

Data Center Unified Computing Design

Volume 1

Version 4.0

Student Guide

Text Part Number: 97-3024-01



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 of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at 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. (1005R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED "AS IS" AND AS SUCH MAY INCLUDE TYPOGRAPHICAL, GRAPHICS, OR FORMATTING ERRORS. 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.



*Students, this letter describes important
course evaluation access information!*

Welcome to Cisco Systems Learning. Through the Cisco Learning Partner Program, Cisco Systems is committed to bringing you the highest-quality training in the industry. Cisco learning products are designed to advance your professional goals and give you the expertise you need to build and maintain strategic networks.

Cisco relies on customer feedback to guide business decisions; therefore, your valuable input will help shape future Cisco course curricula, products, and training offerings. We would appreciate a few minutes of your time to complete a brief Cisco online course evaluation of your instructor and the course materials in this student kit. On the final day of class, your instructor will provide you with a URL directing you to a short post-course evaluation. If there is no Internet access in the classroom, please complete the evaluation within the next 48 hours or as soon as you can access the web.

On behalf of Cisco, thank you for choosing Cisco Learning Partners for your Internet technology training.

Sincerely,

Cisco Systems Learning

Table of Contents

Volume 1

<i>Course Introduction</i>	<i>1</i>
Overview	1
Learner Skills and Knowledge	2
Course Goal and Objectives	3
Course Flow	4
Additional References	5
Cisco Glossary of Terms	7
Your Training Curriculum	8
Five Levels of IT Certification	8
Eight Different Paths	8
Additional Resources	12
<i>Cisco Unified Computing System Solution</i>	<i>1-1</i>
Overview	1-1
Objectives	1-1
<i>Analyzing Data Center Computing Solutions</i>	<i>1-3</i>
Overview	1-3
Objectives	1-3
Data Center Computing Solutions Overview	1-4
Virtualization Is Changing the Data Center Architecture	1-4
Computing Solutions Challenges	1-8
Architectural and Mechanical Specifications	1-10
Physical Security	1-11
Space	1-11
Power	1-12
Cooling	1-13
Increasing Heat Production	1-13
LAN and SAN Implications	1-17
Cloud Computing Overview	1-19
SaaS	1-22
PaaS	1-22
IaaS	1-23
Evaluate Cloud Computing	1-30
Summary	1-38
<i>Identifying a Cisco Unified Computing System Solution</i>	<i>1-39</i>
Overview	1-39
Objectives	1-39
Cisco Data Center Business Advantage Overview	1-40
Cisco Data Center Business Advantage Network	1-53
Cisco Nexus 7000 Series Switches	1-53
Cisco Nexus 5000 Series Switches	1-54
Cisco Nexus 2000 Series Fabric Extenders	1-54
Cisco Nexus 1000V Series Switches	1-54
Cisco Nexus 1010 Virtual Services Appliance	1-55
Cisco Nexus 4000 Series Blade Switches	1-55
Overlay Transport Virtualization	1-65
Cisco Data Center Business Advantage Storage	1-72
Cisco MDS SAN Switches	1-72
Cisco Nexus 5000 Switches	1-74
Cisco Data Center Business Advantage Compute	1-78
Cisco UCS B-Series	1-81
Cisco UCS C-Series	1-81
Cisco UCS Express	1-82
Cisco Data Center Business Advantage Desktop	1-83
Desktop Challenges	1-83

Desktop Evolution Drivers	1-84
Cisco VXi Overview	1-85
Cisco Data Center Services	1-87
Summary	1-90

Evaluating Server Deployment Options 1-91

Overview	1-91
Objectives	1-91
Server Deployment Options	1-92
Cisco Validated Designs	1-100
Microsoft Hyper-V and Microsoft Exchange 2010	1-101
Exchange 2010 on VMware vSphere and NetApp Storage	1-102
Microsoft SharePoint on Hyper-V	1-102
Server Virtualization with VMware vSphere	1-103
Desktop Virtualization with VMware View	1-103
Oracle Real Application Clusters on the Cisco Unified Computing System with EMC CLARiiON	1-104
Storage	1-104
Cisco Unified Computing System and NetApp Solution for Oracle RAC	1-104
Summary	1-105
References	1-105

Defining a Cisco Unified Computing System Solution Design 1-107

Overview	1-107
Objectives	1-107
Design Process	1-108
Design Deliverables	1-119
Summary	1-122
Module Summary	1-123
Module Self-Check	1-125
Module Self-Check Answer Key	1-126

Assess Computing Solution Requirements 2-1

Overview	2-1
Objectives	2-1

Analyzing Performance Characteristics 2-3

Overview	2-3
Objectives	2-3
Performance Characteristics	2-4
Assess Performance Characteristics	2-9
Summary	2-17

Employing Data Center Reconnaissance and Analysis Tools 2-19

Overview	2-19
Objectives	2-19
Reconnaissance and Analysis Tools	2-20
Purpose	2-21
Analysis Phases	2-21
Inventory Collection	2-22
Utilization and Workload Collection	2-22
Assessment Reports	2-24
VMware Capacity Planner and CapacityIQ	2-29
Microsoft Assessment and Planning Toolkit	2-29
Akorri BalancePoint	2-29
Novel PlateSpin Recon	2-30
VKernel Tools	2-30
Network Instruments Observer	2-30
Employ VMware Capacity Planner	2-31
Leverage Options	2-33
Data Collector	2-34
Data Manager	2-34

Information Warehouse	2-34
Data Analyzer	2-34
Dashboard	2-35
Agent-Free Implementation	2-37
Data Collector	2-37
Information Transfer	2-38
Windows Systems	2-39
Linux and UNIX Systems	2-39
Employ VMware vCenter CapacityIQ	2-46
Employ the MAP Toolkit	2-53
Summary	2-61
Module Summary	2-63
Module Self-Check	2-65
Module Self-Check Answer Key	2-66

Course Introduction

Overview

Cisco Data Center Unified Computing Design (DCUCD) v4.0 is a five-day course that teaches you how to design a Cisco Unified Computing System solution for the data center.

The primary focus of the course is on the next-generation data center platform—the Cisco Unified Computing System (UCS). The course also includes information about Cisco Nexus 7000 Series Switches, Cisco Catalyst 6500 Series Switches, the Cisco Nexus 1000V virtual switch, Cisco MDS switches, and VMware vSphere 4, which are all part of a Cisco Unified Computing System solution.

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

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.

Learner Skills and Knowledge

The following knowledge and skills are recommended prerequisites for students attending this course:

- Cisco CCDA® certification
- Cisco Data Center Networking Infrastructure Design Specialist certification
- Cisco Data Center Storage Networking Design Specialist certification
- Familiarity with Linux or Windows system administration
- Familiarity with server virtualization solutions

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-3

Learner Skills and Knowledge (Cont.)

The following Cisco learning offerings are recommended to fully benefit from this course:

- *Designing for Cisco Internetwork Solutions* (DESGN)
- *Data Center Network Infrastructure Design* (DCNID)
- *Designing Cisco Storage Networking Solutions* (DCSNS)
- Virtualization101 Pre-Course Online Training

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-4

Course Goal and Objectives

This topic describes the course goal and objectives.

“To enable engineers to design scalable, reliable, and intelligent Cisco Data Center Unified Computing System solutions based on the Cisco Unified Computing System, the Cisco data center product portfolio, and contemporary operating systems and virtualization solutions”

Cisco Data Center Unified Computing Design v4.0

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-6

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

- Evaluate the Cisco Unified Computing System solution design process in regard to the contemporary data center challenges, Cisco Data Center Business Advantage architectural framework, and components
- Use the reconnaissance and analysis tools to assess computing solution performance characteristics and requirements
- Describe the hardware components of Cisco UCS and select proper hardware for a given set of requirements
- Describe the basic server deployment model of Cisco UCS
- Propose a Cisco Unified Computing System solution management design for a given environment
- Describe the advanced Cisco UCS server deployment model
- Calculate the ROI and TCO for the solution by using the Cisco UCS ROI tool
- Design a migration plan for an existing implementation to a Cisco Unified Computing System solution

Course Flow

This topic presents the suggested flow of the course materials.

Course Flow

	Day 1	Day 2	Day 3	Day 4	Day 5
A M	Course Introduction	Design Cisco Unified Computing System Solution	Design Cisco Unified Computing System Solution (Cont.)	Design Cisco UCS Management (Cont.)	Evaluate Cisco Unified Computing System Solutions
	Cisco Unified Computing System Solution		Design Server Deployment Design Cisco UCS Management	Design Advanced Server Deployment	Plan Migration to Cisco Unified Computing System Solution
	Lunch				
P M	Assess Computing Solution Requirements	Design Cisco Unified Computing System Solution (Cont.)	Design Cisco UCS Management (Cont.)	Design Advanced Server Deployment (Cont.)	Course Wrap-Up

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-8

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.

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.

Cisco Icons and Symbols

	Cisco Router		Cisco Nexus 7000
	Ethernet Switch		Cisco Nexus 2000 Fabric Extender
	Firewall		Cisco Nexus 5000
	Catalyst 6500		Cisco Nexus 1000V VSM
	Virtual Switching System		Cisco Nexus 1000V VEM

© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0-9

Cisco Icons and Symbols (Cont.)

	MDS 9500		Fibre Channel RAID Storage Subsystem
	MDS 9216		Just a Bunch of Disks (JBOD)
	Director Switch		Fibre Channel Tape Subsystem
	Basic Director-Class Fibre Channel Switch		Disk Array
	Fabric Switch		Tape Storage

© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0-10

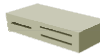
Cisco Icons and Symbols (Cont.)



Blade Server (General)



Cisco UCS 6100 Series
Fabric Interconnect



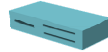
Rack Server (General)



Cisco UCS 5108 Chassis



Server (General)



Cisco UCS C-Series Rack
Server



Cisco UCS
Express

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_\(ITA\)](http://docwiki.cisco.com/wiki/Category:Internetworking_Terms_and_Acronyms_(ITA)).

Your Training Curriculum

This topic presents the training curriculum for this course.



The widely respected IT certification programs that are available through Cisco Certifications bring valuable, measurable rewards to network professionals, their managers, and the organizations that employ them.

Five Levels of IT Certification

Cisco offers five levels of general IT certification: Entry, Associate, Professional, Expert, and Architect, which is the highest level of accreditation within the Cisco Certification program.

Eight Different Paths

There are various tracks—such as Routing and Switching, Network Security, and Service Provider—so that you can match your certification path to your job role or industry.

Cisco Learning Network

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

<https://learningnetwork.cisco.com>



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

Cisco Data Center Specialization Areas

- **Data Center Networking Infrastructure:**
 - Data Center Networking Infrastructure Support Specialist
 - Data Center Networking Infrastructure Design Specialist
- **Data Center Storage Networking:**
 - Data Center Storage Networking Support Specialist
 - Data Center Storage Networking Design Specialist
- **Data Center Unified Computing:**
 - Cisco Data Center Unified Computing Support Specialist
 - Cisco Data Center Unified Computing Design Specialist

© 2011 Cisco Systems, Inc. All rights reserved.

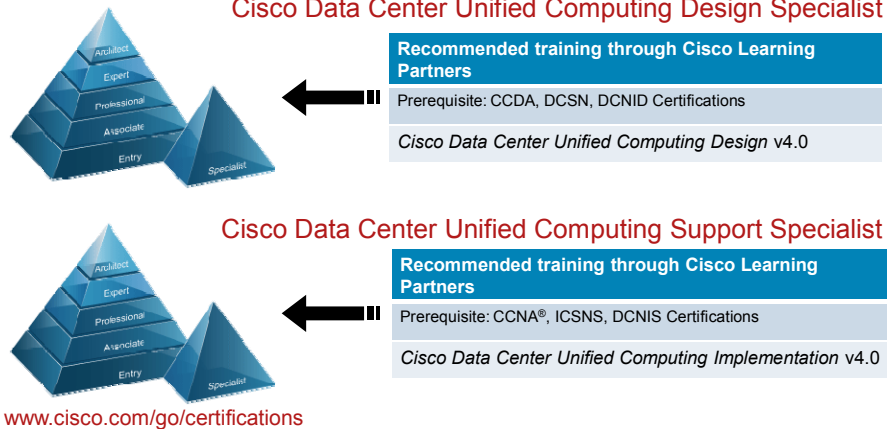
DCUCD v4.0-15

Cisco Specialist certifications demonstrate significant competency in specific technology areas, solutions, or job roles. Individuals who have earned an associate-level certification or higher are eligible to become qualified in these focused areas. With one or more specialist certifications, network professionals can better align their core expertise with current industry needs.

For more information on the Cisco Specialist certifications, visit <http://www.cisco.com/go/certifications>.

Cisco Data Center Unified Computing Specialization Certification Path

Enhance your Cisco Certifications and validate your areas of expertise



The Cisco Data Center Unified Computing Design Specialist certification validates the ability of an engineer to design scalable, reliable, and intelligent data center solutions that are based on the Cisco UCS along with other Cisco data center products, server virtualization software, and server operating systems.

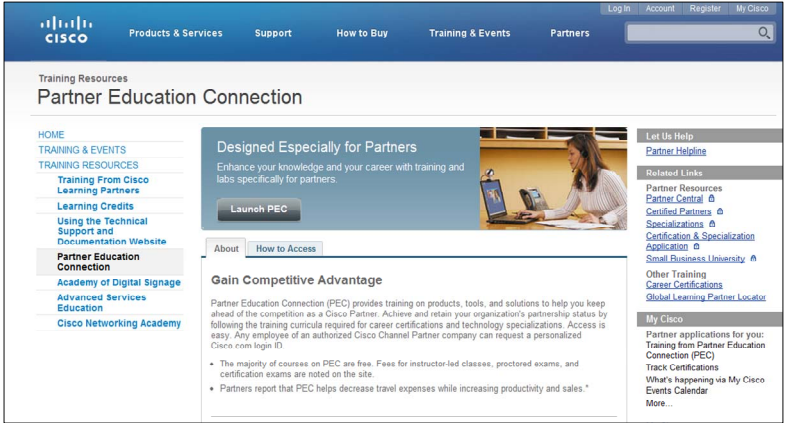
For more information on the Cisco data center-focused certification, visit <http://www.cisco.com/go/certifications>.

Additional Resources

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

Cisco Partner Education Center

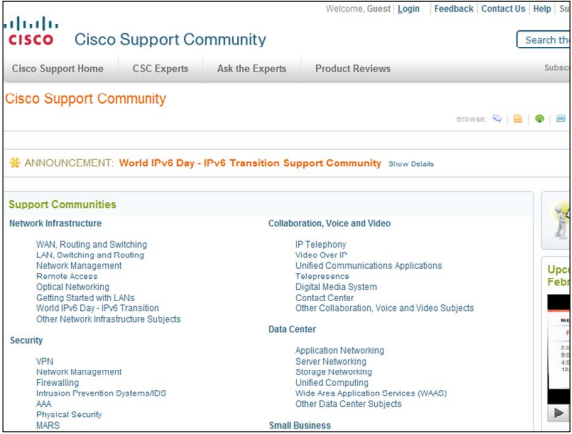
<http://www.cisco.com/go/pec>



© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0-18

Cisco Support Community

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



© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0-19

Cisco NetPro Forums

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

The screenshot shows the Cisco Support Community interface. At the top, there's a navigation bar with links for 'Welcome, Guest', 'Login', 'Feedback', 'Contact Us', and 'Help'. Below this is a search bar. The main header area includes 'Cisco Support Home', 'CSC Experts', 'Ask the Experts', and 'Product Reviews'. The 'NetPro' section is highlighted. On the left, a sidebar lists 'Support Communities' including Home, Network Infrastructure, Wireless, Mobility, Service Providers, Data Center, Collaboration, Voice and Video, Security, Cisco Social, Career Certification, Small Business, and Help. The main content area, titled 'NetPro Communities', displays a table of forum topics with columns for Discussions, Documents, Videos, and Blog Posts.

	Discussions	Documents	Videos	Blog Posts
Network Infrastructure	7	0	0	0
WAN, Routing and Switching	24135	392	0	0
LAN, Switching and Routing	35278	705	1	0
Getting Started with LANs	2033	1	0	0
Network Management	18164	106	2	2
Remote Access	2428	11	0	0
Optical Networking	1051	2	0	0
World IPv6 Day - IPv6 Transition	1	2	0	1

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-20

Cisco Unified Computing System Solution

Overview

This module identifies contemporary data center trends and related challenges, describes the Cisco Data Center Business Advantage architectural framework, and defines the design process for a Cisco Unified Computing System solution.

Objectives

Upon completing this module, you will be able to evaluate the design process for the Cisco Unified Computing System solution in regard to the contemporary data center challenges, the Cisco Data Center Business Advantage architectural framework, and components.

This includes the ability to meet these objectives:

- Analyze the business, technical, and environmental challenges for the contemporary data center solutions
- Provide a high-level overview of the Cisco Business Advantage architectural framework and components within a cloud computing solution
- Evaluate the Cisco UCS deployment options and advantages for the data center solution
- Define the tasks and phases of the design process for the Cisco Unified Computing System solution

Analyzing Data Center Computing Solutions

Overview

This lesson analyzes the business, technical, and environmental challenges for the contemporary data center solutions.

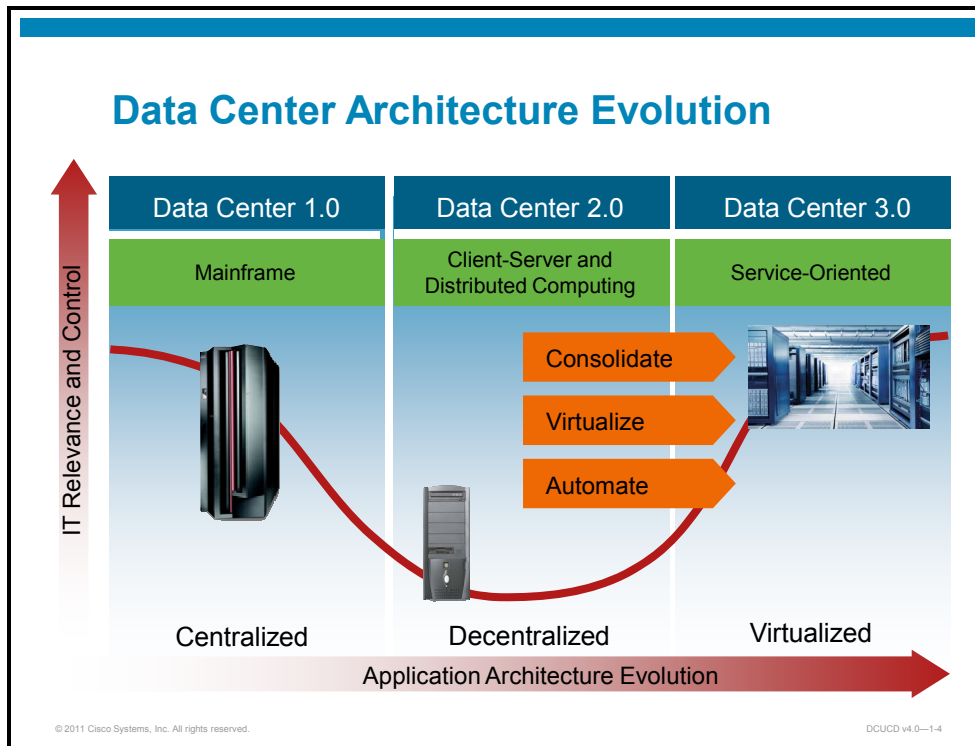
Objectives

Upon completing this lesson, you will be able to meet the following objectives:

- Describe elements of data center computing solutions
- Evaluate business, technical, and environmental challenges for contemporary data center solutions
- Describe cloud computing solutions
- Evaluate cloud computing solutions

Data Center Computing Solutions Overview

This topic identifies solution elements for data center computing.



Data centers have changed and evolved over time.

At first, data centers were monolithic and centralized, employing mainframes and terminals, which the users accessed to perform their work on the mainframe. 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.

Nowadays, with the communication infrastructure being relatively cheaper, and with the computing capacities increased, the 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 much 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).

Virtualization Is Changing the Data Center Architecture

The Cisco Data Center Business Advantage framework brings sequential and stepwise clarity to the data center. Cisco customers are at different stages of this journey.

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

Phase 1: Data Center Networking and Consolidation

Earlier, Cisco entered all the data center transport and network markets—server networking, storage networking, application networking, and security networking. By delivering these types of services as an integrated system, Cisco has improved reliability and validated best-practice architectures, and Cisco has also improved customer success.

Phase 2: Unified Fabric

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

The new atomic unit, or building block, of the data center is not the physical server but is the virtual machine (VM) that 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 market: Cisco Unified Computing. This solution is a simpler, more efficient architecture that extends the life cycle of capital assets, but it 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. It is another choice that simplifies start-up processes and ongoing operations within the virtualized environment, and it can deliver provisioning freedom. This subject will be discussed in more detail later on.

Phase 4: Enterprise-Class Clouds and Utility

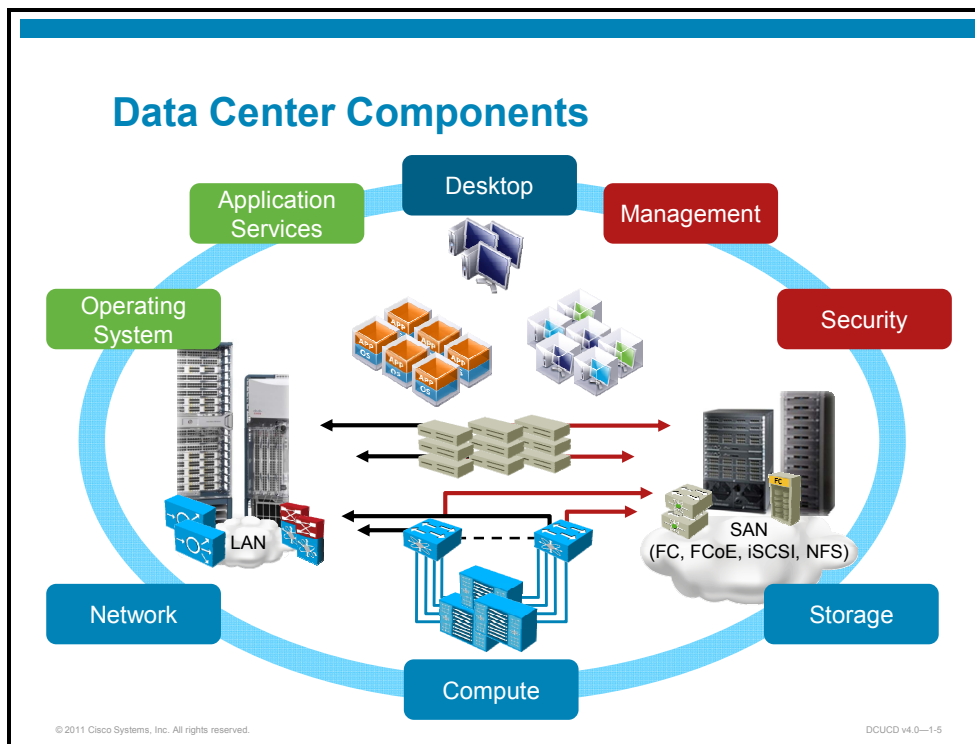
With the integration of cloud technologies, principles, and architectures and the Cisco Unified Computing architecture, workloads can become increasingly portable. Bringing security, control, and interoperability to standalone cloud architectures can enable enterprise-class clouds. The freedom of choice 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: Intercloud Market

A goal of Cisco is to create a market as a new wave of innovation and investment similar to what the industry last saw with the Internet growth of the mid-1990s, but this time the growth should be predicated not on addressing a federation across providers but now on portable workloads. This market extends from the enterprise to the provider, and from the provider to another provider based on available capacity, power cost, proximity, and so on.



Contemporary data center computing solutions encompass multiple aspects and technologies:

- Network technologies and equipment like intelligent switches, multilayer and converged devices, high-availability mechanisms, and Layer 2 and Layer 3 protocols
- Storage solutions and equipment that cover 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 like disk arrays and tape libraries
- Compute technologies and equipment, including general purpose and specialized servers
- Operating system and server virtualization technologies
- Application services technologies and products like load balancers and session enhancement devices
- Management systems that are used to manage network, storage, and compute 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

Other Data Center Aspects

- Physical facility
- IT organization



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v1.0-1-8

Apart from the mentioned aspects and components of the data center solution, there are two very important aspects that also influence how the solution is used, scaled, and defined during its lifetime.

The first aspect is the physical facility, which encompasses all the data center facility physical characteristics that affect the infrastructure. The physical facility includes available power, cooling capacity, physical space and racks, physical security, and fire prevention systems, and so on

The second aspect is the IT organization that defines 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 the IT aspects (typically with the help of external IT partners) or in large companies in the form of multiple departments where each takes care of a subset of the data center infrastructure.

Computing Solutions Challenges

This topic evaluates the business, technical, and environmental challenges for 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, physical space
- Resource utilization, provisioning, repurposing
- Security threats
- Business continuance



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-8

The modern enterprise is being changed by shifting business pressures and operational limitations. While enterprises get ready 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 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 has been rising as well, because more services are concentrated in a single data center. If an attack occurred 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.



The data center is viewed from different perspectives, depending on the position of the person whose view is being expressed.

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, who all have 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.

Facility Challenges

Architectural and mechanical specifications

- Space available
- Load capacity
- Power (electrical) capacity
- Cooling capacity
- Cabling infrastructure

Limited capacities

Compliance and regulations

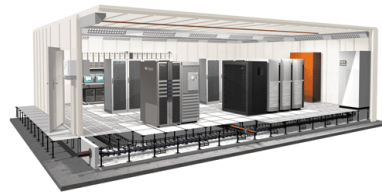
New versus existing solution or facility

Physical security

- Access to the premises
- Fire suppression

Environmental conditions

- Operating temperature
- Humidity level



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-10

The data center facility has multiple aspects that need to be properly addressed (that is, taken into account when the facility is being planned, designed, and built).

Facility capacities are limited and need to be properly designed.

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.

When it comes to 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 are defined as follows:

- How much space is available
- How much load that a floor can bear
- The power capacity that is available for data center deployment
- The cooling capacity that is available
- The cabling infrastructure type and management

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

Physical Security

Finally, 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 deployed optimally.

Although sometimes neglected, the size of the data center has a great influence on cost, life span, 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 nonserver, 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.

Properly sized data center facilities also take into account the placement of equipment—if properly selected, the data center facility can grow when needed. Otherwise, costly upgrades or relocations must be performed.

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

- 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

Power and Cooling Challenges

Power used for the following:

- Servers, storage, network
- Lighting
- Cooling
- Conversion loss

Redundancy

Space-saving servers produce more heat:

- Better computing-to-heat ratio
- More servers deployed

Increased computing and memory power results in more heat.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-11

Power

The power in the data center facility is used to power servers and also storage, network equipment, lighting, and cooling devices (which take up most of the energy), and some power is “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 computing power of the server must work harder, the greater the power that has to be drawn from the AC supply and the greater the 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 be made to accommodate 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.

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 kilowatt (kW) 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 cooling aspect of the facilities must have proper airflow to reduce the amount of heat that is generated by concentrated equipment. Adequate cooling equipment must be used for more 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 value 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 3.412 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 a lot of heat.

In addition, the increasing computing and memory power of a single server results in higher heat production.

Thus, the blade server deployment results in more heat being produced, which requires proper cooling capacity and proper data center design. The solutions that address the increasing heat requirements must be considered when blade servers are deployed within the data center. The design must also take into consideration the cooling that is required for the current sizing of the data center servers, but the design must anticipate future growth, thus also taking into account future heat production.

If cooling is not properly addressed and designed, the result is a shortened equipment life span.

The cooling solution can resolve the increasing heat in these ways:

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

Cabling Challenges

Passive networking infrastructure

- LAN and SAN physical network infrastructure

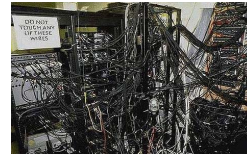
Cabling characteristics:

- Type (copper versus fiber optics)
- Length

Cabling **should not**

- Restrict air flow
- Hinder troubleshooting
- Create unplanned dependencies
- Incur accidental downtime

Wrong



Right



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-12

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 one another 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 offers greater distances while copper-based cabling is ubiquitous 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 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.

The cabling infrastructure must not incur 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 underfloor 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 that are located above the rack for connectivity. 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 proper data center operation and life span. Conversely, poorly designed cabling will incur downtime due to reconfiguration or expansion requirements that were not considered by the original cabling infrastructure shortcomings. Cable management is a major topic in its own right. 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 any new or reconfigured environment in the data center.

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, SAN) attach rates are required.

- Multicore processor deployment affects virtualization and networking requirements.

Virtualization solution infrastructure management

- Management demarcation



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-13

Virtualization, though being the “promised solution” for server-, network-, and space-related problems, brings in 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, brings in increased equipment and network complexity.
- **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 VMs, the information about the VM connection point gets lost. If the information is lost, the configuration of the VM access port does not move with the machine.

Third, by using the virtualization, the servers, network, and storage facilities are under increased loads.

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 properly addressed.
- Multiple Fibre Channel interfaces provide SAN connectivity for storage traffic to allow servers and consequently 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.
- Having multiple adapters increases management complexity. More management effort is put into proper firmware deployment, driver patching, and version management.

Cloud Computing Overview

This topic identifies cloud computing solutions.

Cloud Computing Overview

IT resources and services are **abstracted** from the underlying infrastructure.

They are provided in a **multitenant environment**:

- On-demand provisioning
- Infinite scalability appearance
- Pay-as-you-go billing
- Self-service model

A cloud could be off-premises, hosted model or application hosting, and so on.

© 2011 Cisco Systems, Inc. All rights reserved.DCUCD v4.0—1-15

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-premise, hosted model.

Currently, cloud computing is still new. There are no definite and complete standards yet, because the standards are in the process of being defined.

There are some cloud solutions that are available and offered but not at the scale and ubiquity as anticipated in the future.

One could argue that a component cloud is very much like a virtualized data center. This description is usually true. However, there are some details that differentiate a component cloud from a virtualized data center. One of those details is on-demand billing. Typical to a “cloud” is that resource usage is tracked at a granular level and billed to the customer on a short interval. Another difference could be that, in a cloud, associations of resources to an application are more dynamic and ad hoc than in a virtual data center. However, these differences and definitions are very fragile. Rather than trying to define these differences, it is more useful to focus on the value than on the concepts.

Terms and Characteristics

Cloud types

- Public cloud
- Private cloud
- Federation
- Virtual private cloud
- Open cloud

The solution depends on the vendor.

Common standards have not been defined yet.

Architecture

Characteristics

- Multitenancy and isolation
- Security
- Automation
- Standards

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-16

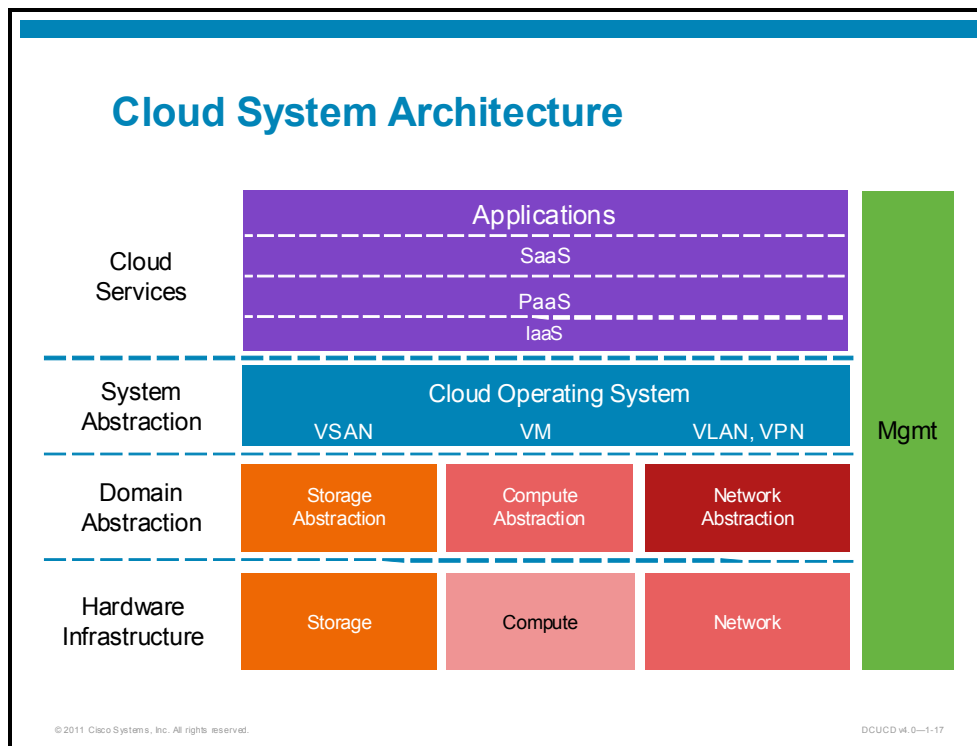
The terms in the figure are used when talking about cloud computing in order to define the type of cloud solution that is used or to be used. These terms are defined in more detail in the following sections.

The concept of cloud services will evolve toward something that ideally hides complexity and allows control of resources, while providing the automation that takes away the complexity.

When talking about cloud computing, the architecture defines the fundamental organization of a system, embodied in its components, their relationship to each other, and the principles governing the design and evolution of the system.

The list here describes the fundamental characteristics of a cloud solution:

- **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 ability is implicitly included in the previous characteristic. 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 also is anything that secures the infrastructure of the cloud itself.
- **Automation:** This feature is an important characteristic that defines how a company that would like to use a cloud-based solution 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.



The system architecture of a cloud is like the architecture of a data center. The cloud has all the components and aspects that a typical data center has.

Cloud Computing Service Categories

Software as a Service (SaaS)

- Applications services are delivered over the network.

Platform as a Service (PaaS)

- Software development frameworks and components are delivered over the network.

Infrastructure as a Service (IaaS)

- Compute, network, and storage services are delivered over the network.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-18

There are different 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. 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.

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

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-19

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:

- Google services and applications like 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

Private Cloud

Enterprise IT infrastructure with cloud computing qualities:

- Self-service
- Pay-as-you-go chargeback
- 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**

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-20

A private cloud is an enterprise IT infrastructure that is managed with cloud computing qualities, such as self-service, pay-as-you-go chargeback, on-demand provisioning, and the appearance of infinite scalability.

Scaling a Private Cloud

Initially

- One or two noncritical application systems

Later

- Combine IT resources under single cloud operating system or management platform
- Provide tens to thousands of applications and services
- Enable new architectures for very large-scale compute activities

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-21

The private clouds are those that will have the following characteristics:

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

Once a company has decided that a private cloud service is good, the private cloud will scale by pooling IT resources under a single cloud operating system or management platform. It can support anywhere from tens to thousands of applications and services.

This arrangement will enable new architectures to target very large-scale activities.

Federation

A federation of clouds links disparate cloud infrastructures with one another:

- Connect and aggregate management and billing
- Allow exchange and leverage of resources

The federation can occur **across data center and organization boundaries** with cloud internetworking.

The federation can provide a unified infrastructure, provisioning, and metering and billing.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-22

Once the private and public clouds are well accepted, the tendency to connect them will form—a federation of clouds.

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

The federation can enable the following 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.

Virtual Private Cloud

A virtual private cloud is a service offering that allows enterprises to **extend their private clouds to include 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.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-23

A virtual private cloud is a service offering that allows enterprises to extend their private clouds in order to include infrastructure services that are provided by the service provider.

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

- Leverage virtual port channel (VPC) services that are offered by 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

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

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-24

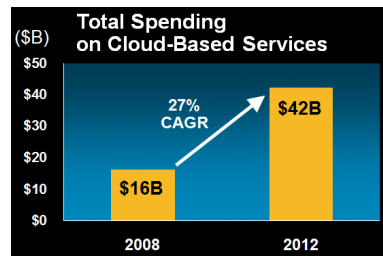
The largest and most scalable cloud computing system—the open cloud—is a service provider infrastructure that allows a federation with a similar infrastructure offered by other providers. Enterprises can choose freely among participants, and service providers can leverage other provider infrastructures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

Evaluate Cloud Computing

This topic evaluates cloud computing solutions.

Cloud Computing Trends

Approximately 50 percent of services spending is defined as SaaS. The five segments of enterprise IT include business applications, application development software, system software, storage, and servers.



Source: IDC, October 2008

© 2011 Cisco Systems, Inc. All rights reserved.

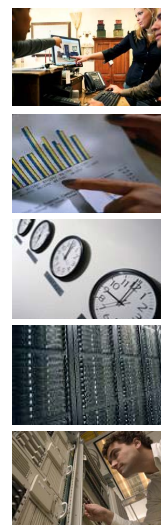
DCUCD v4.0-1-26

Cloud-service fees currently make up 4 percent of enterprise IT spending and are projected to grow to 9 percent of total spending by 2012. Cloud services are projected to grow at 27 percent, or five times the rate of current enterprise IT spending, according to IDC.

According to Gartner Group, by 2012 14 percent of the infrastructure and operations architectures of companies in the Fortune 1000 are expected to be in private cloud systems.

Cloud Computing Aspects

- ✓ Scalability
 - ✓ On-demand availability
 - ✓ Efficient utilization
 - ✓ CapEx to OpEx
 - ✓ Pay-per-use model
 - ✗ No standard provisioning
 - ✗ No standard metering and billing
 - ✗ Limited customer control
 - ✗ Security
 - ✗ Offline operation
- Flexibility
- Cost



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-27

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 just 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 the 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 used physically exist and has no control over how they are used.

- 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.

Open Cloud Benefits

Enterprises choose freely among service providers.

Service providers can leverage a federation cloud to manage exceptional loads.

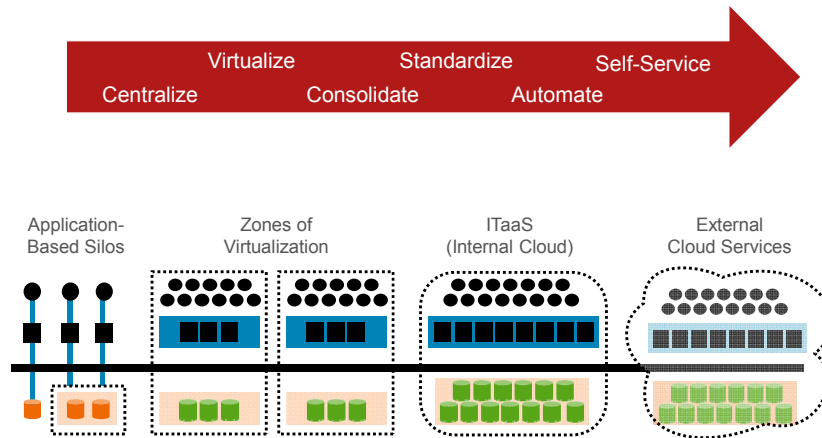
© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-28

Service providers that offer open clouds can provide an open, on-demand marketplace for infrastructure services.

An open-cloud service decouples the one-to-one business relationships with service providers by using intercloud standards and service, and it utilizes standards for consolidated application and service management and billing.

Transition to the Cloud



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-29

The transition to cloud-based computing is an evolution that is characterized by different architectures and steps. The focus should be on helping customers to move to the right side of the figure and build their own dynamic data centers.

The evolution from the current infrastructure to cloud-based computing will proceed in phases. The application silos are where most of the companies are now or they are already using virtualization, deployed in zones like infrastructure applications, per department, and so on

Next, the shared infrastructure, forming an internal cloud, will emerge from the virtualization zones.

Finally, the use of external cloud services will become more popular. The service providers in the cloud market are already on the far right of the scheme.

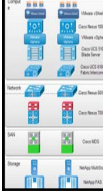
The customers will have more than one architecture (for example, four) and will likely be moving to a more cloud-oriented IT structure.

Cloud computing involves solutions and technologies from multiple vendors:

- **VMware:** Consolidation and agility in the hypervisor
- **Cisco:** Unified compute (stateless compute) and network virtualization, security, and unified fabric
- **NetApp:** Unified storage and virtualized and secure storage (array offload—data protection and cloning)

Cisco Cloud-Enabling Solutions

Network-centered virtualized data center

 Vblock Packages 30 racks reduced down to three racks Provisioning applications in hours instead of weeks	 Secure Multitenancy Securely sharing servers between multiple users and groups without having to add another server
 Cloud-in-a-Box Hundreds of servers in a broom closet with web self-service provisioning and automated application deployment, all in one box	 Virtual Desktop More than 4000 desktops in a single rack Savings of up to 60% or more per PC per year Significant savings in operations

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-30

Cisco already has solutions and designs that were created with cloud computing in mind:

- **Vblock infrastructure packages:** These packages are the integrated best-of-breed packages from Cisco and EMC, together with VMware. They have been engineered, tested, and validated to deliver revolutionary TCO and pervasive virtualization at scale in the most demanding uses.
- **Secure multitenancy:** This validated design is for end-to-end secure multitenancy. It isolates applications across networks, servers, and storage facilities. It also separates confidential information among business units, customers, departments, or security zones.
- **Desktop virtualization:** This solution is an element of the Cisco Virtualization Experience Infrastructure (VXI) system vision and architecture. Cisco VXI leverages the following technology and services: thin clients, the Virtual Desktop Infrastructure (VDI) solution, server virtualization, and compute, storage, and network services.
- **Cloud-in-a-box:** This bundled Cisco UCS solution is provided as a starter kit.

Cisco Virtualized Multi-Tenant Data Center

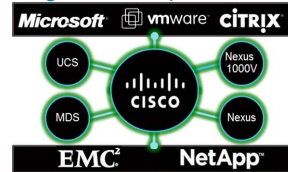
An architectural approach to solving IT challenges

- Tested and validated solutions with prescriptive design guidance

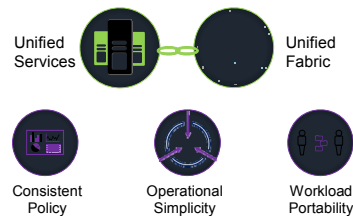
Benefits

- Efficient, flexible, SLA-driven IT infrastructure with automated compliance
- Open, extensible architecture that accelerates application deployment and technology adoption
- Increased IT efficiency and business productivity

Integrated Compute Stacks



Data Center Networking and Services



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-31

The Cisco Virtualized Multi-Tenant Data Center (VMDC) version 2.0 provides design and implementation guidance for enterprises that are planning to deploy private cloud services and for service providers that are building virtual private and public cloud services.

Cisco VMDC v2.0 is a validated architecture that delivers a highly available, secure, flexible, and efficient data center infrastructure. It provides the following benefits:

- **Reduced time to deployment:** It provides a fully tested and validated architecture that accelerates technology adoption and rapid deployment.
- **Reduced risk:** It enables enterprises and service providers to deploy new architectures and technologies with confidence.
- **Increased flexibility:** It provides rapid, on-demand, workload deployments in a multitenant environment due to a comprehensive automation framework with portal-based resource provisioning and management capabilities.
- **Improved operational efficiency:** It integrates automation with a multitenant resource pool (compute, network, and storage), improves asset use, reduces operational overhead, and mitigates operational configuration errors.

Cloud Computing-Based Service Example

IT as a Service (ITaaS)

- Virtual desktop infrastructure
 - Thin clients used to access the desktops
 - Central software installation and configuration
 - Central security management
 - Automatic updates and upgrades
- Software and hardware leasing and maintenance
- Backup services
- Remote and local support



An example of a cloud computing-based service is shown in the figure.

Modern companies in the competitive space need a reliable, functioning, and customized IT system, regardless of their fields of activity. The establishment of an adequate information infrastructure is often a challenge for companies, taking much time and energy, and the process is also cost-intensive.

An IT-as-a-Service (ITaaS) solution would offer complete IT outsourcing to the company whose primary business is not of IT actually being an Everything as a Service (XaaS) solution. ITaaS would offer new, unique, and previously unimaginable business opportunities for service providers.

Such a solution is based on the integration and virtualization of computer resources and support services that are tailored to the user. The company does not need to buy these services and equipment but can simply rent hardware, software, and all the necessary IT services (for example, telephony, Internet access, email, and so on). Each month the company would pay minimal costs that are known in advance.

Summary

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

Summary

- Data centers have to address growing demands despite their operational limitations.
- The challenges for data centers include resource utilization, organizational management, service availability, solution scalability and management, and expense containment.
- Cloud computing provides on-demand resources in a pay-as-you-grow manner.
- Cisco already has solutions and designs that were created with cloud computing in mind, such as Vblock infrastructure packages, secure multitenancy, desktop virtualization, and cloud-in-a-box.

Identifying a Cisco Unified Computing System Solution

Overview

This lesson provides a high-level overview of the Cisco Data Center Business Advantage architectural framework and components within a cloud computing solution.

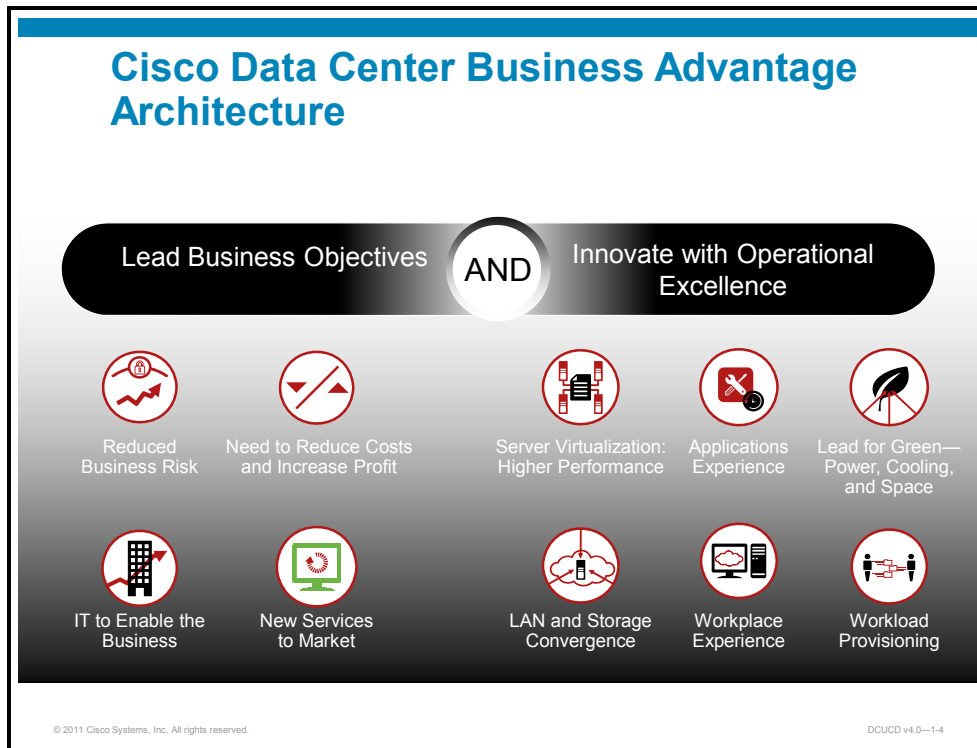
Objectives

Upon completing this lesson, you will be able to meet the following objectives:

- Describe the Cisco Data Center Business Advantage architectural framework
- Describe the Cisco Data Center Business Advantage architectural framework network component
- Describe the Cisco Data Center Business Advantage architectural framework storage component
- Describe the Cisco Data Center Business Advantage architectural framework compute component
- Describe the Cisco Data Center Business Advantage architectural framework desktop component
- Describe Cisco Data Center Services

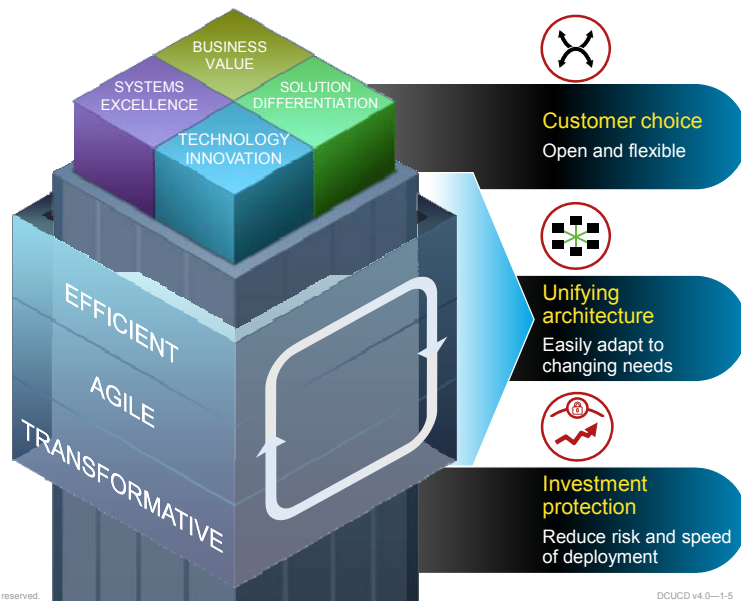
Cisco Data Center Business Advantage Overview

This topic describes the Cisco Data Center Business Advantage architectural framework.



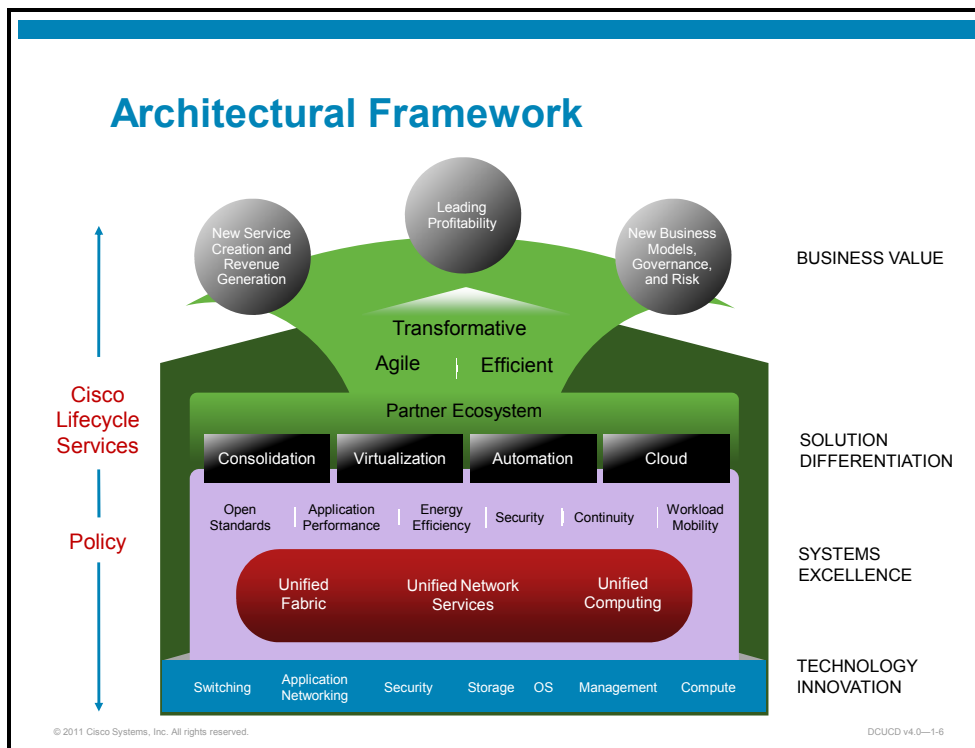
Cisco Data Center Business Advantage represents a fundamental shift in the role of IT into becoming a driver of business innovation. Businesses can create services faster, become more agile, and take advantage of new revenue streams and business opportunities. Cisco Data Center Business Advantage increases 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.

Cisco Data Center Business Advantage Architecture Proposition



The Cisco Data Center Business Advantage framework can deliver an open standard and flexible architecture. This architecture can be adapted easily for changing business and market dynamics.

Cisco Data Center Business Advantage preserves your ability to work with other industry leaders to solve a particular IT or business problem. Because Data Center Business Advantage is built around an ecosystem-centric framework and open solution stacks, you can easily work with your existing data center vendors and the Cisco community of development, design, and deployment partners.



Cisco Data Center Business Advantage 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 Business Advantage 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 a quicker time to market.

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

Benefits

The architectural framework for dynamic networked organizations helps with the following:

- Reduced total cost of ownership
- Accelerated business growth
- Extend the life of current infrastructure life making data center more efficient, agile, and resilient

Cost reduction and increased efficiency leverages

- Infrastructure consolidation
- Energy efficiency
- Business continuance
- Workforce productivity

© 2011 Cisco Systems, Inc. All rights reserved.

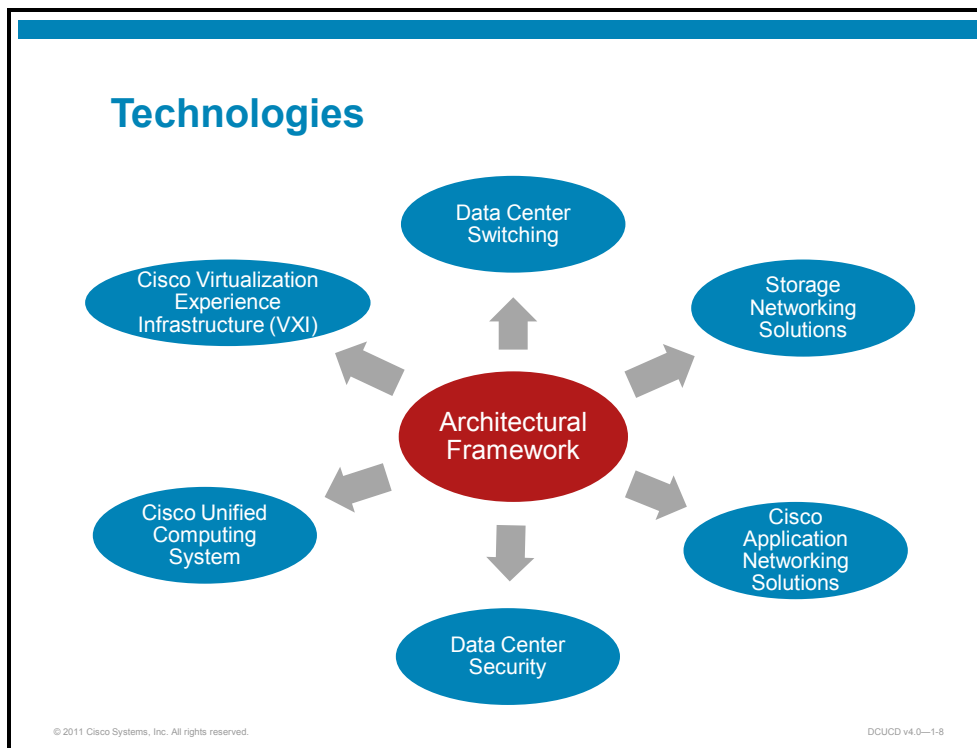
DCUCD v4.0—1-7

The Cisco Data Center Business Advantage 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

Data Center Business Advantage 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 current infrastructure by making your data center more efficient, agile, and resilient



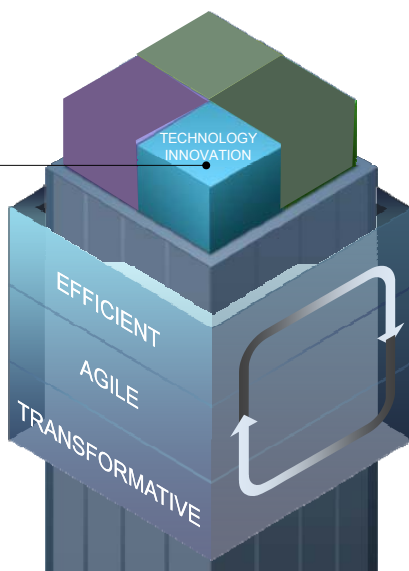
The framework brings together multiple technologies:

- **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.

Components of the Architecture

TECHNOLOGY INNOVATION

- Comprehensive product portfolio
- Standards-based
- Platform choice—dedicated, virtual, appliance
- Significant R&D
- Long-term investment protection

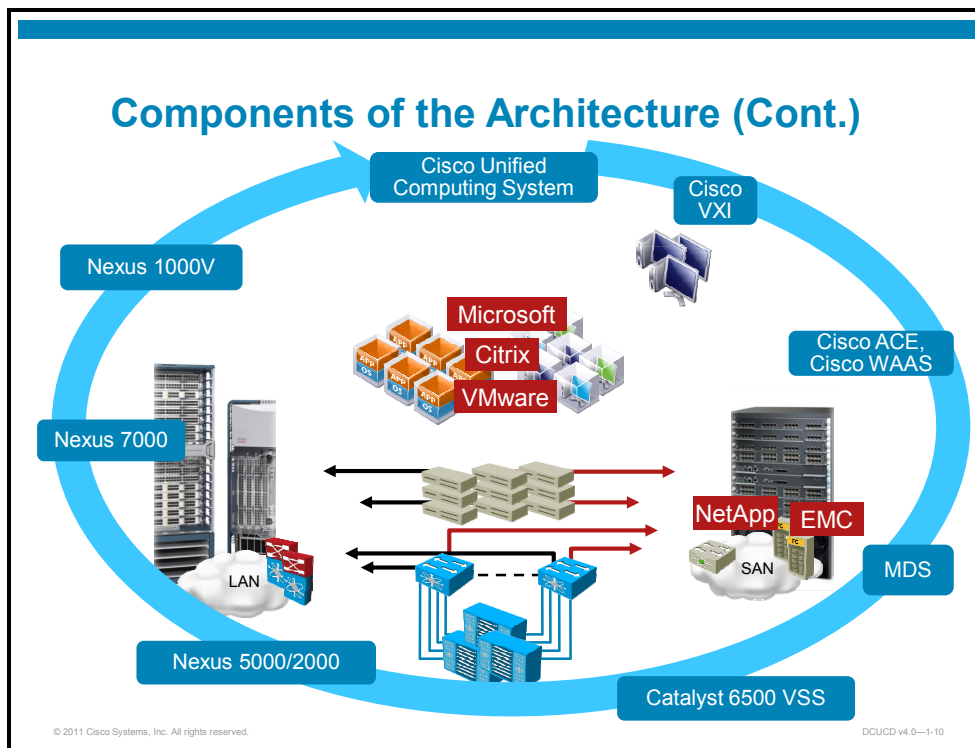


Technology innovations provide the means to transform the data center through continuous standards-based information. Customers are provided with a choice of platforms—appliance, integrated, or virtual—so that they can quickly deliver new capabilities or enter new markets.

Today, IT organizations are being asked to help to cut costs, improve productivity, and provide new ways to interact with customers. Data centers are at the heart of this IT infrastructure and need to be able to adopt new technologies that provide consolidation, virtualization, and automation to help meet customer demands.

IT is central to core business processes, and data centers are taking a central role in the overall success of an organization. A data center must be able to excel in the traditional measures of operational excellence, including system availability, asset utilization, security, and power and cooling efficiency.

In addition, data centers must support a new set of emerging requirements such as business-to-business relationships, regulatory requirements, a global operating environment, and empowered users. Data centers must also be able to adapt quickly and effortlessly to inevitable change.



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.

- **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.

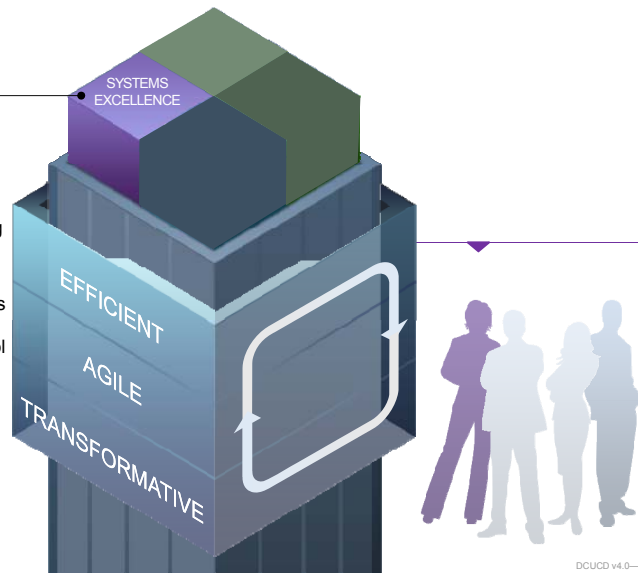
The Cisco UCS provides the following benefits:

- 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

Components of the Architecture (Cont.)

SYSTEMS EXCELLENCE

- Integrated systems simplify design, deployment, and management
- Scalable, flexible building blocks accelerate new service creation
- Open standards and APIs mean that the customer retains choice and control

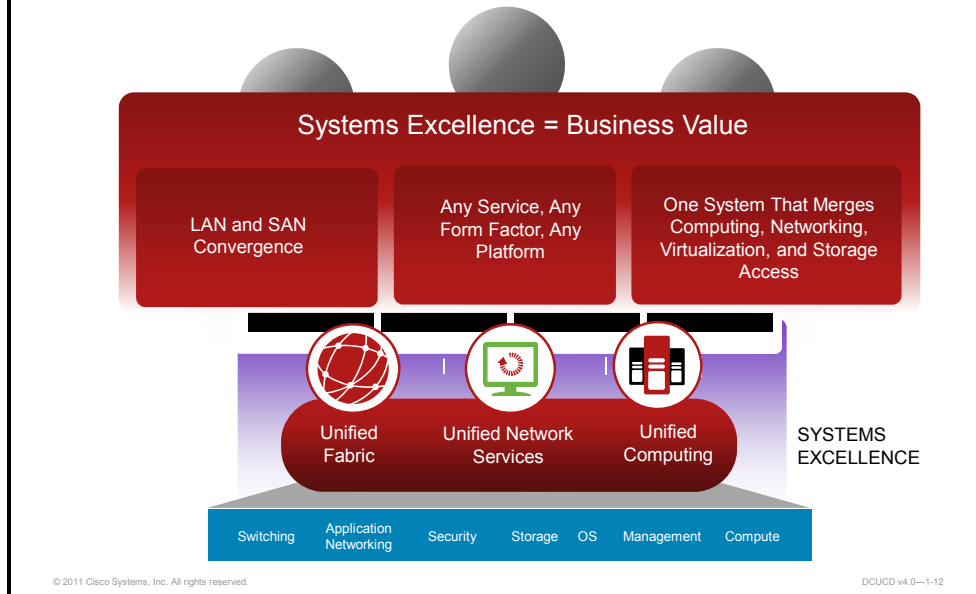


© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-11

Systems excellence is about being able to dramatically reduce complexity, and having a data center that is agile, efficient, and capable of responding to rapidly changing business demands. The Cisco approach for consolidation and virtualization can reduce costs, simplify management, and provide the infrastructure in the data center. With open application programming interfaces (APIs) and standards, scalability and application delivery can be improved.

Components of the Architecture (Cont.)



Unification is the key to the systems excellence layer. Unification can be achieved through LAN and SAN convergence or by bringing together the silos that are traditionally generated from a network storage and compute platform.

The Cisco Unified Fabric integrates storage and data networking to delivery by providing the following:

- Transparent convergence
- Three-dimensional scalability
- Sophisticated intelligent services

The primary benefits of the Cisco Unified Fabric include the following:

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

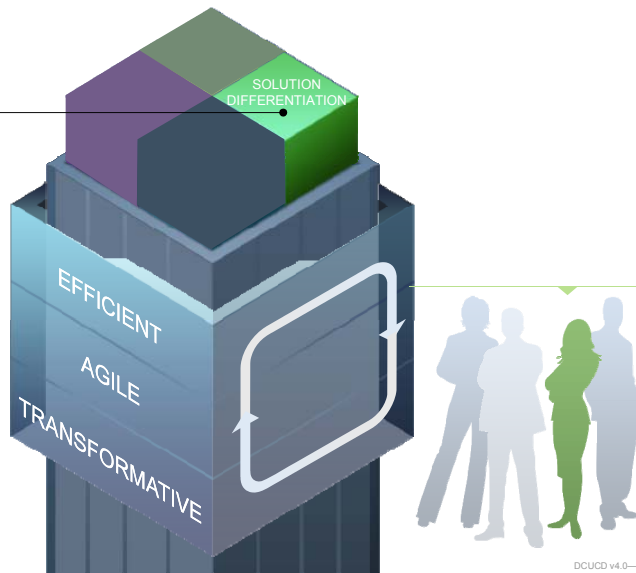
Cisco Unified Network Services can provide a way to unify how the application delivery and security services are provided in the data center network or computing environment. This solution provides consistent and flexible service delivery by using policy-based provisioning across the following:

- **Any service type:** Improving application availability, security, acceleration, and performance monitoring
- **Any form factor:** Supporting network services on dedicated Cisco applications, embedded modules, and VMware virtualized platforms
- **Any environment:** Delivering services from the network core edge or computing edge, or from both

Components of the Architecture (Cont.)

SOLUTION DIFFERENTIATION

- Proven solutions that meet specific needs and that allow rapid provisioning and deployment
- Integrated offers include infrastructure, services, and support
- Open solution stack for choice and flexibility
- Cisco Validated Designs = blueprint for success



© 2011 Cisco Systems, Inc. All rights reserved.

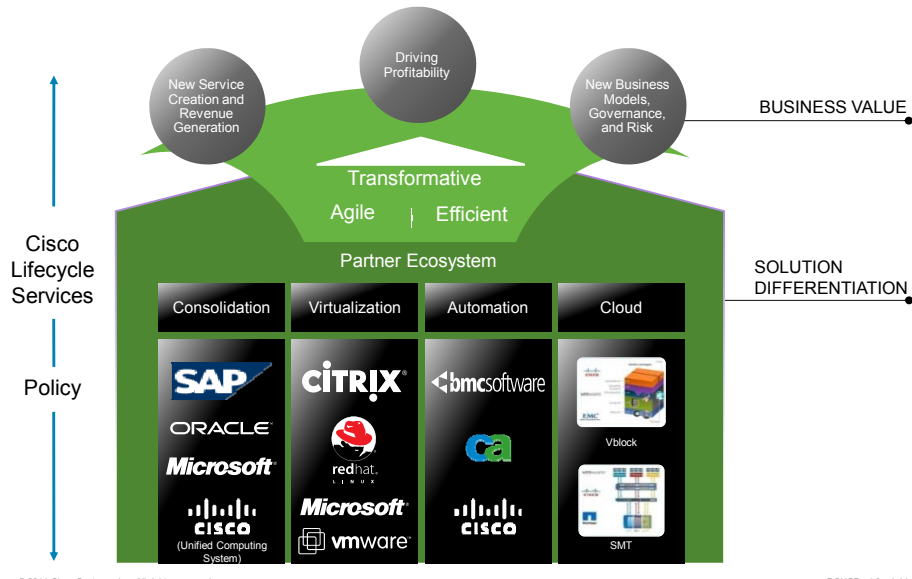
DCUCD v4.0—1-13

Differentiated solutions offer complete holistic offerings to address the IT and business needs of customers together with the Cisco partner ecosystem. These offerings encompass the following:

- Consolidation
- Virtualization
- Automation
- Cloud enablement services
- Partner ecosystem

Cisco has developed validated designs and reference architectures that have been preintegrated and tested to meet the demands of businesses.

Components of the Architecture (Cont.)

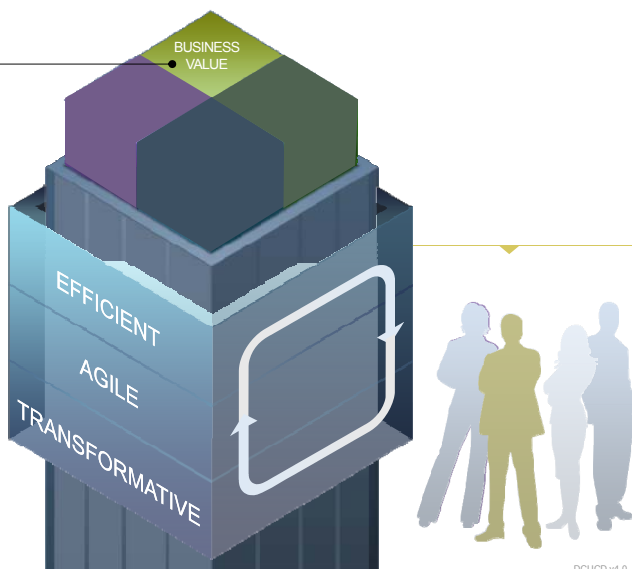


Solution differentiation is a focal part of this framework. The figure shows what is currently available today to provide this differentiation.

Components of the Architecture (Cont.)

BUSINESS VALUE

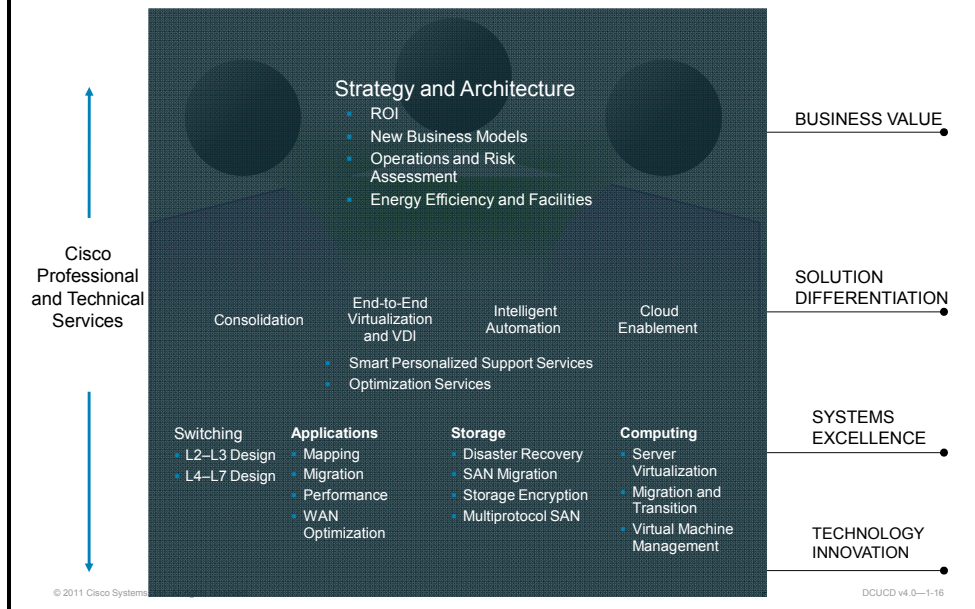
- New IT and business models to lead transformation
- Business agility leading new service creation and accelerating revenue
- Support for new customer growth and lower TCO to improve profitability



The Cisco Data Center Business Advantage architecture provides business value and addresses concerns such as risk mitigation and the ability to respond to the ever-changing demands of businesses.

The Cisco data center vision is evolving to help build an agile and efficient business structure in order to transform the data center from a cost center into a competitive solution.

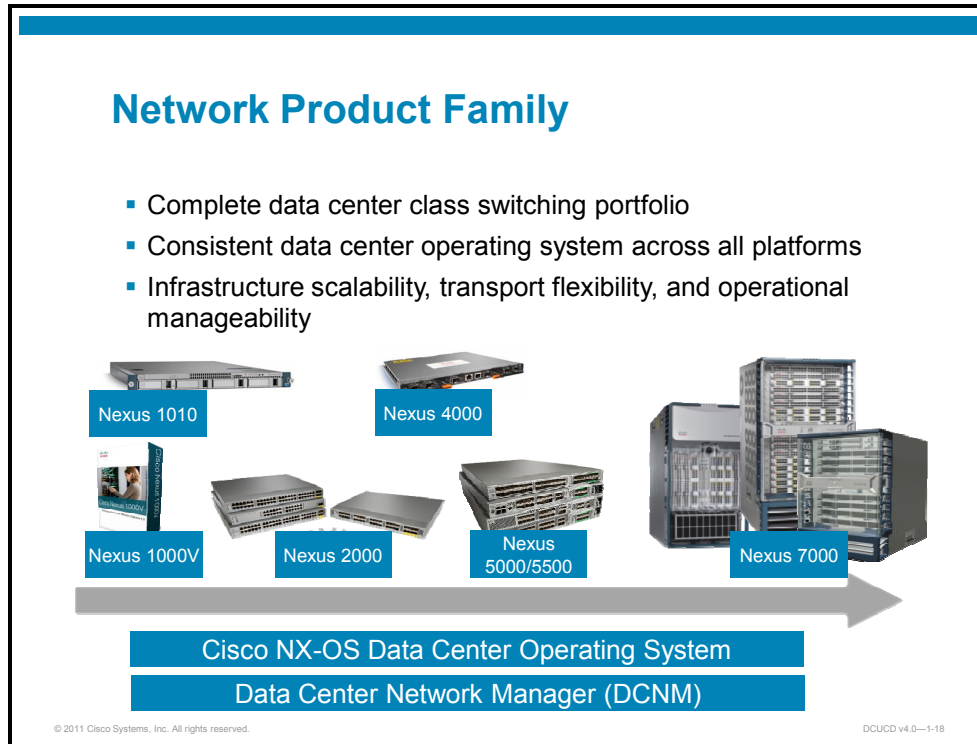
Components of the Architecture (Cont.)



The foundation of technology, systems, and solutions provides businesses with a tangible business value and results. In addition, a comprehensive portfolio of services can help you to plan, build, and run a virtualized data center. At each layer of the framework, there is a corresponding supporting service to help your business get the most from its capital investment.

Cisco Data Center Business Advantage Network

This topic describes the Cisco Data Center Business Advantage architectural framework network component.



Most of the network equipment is capable of resource consolidation and virtualization. Services are integrated among the equipment, having a complete solution offering security of virtualized services and virtualized security devices, application and server load balancing, and traffic load balancing.

The key to the data center network are the switches from the Cisco Nexus product family.

Cisco Nexus 7000 Series Switches

The Cisco Nexus 7000 Series Switches comprise a modular data center-class product line that is designed for highly scalable 10 Gigabit Ethernet networks with a fabric architecture that scales beyond 15 terabits per second (Tb/s). Designed to meet the requirements of the most mission-critical data centers, the switches can deliver continuous system operation and virtualized, pervasive services. The Cisco Nexus 7000 Series is based on a proven operating system, with enhanced features to deliver real-time system upgrades with exceptional manageability and serviceability. Its innovative design is built to support end-to-end data center connectivity, consolidating IP, storage, and interprocess communication (IPC) networks onto a single Ethernet fabric.

Cisco Nexus 5000 Series Switches

Data centers are increasingly filled with dense rack-mount and blade servers that host powerful multicore processors. The rapid increase of in-rack computing density and the increasing use of virtualization software combine to push the demand for 10 Gigabit Ethernet and consolidated I/O: applications for which the Cisco Nexus 5000 Series is an excellent match. With low latency, front-to-back cooling, and rear-facing ports, the Cisco Nexus 5000 Series is designed for data centers that are transitioning to 10 Gigabit Ethernet. The series can also assist with those data centers that are ready to deploy a unified fabric that can manage their LAN, SAN, and server clusters, providing networking over a single link (or dual links for redundancy).

The switch family, using cut-through architecture, supports line-rate 10 Gigabit Ethernet on all ports while maintaining consistently low latency independent of packet size and services enabled. It supports a set of network technologies that is known collectively as IEEE Data Center Bridging (DCB), which increases the 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 switches can connect FCoE to native Fibre Channel, and this function can protect existing storage system investments while dramatically simplifying in-rack cabling.

Cisco Nexus 2000 Series Fabric Extenders

The Cisco Nexus 2000 Series Fabric Extenders comprise a category of data center products that is 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 Megabit Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet, unified fabric, copper and fiber connectivity, rack, and blade-server environments. The platform is ideal to support traditional Gigabit Ethernet while allowing transparent migration to 10 Gigabit Ethernet and to VM-aware unified fabric technologies.

The Cisco Nexus 2000 Series Fabric Extenders behave as remote line cards for a parent Cisco Nexus switch. The fabric extenders are essentially extensions of the parent Cisco Nexus switch fabric, with the fabric extenders and the parent Cisco Nexus switch forming a distributed modular system. This architecture enables physical topologies with the flexibility and benefits of both top-of-rack and end-of-row deployments.

Cisco Nexus 1000V Series Switches

Cisco Nexus 1000V Series Switches are VM access switches that are an intelligent software switch implementation for VMware vSphere environments that are using the Cisco Nexus Operating System (NX-OS). Operating inside the VMware ESX hypervisor, the Cisco Nexus 1000V Series supports Cisco VN-Link server virtualization technology to provide the following benefits:

- Policy-based VM connectivity
- Mobile VM security and network policy
- Nondisruptive operational model for your server virtualization and networking teams

Cisco Nexus 1010 Virtual Services Appliance

The Cisco Nexus 1010 Virtual Services Appliance is a member of the Cisco Nexus 1000V Series Switches. The Cisco Nexus 1010 hosts the Cisco Nexus 1000V Virtual Supervisor Module (VSM) and supports the Cisco Nexus 1000V Network Analysis Module (NAM) Virtual Service Blade to provide a comprehensive solution for virtual access switching. Because the Cisco Nexus 1010 provides dedicated hardware for the VSM, it makes virtual access switch deployment much easier for the network administrator. Because of its support for additional virtual service blades such as the Cisco Nexus 1000V NAM Virtual Service Blade, the Cisco Nexus 1010 is a crucial component of a virtual access switch solution.

With the introduction of the Cisco Nexus 1010, the VSM now has dedicated hardware. Hence, the network administrators can install and configure virtual access switches even more closely to the way they work with physical switches. The dedicated VSM hardware is especially helpful when a data center is powered up, because there is no need to find server resources for the VSM. Thus, the Cisco Nexus 1010 enables network administrators to manage the Cisco Nexus 1000V virtual access switch in the same ways that other physical switches and scale server virtualization deployments are managed.

Cisco Nexus 4000 Series Blade Switches

The Cisco Nexus 4000 Series Blade Switches extend the benefits of the Cisco Nexus Family to blade servers. This type of switch is used for the IBM BladeCenter H and HT chassis and can provide the server I/O solution that is required for high-performance, scale-out, virtualized, and nonvirtualized x86 computing architectures. It is a line-rate, low-latency, nonblocking Layer 2 blade switch for 10 Gigabit Ethernet.

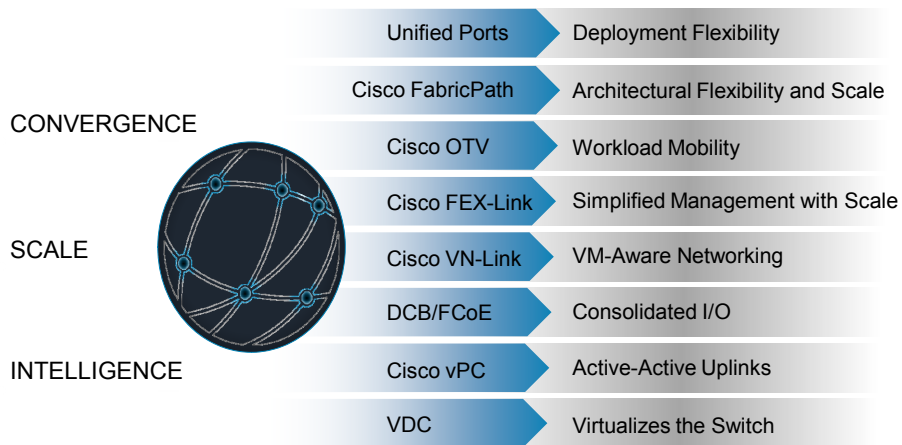
It addresses these challenges:

- Larger numbers of switches and cables within a single chassis
- Implementation of VM mobility
- Higher bandwidth requirements
- Increasing power requirements

Cisco Nexus 4000 switches provide all ports with support for both Gigabit Ethernet and 10 Gigabit Ethernet autonegotiation, for increased investment protection. They are also FCoE switches and they comply with the IEEE DCB specification.

Cisco Unified Fabric

United Fabric consolidates LAN and SAN infrastructure.



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-19

Ethernet is the predominant network choice for interconnecting resources in the data center. It is used often and is well understood by network engineers and developers worldwide, and Ethernet has withstood the test of time against challengers trying to displace it as the popular option for data center network environments. Emerging demands from applications require additional capabilities in networking infrastructures, resulting in deployment of multiple, separate, application-specific networks. It is common for enterprise data centers to deploy an Ethernet network for IP traffic, along with one or two SANs for block-mode Fibre Channel traffic.

Cisco Unified Fabric supports new concepts such as IEEE DCB enhancements that improve the robustness of Ethernet and that reduce the TCO and improves responsiveness by enabling network consolidation.

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 US\$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.

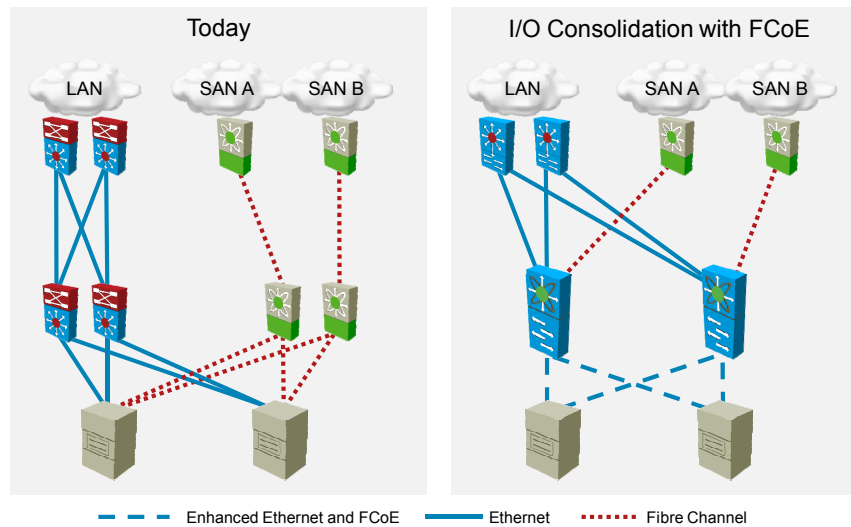
The main advantage of Cisco Unified Fabric is that it offers LAN and SAN infrastructure consolidation. It is no longer necessary to plan for and maintain two completely separate infrastructures. The network comes in as a central component to the evolution of the virtualized data center and to the enablement of cloud computing.

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

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

- **Cisco FabricPath:** Cisco FabricPath is a set of capabilities within the Cisco 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.
- **Cisco Overlay Transport Virtualization (OTV):** Cisco OTV is an industry-first solution that significantly simplifies the extension of 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 to extend the capacities and benefits that are offered by upstream Cisco Nexus switches.
- **Cisco VN-Link:** Cisco VN-Link provides advanced hypervisor switching as well as high-performance hardware switching. It is flexible, extensible, and service-enabled. Cisco VN-Link architecture provides virtualization-aware networking and policy control.
- **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 Systems Interface (SCSI) over IP, or FCoE, or a combination of these technologies, on a converged network.
- **Cisco virtual PortChannel (vPC):** Cisco 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 that will scale high-performance failover characteristics.

I/O Consolidation



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-20

Data centers usually have two different network configurations for LANs and SANs.

- The LAN consists of three layers: Core, Aggregation, and Access. Traditionally, Ethernet and usually Gigabit Ethernet have low bandwidth and relatively high latency. Ethernet frames are dropped if there is congestion, and so Ethernet is regarded as a lossy network.
- The SAN consists of two layers: Core and Edge. Fibre Channel SANs are used to carry the SCSI protocol between servers and storage devices within the payload of Fibre Channel frames. Fibre Channel has relatively high bandwidth with 8-Gb links and low latency at around 20 microsec for Cisco MDS switch store and forward latency. Fibre Channel has a lossless delivery that is provided by its credit-based flow control mechanism, which ensures that frames are never transmitted unless the receiving port provides a credit guaranteeing that it has a buffer available. This mechanism ensures that latency is very low because frames are never dropped so they never need to be recovered.

Nexus 5000 switches, for example, combine the Ethernet access layer switch and the Fibre Channel edge switch to provide a combined FCoE-capable switch that is connected to each server over a Unified I/O 10 Gigabit Ethernet connection. Server virtualization not only increases CPU and memory utilization but also requires much higher bandwidth connections. Unified I/O significantly reduces the number of cables that connect each server to top-of-rack switches while increasing available bandwidth and utilization.

Fibre Channel is a lossless protocol, and therefore Ethernet must provide enhancements to ensure lossless delivery for Fibre Channel frames.

In its initial phase, I/O consolidation is implemented as follows:

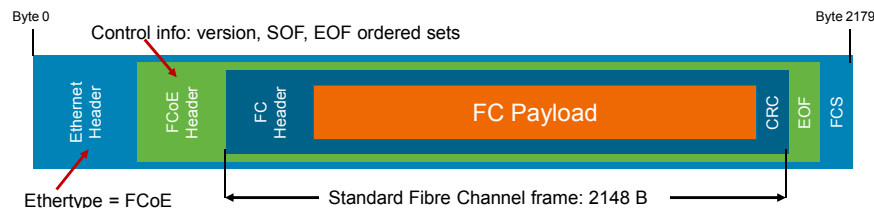
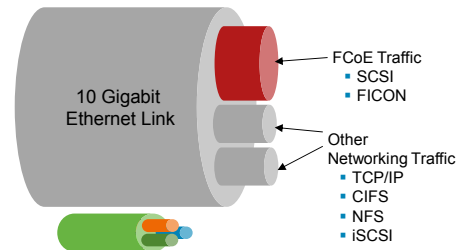
- Reduction of server adapters
- Simplification of access layer and cabling
- Gateway-free implementation, which fits in the installed base of the existing LAN and SAN
- Layer 2 multipathing access

- Lower TCO
- Fewer cables
- Investment protection for LANs and SANs
- Consistent operational model

Fibre Channel over Ethernet

Requirements:

- 10 Gigabit Ethernet
- Encapsulation of complete FC frame onto Jumbo Ethernet frame
- QoS
 - Lossless delivery of FC frames
 - Bandwidth reservation for FC traffic



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-21

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.

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 B 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
- **A 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.

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
Quantized congestion notification (QCN)	802.1au	Provides congestion awareness and avoidance (optional)
Data Center Bridging Exchange Protocol (DCBX)	802.1AB	Protocol that exchanges parameters between DCB devices and that leverages functions provided by LLDP

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-22

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-based flow control (PFC; IEEE 802.1Qbb):** Provides lossless delivery for selected classes of service (CoS)
- **Enhanced transmission selection (ETS; IEEE 802.1Qaz):** Provides bandwidth management and priority selection
- **Quantized congestion notification (QCN; IEEE 802.1au):** Provides congestion awareness and avoidance (optional)
- **Data Center Bridging Exchange Protocol (DCBX; IEEE 802.1AB):** Exchanges parameters between DCB devices and that leverages functions that are provided by LLDP

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: for example, IEEE 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, again 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.

Cisco uses the term Data Center Ethernet to refer to its architecture for a next generation of Ethernet for the data center.

Cisco Data Center Ethernet is a superset of the CEE and DCB proposals, using the same three specifications: PFC, ETS, and DCBX. In addition, Cisco includes Layer 2 multipathing (L2MP), lossless fabric, and congestion notification.

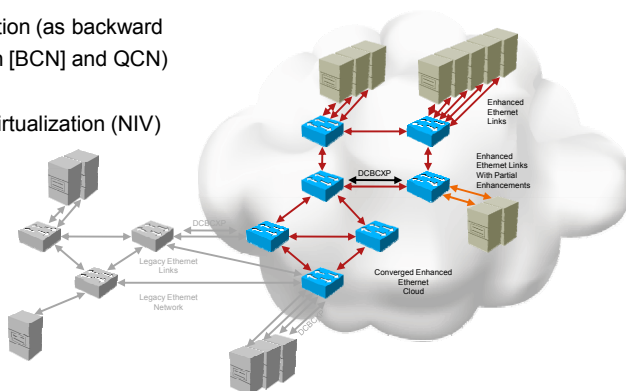
The Data Center Ethernet designation is used when a Cisco product or Cisco partner product includes those specifications and meets the requirements as defined by the Cisco Data Center Ethernet architecture.

DCBX

DCBX is a discovery and capability exchange protocol that is used by devices enabled for Data Center Bridging.

Autonegotiation of capabilities between DCB devices:

- PFC
- ETS
- Congestion notification (as backward congestion notification [BCN] and QCN)
- Logical link down
- Network interface virtualization (NIV)



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-23

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 like 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 QCN)
- Application types and capabilities
- 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)

(See <http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf>.)

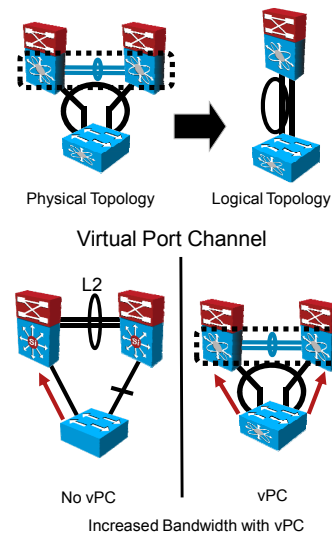
Devices need to discover the edge of the enhanced Ethernet cloud:

- 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 willingness to accept configurations from peers.

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

Virtual Port Channel

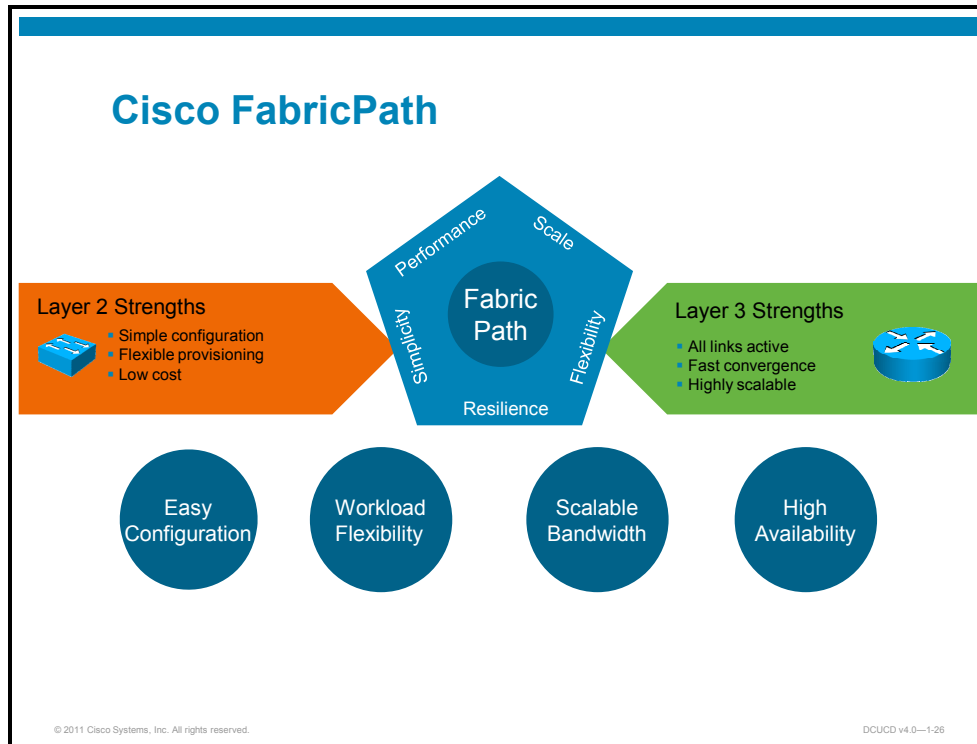
- The vPC is a port-channeling concept that extends link aggregation to two separate physical switches.
- It allows the creation of resilient Layer 2 topologies based on link aggregation.
 - Eliminates the need for STP in the access distribution
- It provides increased bandwidth.
 - All links are actively forwarding
- The vPC maintains an independent control plane.



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-25

A vPC extends the link aggregation from the downstream device to two separate upstream physical switches. These two physical switches are vPC peers, and, to the downstream device, they appear to be a single logical entity. STP now sees the connection between the downstream and upstream devices as a single logical link, meaning that no links are blocked because no loops are identified. This aggregation increases the available bandwidth as all links are actively forwarding. Although the two vPC peers appear as a single logical device to the downstream switch, they remain two separate entities.



Cisco FabricPath is a set of capabilities within the Cisco NX-OS Software that combines the simplicity of Ethernet with the reliability and scalability of Layer 3 routing. Cisco FabricPath provides the capability to build highly scalable Layer 2 multipath networks without the STP. These types of networks are particularly suitable for large virtualization deployments, private clouds, and high-performance computing environments.

Cisco FabricPath provides the following:

- Ease of configuration
- Workload flexibility
- Scalable bandwidth
- High availability

Cisco FabricPath (Cont.)

Single domain:

- Combined compute resources

Increased agility:

- Transparent data center-wide workload mobility

Responsive:

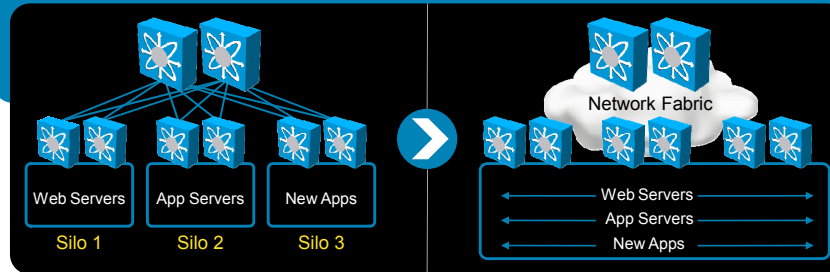
- Virtualized applications move within minutes versus days

CapEx and OpEx savings:

- Maximize resource utilization and simplify IT operations

Multidomain: Siloed

Single Domain: Any app, anywhere



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-27

The benefits of Cisco FabricPath include the following:

■ Simplicity:

- There are minimal commands per switch for ease of configuration and operation.
- Autodiscovery simplifies the addition of new links and switches.
- You can provision additional network bandwidth without impacting the network.

■ Scalability:

- It accommodates both small- and large-scale deployments.
- Equal-cost multipathing across all available paths increases cross-sectional bandwidth.
- The network is scalable up to more than 100 Tb/s of switching capacity.

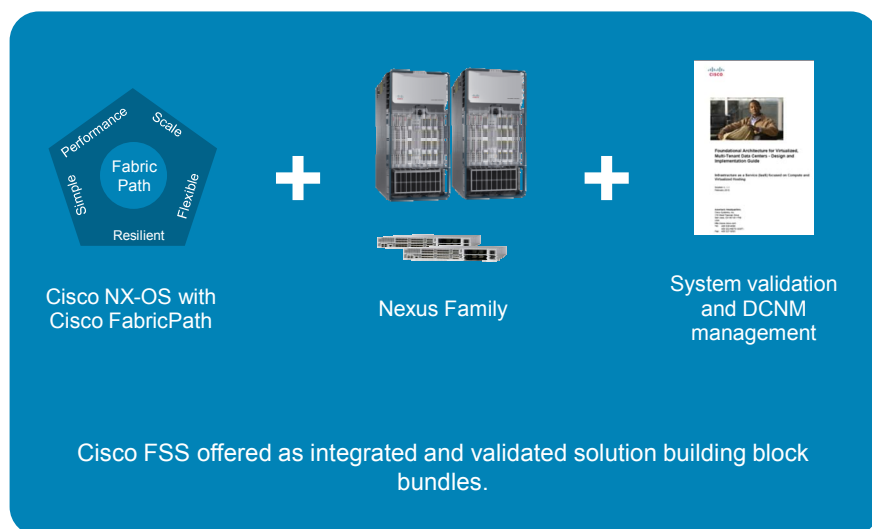
■ Resiliency:

- Multipathing across all links increases availability.
- Eliminating STP results in faster convergence.
- Up to 16 aggregation layer switches minimize the impact of a single switch failure.

■ Flexibility:

- There are multitopology networks that are based on application needs.
- It supports one-, two-, or three-tier architectures.
- The VLAN extensibility is across the data center.

Cisco FabricPath Switching System



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-28

The Cisco FabricPath Switching System (FSS) is a foundation for building scalable fabrics. It combines the benefits of Cisco FabricPath, a critical Cisco NX-OS capability, with the performance advantages of the Cisco Nexus F1 Series Module to build a high-performance system. Cisco FSS is offered as a set of validated solutions that simplifies the construction of data center infrastructures. Cisco FSS can provide optimal bandwidth between any two ports regardless of their physical location. As a result, VLANs can be extended to any port in any switch without any configuration overhead and are no longer restricted to a rack or pod.

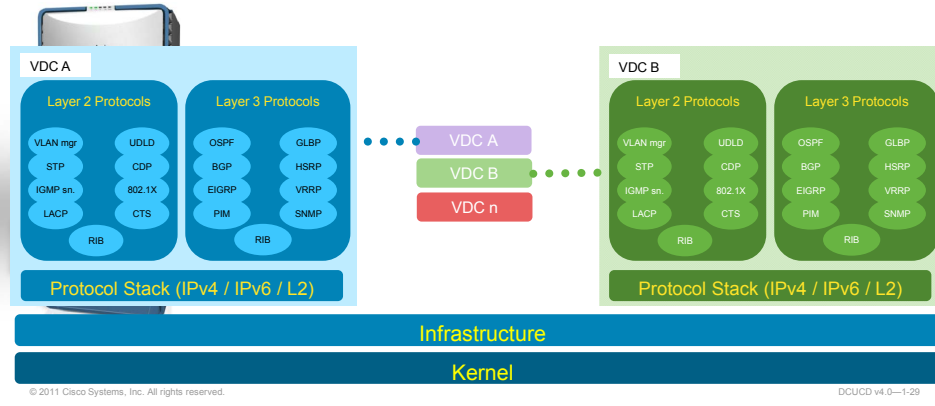
Virtual Device Contexts

Virtual Device Context (VDC)

- Flexible separation and distribution of **software components**
- Flexible separation and distribution of **hardware resources**
- Securely delineated **administrative contexts**

VDCs do not

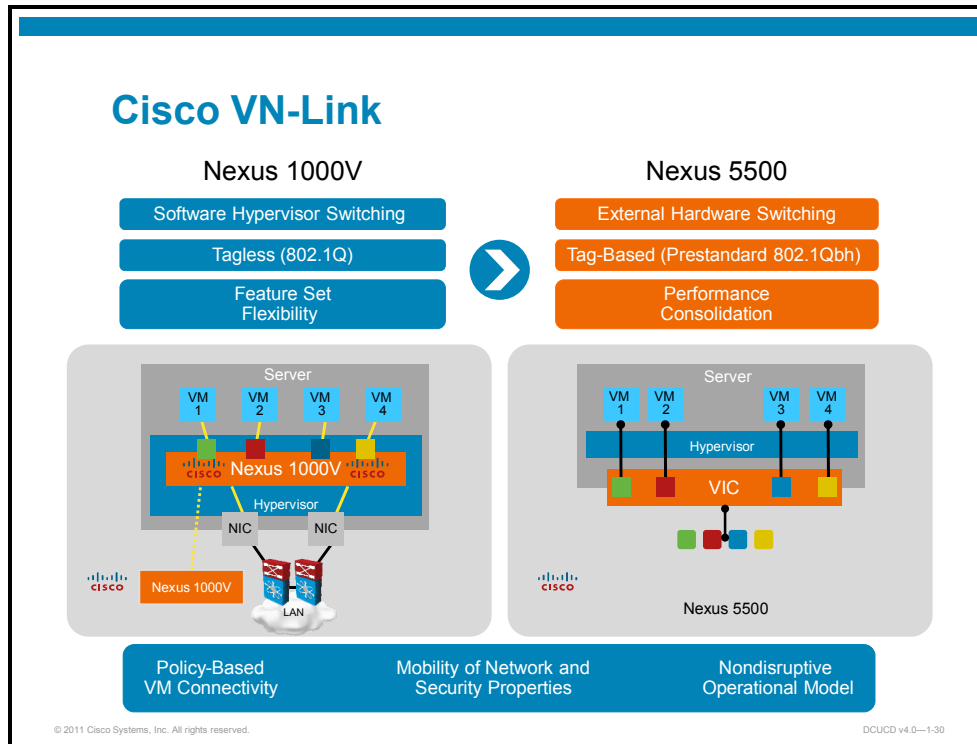
- Have the ability to run different operating system levels on the same box at the same time.
- Have a single "infrastructure" layer that processes hardware programming.



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.



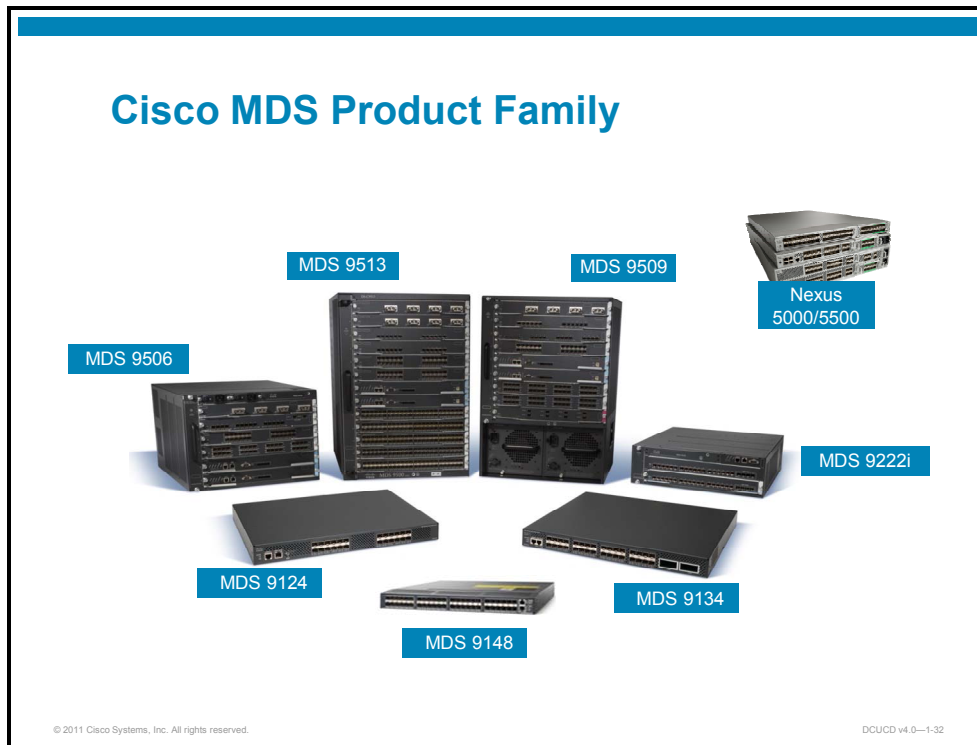
Cisco VN-Link encompasses a number of products and technologies that work together to improve server virtualization strategies:

- **Cisco Nexus 1000V virtual switch:** A software-based switch that was developed in collaboration with VMware. The switch integrates directly with the VMware ESX 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 VN-Link 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 VN-Link provides visibility down to the VM level, simplifying management, troubleshooting, and regulatory compliance.

Cisco Data Center Business Advantage Storage

This topic describes the Cisco Data Center Business Advantage architectural framework storage component.



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 ones.

Multilayer switches are switching platforms with multiple layers of intelligent features, such as the following:

- Ultrahigh 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.

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 new DS-X9304-18K9 line card has eighteen 1-, 2-, or 4-Gb/s Fibre Channel ports and four 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.

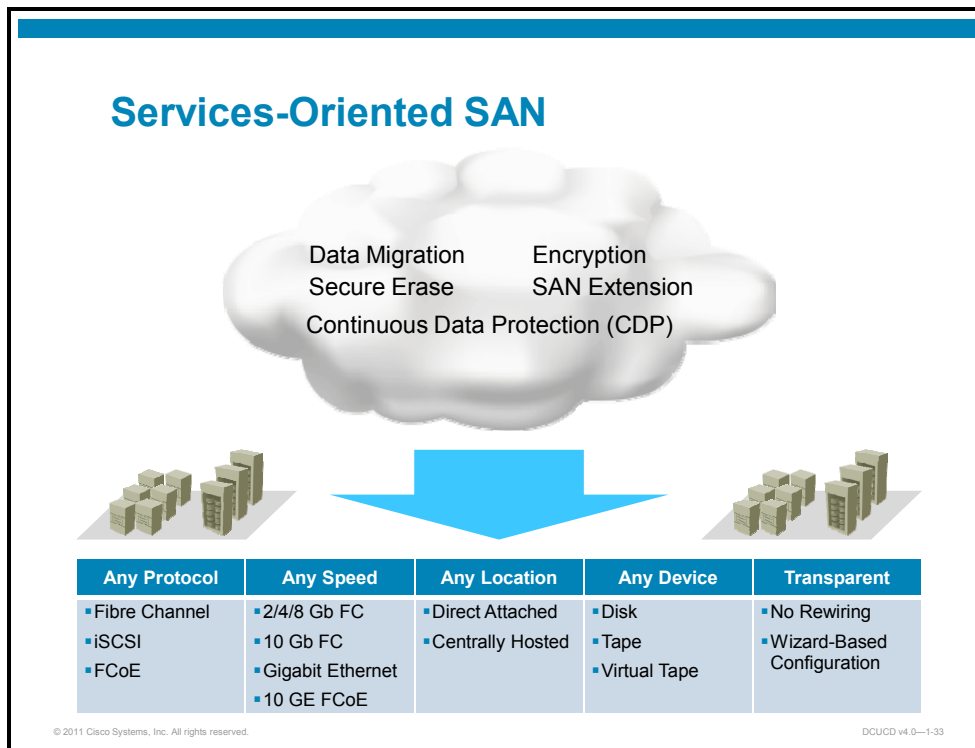
The Cisco MDS 9222i Multiservice Modular Switch uses the 18/4 architecture of the DS-X9304-18K9 line card and includes native support for Cisco SME along with all the features of the Cisco MDS 9216i Multilayer Fabric Switch.

The Cisco MDS 9134 Switch is a 34-port, low-cost Fibre Channel switch providing thirty-two 1-, 2-, or 4-Gb/s ports at full line rate plus two 10-Gb/s ports. New features include the Cisco IOS In-Service Software Upgrade, providing hitless software upgrades, and flexible port licensing, providing port allocation on an as-needed basis for small and medium-sized businesses.

The Cisco MDS 9148 Switch is a new 8-Gb/s Fibre Channel switch providing forty-eight 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.

Cisco Nexus 5000 Switches

The Cisco Nexus 5000 Series Switches are ideal for enterprise-class data center server access layer and smaller-scale, midmarket data center aggregation deployments.

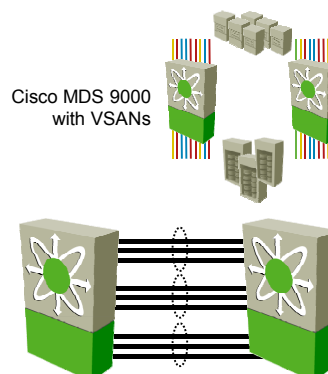


Services-oriented SAN fabrics can transparently extend any of the following SAN services to any device within the fabric:

- 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 like Write Acceleration and Tape Read Acceleration, which reduce overall latency
- Continuous data protection (CDP), which is enabled by the SANTap protocol

Storage and SAN Virtualization

- Different virtualization options
- Storage space virtualization with logical unit numbers
- Fabric virtualization with VSAN
- Inter-VSAN Routing for better scalability
- PortChannels for bandwidth scalability



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-34

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 that can 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 Nexus 7000 Ethernet 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 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 one another by using a hardware-based frame-tagging mechanism on VSAN member ports and enhanced ISL (EISL) links. The EISL link type has been created and includes added tagging information for each frame within the fabric.

The 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, whatever node is connected to a physical port becomes a member of that VSAN port.

Cisco NPV-Enabled Environment

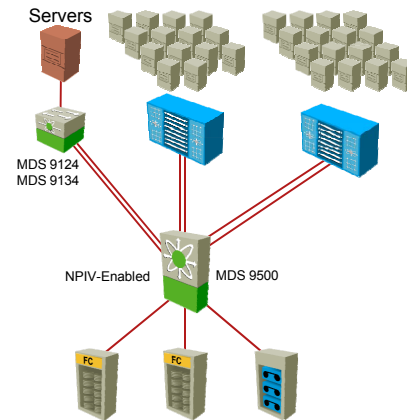
NPV Edge Switches

- Need to enable switch in NPV mode
- Changing to and from NPV mode is disruptive:
 - The switch reboots.
 - The configuration is not saved.
- Supports only F, SD, and NP modes
- Supports 16 VSANs
- Local switching is not supported.
 - 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.



F = fabric mode, SD = SPAN destination mode, NP = proxy N mode
© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-35

The Fibre Channels 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.

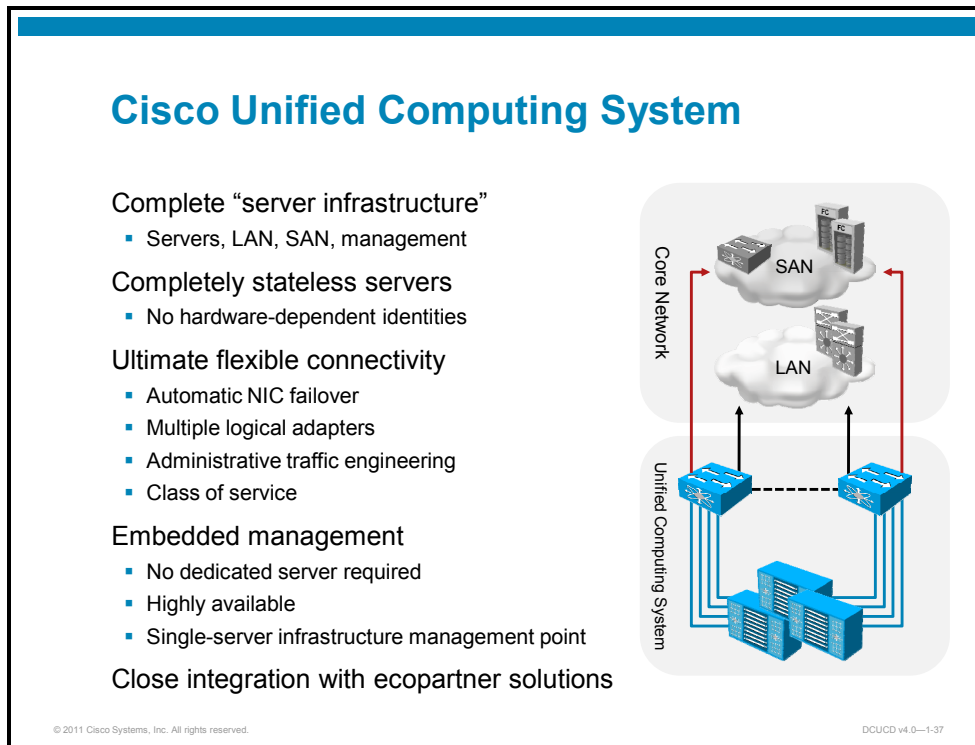
Each Fibre Channel switch is identified by a single domain ID, so effectively there can be no more than 40 switches that are connected together.

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 NP 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.

Cisco Data Center Business Advantage Compute

This topic describes the Cisco Data Center Business Advantage architectural framework compute component.



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

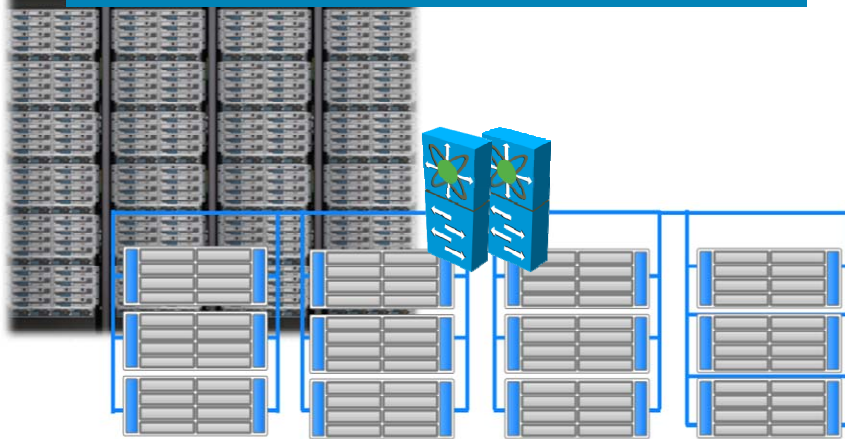
From the perspective of server deployment, the Cisco UCS represents a cable-once, dynamic environment that enables the rapid provisioning of new services. Since the unified fabric is an integral part of the Cisco UCS, fewer cables are required to connect the system components and fewer adapters need 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 also in lower power consumption, which makes the Cisco UCS solution greener. You will achieve better power consumption ratio per computing resource.

Scalability

Up to 160 servers in a single system*



* 20 chassis with the current software release

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-38

The Cisco Unified Computing System offers great scalability because a single use of the system can consist of up to 40 chassis, with each chassis hosting up to eight server blades. This scalability means that the administrator has a single management and configuration point for up to 320 server blades in the future.

Cisco Unified Computing System Key Features

Extended statelessness: Service profiles

Converged networking: Unified Fabric

Enhanced virtualization

Expanded scalability

Simplified management

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-39

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 very 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 what blade you move a server to, the fabric and communications remain the same.

Enhanced Virtualization

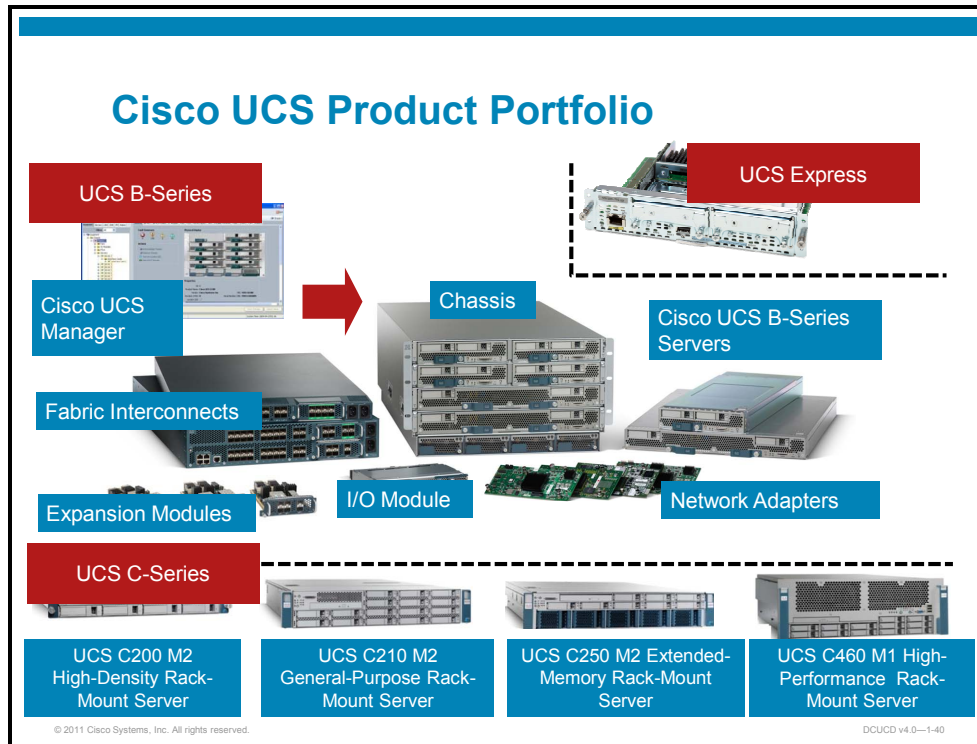
As part of the statelessness, the Cisco UCS is designed to provide visibility and control to the networking adapters within the system. This visibility and control are achieved with the software running on the Cisco UCS and the implementation of the virtual interface card adapters, which, in addition 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 B250 M1 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 command-line interface (CLI), or an XML API.



Cisco UCS B-Series

The Cisco UCS B-Series is composed of the following:

- Two Cisco UCS 6100 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 2104XP I/O Modules (or Fabric Extenders) per chassis.
 - Each chassis can host up to eight server blades (altogether up to 320 server blades).
 - The server is equipped with one or two network adapters.

Cisco UCS C-Series

The 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.

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:

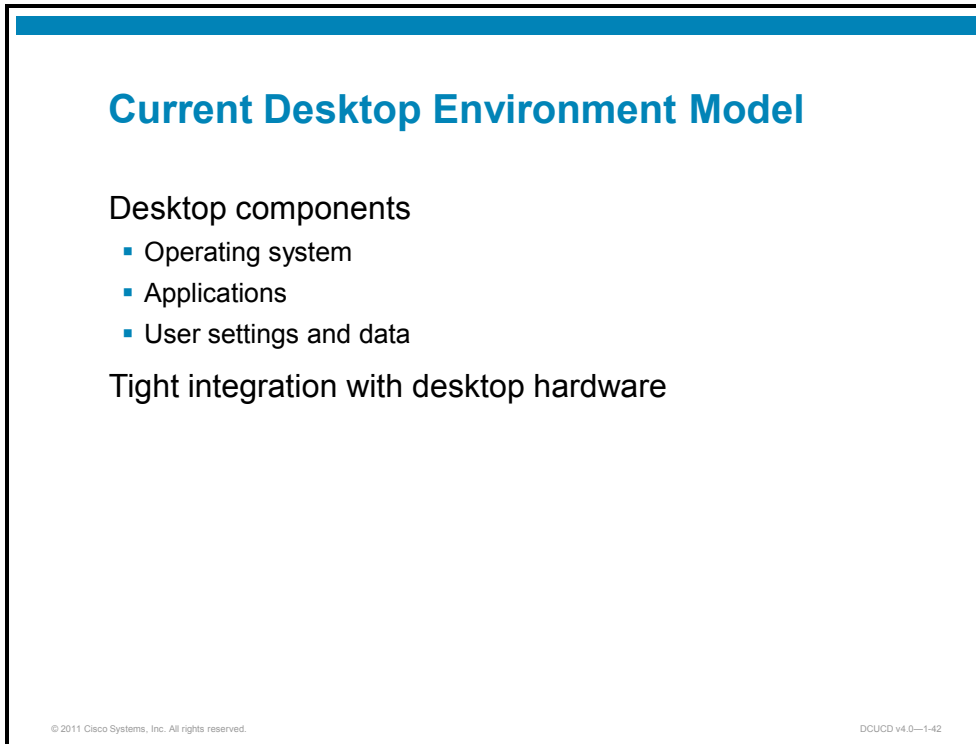
- **Cisco Integrated Services Router Generation 2 (ISR G2):** This router is a single-device network integration that can house all devices such as servers and networking appliances. It uses a Multigigabit Fabric (MGF) backplane switch to connect compute, switching, and other networking devices together without taxing the router CPU.
- **Cisco Service 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.

Cisco Data Center Business Advantage Desktop

This topic describes the Cisco Data Center Business Advantage architectural framework desktop component.



Desktop Challenges

The “thick-client” PC has been the workhorse of desktop computing. It has offered the best available combination of price, performance, and capabilities. However, for many use cases thick-client PCs are not an ideal solution. Among their drawbacks are the following:

- Difficult to manage
- High TCO
- Difficult to protect and secure data
- Inefficient resource utilization

Virtual Desktop Infrastructure

Separates a PC desktop environment from a physical machine by using a client-server model of computing

Replaces desktop hardware and software with virtual components: virtual machine image (VMI)

- Operating system
- Applications
- Data
- User settings

Running user desktops on a central server infrastructure

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-43

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 are running on the server (for each client) and clients can connect to their “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

Cisco Virtualization Experience Infrastructure

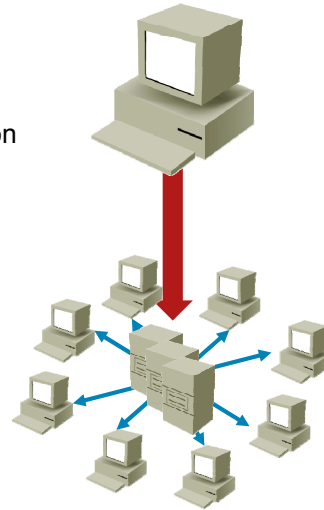
Virtualized end-to-end system

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



© 2011 Cisco Systems, Inc. All rights reserved.

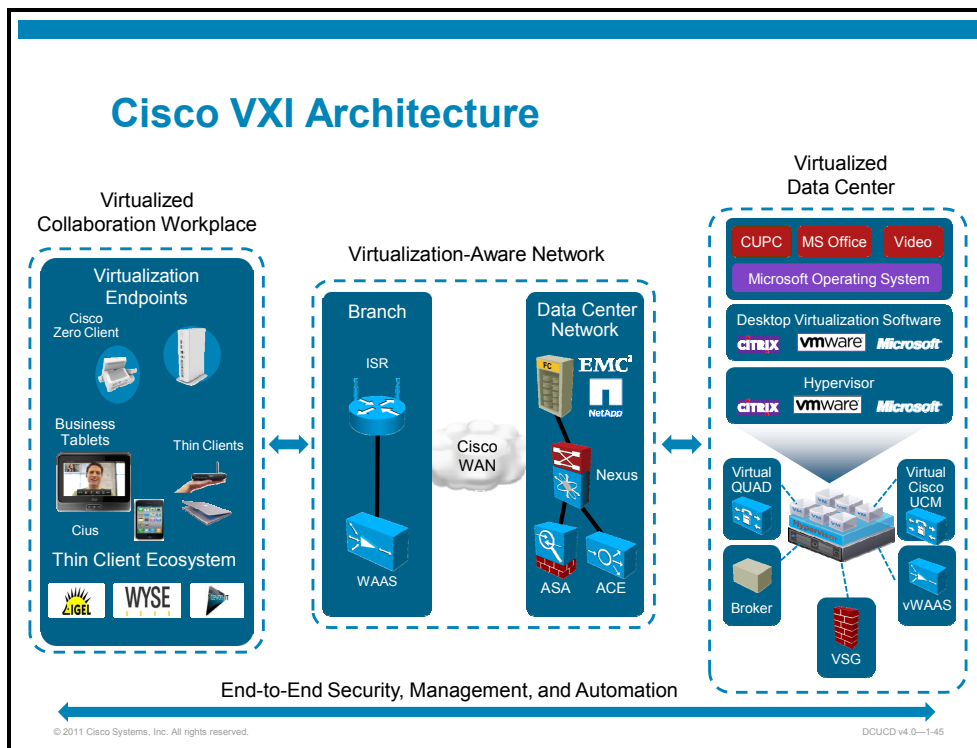
DCUCD v4.0—1-44

Cisco VXi Overview

Cisco VXi is an architectural approach with rich end-to-end services, which deliver 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.



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.

Cisco Data Center Services

This topic describes Cisco Data Center Services.

Services Overview

Strategic IT and Architecture

- End-to-end architecture
- Consolidation and virtualization
- IT operations process
- Business process for infrastructure dependencies and mapping
- Business case and metrics (ROI)

IT Planning and Deployment

- Consolidation and virtualization
- Security integration
- Business continuance planning
- Data center technologies
- Migration to unified fabric, unified computing, and virtualized architecture

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-47

Many enterprise data centers remain complex and siloed environments in which servers and storage equipment are vastly underutilized, cycle times for provisioning new applications are long, and operating costs—especially power and cooling costs—consume an ever-greater percentage of the IT budget. More organizations are turning to virtualization to address these issues. Unfortunately, many virtualization efforts do not deliver the expected results because they focus only on servers, failing to account for storage, network, and other critical resources.

Cisco Data Center Services can help customers consolidate, virtualize, and automate their data centers to meet business goals, increase efficiency, and lower operating expenses.

Even if unified computing is central to your customer virtualization strategy, the data center encompasses a broad range of other critical infrastructures and processes beyond the Cisco Unified Computing System.

Services Overview (Cont.)

Management and Operations

- Rapid problem resolution
- Proactive monitoring
- High availability and operations
- Provisioning

Optimization

- Evolve data center
- Maintain optimal performance
- Increase expertise

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-48

Whether the customer wants services for the entire data center or for specific data center technologies, Cisco offers a broad array of data center services:

- Strategic IT and architecture services to help prepare a strategy for the data center initiative
- IT planning, design, and implementation services to help the customer execute the strategy
- Technical support and operations management services to help maintain the health of the data center
- Optimization services to help maintain a high level of performance as the data center evolves
- Efficiency and facilities services to help adopt a greener IT strategy and design and construct efficient data center facilities

This unique architectural approach provides a unified view of data center resources, empowering the customer to adjust the data center environment for optimal application performance and availability.

Services Benefits

- Help achieve business and technical goals
- Mitigate virtualization risks
- Enhance the solution agility
- Reduce complexity
- Simplify management
- Optimize uptime, performance, and efficiency of the Cisco Unified Computing System
- Improve application performance

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-49

Cisco Data Center Services combine the broad data center expertise of Cisco with the expertise of the Cisco partners to accelerate customer transition to a unified computing architecture. These services help customers with the following objectives:

- Achieve business goals by aligning the data center strategy with unique business and technical objectives
- Speed deployment and mitigate risks of virtualization and other data center projects by applying best practices and methodologies
- Enhance agility by improving the mobility and manageability of traditional and virtual workloads and enabling just-in-time provisioning of resources
- Lower operational costs by identifying opportunities to reduce data center complexity and simplify management
- Optimize the uptime, performance, and efficiency of unified computing systems to help maximize the value of their investment
- Provision new applications more quickly
- Strengthen in-house expertise through knowledge transfer and mentoring
- Improve application performance and availability to meet service level agreements

Summary

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

Summary

- Cisco Data Center Business Advantage is an architectural framework for dynamic networked organizations.
- The framework allows organizations to create services faster, improve profitability, and reduce the risk of implementing new business models.
- A unified fabric uses a single-wire deployment to provide storage and network connectivity.
- The Cisco Unified Computing System is a system that includes compute, storage, and network resources.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-50

Evaluating Server Deployment Options

Overview

This lesson evaluates the Cisco Unified Computing System (UCS) deployment options and advantages for the data center computing solution.

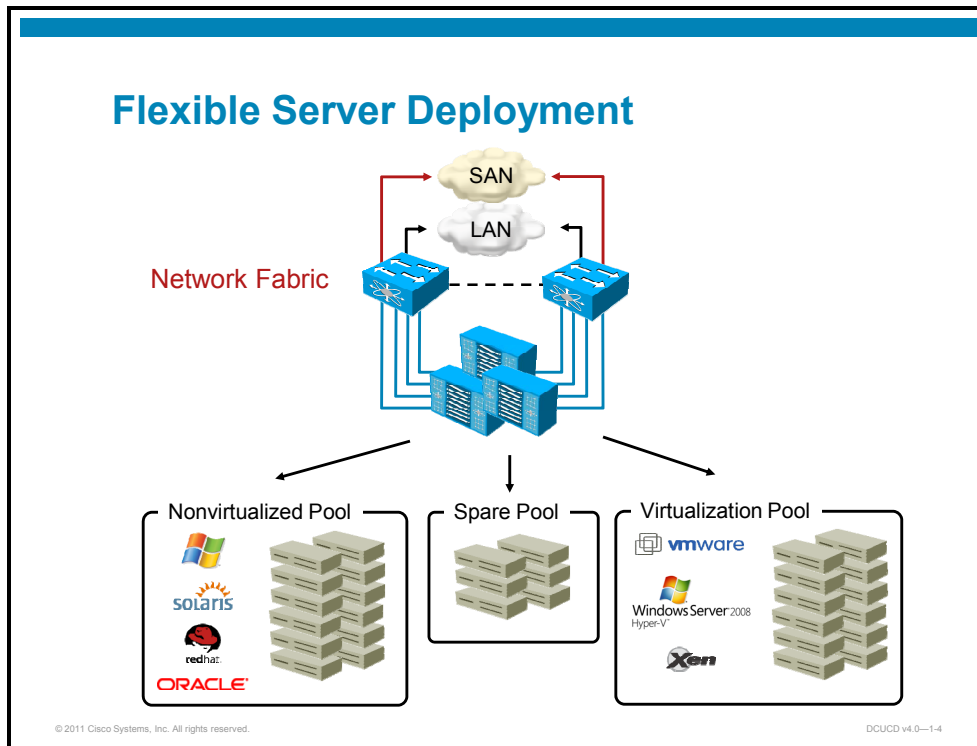
Objectives

Upon completing this lesson, you will be able to identify where and how Cisco UCS can be used. This ability includes being able to meet these objectives:

- Describe the purposes of Cisco UCS
- Describe the Cisco Validated Designs

Server Deployment Options

This topic describes the purposes of the Cisco Unified Computing System.



A single Cisco Unified Computing System can be deployed with a mixture of bare metal operating system installations to virtualized server solutions.

From another perspective, the Cisco UCS can be used for server deployments—which include new servers, upgrading, and repurposing—and for achieving business continuity, which includes high availability and disaster recovery.

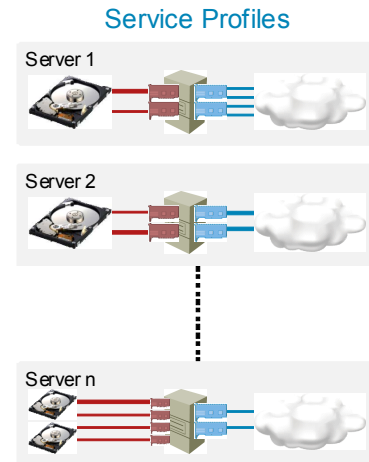
Neither a server deployment nor business continuity is a function that is unique to blade-server markets. Enterprises that have larger server deployments face many of these issues.

Even in smaller environments (that have tens to hundreds of servers), the challenges that are faced in deployment and business continuity can still be addressed by the features that are provided by a Cisco UCS. However, the true advantage of the system is more easily seen on larger server deployments (that have thousands of servers).

The life cycle for servers in any organization is between 18 months to 3 years, which is governed by the advent of new technologies. These technology shifts can happen every 3 to 5 years. The changes in server technologies often follow the life cycle of solutions changes and not so much the changes of the technology.

Simple Server Deployment

- Server, storage, and network administrators
 - Preconfigure multiple servers upfront by using service profiles
- Deploy physical servers as needed over time
 - Plan and preconfigure once
 - “Pay as you grow” nondisruptive incremental deployment
 - Simple server replacement when replacing physical blade



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v1.0-1-5

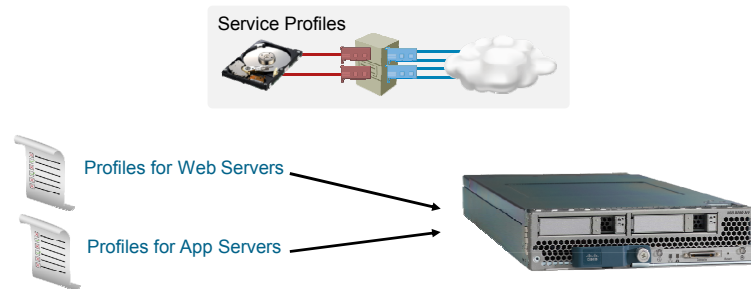
The Cisco UCS platform simplifies server deployments, with service profiles. The service profile is created by a server administrator, and it may possibly use resources that are created by the storage and network administrator, like MAC and world wide name (WWN) pools. Service profiles allow for the following capabilities:

- **Plan and preconfigure once:** By using resource pools, service profiles, and service templates, you can plan and configure for groups of servers all at one time and then, using the profiles, provision new servers at any time.
- **Incremental deployments:** After service profiles are created, adding additional servers is as simple as using a service template to create a new server.
- **Server replacement:** This task is as simple as starting a profile on a new blade after it has been physically replaced.

Dynamic Server Deployment

Apply the appropriate profile to the provision-specific server type:

- Same hardware can be dynamically deployed with different server types
- No need to purchase custom configured servers for specific applications
- Maximized server hardware



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-6

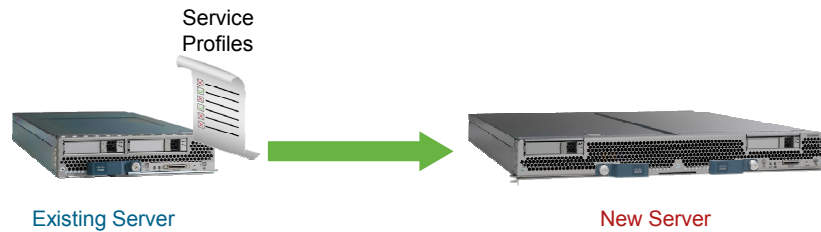
Some of the challenges that are faced by companies that have large numbers of servers are provisioning and server utilization. Typically, when designing a solution, a customer will need to design it for different levels of utilization. In certain solutions, customers may need to purchase and have available the additional server hardware and software, for burst capacity. This hardware, while being underutilized, cannot also be dynamically repurposed; if it is not needed, it must simply occupy space, power, and cooling resources.

Because servers are virtual and you can configure the servers by using profiles, repurposing resources (server hardware) is as simple as disassociating a profile and reassociating a new one. Servers that are not used often can be shut down and repurposed into more highly used servers.

Server Upgrade Within Cisco Unified Computing System

The existing server can be retired or repurposed (created or associated with the appropriate profile).

- Disassociate the service profile from the existing server.
- Assign the service profile to the new server.



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-7

The Cisco UCS enhances server upgrades with service profiles. Simply put, the blade in the system is agnostic and its hardware can be defined with the service profile.

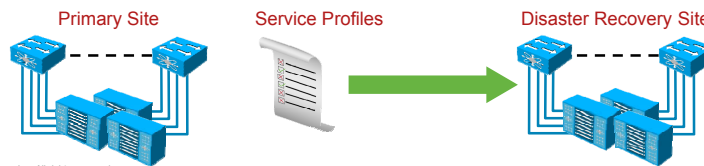
A server in a traditional deployment will be built specifically for a solution that contains adapters, firmware, and boot definitions. If you wish to upgrade such a server, you must again go through the process of reinspecting the server for new hardware and coordinate that process with other administrators. Then, you have to schedule extended time to make the move.

With the Cisco UCS platform, there are two options:

- Change the service profile to boot the server on another upgraded blade.
- Copy the service profile and assign it to the other blade. Then, you can test the new server before pointing users to it.

Server Recovery with Cisco Unified Computing System

- Primary and disaster recovery site
- **Normal operation**
 - Export or import service profiles periodically or on an ad hoc basis
 - Periodic disaster recovery tests at disaster recovery site
 - Servers at disaster recovery site repurposed during normal operations (deployed for testing, development, or quality assurance)
- **Primary site outage**
 - Disaster recovery site: disassociate existing profiles and associate production profiles from backup (might require management IP addressing changes)
 - Other requirements (not in the system) for complete disaster recovery:
 - Data backup
 - IP addressing aspects: overlapping versus changed



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-8

The features of the Cisco UCS can provide aid in a disaster recovery solution. When designing a disaster recovery solution, you must consider the whole infrastructure stack, which includes the following:

- Application failover and dependency matching
- Operating system configuration and compatibility
- Data replication and integrity
- Connectivity matching (LAN and SAN)
- Server compatibility

Of these, the area in which the Cisco UCS can aid significantly is server compatibility. Here are the server compatibility requirements:

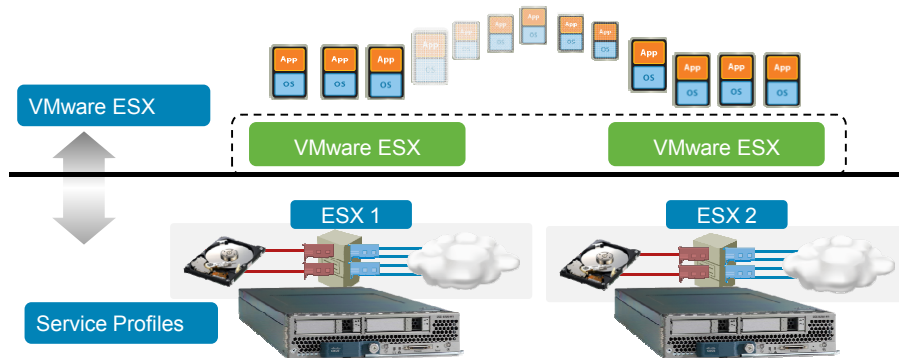
- Configuration matching (adapters)
- Firmware matching
- Server parameters
- LAN or SAN specifics (VLANs and VSANs)

All of these items are configurable in the service profile for a given server, such that when recovering, the system can quickly ensure server compatibility on the recovery system. Without this ability, server incompatibility can completely derail a disaster recovery solution.

Server compatibility is mainly a manual process that leaves much potential for incorrect configuration. The Cisco UCS, with the scope of statelessness and the ability to transparently use profiles over redundant systems, helps with server compatibility issues.

Server Virtualization

- Server virtualization and service profiles are independent of each other.
- The hypervisor is unaware of underlying hardware state abstraction.



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-9

Server virtualization (with VMware vSphere, Microsoft Hyper-V R2, Citrix XenServer, and so on) and the Cisco UCS are complementary products and technologies.

While server virtualization can be used to replace many of the physical servers within an organization, it is still an application and operating system that needs a server platform on which to run. With integration and visibility into network virtualization, the Cisco UCS is uniquely positioned as a preferred platform on which to run server virtualization.

Server virtualization provides a software-based solution that enables server virtualization at a software level. Server virtualization provides a hypervisor that allows virtual machines (VMs) to think that they are accessing a common set of expected hardware.

What the system does is to ensure server hardware compatibility across the entire Cisco UCS. This feature means that the hypervisor will run on any blade in the system, and the Cisco software ensures that the hardware is identically configured, regardless of the blade from which the virtualization server is chosen to run.

VMware vSphere and Cisco Unified Computing System

In a general ESX deployment, one or several blades will be running a VMware ESX hypervisor and will host multiple VMs on the blades. The value propositions make the Cisco UCS ideal for ESX deployments.

- **Expanded-memory blade servers (Cisco UCS B250 M1):** The more memory that is available, the more VMs you can host per blade, allowing your blade to have a very cost-effective approach on a per-VM basis. However, there are other advantages, as well:
 - Reduced hardware costs
 - Reduced licensing (based on the number of cores)
 - Larger memory can manage vMotion better

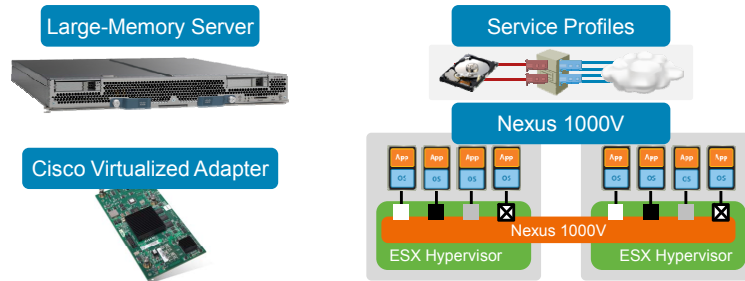
- **Virtualization integration:** The Cisco UCS VIC M81KR adapter provides not only visibility into the VM networking, but can also provide the following:
 - Better security
 - Performance, because virtual adapters can be moved out to be managed at the hardware level, rather than at the hypervisor
 - Improved troubleshooting
- **Cisco Unified Fabric:** This solution provides a single network for connectivity, simplifying your configuration while reducing power and cooling requirements. Virtual network interface cards (NICs) are configured and managed like all other resources through Cisco UCS Manager.
- **Service profiles:** Server statelessness allows for the rapid replication and deployment of servers, which is good for a pay-as-you-grow environment. In addition, server statelessness also provides for the following:
 - **Enhanced disaster recovery:** Eliminates worry in server compatibility for disaster recovery
 - **Easy upgrades and replacement:** Server profiles are combined with vMotion

VMware provides software to enable high availability and disaster recovery for the VMs that are being hosted. This migration can occur by the following means:

- **Live migration across ESX hosts:** This manual intervention can be used for upgrading servers. With the latest version of ESX, service profiles can be used to create identical ESX servers, as needed, to move VMs between. Network and SAN access would be identical, saving time.
- **Policy-based migrations:** Moving a VM from one server to another, based on some type of policy, such as server utilization statistics. All access can be replicated simply through a virtual NIC in service profiles.
- **VM restarts due to failed hosts:** Hosts can be quickly replicated with service profiles, thus relieving failover hosts. Larger memory footprints reduce the impact of increasing the number of VMs during failure.

Virtual Desktop Environment

- Servers for VDI require large amounts of memory.
 - Application virtualization and boot image management apps benefit from caching.
 - Streaming workloads require plenty of CPU cycles.
- The best practices of an advanced architecture require multiple segregated networks.



The virtual desktop environment or virtual desktop infrastructure (VDI) has the following characteristics, which can exploit the benefits of the Cisco UCS:

- VDIs are very CPU- and memory-intensive. The Cisco large-memory server offering is ideal in that the memory is scalable to very large amounts, and the Intel Xeon CPU will be expanding the numbers of cores.
- Some customers may not want to depend on too many VMs at one time. The ability to rapidly deploy servers by using service profiles provides better high availability than traditional servers.
- Application servers and boot servers can benefit from the greater memory available for caching.
- Streaming workloads can cause VMs to be server-bound. Cisco UCS virtualized adapters and adapter profiles provide ease of migration.
- Some architectures require specific network segregation. The Cisco UCS virtualized adapter and VLANs can easily provide this segregation.

Lastly, the Cisco UCS is ideal, because even in an advanced VDI deployment, it can provide for all of the server platforms that are needed, allowing a VDI deployment in a box, or “pod.”

Cisco Validated Designs

This topic describes the Cisco Validated Designs.

Cisco Validated Designs Overview

- Cisco Validated Designs are a set of different solution design guides (blueprints).
 - Integrated solutions that consolidate Cisco UCS, Nexus, ACE, WAAS, and ecopartner products (from Microsoft, EMC, NetApp, VMware, and so on)
- Cisco Validated Designs are tested, verified, and documented designs that have defined the following:
 - Customer requirements
 - Solution architecture
 - Hardware and software components (BOM)
 - Infrastructure topology
 - Descriptions of the components that are used

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-12

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 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 from various partners (such as Microsoft, EMC, NetApp, and VMware).

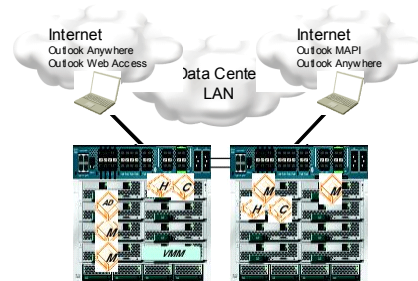
The individual blueprint covers the following aspects:

- Solution requirements from an application standpoint
- Overall solution architecture with all the components that fit together
- Required hardware and software Bill of Materials (BOM)
- Topology layout
- Descriptions of the components that are used and their functionalities

Microsoft Applications and Solutions

Validated design guides that focus on Microsoft products and deployment on the Cisco UCS and other Cisco data center products and technologies:

- Virtualized Microsoft applications
 - Microsoft Hyper-V and Microsoft Exchange 2010
 - Exchange 2010 on VMware vSphere and NetApp storage
 - Microsoft SharePoint on Hyper-V



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-13

A set of the validated designs is focused around the Microsoft solutions and applications.

Microsoft Hyper-V and Microsoft Exchange 2010

This design guide presents an end-to-end solution architecture that demonstrates how enterprises can virtualize their Exchange 2010 environments on the Cisco Unified Computing System. This solution is accomplished by using Microsoft Hyper-V virtualization and NetApp storage with application optimization services provided by the Cisco Application Control Engine (ACE) and Cisco Wide Area Application Services (WAAS).

As enterprises face the critical problem of growing their data centers when they are already reaching maximum capacity for space, cabling, power, cooling, and management, they should consider consolidating and virtualizing their application servers into their data centers. At the top of the list of servers to consolidate and virtualize are those mission-critical enterprise applications that typically require deployment of many physical servers. Exchange 2010 is one such application because it requires four different server roles for a complete system and more often it 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.

Exchange 2010 on VMware vSphere and NetApp Storage

This design and deployment guide demonstrates how enterprises can apply best practices for vSphere 4.0, Exchange 2010, and NetApp fiber-attached storage (FAS) arrays in order to run virtualized and native installations of Exchange 2010 servers on the Cisco UCS. vSphere and Exchange 2010 are deployed in the Cisco Data Center Business Advantage architecture, which provides redundancy and high availability, network virtualization, network monitoring, and application services. The Cisco Nexus 1000V virtual switch is deployed as part of the vSphere infrastructure and is integrated with the Cisco Network Analysis Module appliance to provide visibility and monitoring into the virtual server farm. This solution was validated within a multisite data center scenario with an Exchange deployment that was created by using the Microsoft Exchange 2010 Load Generator to simulate realistic user loads. The goal of the validation was to verify that the combination of the Cisco UCS, NetApp storage, and network link sizing was sufficient to accommodate the Load Generator user workloads. The Cisco Global Site Selector (GSS) provides site failover in this 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 WAAS.

Microsoft SharePoint on Hyper-V

This document is intended for enterprises that are 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 by using the Cisco UCS and Cisco ACE and Cisco WAAS technologies. The document provides guidance to engineers who are interested in deploying Cisco Data Center 3.0 architectures including network, compute, and application services. The design options and implementation details that are provided are intended to be a reference for an enterprise data center.

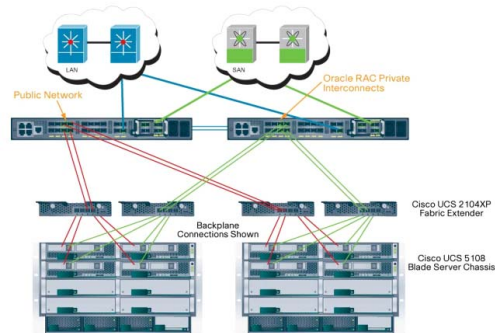
Validated design guides that focus on VMware solutions:

A set of the validated designs is focused on VMware server and desktop solutions.

Oracle Solutions

Validated design guides that focus on Oracle solutions:

- Cisco Unified Computing System with EMC CLARiiON
- Cisco Unified Computing System with NetApp storage



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-15

Oracle Real Application Clusters on the Cisco Unified Computing System 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, which are 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 today, including the need for a simplified deployment and operation model, high performance for Oracle RAC software, and lower total cost of ownership (TCO). The document introduces the Cisco UCS and provides instructions for implementing it; the document concludes with an analysis of the cluster performance and reliability characteristics.

Cisco Unified Computing System and NetApp Solution for Oracle RAC

This Cisco Validated Design describes how the Cisco UCS can be used with NetApp FAS systems to implement a decision-support system (DSS) or online transaction processing (OLTP) database using an Oracle RAC system. The Cisco UCS provides the compute, network, and storage access components of the cluster, which are 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 today, including the need 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.

Summary

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

Summary

- The Cisco Unified Computing System enables flexible server deployment, dynamic server provisioning, and simplified server upgrades within systems.
- Server virtualization, VDI environments, and memory-intensive applications benefit from the large memory support and adapter virtualization of the Cisco Unified Computing System.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-16

References

For additional information, refer to these resources:

- Design Zone for Unified Computing at http://www.cisco.com/en/US/netsol/ns1071/networking_solutions_sub_program_home.html
- “Data Center Designs: Application Networking—Cisco UCS and Application Delivery for Microsoft Hyper-V Virtualization of Exchange 2010 with NetApp Storage” at http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/App_Networking/hypervexchange.html
- “Data Center Designs: Application Networking—Microsoft Exchange 2010 with VMware VSphere on Cisco Unified Computing System with NetApp Storage” at http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/App_Networking/Exchange_VSphere_UCS_NetApp.html

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.

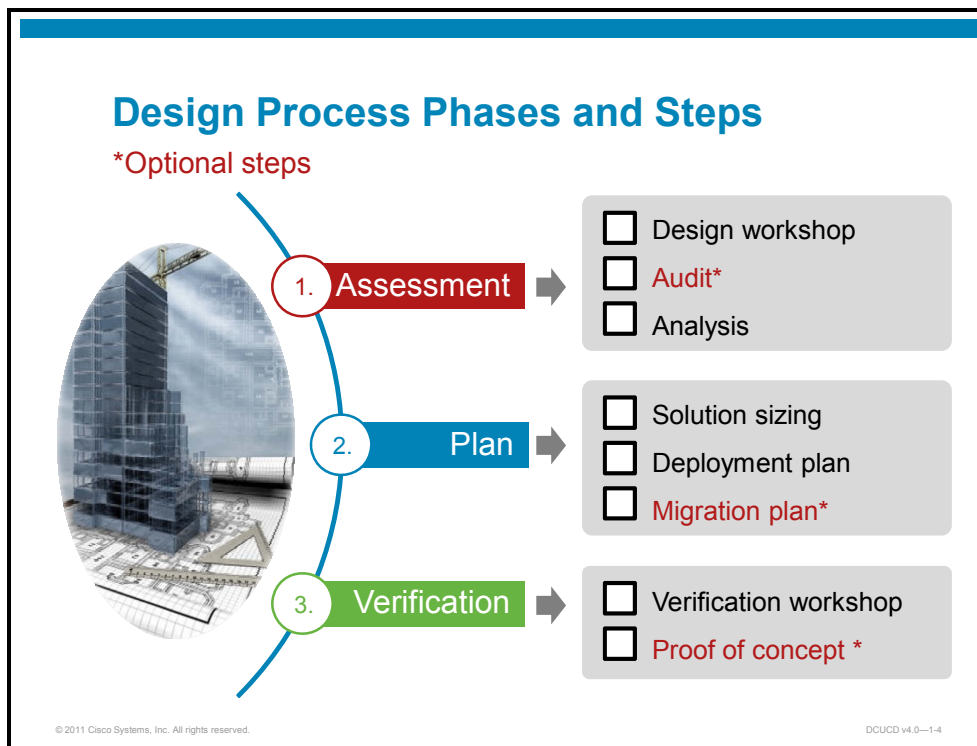
Objectives

Upon completing this lesson, you will be able to meet the following objectives:

- Describe the design process for the Cisco Unified Computing System solution
- Assess the deliverables for the Cisco Unified Computing System solution

Design Process

This topic describes the design process for the Cisco Unified Computing System solution.



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 does meet the 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 the 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 the project (for example, new deployment versus migration).

To track the design process progress and completed and open actions, the checklist in the figure can aid the effort.

Assessment Phase: Design Workshop

Conduct workshop with customer

- Who should be involved?
- One or many iterations

Design workshop agenda

- Define the business and technical goals
- Select the design process phases
- Identify the data center technologies that are included
- Project type
- Identify the requirements and limitations
- Collect relevant information

Mandatory

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-5

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 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 consequently you should ensure that the project is successful. With the list of goals, 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 likewise 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 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. They are new deployments or the migration of existing solutions.
- **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 proper reconnaissance and analysis tools.

The workshop can be conducted in person or it can be virtual 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 on which the design can be created.

People Involved

Design workshop rationale—**cover all relevant aspects of design**

Talk to all relevant customer personnel:

- Facility administrators
- Network administrators
- Storage administrators
- Server and application administrators
- Security administrators

© 2011 Cisco Systems, Inc. All rights reserved.

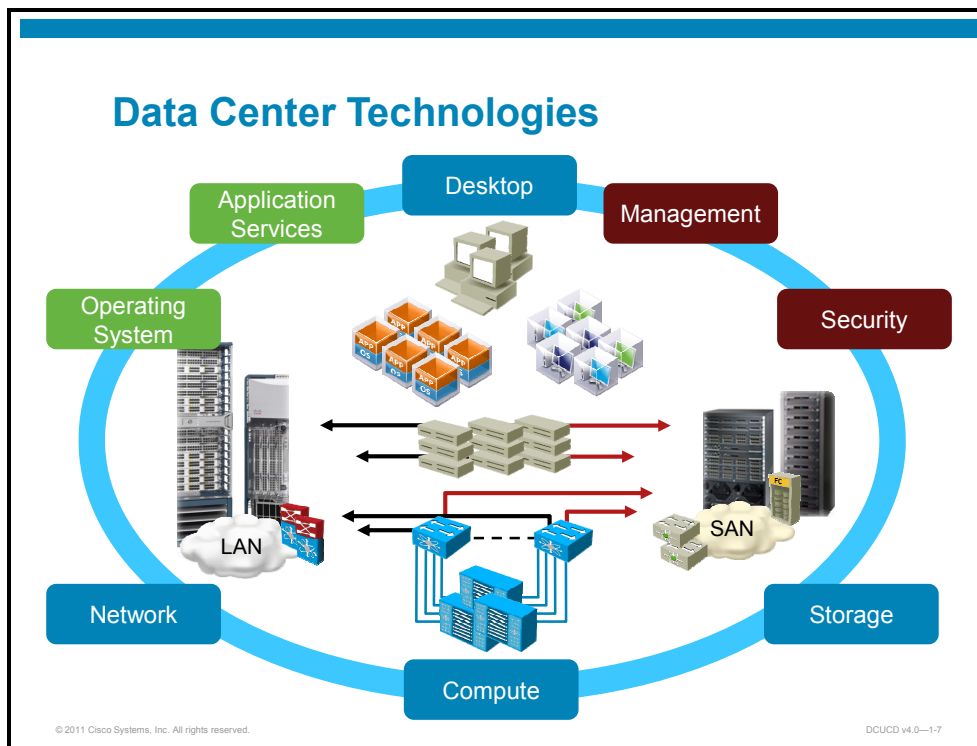
DCUCD v4.0-1-6

It is very important to gather all the relevant people in the design workshop to cover all the aspects of the solution (the design workshop can be a multiday event).

The Cisco Unified Computing System 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 Unified Computing System 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.
- **Storage administrators:** Here, the relevant information encompasses storage capacity (available and used), storage design and redundancy mechanisms (logical unit numbers [LUNs], Redundant Array of Independent Disks [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 for 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 demilitarized zone [DMZ] and private segments). The security policy also defines the control of equipment administrative access and allowed and restricted services (for example, Telnet versus Secure Shell [SSH]), and so on.



The Cisco Data Center Business Advantage architectural framework encompasses technologies that address network, storage, compute, operating system, application services, management, and security aspects of the Cisco Unified Computing System solution.

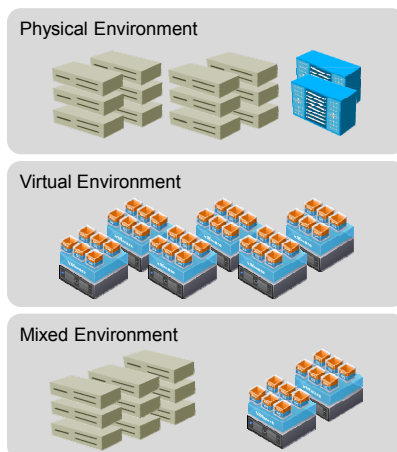
Because each project is different, in the design workshop the solution architect should define the aspects and technologies that should be part of the design with the customer.

The technologies should be paired with relevant equipment when discussing needs with the customer 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.

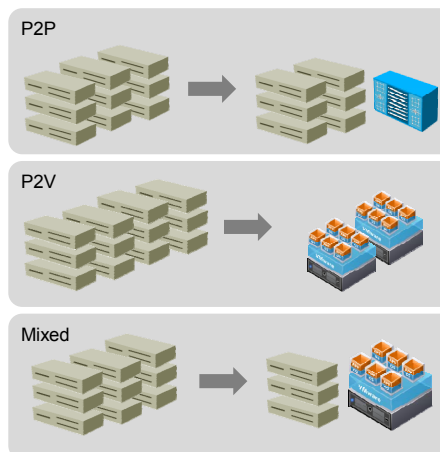
Thus, the attendees in the design workshop should discuss the Cisco Nexus switching portfolio from Nexus 7000 to Nexus 1000V, the UCS B-Series and C-Series products, Cisco MDS SAN switches, application service equipment like Cisco Application Control Engine (ACE) and Cisco Wide Area Application Services (WAAS), hypervisor integration with VMware vSphere, management applications like Cisco UCS Manager and Cisco Data Center Network Manager (DCNM), and so on

Project Type

New Deployment



Migration



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-8

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 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 life cycle and needs to be replaced with a new solution. Here, 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. Environments that may be chosen are 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 to virtual or mixed environment, or even a migration from one virtual environment to another.

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

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-9

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 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, 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).

Assessment Phase: Analysis

Baseline and optimize the requirements:

- **Compute**—memory, CPU, storage
- **Network**—throughput, redundancy
- **Storage**—space, features, IOPS
- **Operating system and applications** specifics

Set limitations

Mandatory

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-10

The analysis is the last part of the assessment phase. The solution architects must review all the collected information and then select only the important details.

The architects must baseline and optimize the requirements, which 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.

Plan Phase: Activities Checklist

- ☐ **Solution sizing**
 - ☐ Size the solution
 - ☐ Select LAN and SAN equipment
 - ☐ Calculate environmental characteristics
 - ☐ Create BOM
- ☐ **Deployment plan**
 - ☐ Physical deployment
 - ☐ Server deployment
 - ☐ LAN and SAN integration
 - ☐ Administration and management
- ☐ **Migration plan**
 - ☐ Prerequisites
 - ☐ Migration and rollback procedures
 - ☐ Verification steps

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-11

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 has to be sized and it must be decided whether B-Series or C-Series equipment should be used. This decision can be a complex process where 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 Nexus and 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.
 - Last but not least, the Bill of Materials (BOM), which is a detailed list of the equipment parts, needs to be created. The BOM does not include only the Cisco UCS or Nexus products but also all the necessary patch cables, power inlets, and so on.
- **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). Details about these designs are discussed later in the course.

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



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-12

Once the design phase is completed, the solution must be verified and approved by the customer. This approval is typically received 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.

Design Deliverables

This topic assesses the deliverables of the Cisco Unified Computing System solution.

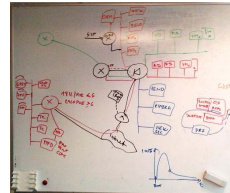
Deliverables: Assessment Phase

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



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-14

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:** This 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 the information that is gathered in the assessment phase.

Deliverables: Plan Phase

High-level design (HLD) document

- Conceptual design description
- High-level equipment list

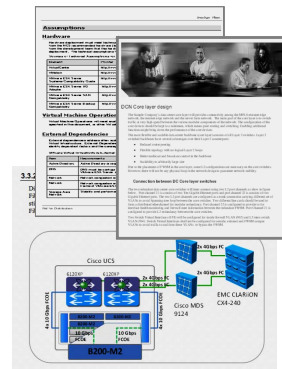
Detailed design (DD) document

- Detailed description of the design
- Bill of Materials (BOM)

Site requirements specification (SRS)

Site survey form (SSF)

Migration plan (MP) document



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-1-15

The design phase is the most document-intensive phase. It should result in the following documentation:

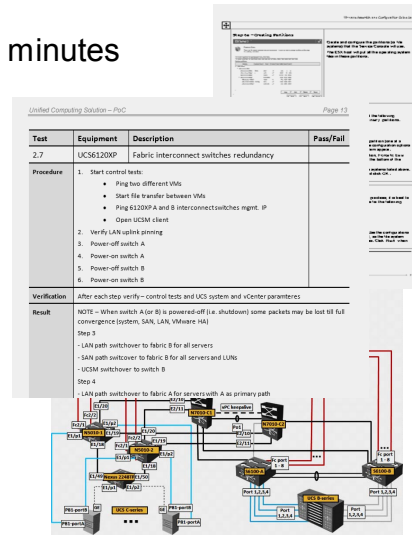
- **High-level design (HLD):** This document describes the conceptual design of the solution, such as the solution components, which equipment is used (not detailed information), how the high-availability mechanisms work, how the business continuance is achieved, and so on.
- **Detailed design (DD) (also known as a low-level design [LLD]):** This document describes the design in details, such as the detailed 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 (SSF):** 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.
- **Migration plan:** This document is necessary when the project is a migration and it must have at least 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

Deliverables: Verification Phase

Verification workshop meeting minutes

Proof of concept document

- Required resources
 - Hardware and software
 - Power, cooling, and cabling
- Lab topology
- List of tests with expected results



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 off on.

Second, 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 and it is for the equipment that will be used in the proof of concept. Apart from that, 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.

Summary

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

Summary

- The design of a Cisco Unified Computing System solution comprises several phases. Of these phases, analysis, sizing, and deployment design are necessary.
- The analysis phase should include key IT personnel: server, network, storage, application, and security professionals.
- Design deliverables are used to document each of the solution phases.

Module Summary

This topic summarizes the key points that were discussed in this module.

Module Summary

- The Cisco Data Center Business Advantage is an architectural framework for dynamic networked organizations, and it allows organizations to create services faster, improve profitability, and reduce the risk of implementing new business models.
- The Cisco Data Center Business Advantage framework includes technologies for switching, application networking, security, storage, operating system, management, and compute aspects of the data center.
- The Cisco Unified Computing System solution design process consists of assessment, plan, and verification phases.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—1-1

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: Analyzing Data Center Computing Solutions)
- A) virtualization support
 - B) available power
 - C) business demands
 - D) organizational structure
 - E) space
- Q2) Which three options are currently the benefits of cloud computing? (Choose three.) (Source: Analyzing Data Center Computing Solutions)
- A) on-demand availability
 - B) standard provisioning
 - C) efficient utilization
 - D) pay per use
 - E) customer control
- Q3) Which data center component is a network load balancer? (Source: Analyzing Data Center Computing Solutions)
- A) network
 - B) compute
 - C) security
 - D) application services
- Q4) Which key Cisco UCS feature is enabled by service profiles? (Source: Identifying a Cisco Unified Computing System Solution)
- A) large memory footprint
 - B) unified fabric
 - C) extended statelessness
 - D) separate management domain
 - E) integrated power management
- Q5) Which three features make the Cisco UCS a natural selection for VDI deployments? (Choose three.) (Source: Evaluating Server Deployment Options)
- A) server recovery
 - B) support for large memory
 - C) service profiles
 - D) fast server repurpose
 - E) virtualized adapters
- Q6) In which cloud computing service category is the Cisco VMDC? (Source: Defining a Cisco Unified Computing System Solution Design)
- A) virtual private cloud
 - B) IaaS
 - C) SaaS
 - D) PaaS

Module Self-Check Answer Key

- Q1) B, E
- Q2) A, C, D
- Q3) D
- Q4) C
- Q5) B, C, E
- Q6) B

Assess Computing Solution Requirements

Overview

This module identifies the relevant network, storage, and compute historical performance characteristics, and it discusses the relevant characteristics and functions of the reconnaissance and analysis tools for the Cisco Unified Computing System solution.

Objectives

Upon completing this module, you will be able to use the reconnaissance and analysis tools to assess the performance characteristics and requirements of the computing solutions. This includes the ability to meet these objectives:

- Assess the performance characteristics and requirements for the computing solution
- Describe the use of the reconnaissance and analysis tools to examine performance characteristics of the computing solution

Analyzing Performance Characteristics

Overview

This lesson assesses the performance characteristics and requirements for the computing solution.

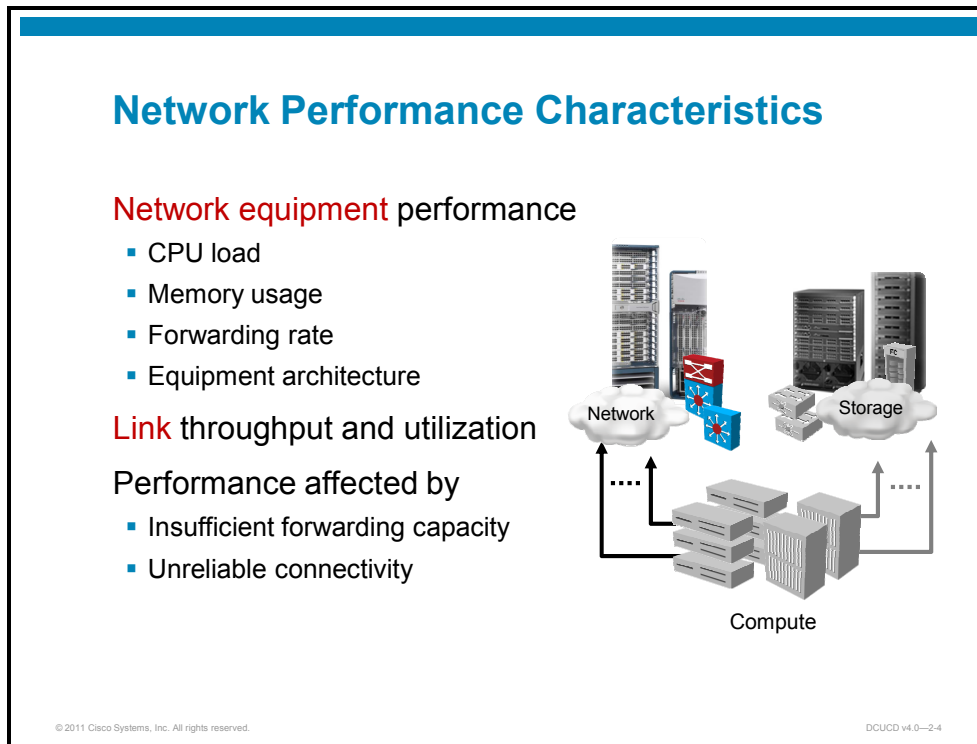
Objectives

Upon completing this lesson, you will be able to meet the following objectives:

- Describe the performance characteristics of the computing solution
- Assess the performance characteristics for the relevant operating systems and hypervisors

Performance Characteristics

This topic introduces the performance characteristics of the computing solution.



The network performance characteristics are defined by the network device performance and connections throughput.

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.

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.

DCUCD v4.0—2-5

<https://t.me/learningnets>

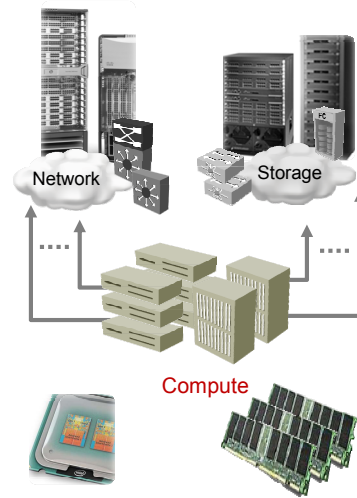
Server Performance Characteristics

CPU characteristics

- Speed
- Number of cores and concurrent threads
- Affected by operating system
- Average and maximum CPU utilization

Memory characteristics

- Size
- Access speed
- Maximum operating system addressable memory
- Internal server bus architecture
- Average and peak memory usage



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-6

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 denominator 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.

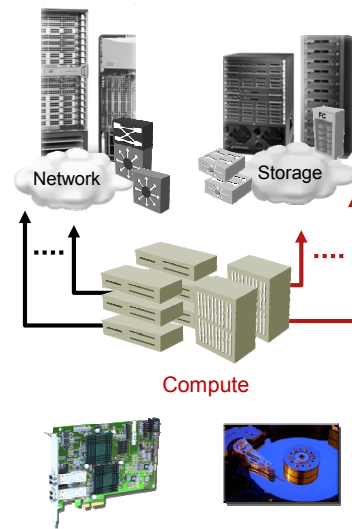
Server Performance Characteristics (Cont.)

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 and access time



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-7

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.

Application Performance

Application performance depends on the following:

- Server performance
- Application code that is used
- Application protocol

Chatty protocols incur additional processing time, thus limiting application responsiveness.

[illegible]

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-8

The performance characteristics of applications are affected by all the server components plus by the applications themselves. Applications can be written in a suboptimal way, resulting in a slow response—even if the underlying server performance is sufficient.

Assess Performance Characteristics

This topic assesses the performance characteristics for relevant operating systems and hypervisors.

Microsoft Windows Server 2003

32-bit and 64-bit operating systems for x86 platform:

- Scalable high availability
- Direct SAN access
- Resource management
- Terminal services session directory

Version	32-bit	64-bit
Windows Server 2003 SP2 Datacenter Edition	128 GB*	2 TB
Windows Server 2003 SP2 Enterprise Edition	64 GB	2 TB
Windows Server 2003 R2 Datacenter Edition	128 GB**	1 TB

*64 GB with 4-GB tuning

**16 GB with 4-GB tuning

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-10

Windows Server 2003 R2 Datacenter Edition provides support for the following:

- 2 TB of RAM (64-bit version, service pack 2 [SP2])
- 1 TB of RAM (64-bit version, service pack 1 [SP1])
- 128 GB of RAM (32-bit version)

The multiprocessor servers can scale up to 64 processors for the 64-bit versions and 32 processors for the 32-bit version.

The Windows Server 2003 incorporates several improvements over its predecessor, Windows Server 2000:

- Intel Hyper-Threading Technology
- Symmetric multiprocessing (SMP): 64-way (64-bit versions) and 32-way (32-bit version)
- Non-Uniform Memory Access (NUMA)
- Cluster service for up to eight node clusters
- Windows Sockets: Direct access for SANs
- Terminal Services Session Directory
- Windows System Resource Manager (WSRM)

Microsoft Windows Server 2003 (Cont.)

Version	32-bit	64-bit
Windows Server 2003 Datacenter Edition	128 GB*	512 GB
Windows Server 2003 R2 Enterprise Edition	64 GB*	1 TB
Windows Server 2003 Enterprise Edition	32 GB*	64 GB
Windows Server 2003 R2 Standard Edition, Standard Edition SP1	4 GB	32 GB
Windows Server 2003 Standard Edition	4 GB	16 GB
Windows Small Business Server 2003	4 GB	n/a
Windows Server 2003 Web Edition	2 GB	n/a

***16 GB with 4-GB tuning**

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0--2-11

The table in the figure lists the maximum amount of memory that is supported by certain Windows Server 2003 versions.

The 64-bit version supports more memory than the 32-bit version.

All processes that are running under 32-bit Windows gets virtual memory addresses going from 0 to 4 GB, no matter how much RAM is actually installed on the computer.

In the default Windows operating system configuration, 2 GB of this virtual address space is designated for the private use of each process and the other 2 GB is shared between all processes and the operating system. Normally, applications (such as Notepad, Word, Excel, and Acrobat Reader) use only a small fraction of the 2 GB of private address space.

The operating system only assigns RAM page frames to virtual memory pages that are in use. The Physical Address Extension (PAE) is the feature of the Intel 32-bit architecture that expands the physical memory (RAM) address to 36 bits. PAE does not change the size of the virtual address space, which remains at 4 GB. PAE adjusts only the amount of actual RAM that can be addressed by the processor. The translation between the 32-bit virtual memory address that is used by the code running in a process and the 36-bit RAM address is managed automatically and transparently by the computer hardware according to the translation tables that are maintained by the operating system. Any virtual memory page (a 32-bit address) can be associated with any physical RAM page (a 36-bit address).

At some point, the resource that is in critical short supply cannot be increased and that an architectural limit has been reached. Here are some commonly reported architectural limits in Windows:

- 2 GB of shared virtual address space for the system
- 2 GB of private virtual address space per process
- 660-MB system path terminating equipment (PTE) storage
- 470-MB paged pool storage
- 256-MB nonpaged pool storage

- 1-GB system cache
- 16,000-GB page file size

These limits apply to Windows Server 2003 specifically but also apply to Windows XP and Windows 2000. Significant changes were made to Windows Server 2003 to reduce the likelihood that these architectural limits are reached. For example, some processes that were in the kernel have been moved to nonkernel processes to reduce the amount of memory that is used in the shared virtual address space.

Microsoft Windows Server 2008

32-bit and 64-bit operating systems for x86 and Itanium platforms:

- Power management
- HAL
- Plug-and-play feature
- Security features

Version	32-bit	64-bit
Windows Server 2008 Datacenter (complete or core installation)	64 GB	2 TB
Windows Server 2008 Enterprise		
Windows Server 2008 for Itanium-based systems	n/a	2 TB
Windows Server 2008 Standard	4 GB	32 GB
Windows Web Server 2008		

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-12

Microsoft Windows Server 2008 features several kernel enhancements in key areas:

- Improved memory and heap management
- Management mechanisms, including the registry, the services model, and the new Microsoft Windows Hardware Error Architecture (WHEA)
- Security features, including kernel patch protection, code signing and code integrity, and support for protected processes
- Other application support mechanisms, including the dynamic link library (DLL) loader and the thread pool
- Support for hardware innovation, including dynamic partitioning, Advanced Configuration and Power Interface (ACPI) 2.0, Peripheral Component Interconnect (PCI), PCI Express (PCIe), and boot environment enhancements
- Power management enhancements
- Support additions for the “plug-and-play” feature (which means that the product will work as soon as it is installed)
- Hardware abstraction layer (HAL)
- Support for Extensible Firmware Interface (EFI) run-time calls

The table in the figure lists the maximum memory size that is supported by each version of Windows Server 2008.

Red Hat Enterprise Linux

32-bit and 64-bit operating systems for various platforms:

- CPU scheduler
- IPv4 and IPv6 enhancements
- SMP locking enhancements
- High-performance pipe implementation

Type	Version 3	Version 4	Version 5
X86 32-bit	64 GB*	64 GB*	16 GB**
IA 64/Itanium2	128 GB	2 TB	1 TB/1024 TB
AMD 64/Intel 64	128 GB	256 GB/1 TB	256 GB/1 TB
PPC/Power	64 GB	128 GB/1 TB	512 GB/1 TB
S390/System Z	64 GB	64 GB	64 GB

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-13

Red Hat Enterprise Linux can support various platforms, even those that are not based on Intel x86 as shown in the table in the figure.

*The SMP kernel supports a maximum of 16 GB of main memory. Systems with more than 16 GB of main memory use the HUGEMEM kernel. In certain workload scenarios, it may be advantageous to use the HUGEMEM kernel on systems with more than 12 GB of main memory.

**The x86 HUGEMEM kernel is not provided in Red Hat Enterprise Linux 5.

Red Hat Enterprise Linux 4

The Red Hat Enterprise Linux 4 kernel offers improvements over earlier kernels, including many enhanced algorithms and features:

- **Generic logical CPU scheduler:** Manages multicore and hyperthreaded CPUs
- **Object-based reverse-mapping VM:** Improves performance in memory constrained systems
- **Read-copy-update:** SMP algorithm optimization for operating system data structures
- **Multiple I/O schedulers:** Selectable based on the application environment
- **Enhanced SMP and NUMA support:** Improved performance and scalability for large servers
- **Network interrupt mitigation (NAPI):** Improved performance for heavy network loads

Red Hat Enterprise Linux 5

Red Hat Enterprise Linux 5 is based on the Linux 2.6.18 kernel and includes many enhancements over prior releases. These include the following:

- Per-queue switchable I/O schedulers
- High-performance pipe implementation
- IPv4 and IPv6 performance enhancements

- Support for shared cache and multicore systems
- Kernel SMP locking enhancements for improved scalability and performance
- Improved performance and scalability for memory, CPUs, I/O, and networking

VMware Server Virtualization

VMware vSphere 4 **host scalability**

- 256 VMs
- 64 vCPUs
- 512-GB memory

VMware **VM scalability**

Characteristic	Typical App.	ESX 3.5	ESX 4
CPU	1 to 2 CPUs	4 vCPUs	8 vCPUs
Memory	Peak <4 GB	64 GB per VM	255 GB per VM
Network	<2.4 Mb/s	9 Gb/s	40 Gb/s
IOPS	Peak <100	100,000	200,000+

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-14

VMware vSphere 4 Hypervisor Scalability

The ESX host and VM characteristics govern the VMware solution scalability.

The scalability characteristics of the ESX host define the performance and consolidation rates that are available for each physical server, and the scalability characteristics of the VM define the amount of workload that a VM can manage.

VMware ESX version 4 now supports in real time the additions of resources like memory, CPU, network connectivity, and storage. With this ability, the VMs no longer need to be shut down in order to make configuration changes.

The following details apply:

- The scalability of the ESX host is limited to the following:
 - 512-GB memory
 - 256 virtual CPUs per host
 - 256 VMs
- The scalability of the VM is limited to the following:
 - 8 virtual CPUs with virtual SMP
 - 255-GB memory

Note To be able to use large memory capacities, the ESX has the 64-bit VM kernel.

VMware vSphere 4 VM Scalability

VMware ESX 4 is the next generation hypervisor, which extends ESX 3.5 performance with the following characteristics:

- Support for VMs with up to 8 virtual CPUs
- Up to 255 GB of memory that is assigned per VM

- Sustained 40 Gb/s of network I/O
- Up to 200,000 I/O operations per second

The VMware hypervisor—ESX—uses intelligent mechanisms to address virtualization, such as a purpose-built scheduler, the use of hardware-assisted technologies for optimized CPU access, transparent page sharing and ballooning for effective memory usage, and hardware-assisted memory management for low overhead.

The ESX 4 host can be scaled to support 64 cores and up to 512 GB of physical memory (double the previous limit) to be able to support very large servers and to run them efficiently.

Virtual SMP enables the VMs to use multiple physical processors. This ability means that an administrator can assign multiple virtual CPUs to a VM.

From a management perspective, the VMware ESX host can be configured by using a specific client (for example, using Secure Shell [SSH] to access a Linux-like command-line interface [CLI]). The CLI and the commands that are available can be used for virtual switch configuration, network interface card (NIC) teaming, and so on. Typically, the VMware ESX host is configured with the VMware vCenter server (formerly VMware VirtualCenter).

VMware vCenter 4 Characteristics

When the size of the virtual infrastructure grows—that is, more ESX hosts are added to the virtual infrastructure—it becomes difficult to keep up with the management of these hosts. The VMware vCenter server is the solution for virtual infrastructure management, having the following characteristics:

- Windows-based application
- Resource management for ESX and ESXi hosts and VMs
- VM deployment and template management
- Task scheduling, statistics, logging, alarms, and events monitoring

Although VMware vCenter is not required for ESX host and VM deployment, it is required for some advanced services, such as vMotion, high availability, fault tolerance, and Distributed Resource Scheduler (DRS).

The VMware vCenter server can be installed on a physical server or VM. In either case, the requirements are the following:

- Processor = 2 GHz or faster
- Memory = 2 GB
- Disk space = minimum of 1 GB
- Network adapter = recommended 1 Gigabit Ethernet
- Operating system (32- or 64-bit) = Windows Server 2003 with SP1 or SP2, Windows Server 2003 R2, Windows Server 2008
- Database server = Oracle 10g, Oracle 11g, Microsoft SQL Server 2005 Express Edition (comes with vCenter Server), Microsoft SQL Server 2005 with SP2, Microsoft SQL Server 2008

Because vCenter is an important part of the VMware solution, it is vital to plan proper availability.

Summary

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

Summary

- The key server performance characteristics that influence application performance are CPU and memory utilization, network throughput, and storage space and access speed.
- Operating system requirements and characteristics typically increase from version to version.
- There is less support for memory per server in 32-bit operating systems than in 64-bit operating systems.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-15

Employing Data Center Reconnaissance and Analysis Tools

Overview

This lesson describes the use of reconnaissance and analysis tools to examine the performance characteristics of the given computing solution.

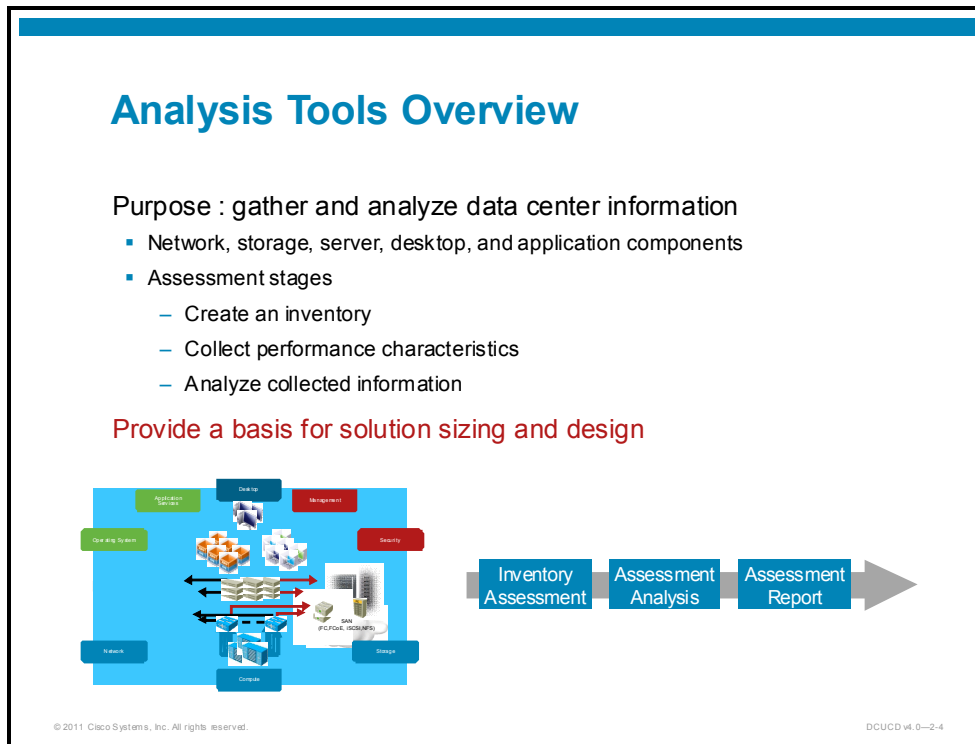
Objectives

Upon completing this lesson, you will be able to meet the following objectives:

- Identify characteristics and functions of the reconnaissance and analysis tools
- Evaluate the VMware Capacity Planner for computing solution analysis
- Evaluate the VMware vCenter CapacityIQ for computing solution analysis
- Evaluate the MAP Toolkit for computing solution analysis

Reconnaissance and Analysis Tools

This topic identifies the characteristics and functions of the 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 beforehand.

Because the data center consists of so many components and there are so many aspects, 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 complete 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).

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 is 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

Using Compute Analysis Tools

Assessment activities

1. **Create an inventory.**
 - Count and classify servers, desktops, and so on.
 - Physical and virtual resources
 - CPU, memory, storage space, and network characteristics
 - Services and applications running
2. **Assess the current utilization and workload.**
 - Run a collection engine **over time**.
 - Collect resource utilization ratios.
 - CPU, memory, storage space
 - Network throughput, **storage I/O**

The screenshot displays two panels from the DCUCD v4.0-2.6 tool. The top panel, titled 'pod-1-desktop Discovered Applications', shows a list of installed applications with columns for Product Name, Version, and Status. Applications listed include Java(TM) 6 Update 6, Microsoft .NET Framework 3.5, Microsoft Visual J++ 6.0, MSXML 6.0, PostgreSQL 8.2, VMware Tools, and VMware vCenter Client. The bottom panel, titled 'pod-1-desktop Discovered Services', shows a list of running services with columns for Service Name, Version, and Status. Services listed include .NET Runtime Optimization Service v2.0.50, Alster, Application Experience Lookup Service, Application Layer Gateway Service, Application Management, ASP.NET State Service, Automatic Updates, Background Intelligent Transfer Service, Cisco DCNM Server, ClipBook, COM+ Event System, COM+ System Application, Computer Browser, and Civiologic Services.

Product Name	Version	Status
ActivePerl 5.8.8 Build 822		
Java(TM) 6 Update 6		
Microsoft .NET Framework 3.5		
Microsoft .NET Framework 3.5		
Microsoft Visual J++ 6.0		
MSXML 6.0		
PostgreSQL 8.2	ESX-1	Success
VMware Tools	ESX-2	Success
VMware vCenter Client	ESX-3	Success

Service Name	Version	Status
.NET Runtime Optimization Service v2.0.50		Success
Alster		Success
Application Experience Lookup Service		Success
Application Layer Gateway Service		Success
Application Management		Success
ASP.NET State Service		Success
Automatic Updates		Success
Background Intelligent Transfer Service		Success
Cisco DCNM Server		Success
ClipBook		Success
COM+ Event System		Success
COM+ System Application		Success
Computer Browser		Success
Civiologic Services		Success

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

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 be 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.

- **Storage space utilization level:** This level shows how storage space requirements have grown (or have been reduced, which is 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 way over the IOPS available to the machine, that 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 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.

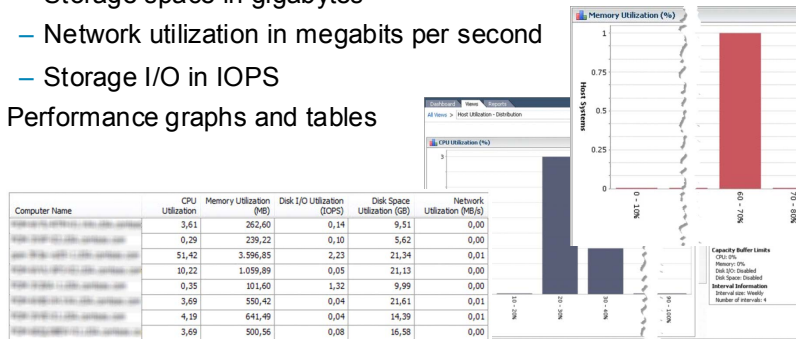
The recommended practice is to keep the collection running for at least one month, which on average gives desired results. But the interval really depends on the customer business process of the person who is conducting the analysis with the tool, and the person who uses the tool should understand the processes, activities, and applications and the ways that they are used.

Using Compute Analysis Tools (Cont.)

Postassessment activities

3. Review the assessment reports.

- Minimum to peak to 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



The postassessment activities are the third stage of the analysis and they 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 with the tool, which includes reviewing plain inventory data to 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.

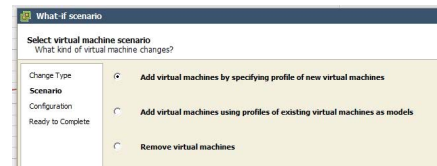
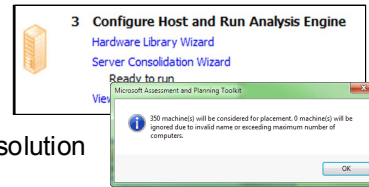
The data that is collected in the previous steps is usually saved in some sort of database that can be queried using prebuilt or custom-made queries to extract the information that is relevant to the person who is conducting the analysis.

Using Compute Analysis Tools (Cont.)

Postassessment activities

4. Help with the solution planning.

- Provide benchmarking based on references
- Plan capacity optimization
 - Server consolidation
 - Consolidation with virtualization
 - Planning per ecopartner specific solution
- Perform “what-if” analysis
 - Add or remove resources and demonstrate changes to the utilization/workload



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v1.0-2-7

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 use large knowledge databases to do the benchmarking of the environment and to present a report that proposes the capacity and infrastructure optimization, or 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 with new environment characteristics
- Server consolidation with virtualization getting even more aggressive consolidation ratios along with the information about the virtualization environment requirements like licensing

Another nice 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 to 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.

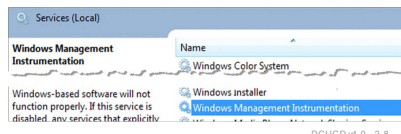
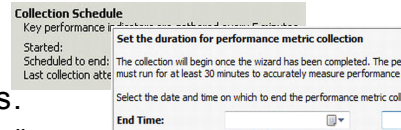
Preparing to Run Analysis

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



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-8

Before the analysis is conducted, some preparation has to be done and some 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 who is 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 before, this evaluation is important from the analysis perspective to get the relevant data. If the interval is too small it might not collect all the peak utilization times and levels, which could lead to an undersized solution if it would be based on the outcome of such analysis. The data center architect should understand the customer business process and perhaps the applications that are used in order to be able 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 to gather relevant data.

Second, in order to be able to collect the data, the data center architect should know what means the analysis tools uses to gather the data.

- The tool usually needs administrative access 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 might need an agent to be installed on the systems that are being surveyed. Typically, the tools are agentless now, so there is no need for intervention and software installation on the surveyed systems.

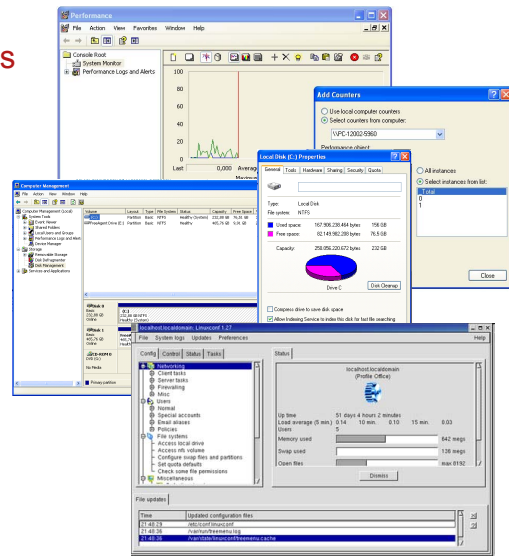
Operating System Embedded Tools

Microsoft Windows tools

- Computer management
 - Disk management
 - Volume properties
- Administrative tools
 - System performance
- Task manager

Linux embedded tools

- Linuxconf
- Webmin



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-9

The operating systems typically have built-in tools that can be used for a 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 it 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 that, you can access this tool from any user account as long as you know the root password.

Available Analysis Tools

VMware

- Capacity Planner
- CapacityIQ



Microsoft Assessment and Planning Toolkit



Akorri BalancePoint

Novell PlateSpin Recon



VKernel

- Inventory
- Capacity Analyzer
- Optimization Pack



Network Instruments Observer



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0--2-10

To conduct a more thorough analysis, the data center architect should use the purpose-built analysis and reconnaissance tools. Some of the representatives are mentioned in the following paragraphs.

VMware Capacity Planner and CapacityIQ

There are two tools available from VMware: the Capacity Planner, which is a previrtualization tool that is used mainly to analyze the pre-VMware virtualization environments. The second tool is the 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

Similar to VMware Capacity Planner and CapacityIQ is the Microsoft Assessment and Planning Toolkit (MAP). More information is provided in the following topic.

Akorri BalancePoint

Akorri BalancePoint 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 make sure that their data center server and storage virtual and physical systems deliver the best application performance possible.

BalancePoint can help reduce operations and infrastructure costs by using servers and storage more efficiently and reducing 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. Deeply buried contention, hotspots, and bottlenecks can be found. In addition, when problems are found, it provides direct automatic troubleshooting analysis.

- Optimize application performance and resource utilization. You can manage IT as a business with performance indicators that indicate the optimal balance between application requirements and resource capabilities.

The main capabilities of BalancePoint derive from three key components:

- **ScanPoint:** This component is an agentless discovery and data collection that provides performance and utilization data about the operating system and VMware, databases, servers, and storage infrastructure resources. This data is stored in a self-managing internal performance database that is mined for historical patterns, performance trends, and modeling baselines.
- **ViewPoint:** This component provides unique performance topology visualizations that show the performance impact on data center applications that run through all the physical and logical layers of the IT infrastructure, including server and storage virtualization. These topological views are automatically built from ScanPoint data, and they are color-coded with the BalancePoint assessment of performance problems.
- **GuidePoint:** This component provides actionable recommendations and analyses that are based on the Akorri Cross-Domain Analytics technology. These mathematically advanced algorithms provide intelligent alerting and proactive service management tools to help with remediation, optimization, and planning.

Novel PlateSpin Recon

PlateSpin, now part of Novell, offers reconnaissance tools that can be used for tasks similar to those performed by VMware Capacity Planner and Akorri BalancePoint.

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 Tools

VKernel offers multiple tools 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.

Employ VMware Capacity Planner

This topic evaluates the VMware Capacity Planner for computing solution analysis.

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 are required?
- What is the number of VMware infrastructure licenses that are required?



© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0—2-12

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.

Capacity Planner Functions

Discovery and collection of **inventory and performance metrics**

- Collect server and desktop inventory data
- Measure and benchmark performance metrics
- Create system utilization profiles (peak utilization)
- Run consolidation scenarios and “what-if” analyses
- Trend and predict future virtualized workloads

Leverage options

- **Consolidation estimates**—optimized for presales sizing estimates
- **Capacity assessments**—optimized for deep analysis of customer data center environment

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-13

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:

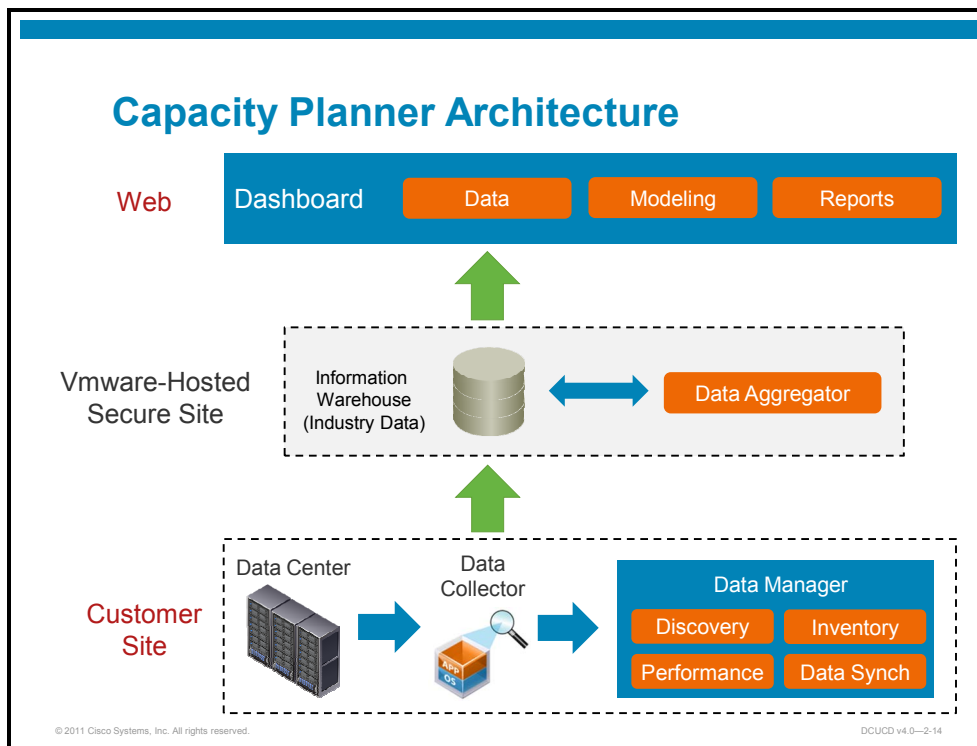
- Comprehensive collection
 - Simple collector
 - Supports up to 500 systems
 - Easy to scale by adding collectors as needed
 - Metrics reconciliation in the dashboard
 - Easy administration
 - No maintenance
 - Remote monitoring
 - Robust collection
 - Support for many operating systems
 - Inventory
 - 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

- 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.



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, for scenario modeling, and for 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.

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
Processor	At least 1-GHz CPU
Memory	At least 512 MB
Storage space	At least 250 MB
Settings	English language Firewall deactivated

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-15

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 complementary.

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, or 2003
- At least 1-GHz CPU
- At least 512 MB of RAM
- At least 250 MB 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 reserved.

Technical Considerations

Agentless discovery

- Active Directory, IP scanning, DNS queries, and NetBIOS

Data Collector—installed on own system

- Up to 500 systems monitored per individual instance supported
- Use multiple instances if necessary

Operating System	Data Source
Windows 3.1 or higher	WMI, Registry and Perfmon API calls
Linux or UNIX	SSH to run standard system utilities, SCP to collect data

Transmitting data to the Information Warehouse

- CSV files uploaded with HTTPS by using SSL encryption
- Synchronization interval (one-hour default)

SCP = Secure Copy
© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-16

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, 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 Win 3.1 or higher; 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 NT 3.51, NT 4.0, 2000, and 2003; Red Hat Linux 8 and 9 and Enterprise Linux 3 and 4; SUSE Linux 8, 9, 10, and Enterprise Server 9; Hewlett-Packard UNIX (HP-UX) 10.x, 11.0, 11.11, 11.22 (PA-RISC) and 11.23 (Itanium); and Solaris 7, 8, 9, and 10 (Scalable Processor Architecture [SPARC]) and 9 and 10 (x86) operating environments.

Information Transfer

When gathered, the data is uploaded to the Information Warehouse in comma-separated (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 time, the data is synchronized in intervals (by default once every hour).

Technical Considerations (Cont.)

Requirements for systems being analyzed

Windows

- Global, domain, or local **administrative account**
- Network connectivity—**TCP/UDP ports 135, 137–139, 445** must be open
- Enabled **Windows Management Instrumentation (WMI)**

Linux or UNIX

- **Root** access
- Network connectivity—**access to port 22**
- **SSH daemon** running

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0–2-17

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 system. 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/User Datagram Protocol (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 its SSH connection.

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.**

- CSV 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.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0–2-18

The Capacity Planner Data Collector systematically discovers domains and potential target systems in those domains, then it 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).

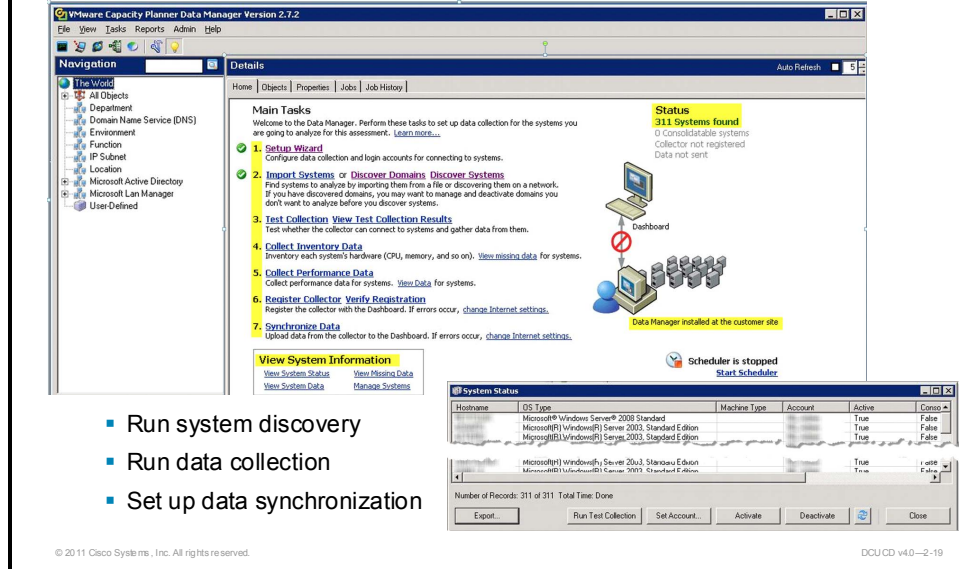
After collecting the inventory and performance data, Capacity Planner makes the data anonymous, then transmits the data over a secure connection to the Information Warehouse where the Data Analyzer aggregates it. The Data Collector sends the data to the VMware data center in CSV files via an HTTPS connection 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 contain domain names and server names.

The data is retained and available until it is archived after one year.

Conducting Analysis: Data Manager



- Run system discovery
- Run data collection
- Set up data synchronization

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

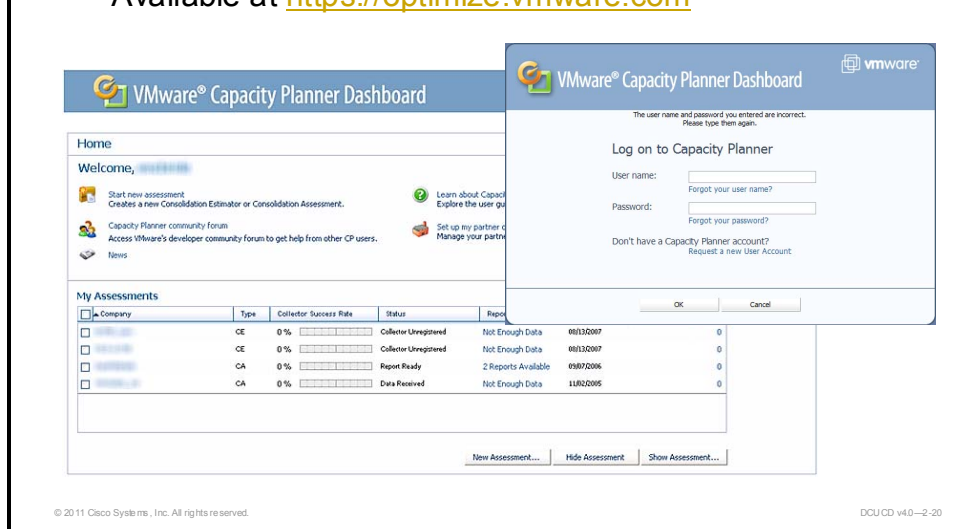
The fact that the Data Collector discovers a system or node in the network does not mean that inventory or performance data must be collected from that system or node. Likewise, a node that is inventoried might not have performance data collected from it. The number of discovered nodes is often greater 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 method 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.

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. After that, the Capacity Planner synchronizes data automatically every hour. A custom time interval or even a manual setting can be used for subsequent synchronizations.

Conducting Analysis: Dashboard

Available at <https://optimize.vmware.com>



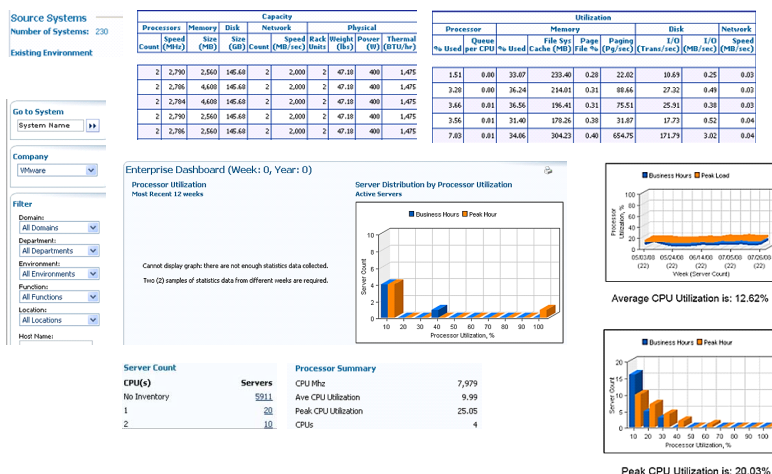
Once the assessment phases are complete, the architect can log in to <https://optimize.vmware.com> with a username and password that are obtained after the two-day training.

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.

Conducting Analysis: Dashboard

Review the existing inventory and performance data.



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.

Conducting Analysis: Scenarios

- Create and run a scenario.

Scenario Options

Cross Boundaries: Departments, Environments, Functions, Locations, Operating System

Merge Architectures: Merge x86, IA64 and AMD64 **Virtualization:** ESX Server 3.x; **Redeployment:** Re-deploy to new hardware only

New Hardware Configuration

	Capacity									
	Processors		Memory		Disk		Network		Physical	
Make/Model	Count	Speed (MHz)	Word Length	Size (MB)	Size (GB)	I/O (MB/sec)	Speed (Mbps)	Rack Units	Weight (lb.)	Power (W) (BTU/hr)
Template/Server	4	3,060	32	32,768	600	320	100	4	1,000	40

Max Load Thresholds

Utilization Limits									
Processor		Memory			Disk		Network		
% Used	Queue per CPU	% Used	Cache (MB)	Page File %	Paging (Pg/sec)	I/O (MB/sec)	Trans (sec)	I/O (MB/sec)	Speed (Mbps)
50	4	90	572	100	30,000	1,000	500	1,000	1,000

- Review the recommended server consolidations.

Optimization Opportunities

Consolidation: 14
Anomalies: 0
Alerts: 0
Trend Deviations: 0

Conservative

Make: VMware
Model: VMware Moderate Hardware Server
CPU: 4
Memory: 16 GB

Aggressive

Make: VMware
Model: VMware Aggressive Hardware Server
CPU: 8
Memory: 32 GB

Before Virtualization		With VMware Virtualization						
Total Servers	Eligible Servers	Consolidation Scenario and Platform	ESX Hosts	ESX CPU Utilization	ESX Memory Utilization	Average Memory Per VM	Racks Saved	Eligible Server Consolidation Ratio
23	13	Conservative	5	10 %	20 %	5.75 G	1	62 %
23	13	Aggressive	5	6 %	16 %	11.25 G	0	35 %

© 2011 Cisco Systems, Inc. All rights reserved.

DCU CD v4.0--2-22

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 pre- and postvirtualization server quantities.

The architect can also create and run different scenarios. For this option, these parameters of the scenario have to be entered:

- Processor architecture, virtualization type, deployment type (new versus reused hardware)
- Hardware configuration for new servers if used

Employ VMware vCenter CapacityIQ

This topic evaluates VMware vCenter CapacityIQ for computing solution analysis.

VMware vCenter CapacityIQ Overview

Analyze, forecast, and plan virtualized data center or desktop capacity

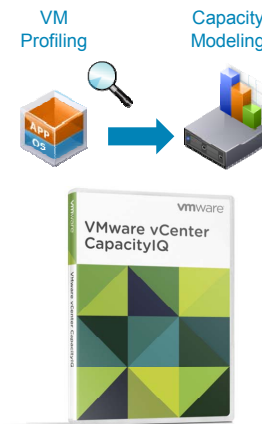
Postvirtualization tool

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?



© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0—2-24

VMware vCenter provides a set of management vServices that greatly simplify 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.

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?

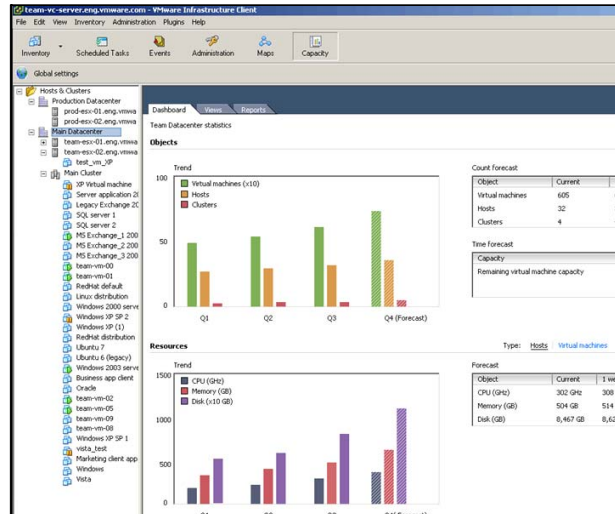
© 2011 Cisco Systems, Inc. All rights reserved.

DCU CD v4.0--2-25

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?

CapacityIQ: Examining Current Resources

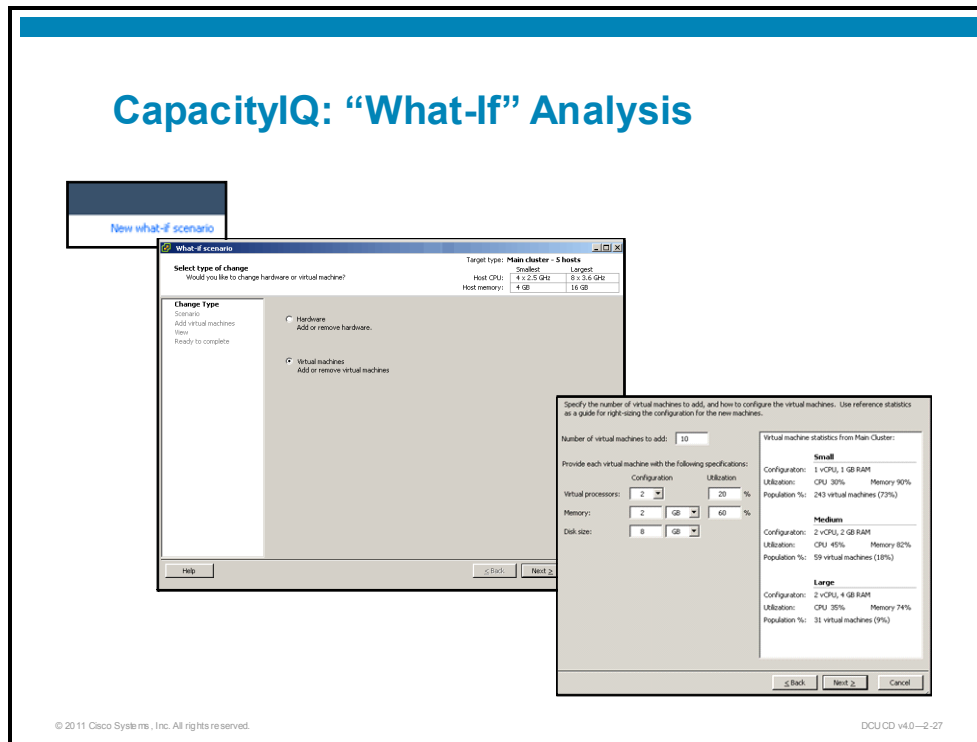


© 2011 Cisco Systems, Inc. All rights reserved.

DCU/CD v4.0—2-26

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.

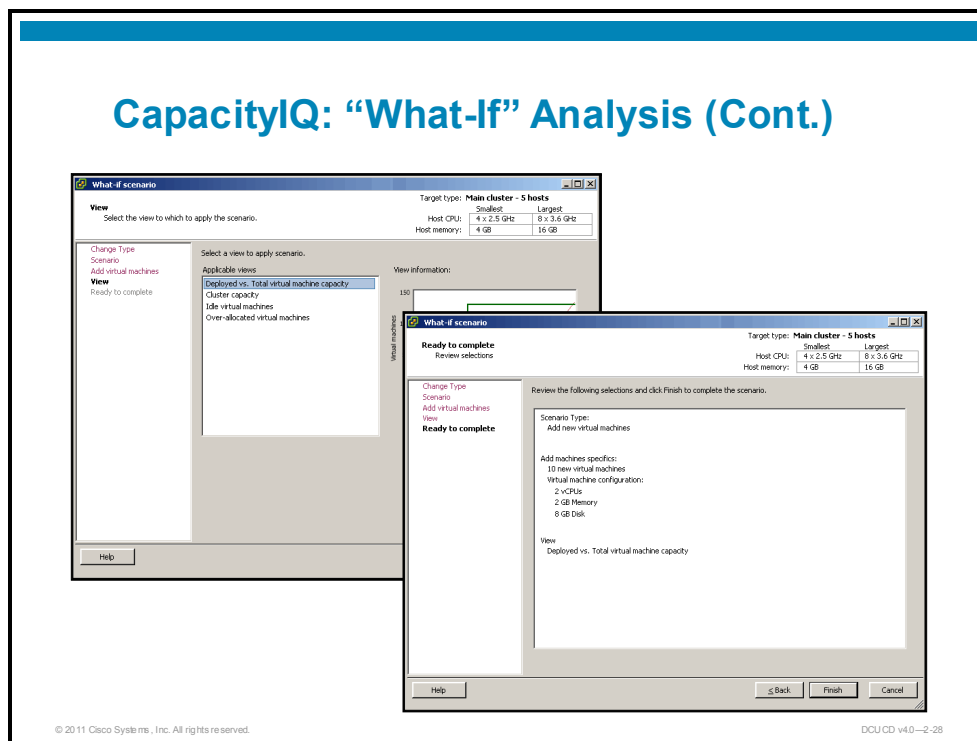
CapacityIQ: “What-If” Analysis



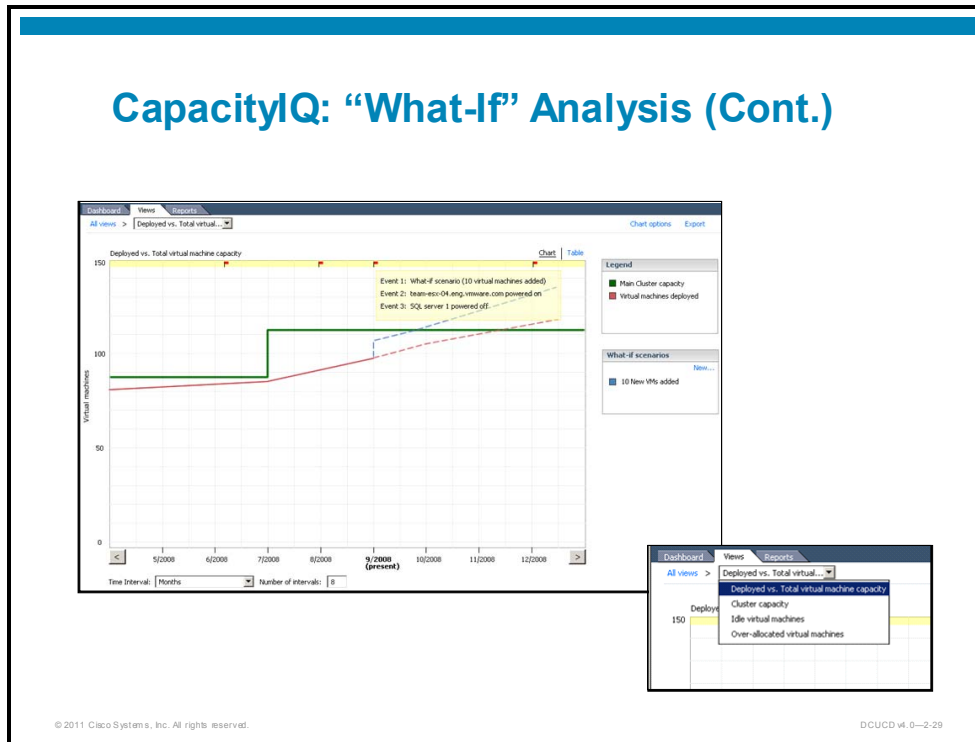
Second, a “what-if” analysis can be started. For example, what happens if 10 new VMs are added?

- 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 kind 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.

CapacityIQ: “What-If” Analysis (Cont.)



Choose how you would like to view the results, review your selections, and click **Finish** to complete the scenario.

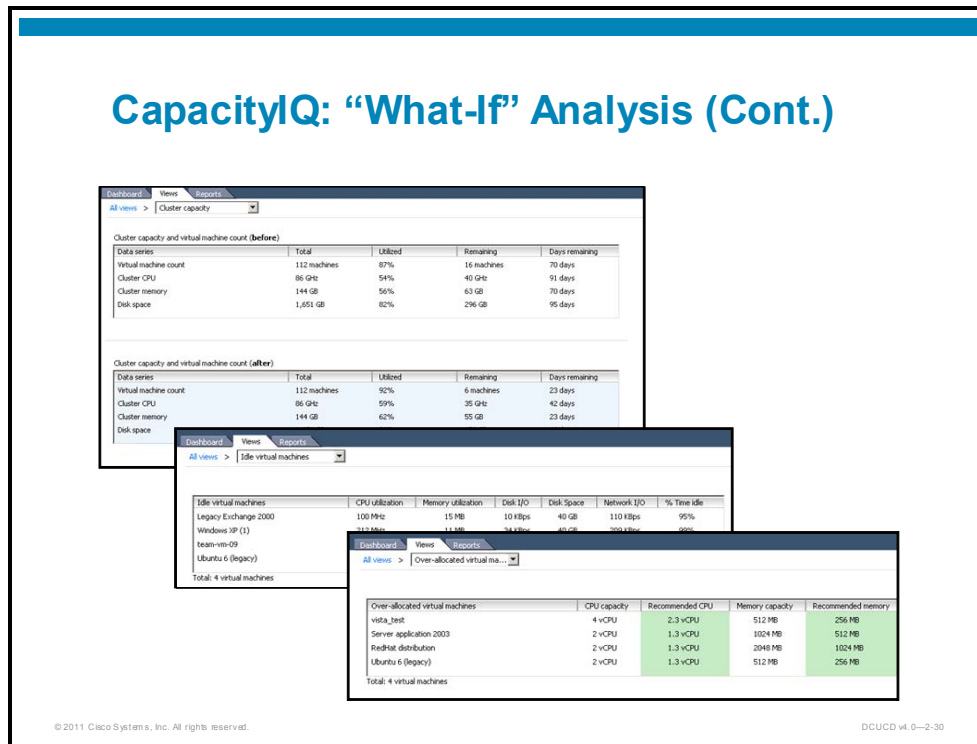


The “what-if” scenario result for the deployed versus total VM capacity (in the figure) shows that, at the current provisioning rate, capacity will run out in 60 days (this rate is represented by the red line).

If 10 new VMs are deployed today, capacity would run out in 23 days.

You can choose alternate views to examine additional information.

CapacityIQ: “What-If” Analysis (Cont.)



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 >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.

CapacityIQ: “What-If” Analysis Results

Understand the current capacity usage

- How much capacity is being used right now?
 - 97 VMs on a cluster that is 87 percent complete

Forecast future capacity needs

- How many more VMs can be added?
 - 16 more VMs can be added.
 - If the trend continues, new capacity must be added in 70 days.

Predict the impact of capacity changes

- What happens to capacity if more VMs are added?
 - Adding 10 more VMs depletes cluster capacity in 23 days.

Maximize the utilization of existing capacity

- How much capacity can be reclaimed?
 - There are four idle and four overallocated VMs. You can reclaim 2 GB of memory.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-31

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.

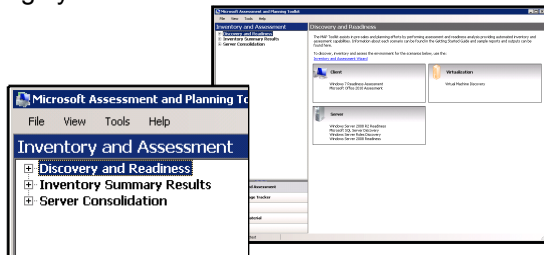
Employ the MAP Toolkit

This topic evaluates the MAP Toolkit for computing solution analysis.

MAP Toolkit Overview

Inventory, assessment, and reporting tool

- To assess current IT infrastructure
- Determine the right Microsoft technologies (optional)
- Agentless
- Version 5.0
- Create an inventory of computer hardware, software, and operating systems



© 2011 Cisco Systems, Inc. All rights reserved. DCUCD v4.0--2-33

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 networkwide 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, 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 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.

For server consolidation and virtualization through technologies such as Hyper-V and Virtual Server 2005 R2, 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



MAP can inventory the following technologies:

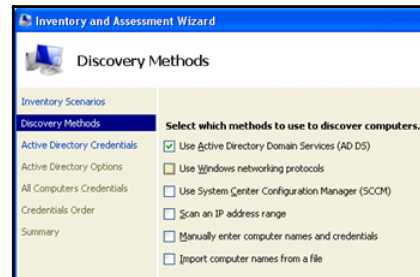
- Windows desktop operating systems, which include Windows 7, Vista, and XP Professional
- Office 2010 and previous versions
- Windows Server versions including 2008 or 2008 R2, 2003 or 2003 R2, 2000 Professional, or 2000 Server
- VMware ESX/ESXi and VMware Server
- Linux variants
- SQL Server 2008

MAP uses WMI, Active Directory Directory Services (DS), 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.

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



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-35

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 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 small number of specific computers.

MAP Advanced Functionality

Advanced analysis for the most widely used Microsoft technologies:

- Microsoft Office 2010 readiness assessment
- SQL Server discovery and assessment for consolidation
- Windows 2000 migration assessment

Software usage tracking for the following:

- Windows Server
- SharePoint Server
- System Center Configuration Manager
- Exchange Server
- SQL Server

Heterogeneous IT environment inventory

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0--2-36

MAP version 5.5 and higher include the following analysis for simplified IT infrastructure planning:

- Heterogeneous IT environment inventory
- Software usage tracking for Windows Server, SharePoint Server, System Center Configuration Manager, Exchange Server, and SQL Server
- Microsoft Office 2010 readiness assessment
- SQL Server discovery and assessment for consolidation
- Windows 2000 migration assessment
- Windows 7 readiness assessment
- Windows Server 2008 readiness assessment
- Windows Server 2008 R2 readiness assessment
- Virtual machine discovery
- Microsoft SQL Server discovery
- Computer and product inventory
- Server consolidation
- MySQL, Oracle, and Sybase discovery
- Internet Explorer assessment
- Windows Azure Platform assessment

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
- Microsoft Office Excel and Word 2007 SP2 or 2010
- .NET Framework 3.5SP1 (3.5.30729.01)
- Windows Installer 4.5

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-37

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 with Service Pack 2 or later
 - Windows Server 2003 R2
 - Windows Server 2003 with Service Pack 1 or later

- 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

Step 1. Collect the inventory data.

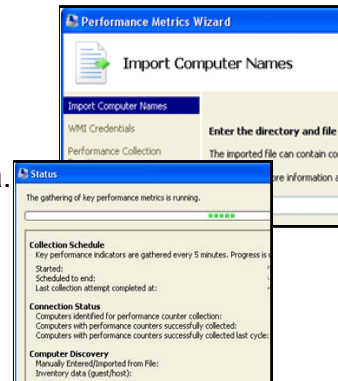
- Run the **Inventory and Assessment Wizard**.

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.



© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0-2-38

The MAP Toolkit can be used for analysis and reconnaissance in three steps:

- Running the inventory discovery
- Running the performance data collection
- Reviewing the collected data and 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.

Summary

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

Summary

- Reconnaissance and analysis tools are used to collect and analyze the relevant historical performance characteristics for migration projects.
- Advanced tools provide server virtualization scenarios that help to identify underutilized resources and hardware specifications for successful server consolidation.
- VMware Capacity Planner collects comprehensive resource utilization data in heterogeneous environments, and it compares this data to industry-standard reference data to provide analysis and decision support modeling.
- The MAP Toolkit also provides usage information and readiness assessment for the most widely used Microsoft technologies (SQL Server, Windows 2000 migration, Windows 7, Office 2010, Hyper-V, and so on).

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0--2-39

Module Summary

This topic summarizes the key points that were discussed in this module.

Module Summary

- Reconnaissance and analysis tools can collect information about the resource utilization levels and inventory of the existing environment.
- 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.
- VMware CapacityIQ is a postvirtualization tool that is used to analyze, forecast, and plan the virtual infrastructure.

© 2011 Cisco Systems, Inc. All rights reserved.

DCUCD v4.0—2-1

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 of the following affects actual server performance? (Source: Analyzing Performance Characteristics)
- A) application chattiness
 - B) operating system management
 - C) CMOS
 - D) BIOS
- Q2) Which tool can be used to evaluate requirements for converting physical Windows 2003 servers into virtual machines? (Source: Employing Data Center Reconnaissance and Analysis Tools)
- A) VMware CapacityIQ
 - B) VMware Capacity Planner
 - C) Webmin
 - D) CLI
- Q3) Which two of the following are required to conduct an analysis with the Capacity Planner? (Choose two.) (Source: Employing Data Center Reconnaissance and 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
- Q4) What is required by the MAP Toolkit to examine Windows servers in the existing environment? (Source: Employing Data Center Reconnaissance and 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

Module Self-Check Answer Key

- Q1) A
- Q2) B
- Q3) B, C
- Q4) E