024



Expansion module with 16 unified ports

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco UCS 6248UP

The Cisco UCS 6248UP is a fabric interconnect that doubles the switching capacity of the data center fabric to improve workload density (from 520 Gb/s to 1 Tb/s), reduces end-to-end latency by 40 percent to improve application performance (from 5.2 usec to 3.2 usec), and provides flexible unified ports to improve infrastructure agility and transition to a fully converged fabric.

- 48 ports in a single RU:
  - Has 32 fixed ports on the base switch
  - 16 optional ports on the expansion module
- Redundant front-to-back airflow
- Has dual power supplies for both AC and DC -48V. The power consumption of the fabric interconnect itself is ~half gen-1 fabric interconnects
- All ports on the base and expansion module are “Unified Ports”
- Each of these ports can be configured as Ethernet/FCoE or native Fibre Channel
- Depending on optics, these can be used as SFP 1 Gigabit Ethernet, SFP+ 10 Gigabit Ethernet, Cisco 8/4/2 Gb and 4/2/1 Gb Cisco Fibre Channel.

Cisco UCS 6296UP

The Cisco UCS 6296UP 96-Port Fabric Interconnect is a 2-RU 10 Gigabit Ethernet, FCoE and native Fibre Channel switch offering up to 1920 Gb/s throughput and up to 96 ports. The switch has 48 1/10 Gb/s fixed Ethernet, FCoE and Fiber Channel ports and three expansion slots.

- 96 ports in 2 RU
  - 48 fixed ports
  - additional 48 ports available through three expansion modules
- Redundant front-to-back airflow

https://t.me/learningnets

- Has dual power supplies for both AC and DC -48V. The power consumption of the Fabric Interconnect itself is ~half gen-1 Fabric Interconnects
- All ports on the base and expansion module are “Unified Ports”
- Each of these ports can be configured as Ethernet/ FCoE or native Fibre Channel
- Depending on optics, these can be used as SFP 1 Gigabit Ethernet, SFP+ 10 Gigabit Ethernet, Cisco 8/4/2 Gb and 4/2/1 Gb Cisco Fibre Channel.

IOM Comparison

IOM generations

- Gen 1
- Gen 2





	IOM 2204XP	IOM 2208XP
IOM Uplinks	4x 10GE FCoE capable	8x 10GE FCoE capable
IOM EtherChannel	Yes	Yes
Server Ports - Half-width Slot	2x 10G DCB	4x 10G DCB
Server Ports - Full-width Slot	4x 10G DCB	8x 10G DCB
SFP+	10G SR/LR/CU (1,3,5 twinax) 10G FET MMF	10G SR/LR/CU (1,3,5 twinax) 10G FET MMF

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco UCS 2208 IOM

The Cisco UCS 2208XP is a chassis IOM that doubles the bandwidth to the chassis to improve application performance and manage workload bursts (from 80 Gb to 320 Gb to the blade). The Cisco UCS 2208XP Fabric Extender has eight 10 Gigabit Ethernet, FCoE-capable, Enhanced Small Form-Factor Pluggable (SFP+) ports that connect the blade chassis to the fabric interconnect. Each Cisco UCS 2208XP has thirty-two 10 Gigabit Ethernet ports connected through the midplane to each half-width slot in the chassis. Typically configured in pairs for redundancy, two fabric extenders provide up to 160 Gb/s of I/O to the chassis.

Cisco UCS 2204 IOM

The Cisco UCS 2204XP Fabric Extender (Figure 4) has four 10 Gigabit Ethernet, FCoE-capable, SFP+ ports that connect the blade chassis to the fabric interconnect. Each Cisco UCS 2204XP has 16 10 Gigabit Ethernet ports connected through the midplane to each half-width slot in the chassis. Typically configured in pairs for redundancy, two fabric extenders provide up to 80 Gb/s of I/O to the chassis.

https://t.me/learningnets

Cisco UCS B-Series VIC Adapter Comparison

VIC adapter generations

- Gen 1
- Gen 2





	VIC 1240	VIC 1280	M81KR VIC
Total Interfaces	256	256	128
Interface Type	Dynamic	Dynamic	Dynamic
Ethernet NICs	0-256	0-256	0-128
FC HBAs	0-256	0-256	0-128
VM-FEX	HW/SW	HW/SW	HW/SW
Failover Handling	Hardware, no driver needed	Hardware, no driver needed	Hardware, no driver needed
Form Factor	Modular LOM	Mezzanine	Mezzanine
Network Throughput	40-80* GB	80 GB	20 GB
Srv. Compatibility	M3 blades	M1/M2 blades	M1/M2 blades

* With use of Port Expander Card for VIC 1240 in the optional mezzanine slot

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco UCS VIC M81KR, 1240, 1280, P81E Adapters

The Cisco UCS VIC supports NIC virtualization either for a single operating system or for VMware vSphere. The number of virtual interfaces supported on an adapter depends on the number of uplinks between the IOM and the fabric interconnect, as well as the number of interfaces in use on other adapters sharing the same uplinks.

https://t.me/learningnets

Cisco UCS B-Series 4-Interface Adapter Comparison

Qlogic/Emulex CNA Generations

- Gen 1
- Gen 2
- Gen 3







	M71KR Q/E	M72KR Q/E	M73KR Q/E
Total Interfaces	4	4	4
Interface Type	Fixed	Fixed	Fixed
Ethernet NICs	2	2	2
FC HBAs	2	2	2
VM-FEX	Software	Software	Software
Failover Handling	Hardware, no driver needed	Software, bonding driver	Software, bonding driver
Form Factor	Mezzanine	Mezzanine	Mezzanine
Network Throughput	20 GB	20 GB	20 GB
Srv. Compatibility	M1/M2 blades	M1/M2 blades	M3 blades

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco UCS CNA M71KR/M72KR, UCS 82598KR-CI Adapter

You can create up to two virtual network information cards (vNICs) using either the Cisco UCS CNA M71KR/M72KR or the Cisco UCS 82598KR-CI adapters.

With all but M71KR, you must match the physical setting (first adapter goes to Fabric A, second to Fabric B), and you cannot choose failover.

With Cisco UCS CNA M71KR, each vNIC is associated with a particular fabric, but you can enable failover.

https://t.me/learningnets

Cisco UCS B-Series 2-Interface Adapter Comparison

CNA Generations

- Gen 1
- Gen 2
- Gen 3







	Intel M61KR-I CNA	Broadcom M51KR-B 57711	Broadcom M61KR-B 57712
Total Interfaces	2	2	2
Interface Type	Fixed	Fixed	Fixed
Ethernet NICs	2	2, iSCSI ToE	2, iSCSI ToE
FC HBAs	Future	0	0
VM-FEX	Software	Software	Software
Failover Handling	Yes	Software, bonding driver	Software, bonding driver
Form Factor	Mezzanine	Mezzanine	Mezzanine
Network Throughput	20 GB	20 GB	20 GB
Srv. Compatibility	M1/M2 blades	M1/M2 blades	M3 blades

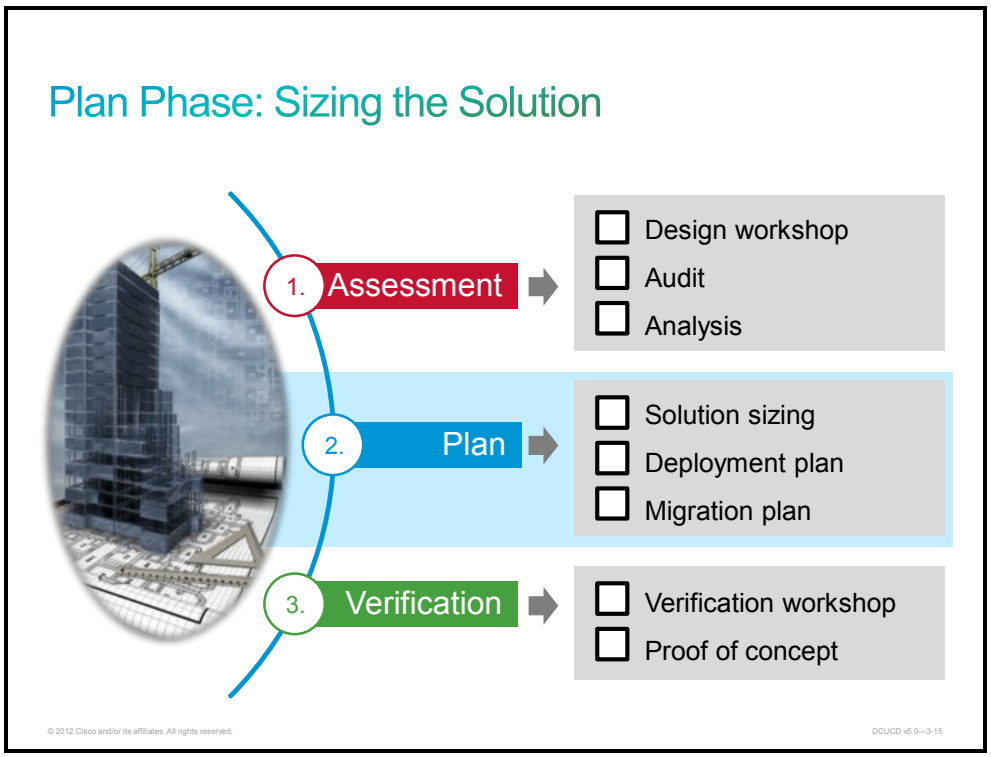
© 2012 Cisco and/or its affiliates. All rights reserved.

The table compares the two-interface adapters available for B-Series.

https://t.me/learningnets

Size the Desktop Virtualization Solution—Plan

This topic describes an example of gathering requirements for a given VMware View desktop virtualization solution.



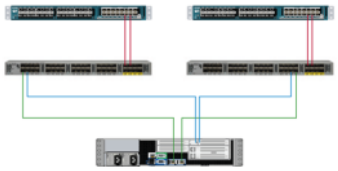
The second step in the design is plan. Sizing the solution actually means selecting the proper hardware.

https://t.me/learningnets

Design for “Physical” Applications

Requirements

- Application requirements fit memory, CPU, disk resources of two servers
- Requires LAN and SAN connectivity
  - Some data will be stored on central drive arrays – dedicated LUNs
  - LAN for data traffic
- Should be integrated with virtual server infrastructure as much as possible



Repurpose of existing C-Series

- No upgrade required for CPU, memory, local disks
- Upgrade required
  - Currently on-board GE used => add Cisco UCS P81 VIC
- Integration with new UCS => add Nexus 2232PP

© 2012 Cisco and/or its affiliates. All rights reserved.

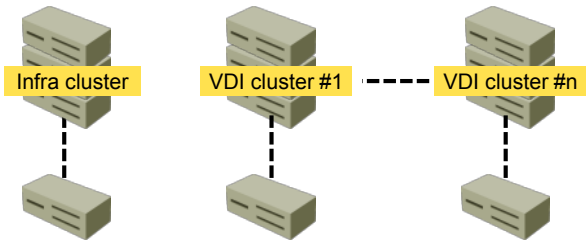
The infrastructure for the physical applications will be repurposed C-Series servers, which are currently used by the software development department. They have enough resources to fit the requirements for the physical applications; the only thing that will be added is the Cisco Nexus 2232PP FEXs to integrate them into B-series infrastructure.

https://t.me/learningnets

High-Level Design for Virtual Infrastructure

Compute design – multiple vSphere clusters

- Infrastructure cluster
  - for server VMs (infrastructure, applications, etc.)
  - for development VMs
  - for VDI VSES
- Multiple VDI clusters – for VDI VMs
  - NOTE: View limitation – 8 hosts per cluster



© 2012 Cisco and/or its affiliates. All rights reserved.

The server virtual infrastructure will be divided into two types of clusters:

- Infrastructure that will host infrastructure virtual machines (VMs), applications VMs, software development VMs
- Virtual desktop infrastructure (VDI) that will host VDI desktops

This design choice is recommended practice for deploying mixed VDI and server virtualization environment. This way the server VMs which host applications that are accessed by multiple users but do not compete for the resources with virtual desktops.

There is a VMware View limitation of the ESXi cluster size; there can be only eight hosts per VDI cluster.

https://t.me/learningnets

Compute Design for Infrastructure Cluster

Infrastructure cluster

- Memory = 1036+ GB
- CPU = 87+
- vSphere Ent.Plus = 192GB per hosts

Parameter	Memory	CPU
Server VMs with FT	400+ GB	20+
Server VMs with HA	500+ GB	50+
Devel.VM with HA	120 GB	12,5
VDI VSes	16+ GB	4.5+

Server type selected – UCS B200 M3 – 6 hosts

- Memory calculation => 1036+/192 per host = 5,39 servers
- CPU ratio => 87/2 sockets/ 8cores = 5,44 servers

Component	Part	Qty	Comment
Processor	Intel Xeon 2.40 GHz E5-2665/115W 8C/20MB Cache/DDR3 1600MHz	2	
DIMM	8GB DDR3-1600-MHz RDIMM/PC3-12800/dual rank/1.35v	24	Licensing and fault domain size observed + performance/value ratio
Flash	4GB Flash USB Drive	1	For ESXi
Network adapter	VIC 1280 dual 40 GB capable	1	Ethernet + FCoE

© 2012 Cisco and/or its affiliates. All rights reserved.

The design for the infrastructure cluster is detailed in the above tables.

The memory and CPU requirements are calculated from the requirements gathered and are updated with the agreed oversubscription levels.

The selection of the VMware vSphere license pack also governs the total amount of memory per server blade.

https://t.me/learningnets

Compute Design for VDI Clusters

VM	Memory
Memory	2500* 1.5GB (oversub 1:1.5)
CPU	2500*1 (oversub 1:7)
HA level	Max. 12.5%
Total mem.	2812.5+GB
Total CPU	401+

VDI cluster boundaries

- Max. 8 hosts per cluster (7+1)
- Limit for fault domain = 120 VMs per host
  - Memory required = 120*1.5/1.5 = 120GB
  - CPU required = 120/7 = 17
  - For 2500 desktops = 21 hosts (+ HA host per cluster)
    - 3 clusters (7+1 hosts per cluster) = 24 hosts

Server selected – UCS B230 M2

Component	Part	Qty	Comment
Processor	2 GHz E7-2850 130W 10C /24M Cache	2	20C per server
DIMM	2X8GB DDR3-1333-MHz RDIMM/PC3-10600/dual rank/x2/1.35v	8	Fault domain size observed + performance/value ratio
Disk	None		Boot from SAN
Network adapter	VIC 1280 dual 40 GB capable	1	Ethernet + FCoE

© 2012 Cisco and/or its affiliates. All rights reserved.

Depending on the processor selected and the size of the RAM, plus the version of the ESXi hypervisor and View, different consolidation numbers can be achieved. The general rule is that the newer the software version, the faster the CPU, and the bigger the memory, the more VDI VMs can be run on a single blade server.

Note Running many VMs on a single blade creates a large fault domain.

The design for the VDI clusters follows these rules:

- VDI cluster will have a maximum of eight blades.
- Seven blades are used to run the workload; the eighth one is for redundancy to meet the 12.5 percent of resources reserved for host failures.
- The required memory and CPU cores are calculated from the total amount and design decision to host up to 120 virtual desktops per blade. This does *not* create a too-large fault domain._

https://t.me/learningnets

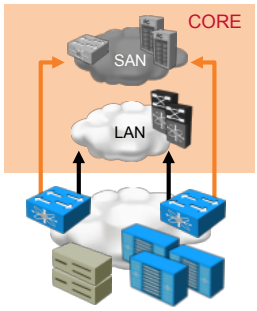
System Design

Servers

- Infra cluster = 6x B200 M3
- VDI clusters = 24x B230 M2
- Physical app = 2x C200 M2

Blade chassis

- 4x 5108 with IOM 2208
  - 4 uplinks per IOM
  - Future proven + able to scale



Connectivity per Fabric Interconnect

Role	Qty	Comment
IOM link	16	
LAN uplink	4x 10GE 1x 1GE	Internal LAN DMZ segment (View Security Servers)
FEX link	1x 10GE	C 200 M2 integration
SAN uplink	4x 8G FC	

- UCS 6248 UP fabric interconnect
  - 22x 10GE links
  - 4x 8G FC links

© 2012 Cisco and/or its affiliates. All rights reserved.

Lastly, the design for the whole system needs to be completed—the selection of fabric interconnects, IOMs, chassis, and links.

Overall system for this example is straight forward:

- Two 6248UP fabric interconnects are used for dual fabric design.
- To install the 30 server blades, four 5108 chassis are used. This leaves two slots empty for future expansion.
- The Cisco Nexus 2232PP is used to integrate the existing C200 M2 series servers in Cisco UCS Manager.
- To perform the integration of the C200 M2, the two servers need to be upgraded with the Cisco UCS P81 VIC network adapter.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.

Summary

- The Cisco UCS B-Series system is typically designed in a dual-fabric manner.
- The Cisco UCS solution sizing for VDI deployment depends significantly on the type of the workload—that is, worker type.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Lesson 3

Planning Unified Computing Deployment

Overview

When the data center architect completes the design of the Cisco Unified Computing System (UCS) solution—following the sizing of the equipment—the physical deployment plan needs to be prepared. It is vital to observe the data center facility, server room, and rack environmental characteristics like power, cooling, space, loading, and cabling.

This lesson discusses the physical deployment aspects for Cisco UCS solutions and presents the Cisco UCS Power Calculator tool.

Objectives

Upon completing this lesson, you will be able to create a physical deployment plan and use Cisco UCS Power Calculator tool to create one. You will be able to meet these objectives:

- Recognize the Cisco UCS Power Calculator tool
- Propose a physical deployment plan

https://t.me/learningnets

Cisco UCS Power Calculator Tool

This topic introduces the Cisco UCS Power Calculator tool.

Cisco UCS Power and Cooling Requirements

Cisco UCS Power Calculator tool

- http://express.salire.com/Go/Cisco/Cisco-UCS-Power-Calculator.aspx

Enter Cisco UCS configuration

- B-Series and/or C-Series blades, chassis, fabric interconnect

Power and cooling per

- Idle
- 50% load
- Maximum load

Weight



© 2012 Cisco and/or its affiliates. All rights reserved.

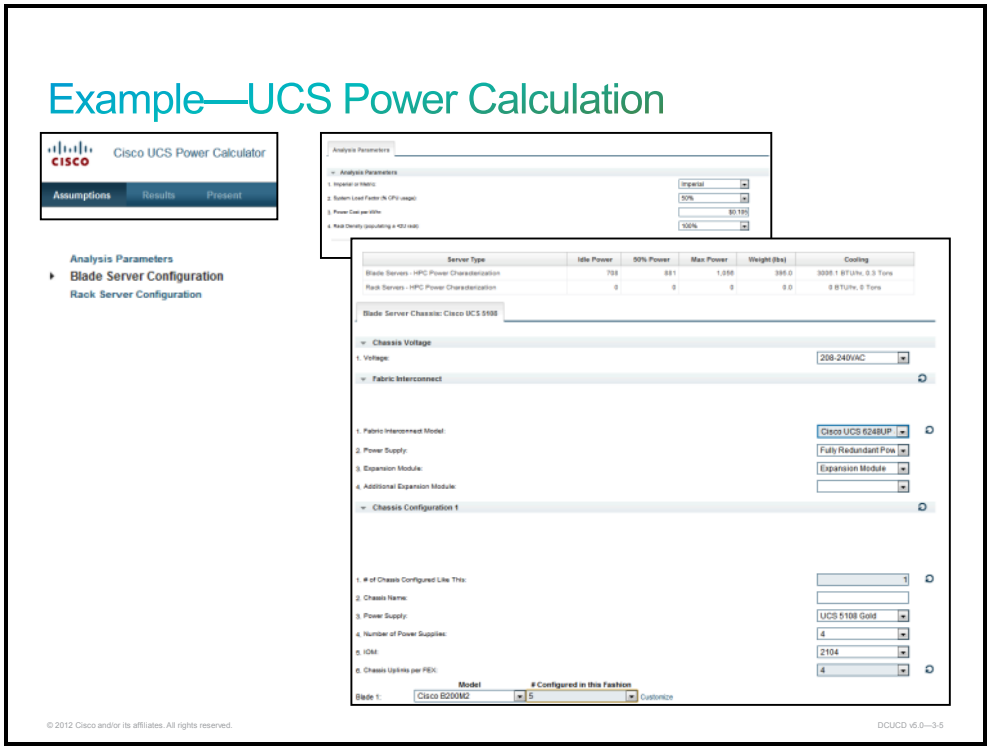
When designing physical deployment for Cisco UCS, the power consumption has to be calculated for the solution. This calculation is necessary because using one or two processors, disk drives, or DIMMs of different speed and size will result in different power consumption.

The Cisco UCS Power Calculator tool enables the designer to calculate exact values for the environmental parameters of the designed Cisco UCS solution—individual chassis, fabric interconnect switches, and B-Series and C-Series servers:

- Power consumption required and cooling capacity for idle, 50 percent, and maximum load
- Weight

Note The tool is available to Cisco partners at http://express.salire.com/Go/Cisco/Cisco-UCS-Power-Calculator.aspx.

https://t.me/learningnets



The Cisco Power Calculator tool is used to calculate the characteristics for this equipment:

- Cisco UCS C-Series rack-mount servers
- Cisco UCS B-Series blade servers and chassis
- Cisco UCS 6200UP or older 6120XP Fabric Interconnect switches

To produce the report on the required power, cooling capacity, and weight, the design needs to contain hardware that can provide the solution with the proper characteristics (that is redundancy, quantities, and so on).

https://t.me/learningnets

Physical Deployment Design

- 6-foot rack = 42 RU – 13 RU free per rack
- Available power = 3 kW per rack

Component	Selection
Server type	C460 M2 rack-mount server
Size	4 RU
Processor	Intel Xeon X7560 130W
Processor quantity	4
DIMM size	8 GB
DIMM quantity	32
RAID controller	—
Network adapter	Cisco UCS P81 VIC
Network adapter quantity	1
Local disk type	500-GB SATA 7200 rpm
Local disk quantity	1
Power supply	2x 850-W power supply unit
LAN physical cabling per server	2x 5-m twinax cables
SAN physical cabling per server	0
Server quantity	6



Component	Value
Idle power W	3875
50% load power W	5237
Maximum power W	6642
Weight	552 lb (237 kg)
Idle power BTU	13218
50% load power BTU	17856
Maximum power BTU	22644

© 2012 Cisco and/or its affiliates. All rights reserved.

The VMware small solution in this example is composed of C460 Series rack-mount servers with the following characteristics.

Component	Selection
Server type	C460 M2 rack-mount server
Size	4 RU
Processor	Intel Xeon X7560 130 W
Processor quantity	4
DIMM size	8 GB
DIMM quantity	32
RAID controller	—
Network adapter	Cisco UCS P81 VIC
Network adapter quantity	1
Local disk type	500 GB SATA 7200 rpm
Local disk quantity	1
Power supply	2x 850 W power supply unit
LAN physical cabling per server	2x 5-m twinax cables
SAN physical cabling per server	0
Server quantity	6

Note This is an example of using a power calculator tool in a solution with C-Series servers only; usually there is additional equipment.

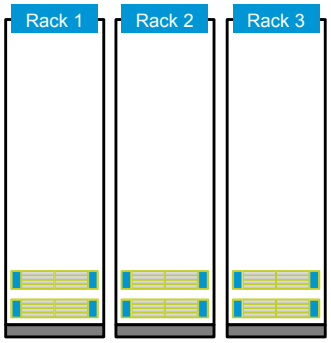
https://t.me/learningnets

The solution needs to be installed in the existing facility where 3 kW and 13 rack units (RU) of space per rack are available.

Physical Deployment Design (Cont.)

- Total power = 6,642 W

Parameter	Value
Server size	4 RU
Required power per server	1107 W
Servers per rack deployed	2
Power consumed by servers per rack	2214 W
Required rack quantity	3



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Power Calculator tool, which is used to calculate power consumption of the solution, gives the following numbers per configured server.

Component	Selection
Server type	C460 M2 rack-mount server
Server quantity	6
Idle power W	3875
50% load power W	5237
Maximum power W	6641
Weight	552 lb (237 kg)
Idle power BTU	13218
50% load power BTU	17856
Maximum power BTU	22644

https://t.me/learningnets

Physical Deployment Plan

The calculation has considered the following parameters:

- The C460 size is 4 RU.
- The maximum power consumption per server for a given solution is 1107 W. The maximum power consumption has been taken into account to prevent server outages due to sudden load increases.
- With the given constraints—3 kW per rack and 13 RU of free space—two servers per rack can be installed.
- Three racks are required to install the servers.

Parameter	Value
Server size	4 RU
Required power per server	1107 W
Servers per rack deployed	2
Power consumed by servers per rack	2214 W
Required rack quantity	3

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Create a Physical Deployment Plan

This topic describes how to propose a physical deployment plan.

Physical Deployment Considerations

- Rack capacity—chassis density per rack cabinet
  - Usable space
  - Loading
  - Power and cooling—affected by chassis requirement
- Rack density per array (row)
- Fabric interconnect physical placement
  - Depends on racks and chassis density
  - Governs selection of cabling

Cabling Type	Distance (m)	Placement
Copper (twinax)	1, 3, 5, 7, 10	MoR
Optical (short reach)	82	EoR
Optical (short reach)	300	Multirow

Standard rating
1 W = 3.41 BTU/hr



© 2012 Cisco and/or its affiliates. All rights reserved.

An important part of any data center solution is the physical deployment design. This design includes the following:

- Sizing the racks:
  - Determining how much space is available per rack
  - Defining the amount of equipment per rack
  - Calculating the power requirements per rack
  - Determining the number of power supplies per Cisco UCS 5108 chassis to achieve the required redundancy level
  - Calculating the heat dissipation per rack
- Sizing the array, which includes determining how many rack cabinets are required per Cisco UCS cluster.
- Fabric interconnect placement, which includes determining where to put the Cisco UCS 6200UP Fabric Interconnect switches, and which cables to use for the I/O module (IOM)-to-fabric interconnect physical connectivity.

The table in the figure summarizes the cable lengths, which depend on the media type.

When calculating the heat dissipation, you can use the standard energy-to-BTU rating (1 W = 3.41 BTU/hr).

https://t.me/learningnets

Rack Capacity Design Example

Space per Rack

Rack type	Usable RU	Chassis per Rack
6-foot rack	42	Up to 7
7-foot rack	44	Up to 7

Power Requirements per Chassis*

	Power Consumption Under Load (W)
Chassis	417
Per Blade	205
4-Blade Configuration	1237
8-Blade Configuation	2057

* Using two Intel Xeon E5540, 16 GB RAM, two 72-GB drives

Power Requirements per Rack

Chassis per Rack	4-Blade Conf. (W)	8-Blade Conf. (W)
2	2474	4114
3	3711	6171
4	4948	8228
5	6184	10,285
6	7422	12,342
7	8659	14,399

N+1 Power Supply Redundancy

Blade Quantity	PS Quantity
4	42
8	44

© 2012 Cisco and/or its affiliates. All rights reserved.

To calculate the available space per rack, the designer needs information about the size of the rack cabinets that are going to be used. Typically, the RU is used for this purpose. Common rack sizes used are as follows:

- 6 feet, where 42 RU of space is available
- 7 feet, where 44 RU of space is available

With these two options, up to seven Cisco UCS 5108 chassis (6 RU) can be installed per rack.

Secondly, the power requirement per rack is also important in this design because the power is limited.

Based on the power requirements per chassis, the designer can calculate per-rack power requirements for a different number of chassis in the rack as well as the heat dissipation (using the standard energy-to-BTU rating).

Note The power requirements per chassis table lists the values measured for the blade using two Intel Xeon E5540 processors, 16 GB RAM, and two 72 GB disks. The actual numbers may differ in your case.

https://t.me/learningnets

Array Capacity Design

- High-density rack
  - More power per rack—fewer racks per array
- Low-density rack
  - Less power per rack—more racks per array
- Cabling consideration
  - Chassis density per rack
  - Number of uplinks per chassis
- Rack density per Cisco UCS 6200UP FI



3 Chassis per Rack

IOM Uplinks	Uplinks per Rack	Racks per UCS 6248UP Cluster	Racks per UCS 6296UP Cluster
One 10 GE	6	Up to 16	Up to 32
Two 10 GE	12	Up to 8	Up to 16
Four 10 GE	24	Up to 4	Up to 8

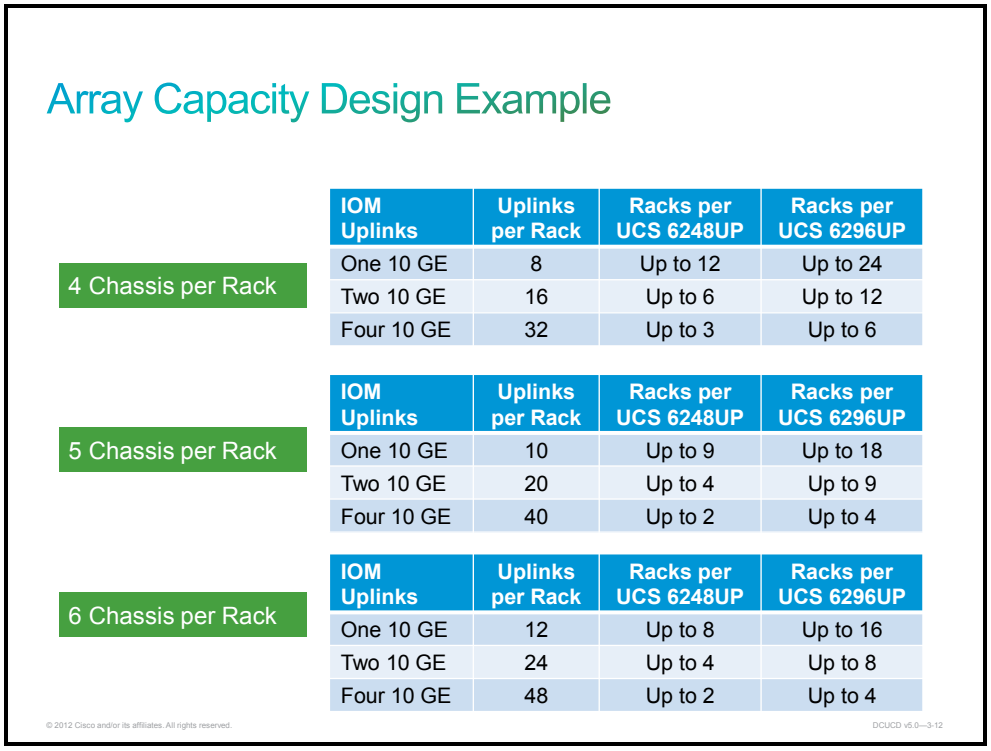
GE = Gigabit Ethernet

© 2012 Cisco and/or its affiliates. All rights reserved.

Next, the rack density per Cisco UCS cluster can be determined. The actual number varies case by case, but in general, it depends on the number of chassis that are supported by the Cisco UCS cluster.

The table in the figure lists an example where up to three chassis per rack are installed. The calculation is done for the maximum number of racks per Cisco UCS 62248UP or 6296UP Fabric Interconnect switches. Note that the number of racks that are calculated in the example also depends on the number of IOM uplinks.

https://t.me/learningnets



When designing the physical deployment, a different number of chassis can be put into a single rack. The three tables seen in the figure list the calculation for four, five, and six chassis per rack design where one, two, or four IOM-to-fabric interconnect uplinks are used.

https://t.me/learningnets

Physical Deployment Example

Cisco UCS 6248UP Cluster

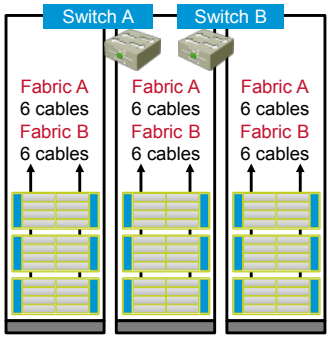
- Total power = 19,213 W
- Total heat = 66,516.33 BTU/hr

	Value
Blades* per chassis	8
Chassis per rack	3
Uplinks per IOM	2
Racks per cluster	3
Ports per fabric	18

Cabling Type	Distance (m)	Placement
Copper (twinax)	1, 3, 5, 7, 10	MoR

	Value
Power per rack	6171 W
Heat dissipation	21,043.11 BTU/hr
Power per UCS 6248UP	350 W
Heat dissipation	1193.5 BTU/hr

* Using Intel Xeon E5540, 16 GB RAM, two 72-GB drives



© 2012 Cisco and/or its affiliates. All rights reserved.

This example describes the physical deployment design using the following input information:

- Chassis with eight blades are used.
- Consumption of a chassis was determined by preliminary measurement and is 205 W per blade server and 417 W per chassis.
- You need to install nine chassis; you can use three 42 RU rack cabinets.
- From an individual chassis, two IOM-to-fabric interconnect uplinks per IOM are used.
- The available power per rack is 10 kW.
- The measured power consumption per Cisco UCS 6200UP Fabric Interconnect switch is 350 W.

Based on the input information, the following design has been proposed:

- Install up to three chassis per rack.
- A single rack requires 6171 W of power (counting only the chassis) and produces 21,002.19 BTU/hr of heat.
- Two fabric interconnect switches will be placed in the middle rack cabinet.
- That rack will need an extra 700 W of power and produce an extra 1193.5 BTU/hr of heat.

Note Installing three chassis plus two fabric interconnect switches in a rack achieves the 10 kW power limit per rack.

- The 5-m twinax cables will be used between the IOM and fabric interconnects.

In total, the whole cluster at maximum load would need 19,213 W of power and produce 66,516.33 BTU/hr of heat.

The power requirements per chassis may differ in your case.

https://t.me/learningnets

Physical Deployment Design Example

- 6-foot rack = 42 RU
- Available power = 10kW per rack
- N+1 power supply redundancy required

	Power under Load [W]
Servers	471
PS-class1 Server	125
PS-class2 Server	125
VS-class1 Server	256
Cisco UCS 6296UP	480



Chassis	Server	Power [W]	Qty.	PS Qty.
BC-class1	6x PS-class1	1221	2	2
BC-class1	6x PS-class1 1x PS-class2	1346	1	2
BC-class1	8x PS-class2	1471	2	2
BC-class2	5x VS-class1	1751	3	2

© 2012 Cisco and/or its affiliates. All rights reserved.

Here is the input information for the physical deployment design:

- The server and chassis design, chassis population, and quantities are known from the Cisco UCS design phase.
- The power consumption was determined in a preliminary measurement and is as follows:
  - 471 W for chassis
  - 125 W for PS-class1 server
  - 125 W for PS-class2 server
  - 256 W for VS-class1 server
  - 480 W for Cisco UCS 6296UP Fabric Interconnect switch.
- The data center will be equipped with 42-RU rack cabinets.
- The maximum amount of power per rack will be 10 kW.
- The Cisco UCS 5108 Server Chassis redundancy must respect the N+1 redundancy scheme.

From the input information, you can calculate the power requirements per chassis class in respect to the number of servers installed. From this information, you can then also determine the number of required power supplies per chassis.

Note The measured power consumption per server, chassis, and fabric interconnect might be different on a case-by-case basis.

https://t.me/learningnets

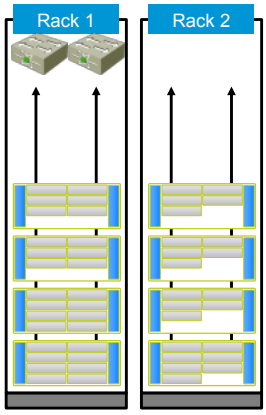
Physical Deployment Design Example (Cont.)

Rack 1	Qty.	Server
BC-class1	2	6x PS-class1
BC-class1	2	8x PS-class2

Cisco UCS 6296XP Cluster

- Total power = 12,943 W
- Total heat = 44,136 BTU/hr

Rack 2	Qty.	Server
BC-class1	1	6x PS-class1 1x PS-class2
BC-class2	3	5x VS-class1



	Rack 1	Rack2
Power	6344 W	6599 W
Heat Dissipation	21633 BTU/hr	22503 BTU/hr

Cabling Type	Length	Placement
Copper SFP+	5 m	MoR

© 2012 Cisco and/or its affiliates. All rights reserved.

When you have determined the individual chassis power requirements, you can then proceed with the design by populating the rack cabinets and determining the number of required racks.

The following design has been proposed:

- Two racks will be used for equipment setup.
- Rack 1 will be populated with four BC-class1 chassis, housing six PS-class1 servers per chassis, and eight PS-class2 servers per chassis.
- Rack 2 will be populated with one BC-class1 chassis (with six PS-class1 and one PS-class2 servers) and three BC-class2 chassis (with five VS-class1 servers).
- The Cisco UCS 6296UP Fabric Interconnect switches will be placed in the first rack.
- The twinax cabling (5 m) will be used for IOM-to-fabric interconnect connectivity.
- The proposed physical setup requires approximately 6.5 kW per rack, which is below 10 kW per rack.

https://t.me/learningnets

Summary

This topic summarizes the primary points that were discussed in this lesson.

Summary

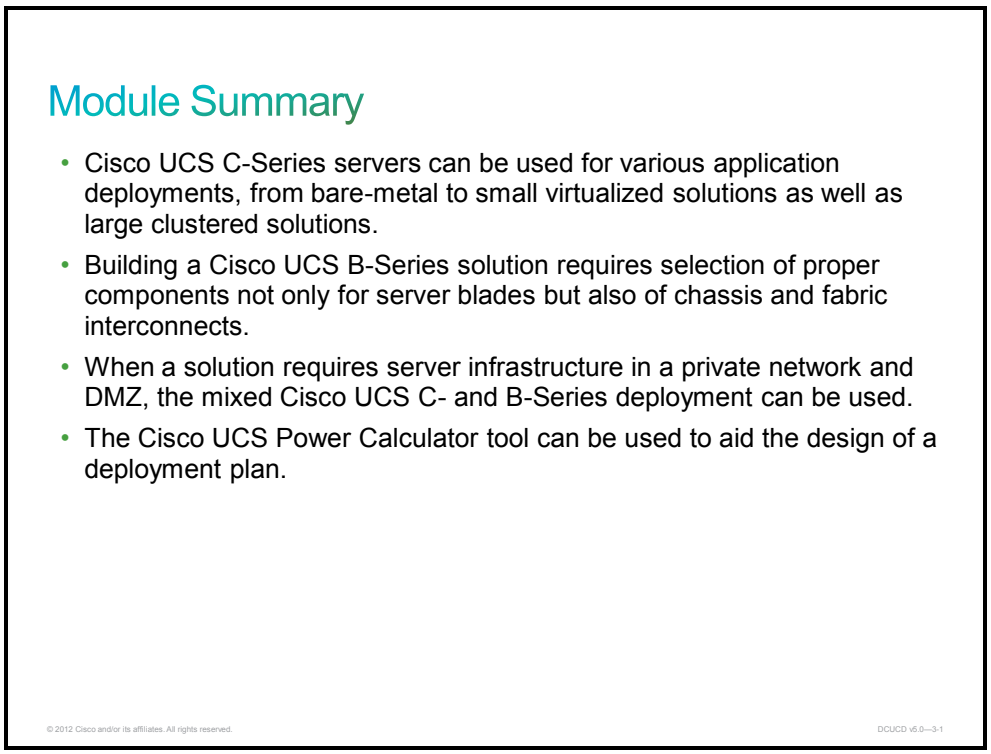
- The Cisco UCS Power Calculator tool is used to calculate the power consumption, required cooling capacity, and weight of the designed Cisco UCS solution.
- Physical deployment design is necessary to determine rack and array densities with regard to available power and space.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Module Summary

This topic summarizes the primary points that were discussed in this module.



References

For additional information, refer to this resource:

- Cisco web pages for Cisco partners at https://communities.cisco.com/community/partner/datacenter

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) What is required to integrate the Cisco UCS C-Series with Cisco UCS Manager? (Source: Sizing the Cisco UCS C-Series Server Solution)

A) Cisco IMC version 1.3
B) EHM
C) Cisco Nexus 2232PP
D) Fibre Channel HBA

Q2) Which two Cisco UCS C-Series components can affect application licensing? (Choose two.) (Source: Sizing the Cisco UCS C-Series Server Solution)

A) network adapter
B) CPU
C) memory size
D) memory speed

Q3) Which network adapter can be used to create 16 Ethernet NICs and two Fibre Channel HBAs? (Source: Sizing the Cisco UCS C-Series Server Solution)

A) P81 VIC adapter
B) Broadcom 5709 adapter
C) QLogic CNA
D) Emulex CNA

Q4) Which option is used to connect the Cisco UCS 5108 chassis to the 6248UP? (Source: Sizing the Cisco UCS B-Series Server Solution)

A) M81KR VIC adapter
B) IOM 2204
C) Cisco IMC
D) Cisco UCS Manager

Q5) What is the maximum number of IOM uplinks that can be used to connect IOM 2208 to 6248UP? (Source: Sizing the Cisco UCS B-Series Server Solution)

A) 1
B) 4
C) 16
D) 8

Q6) What kind of physical connectivity topology can be used to connect 6296UP fabric interconnects and chassis with 1-m twinax cables? (Source: Planning Unified Computing Deployment)

A) ToR
B) MoR
C) EoR
D) direct attachment

https://t.me/learningnets

Module Self-Check Answer Key

Q1) C

Q2) B, C

Q3) A

Q4) A

Q5) D

Q6) A

© 2012 Cisco Systems, Inc.

https://t.me/learningnets