



Server Consolidation Using Cisco Unified Computing System and EMC CLARiiON Storage

Benefits of Cisco Extended Memory Technology for Consolidation of Legacy Microsoft SQL Database Servers Using VMware vSphere 4

White Paper

June 2010, Revision 1.0



Contents

1 Executive Summary	3
2 Introduction	4
3 Motivation for Server Consolidation	5
The Data Center Challenge	5
Technology Trends	5
Virtualization	5
Multicore Processors	6
Power and Cooling	6
4 The Solution: Cisco Unified Computing System With EMC CLARiiON	8
The Cisco Unified Computing System	8
Innovations That Support Business Benefits	8
Cisco Unified Computing System Components	9
Cisco UCS 6100 Series Fabric Interconnects	9
Cisco UCS 2100 Series Fabric Extenders	9
Cisco UCS Network Adapters	10
Cisco UCS Manager	10
Cisco UCS 5100 Series Blade Server Chassis	10
Cisco UCS B200 M1 and M2 Blade Servers	10
Cisco UCS B250 M1 and M2 Extended Memory Blade Servers	10
Cisco Extended Memory Technology	11
EMC CLARiiON	12
5 Server Consolidation Case Study	14
Case Study-Microsoft SQL Server Using VMware vSphere	14
Scenario	14
Workload	14
Details of the System Under Test	15
Cisco UCS Platform Configuration Details	16
Setting Up and Configuring EMC Storage	16
Performance Results	17
Legacy Server Details	17
6 Summary Comparison	21
7 Conclusion	22
8 Appendix	23

1 Executive Summary

Server consolidation has gathered a great deal of momentum in the marketplace. Companies of all types are struggling with increasing infrastructure demands and changing business practices while often being constrained by smaller IT budgets. As a result, consolidation is increasingly becoming a central theme in many companies' IT strategies. These trends are accelerated by the challenging economic environment.

Consolidation is performed for many reasons; however, the main factors include the desire to reduce costs, increase IT efficiency, and improve service levels. Advances in server and storage technologies are presenting IT users with an opportunity to extend the scope of their consolidation activities further into their IT infrastructure. In particular, mainstream adoption of virtualization technologies, multicore processors, and better energy efficiency techniques creates new and deepening consolidation opportunities in the marketplace.

Even though virtualization has created a market transition, customers are facing challenges in achieving the full benefits of the technology. IT departments are constantly working against existing rigid, inflexible hardware platforms. To address these challenges and also as a natural next step in the Cisco® Data Center 3.0 vision, Cisco has developed the Cisco Unified Computing System™ (UCS). The Cisco Unified Computing System is the next-generation data center platform that unites compute, network, and storage access resources. The platform, optimized for virtual environments, is designed using open industry-standard technologies and aims to reduce total cost of ownership (TCO) and increase business agility. Cisco UCS servers are powered by Intel® Xeon® processors.

While much of the value of the Cisco Unified Computing System is derived from its architecture and integration of network, server, and storage resources and management, Cisco has also pioneered breakthrough hardware capabilities, among them the capability to expand the memory capacity of a standard Intel Xeon two-socket server to 48 DIMM slots, far beyond the commonly accepted industry standard of twelve 18-DIMM slots. This expansion is designed to increase performance and capacity for the most demanding virtualization and large-data-set applications. By enabling denser consolidations, these large-memory servers improve the server price-to-performance ratio, which ultimately yields a quicker and higher return on investment (ROI) with a lower TCO.

This document demonstrates the benefits of the Cisco UCS platform with EMC CLARiiON for server consolidation with virtualization. It discusses consolidation of older Microsoft SQL database servers onto a Cisco UCS platform using VMware vSphere 4.0. The document compares the Cisco UCS B200 and B250 platforms, and results show that the Cisco UCS B250 M1 and M2 Extended Memory Blade Server platforms can consolidate more than twice the number of virtual machines as a standard server platform using the same processor. The study was done for both the M1 and M2 generations of Cisco UCS B-Series Blade Servers.

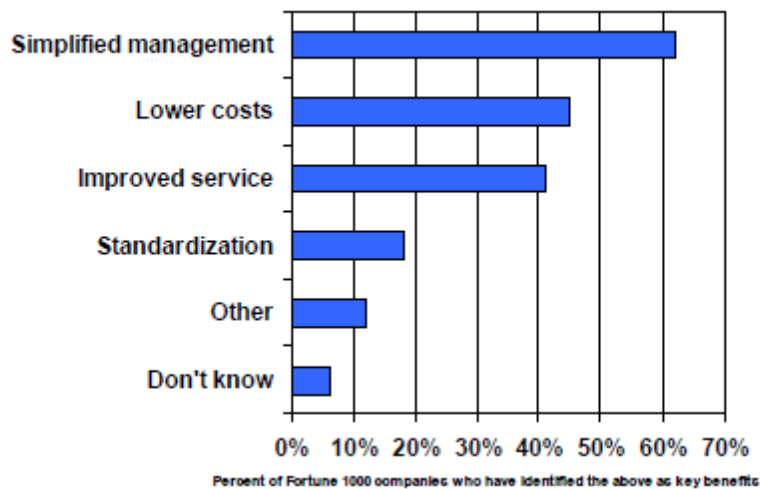
2 Introduction

Server consolidation provides a solution to reduce capital expenditures (CapEx) and operating expenses (OpEx) associated with running servers. The main reasons that companies undertake server consolidation are to simplify management by reducing complexity and eliminating server sprawl; to reduce costs, such as hardware, software, staffing, and facilities costs; and to improve service.

Server consolidation is not a new phenomenon and has been occurring across all sizes of organizations and industries for many years. But recent trends in the server industry such as virtualization, multicore processors, and advanced power management technologies have rapidly accelerated server consolidation because they are delivering exceptional CapEx and OpEx savings to corporate IT.

Figure 1 shows the main reasons cited for undertaking server consolidation projects according to a Forrester report on Fortune 100 companies.

Figure 1. Main Reasons for Server Consolidation



As with other parts of business, the success of an information technology consolidation effort is often measured by cost reduction. However, it is clear from the Forrester study that other dimensions, such as business agility, improved responsiveness, and lower management overhead, are critical as well.

The rest of this document details the server consolidation study that was performed on the Cisco UCS platform with EMC CLARiiON storage. VMware vSphere was used as the hypervisor, and a Microsoft SQL Server-based database application was used as the workload.

3 Motivation for Server Consolidation

The Data Center Challenge

Data powers essentially every operation in a modern enterprise, from keeping the supply chain operating efficiently to managing relationships with customers. The past decade has seen exceptional data growth, resulting in larger and increasingly more complex data centers. IT departments faced with this trend over the years have had to deal with many challenges:

- **Increasing TCO:** Updating, maintaining, and operating a data center is costly. A major concern is how to lower the TCO and increase ROI effectively and efficiently.
- **Server proliferation:** With the increasing capabilities, easy deployment, and low cost of servers, server proliferation has occurred in most data centers. This proliferation has resulted in server sprawl, adding to the overall cost and inefficiency of data centers.
- **Poor server utilization:** Older servers are underutilized, resulting in increased costs and affecting support and system, power, and cooling maintenance.
- **Licensing and contract issues:** Most IT departments are challenged to oversee all the many server and application licenses and contracts, each with its own expiration date. Managing a data center is difficult while also ensuring compliance.
- **Power and cooling:** Due to the significant increase of servers in a data center, power and cooling costs have increased dramatically. Other associated challenges are server overload, overheating, and loss of redundancy.
- **Data center management:** Data centers have become complex environments with numerous management points. Most current blade systems have separate power and environmental management modules, adding cost and management complexity. Ethernet network interface cards (NICs) and Fibre Channel host bus adapters (HBAs), whether installed in blade systems or rack-mount servers, require configuration and firmware updates. Blade and rack-mount server firmware must be maintained, and BIOS settings must be managed for consistency. As a result, data center environments have become more difficult and costly to maintain, while security and performance may be less than desired. Scaling out IT infrastructure using these systems is costly in terms of the number of I/O interfaces that each chassis must support, the power and cooling required, the administrative and management overhead of individual blade servers, and the business agility lost due to delayed deployment times.

Technology Trends

Data centers have grown exponentially over recent years, and technology has been developed to address many of the critical challenges faced by IT departments today. Virtualization, multicore processors, and advances in power and cooling management have dramatically reduced overall TCO while increasing ROI.

Virtualization

Data centers have enthusiastically adopted virtualization since its introduction in 2001. Virtualization delivers exceptional CapEx savings on hardware, power, cooling, and space through server consolidation. Over the years, virtualization has matured and advanced and now delivers operating efficiencies while propelling data center transformation.

Consequently, IT departments around the world have moved beyond data center and infrastructure consolidation to virtualization. According to a study published by VMware, this trend is continuing as the majority of IT executives are planning to virtualize more than 50 percent of their computing infrastructure within the next 2

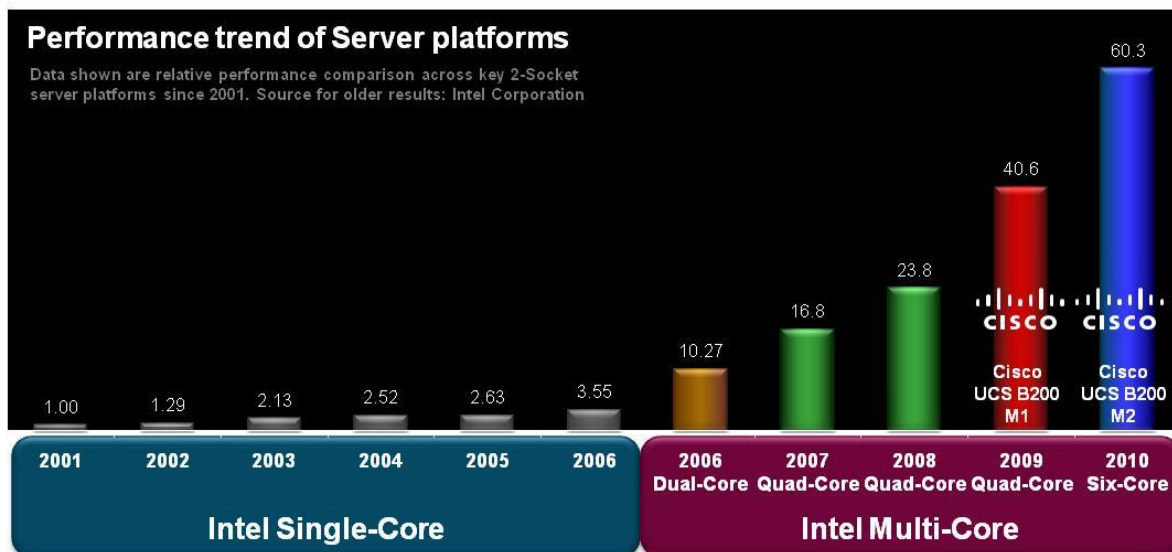
years; 70 percent of these executives are implementing virtual machine mobility for better load balancing and business continuity. The average enterprise that deploys virtualization can expect an ROI of more than 400 percent and payback within 12 months. This payback is largely the result of server consolidation, increased utilization, and higher availability.

Multicore Processors

Data centers are increasingly deploying dense, multicore, virtual machine-intensive servers. Multicore technology harnesses multiple processors onto one semiconductor, effectively increasing their collective capabilities. This approach offers a huge advantage, since most processors can handle only one thread at a time. Multicore technology improves data center system efficiency and application performance for computers running multiple applications at the same time.

With multicore technology, data centers can achieve new levels of energy-efficient performance, dividing the power typically required by a higher-frequency single-core processor with equivalent performance. Figure 2 shows the trend and effect of multicore processors.

Figure 2. Performance Trends for Server Platforms



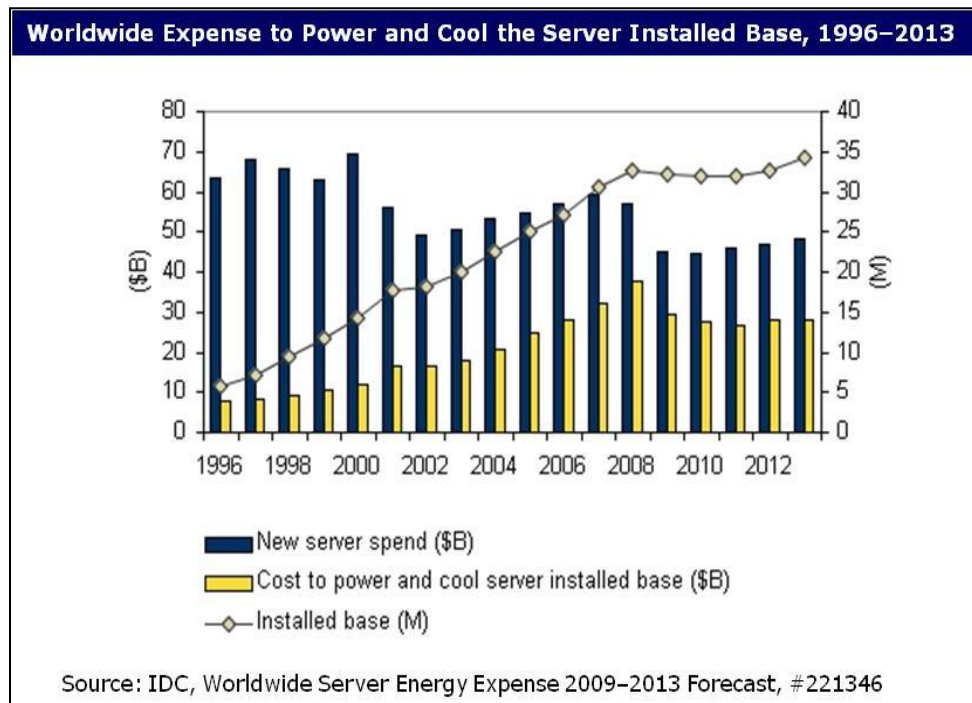
As the figure shows, the first half of the decade (2001 to 2006) saw a 3X growth in the compute power of servers, and the second half of the decade (2006 to 2010) saw a 15 to 20X growth, all propelled by multicore technology.

Power and Cooling

For many years now, IT departments have experienced major growth; server sprawl is increasing system management costs and outstripping available data center space and power and cooling capabilities. A recent study found that 42 percent of data center owners said they would exceed their power capacity within the next 1 or 2 years, and 39 percent said they would exceed their cooling capacity in the same time frame. Energy efficiency is becoming a critical issue in the data center.

A study by IDC shows that the cost to power and cool servers has increased from about 10 percent of new server spending to about 60 percent of new server spending, quantifying the severity of the challenge (Figure 3). In 2009, this cost was about US\$29 billion.

Figure 3. Power and Cooling Expenses



Data centers consume more power than any other IT environment, and the costs of providing power, cooling, and supporting facilities are increasing rapidly. Some analysts predict that in the coming year, energy costs will be second only to payroll for many large enterprises.

Today's power and cooling solutions can help data centers reduce the budgetary impact of rising power costs and corporate sustainability and compliance initiatives, thus enhancing the IT department's ability to support sustainable business growth.

The virtualized approach can increase service density and efficiency for networks, servers, and storage. This approach combines the manageability and energy and space savings of a centralized infrastructure with the flexibility of a locally distributed system to:

- Increase the operational efficiency and utilization of service-delivery infrastructure
- Reduce power consumption and cooling requirements
- Eliminate underutilized and low-value infrastructure

These solutions help eliminate the need for duplicate IT infrastructure and parallel storage and computer networks, significantly reducing cabling and switching costs.

4 The Solution: Cisco Unified Computing System With EMC CLARiiON

The Cisco Unified Computing System

The Cisco Unified Computing System is a next-generation data center platform that unites compute, network, and storage access. The platform, optimized for virtual environments, is designed using open industry-standard technologies and aims to reduce TCO and increase business agility. The system integrates a low-latency, lossless 10 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. The system is an integrated, scalable, multichassis platform in which all resources participate in a unified management domain.

The Cisco Unified Computing System represents a radical simplification of the traditional blade server deployment model, with simplified, stateless blades and a blade server chassis that is centrally provisioned, configured, and managed by Cisco UCS Manager. The result is a unified system that significantly reduces the number of components while offering a just-in-time provisioning model that allows systems to be deployed or redeployed in minutes rather than hours or days.

The Cisco Unified Computing System is designed to deliver:

- Reduced TCO at the platform, site, and organizational levels
- Increased IT staff productivity and business agility through just-in-time provisioning and mobility support for both virtualized and nonvirtualized environments
- A cohesive, integrated system that is managed, serviced, and tested as a whole
- Scalability through a design for up to 320 discrete servers and thousands of virtual machines, and the capability to scale I/O bandwidth to match demand
- Industry standards supported by a partner ecosystem of industry leaders

Innovations That Support Business Benefits

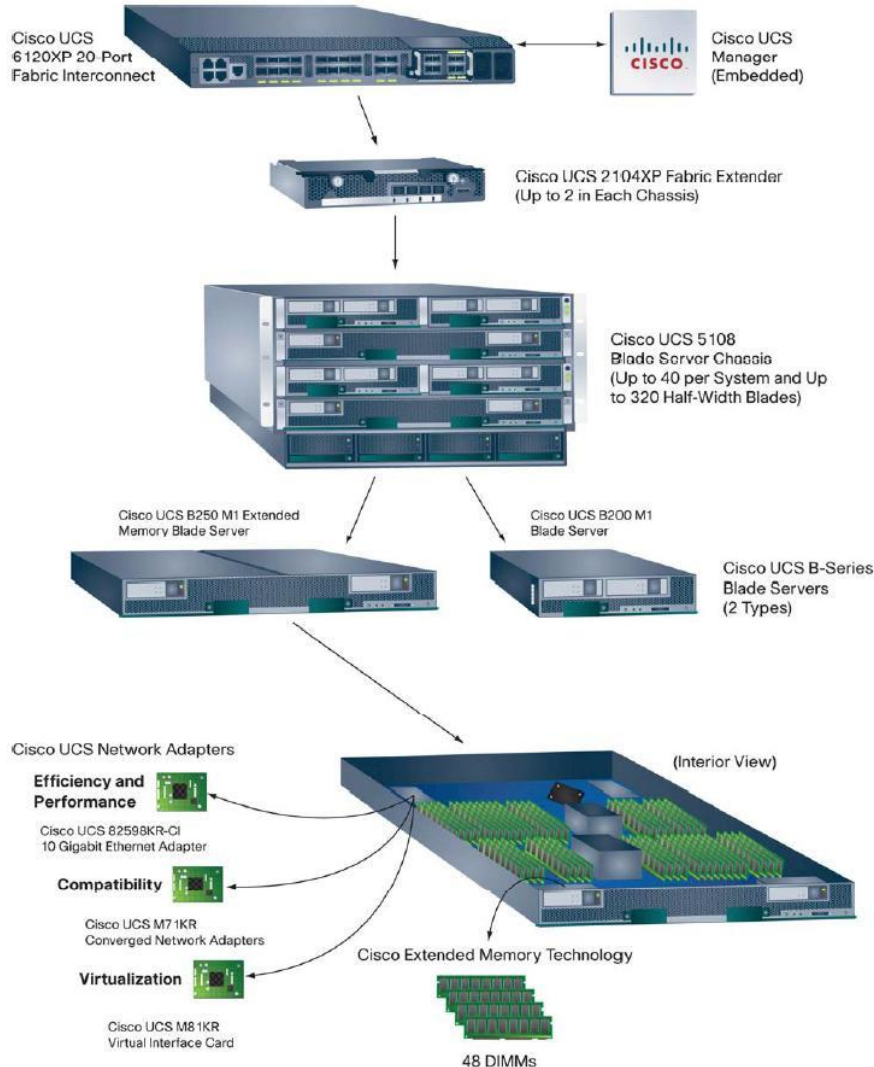
Each of the system's business benefits is supported by a rich set of technical innovations that contributes to this first implementation of the Cisco unified computing vision:

- Embedded system management through Cisco UCS Manager
- Just-in-time provisioning with service profiles
- Unified fabric using 10 Gigabit Ethernet
- Cisco VN-Link virtualization support
- Cisco Extended Memory Technology
- State-of-the-art performance using Intel Xeon Processors
- Energy-efficient platform design

Cisco Unified Computing System Components

Figure 4 shows the components of the Cisco Unified Computing System.

Figure 4. Cisco Unified Computing System Components



Cisco UCS 6100 Series Fabric Interconnects

The Cisco UCS 6100 Series Fabric Interconnects are a family of line-rate, low-latency, lossless, 10-Gbps Ethernet interconnect switches that consolidate I/O within the system. Both 20-port one-rack-unit (1RU) and 40-port 2RU versions accommodate expansion modules that provide Fibre Channel and 10 Gigabit Ethernet connectivity.

Cisco UCS 2100 Series Fabric Extenders

The Cisco UCS 2100 Series Fabric Extenders bring unified fabric into the blade-server chassis, providing up to four 10-Gbps connections each between blade server and the fabric interconnect, simplifying diagnostics, cabling, and management.

Cisco UCS Network Adapters

Cisco UCS network adapters offer a range of options to meet application requirements, including adapters optimized for virtualization, converged network adapters (CNAs) for access to unified fabric and compatibility with existing driver stacks, Fibre Channel HBAs, and efficient, high-performance Ethernet adapters.

Cisco UCS Manager

Cisco UCS Manager provides centralized management capabilities, creates a unified management domain, and serves as the central nervous system of the Cisco Unified Computing System. Cisco UCS Manager is embedded device-management software that manages the entire system as a single logical entity through an intuitive GUI, command-line interface (CLI), or XML API. Cisco UCS Manager implements role- and policy-based management using service profiles and templates. This construct improves IT productivity and business agility. Now infrastructure can be provisioned in minutes instead of days, shifting IT's focus from maintenance to strategic initiatives.

Cisco Unified Computing System resources are abstract in the sense that their identity, I/O configuration, MAC addresses and World Wide Names (WWNs), firmware versions, BIOS boot order, and network attributes (including quality-of-service [QoS] settings, access control lists [ACLs], pin groups, and threshold policies) all are programmable using a just-in-time deployment model. The manager stores this identity, connectivity, and configuration information in service profiles that reside on the Cisco UCS 6100 Series Fabric Interconnects. A service profile can be applied to any resource to provision it with the characteristics required to support a specific software stack. A service profile allows server and network definitions to move within the management domain, enabling flexibility in the use of system resources.

Cisco UCS Manager also offers role-based management that helps organizations make more efficient use of their limited administrator resources.

Cisco UCS 5100 Series Blade Server Chassis

The Cisco UCS 5100 Series Blade Server Chassis is a logical part of the Cisco Unified Computing System fabric interconnects, adding no management complexity to the system.

The Cisco UCS 5108 Blade Server Chassis fits on a standard rack and is six RUs high and physically houses blade servers and up to two Cisco UCS 2100 Series Fabric Extenders. It also houses eight cooling fans and four power supply units. The cooling fans and power supplies are hot swappable and redundant. The chassis requires only two power supplies for normal operation; the additional power supplies are for redundancy. The highly efficient (in excess of 90 percent) power supplies, in conjunction with the simple chassis design, which incorporates front-to-back cooling, makes the Cisco Unified Computing System very reliable and energy efficient.

Cisco UCS B200 M1 and M2 Blade Servers

The Cisco UCS B200 M1 and M2 Blade Servers balance simplicity, performance, and density for production-level virtualization and other mainstream data center workloads. The servers are half-width, two-socket blade servers.

The first-generation Cisco UCS B200 M1 uses the Intel Xeon processor 5500 series and the next generation Cisco UCS B200 M2 uses the Intel Xeon 5600 processor.

Each Cisco UCS B200 server uses a CNA for consolidated access to the unified fabric. This design reduces the number of adapters, cables, and access-layer switches needed for LAN and SAN connectivity.

Cisco UCS B250 M1 and M2 Extended Memory Blade Servers

The Cisco UCS B250 M1 and M2 Extended Memory Blade Servers use the patented Cisco Extended Memory Technology. This Cisco technology provides more than twice as much industry-standard memory (384 GB) as traditional two-socket servers, increasing performance and capacity for demanding virtualization and large-dataset workloads. Alternatively, this technology offers a more cost-effective memory footprint for less-demanding workloads.

The first-generation Cisco UCS B250 M1 uses the Intel Xeon processor 5500 series and the next generation Cisco UCS B250 M2 uses the Intel Xeon 5600 processor series.

Each Cisco UCS B250 uses two CNAs for consolidated access to the unified fabric. This design reduces the number of adapters, cables, and access-layer switches needed for LAN and SAN connectivity.

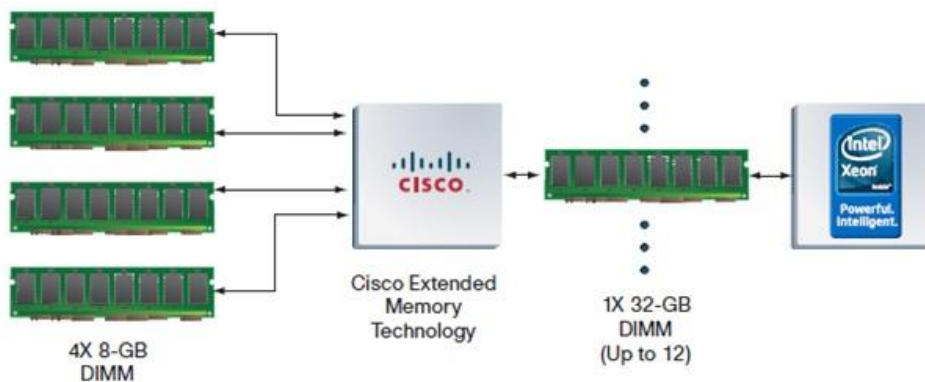
Cisco Extended Memory Technology

Modern CPUs with built-in memory controllers support a limited number of memory channels and slots per CPU. Virtualization software, to run multiple OS instances, demands large amounts of memory, and CPU performance is outstripping memory performance, leading to memory bottlenecks. Even some traditional nonvirtualized applications demand large amounts of main memory: database management system performance can be improved dramatically by caching database tables in memory, and modeling and simulation software can benefit from caching more of the problem state in memory.

To obtain a larger memory footprint, most IT departments are forced to upgrade to larger, more-expensive four-socket servers. CPUs that can support four-socket configurations typically cost more, require more power, and entail higher licensing costs.

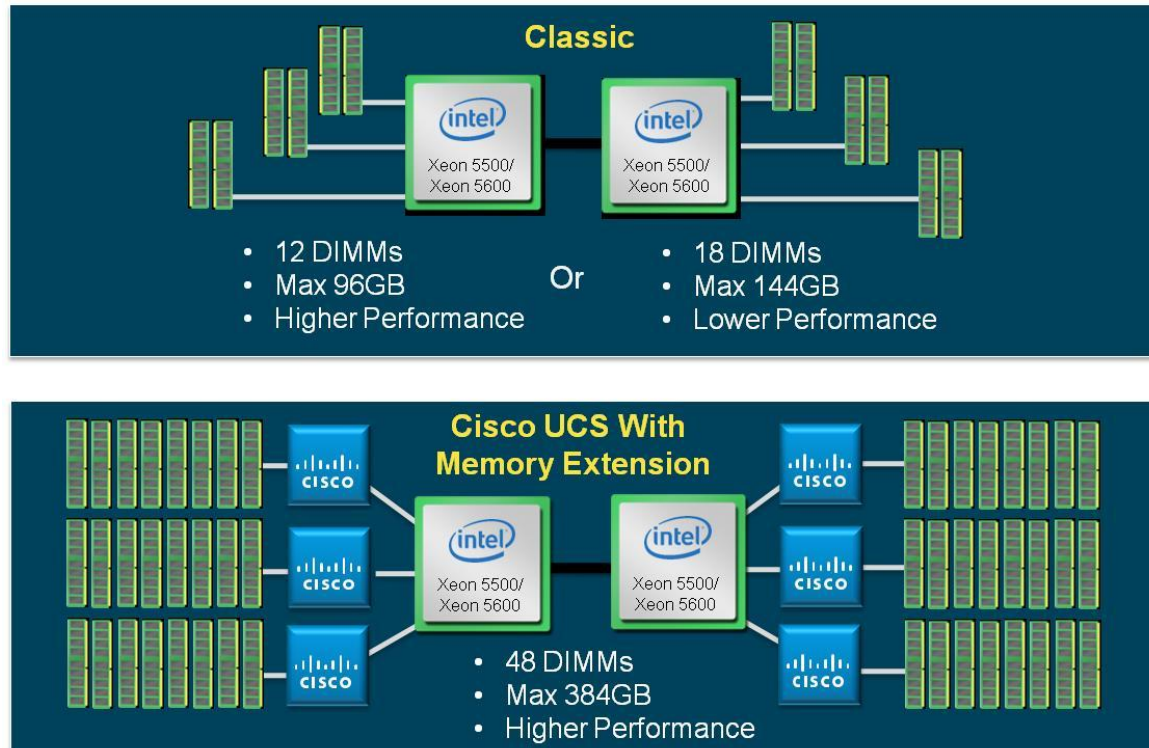
Cisco Extended Memory Technology expands the capabilities of CPU-based memory controllers by logically changing the main memory while still using standard DDR3 memory. The technology makes every four DIMM slots in the expanded memory blade server appear to the CPU's memory controller as a single DIMM that is four times the size. For example, using standard DDR3 DIMMs, the technology makes four 8-GB DIMMS appear as a single 32-GB DIMM (Figure 5).

Figure 5. Cisco Extended Memory Technology Architecture



Cisco UCS Extended Memory Technology provides flexibility between memory cost and density (the number of machines required to manage a particular set of data). This extended memory uses a high-performance, ultra-fast technology that is implemented in its ASIC to allow 48 memory modules (DIMMs) to be addressed at high speed. The total memory address space per blade jumps to 384 GB at 1333 MHz compared to 96 GB at 1333 MHz, or 144 GB at 800 MHz, on alternative hardware provided by other x86-based two-socket server vendors, which can use up to 18 memory modules (DIMMs), as shown in Figure 6.

Figure 6. Compare the Classic Model with Cisco Unified Computing System with Memory



This patented technology allows the CPU to access more industry-standard memory than ever before in a two-socket server:

- For memory-intensive environments, data centers can better balance the ratio of CPU power to memory and install larger amounts of memory without having the expense and energy waste of moving to four-socket servers simply to have a larger memory capacity. With a larger main memory footprint, CPU utilization can improve because of fewer disk waits on page-in and other I/O operations, making more effective use of capital investments and more conservative use of energy.
- For environments that need significant amounts of main memory but which do not need a full 384 GB, smaller-sized DIMMs can be used in place of 8-GB DIMMs, with resulting cost savings.

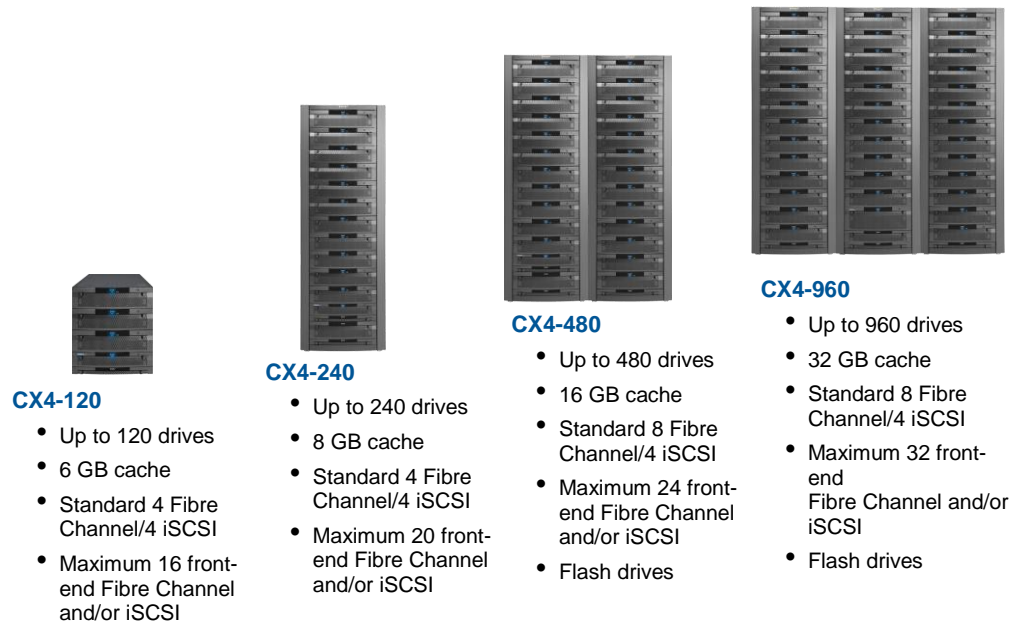
EMC CLARiiON

The EMC CLARiiON CX4 series (Figure 7) with UltraFlex technology is based on a new breakthrough architecture and extensive technological innovation, providing a mid-range solution that is highly scalable, meeting the price points of most mid-range customers. The unique modularity of the UltraFlex technology allows you to use a combination of protocols within a single storage system, providing online-expandable connectivity options. It also includes new levels of ease of use, making the CLARiiON CX4 easy to install, manage, and scale. The CLARiiON CX4 is the fourth-generation CLARiiON CX series and continues EMC's commitment to enhancing customers' investments in CLARiiON technology by helping ensure that existing resources and capital assets are optimally utilized as customers adopt new technologies. The innovative technologies in the CLARiiON CX4 include fully automated storage tiering and support for the latest generation of disk-drive technologies, such as flash drives.

The CLARiiON CX4 series introduces thin logical unit number (LUN) technology that builds on CLARiiON virtual LUN capabilities and transparently integrates with CLARiiON management and replication software. With CLARiiON Virtual Provisioning, you can choose between traditional LUNs, metaLUNs, and thin LUNs. The capability to nondisruptively migrate data to different LUN and disk types allows you to deploy the best solution without incurring downtime. CLARiiON Virtual Provisioning enables organizations to reduce costs by increasing

utilization without overprovisioning of storage capacity, simplifying storage management and reducing application downtime.

Figure 7. EMC CLARiiON CX4 Series



EMC CLARiiON also provides the Navisphere® Management Suite, which is a suite of tools that allows centralized management of CLARiiON storage systems. Navisphere provides a centralized tool to monitor and configure CLARiiON storage. The Navisphere Management suite includes Navisphere Manager, which has a web-based UI, Navisphere Secure Command-Line Interface (CLI). CLARiiON provides functional capabilities such as point-in-time local replicas and remote replication options for business continuity using the Navisphere Manager tool. Navisphere Management Suite also includes EMC Navisphere QoS Manager, Navisphere Analyzer, SnapView™, SAN Copy™, and MirrorView™.

5 Server Consolidation Case Study

Case Study-Microsoft SQL Server Using VMware vSphere

This section details the sample scenario implemented for this solution. The solution compared relative Microsoft SQL Server 2008 performance of the Intel Xeon Processor X5680 and the Intel Xeon Processor X5570 on the Cisco UCS B250 M1 and M2 Extended Memory Blade Servers, and the Cisco UCS B200 M1 and M2 Blade Servers, using VMware vSphere 4.

Scenario

To demonstrate server consolidation a typical 3- to 5-year-old server was used as the baseline platform. This server is assumed to be running in physical mode (nonvirtualized) and to be underutilized (about 20 percent CPU utilization). The workload chosen was an online-transaction-processing (OLTP) database server running a Microsoft SQL Server database. VMware vSphere 4.0 was used as the virtualization software. On the Cisco UCS platform, virtual machines with throughput equivalent to that of the baseline platform are added until the server can no longer add virtual machines. This value represents the consolidation ratio. The following are the platforms used:

- Older server: HP ProLiant DL380 G4 server with single-core Intel Xeon processor 3.60-GHz
- New servers:
 - Cisco UCS B200 M1 with Intel Xeon X5570 CPU and 96 GB of memory
 - Cisco UCS B250 M1 with Intel Xeon X5570 CPU and 192 GB of memory
 - Cisco UCS B200 M2 with Intel Xeon X5680 CPU and 96 GB of memory
 - Cisco UCS B250 M2 with Intel Xeon X5680 CPU and 384 GB of memory
- EMC® CLARiiON® CX4-240 was used as the storage platform

Workload

This workload implemented DVD Store Version 2 (DS2) and ran on virtual machines simultaneously for 30 minutes. DS2 is an open source simulation of an online e-commerce DVD store in which customers log in, browse, and order products. DS2 has database and web server components and includes driver programs; these driver programs place heavy loads on the components.

To stress the database components, a driver program was implemented. Each blade server ran multiple Microsoft SQL Server 2008 workloads: one per virtual machine. One copy of the DS2 database was on each virtual machine, with a 10-GB database for each virtual machine. By combining multiple virtual machines, all running a CPU, memory, and disk-intensive workload, the test placed a heavy load on both the server and the storage. Thus, it measured the performance of the blade server, especially as it relates to an active OLTP environment. The main DS2 metric is orders per minute (OPM), which the driver program calculates and reports through the Microsoft Windows Performance Monitor utility on the client machines. The output from the driver program was used to record OPM.

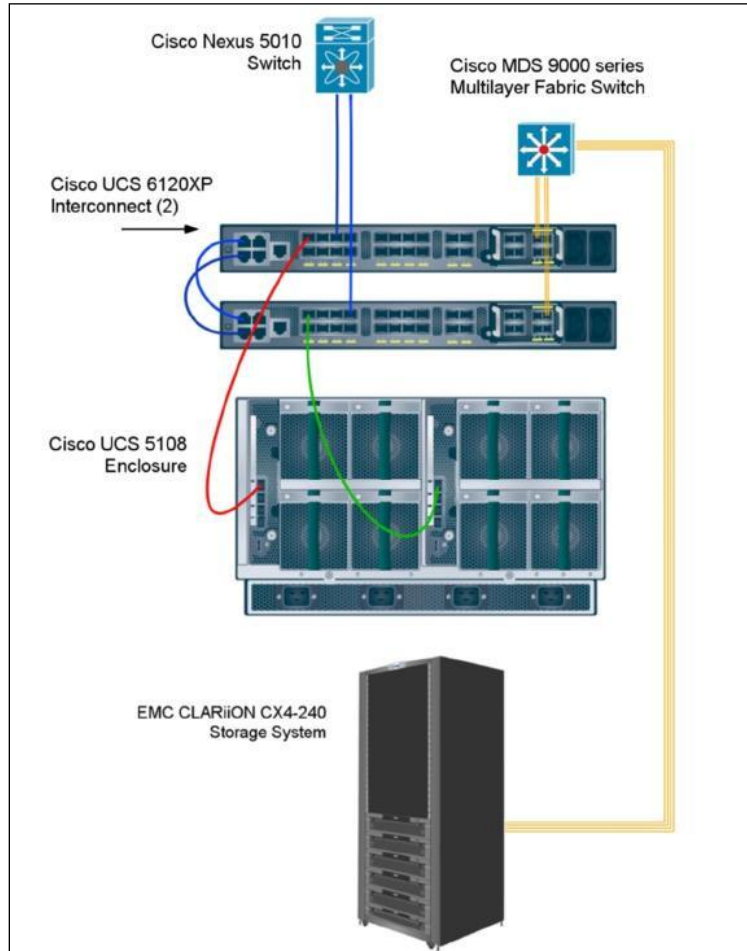
When DS2 is running, simulated customers log in; browse movies by actor, title, or category; and purchase movies. Each DS2 order consists of a customer log in, a number of searches for movies, and a purchase. The title and actor searches use full-text searching, and the other customer actions exercise a wide range of database functions.

To isolate and test database server performance, the included driver application ran on client machines, through its CLI. The default DS2 parameters and setup configuration were implemented. Each client machine ran a single instance of DS2 with 32 threads. This setup simulated a heavily loaded environment; the load-generating client virtual machines ran with 0.3 second latency, sending requests as quickly as the blade servers could handle them.

Details of the System Under Test

Figure 8 shows the systems under test (SUT).

Figure 8. Reference Platform



Cisco UCS Platform Configuration Details

Table 1 shows the Cisco UCS platform compared in the study.

Table 1. Cisco Unified Computing System Reference Configuration

Hardware Specification	Cisco UCS B200 M1 Blade Server	Cisco UCS B250 M1 Extended Memory Blade Server	Cisco UCS B200 M2 Blade Server	Cisco UCS B250 M2 Extended Memory Blade Server
CPU	Intel Xeon X5570	Intel Xeon X5570	Intel Xeon X5680	Intel Xeon X5680
CPU speed	2.93 GHz	2.93 GHz	3.33 GHz	3.33 GHz
Number of processors	2	2	2	2
Number of cores per processor	4	4	6	6
Number of hardware threads per core	2	2	2	2
Memory type	DDR3 PC3-10600R	DDR3 PC3-10600R	DDR3 PC3-10600R	DDR3 PC3-10600R
Total memory	96 GB (12 x 8 GB)	384 GB (48 x 8 GB)	96 GB (12 x 8 GB)	384 GB (48 x 8 GB)
Maximum supported memory	96 GB	384 GB	96 GB	384 GB

Table 2 shows the settings used on virtual machines on all four test servers. In this test, in general, the more RAM the blade server can hold, the more virtual machines the blade server can run. Thus, the Cisco UCS B250 M2, with VMware vSphere and the Intel Xeon Processor X5680 and the Cisco UCS B250 M1, with VMware vSphere and the Intel Xeon Processor X5570 were able to run more virtual machines with their 384 GB of RAM than the Cisco UCS B200 M2, with VMware vSphere and the Intel Xeon Processor X5680 and the Cisco UCS B200 M1, with VMware vSphere and the Intel Xeon Processor X5570 with their 96 GB of RAM.

Table 2. Virtual Machine Configuration Used in the Tests

Virtual Machine Specifications	Virtual Machine Settings
Virtual CPU (vCPU)	1
Memory	8 GB
Virtual NIC (vNIC) type	VMXNET 3
Number of virtual disks	3 (OS, SQL logs, and SQL database)

Setting Up and Configuring EMC Storage

An EMC CLARiiON CX4-240 Fibre Channel SAN was used for testing. The SAN had two storage processors: SP-A and SP-B. The integrated dual-port HBA was used in the Cisco UCS B250 M1. One HBA port was cabled to each storage processor (SP-A and SP-B) on the SAN to balance the load between storage processors. Five disk enclosures were used on the SAN; with one exception, each enclosure had 15 disks. RAID 10 was designated for the virtual machine OS LUN and for all other RAID group configurations.

One RAID group was created on the enclosure that had only 12 disks. Then one LUN was created on this RAID group for the virtual hard drives that would hold the operating systems for the virtual machines. Eight RAID groups were created for Microsoft SQL Server database data. Each of these eight RAID groups was composed of six disks. Then one 256-GB LUN was created in each RAID group, for a total of eight LUNs dedicated to virtual machines to hold Microsoft SQL Server database data. The EMC metaLUN feature was used, and four LUNs were assigned to metaLUN-1, and four LUNs were assigned to metaLUN-2. To balance processing load and disk

transfers, metaLUN-1 was assigned to SP-A, and metaLUN-2 was assigned to SP-B. Virtual machine database data was evenly distributed across the two metaLUNs.

Two more RAID groups were created for Microsoft SQL Server log data. Each of these RAID groups was composed of four disks. One 536-GB LUN was created in each RAID group, for a total of two LUNs dedicated to virtual machines to hold Microsoft SQL Server log data. One log LUN was assigned to each storage processor. To balance processing load and disk transfers, logLUN-1 was assigned to SP-B, and logLUN-2 was assigned to SP-A.

Performance Results

Legacy Server Details

For the reference score, the OPM value on a 3-year-old HP legacy server running a typical workload with less than 20 percent processor utilization was used. The older HP server was a dual-socket server with two Intel Xeon 3.6-GHz processors, 4 GB of RAM, and two onboard NICs using one tray of external storage. The reference score was used as the minimum acceptable average score for the virtual machines. A baseline score of 5135 OPM was used, while the older HP server ran Microsoft Windows Server 2003 with Microsoft SQL Server 2005. An older version of Microsoft Windows Server and SQL Server were used on the older server since a 3- to 5-five-year old server would run an older version of the operating systems and Microsoft SQL Server.

Figure 9. Cisco UCS B200 M1 Blade Server Replacing 13 Legacy Servers

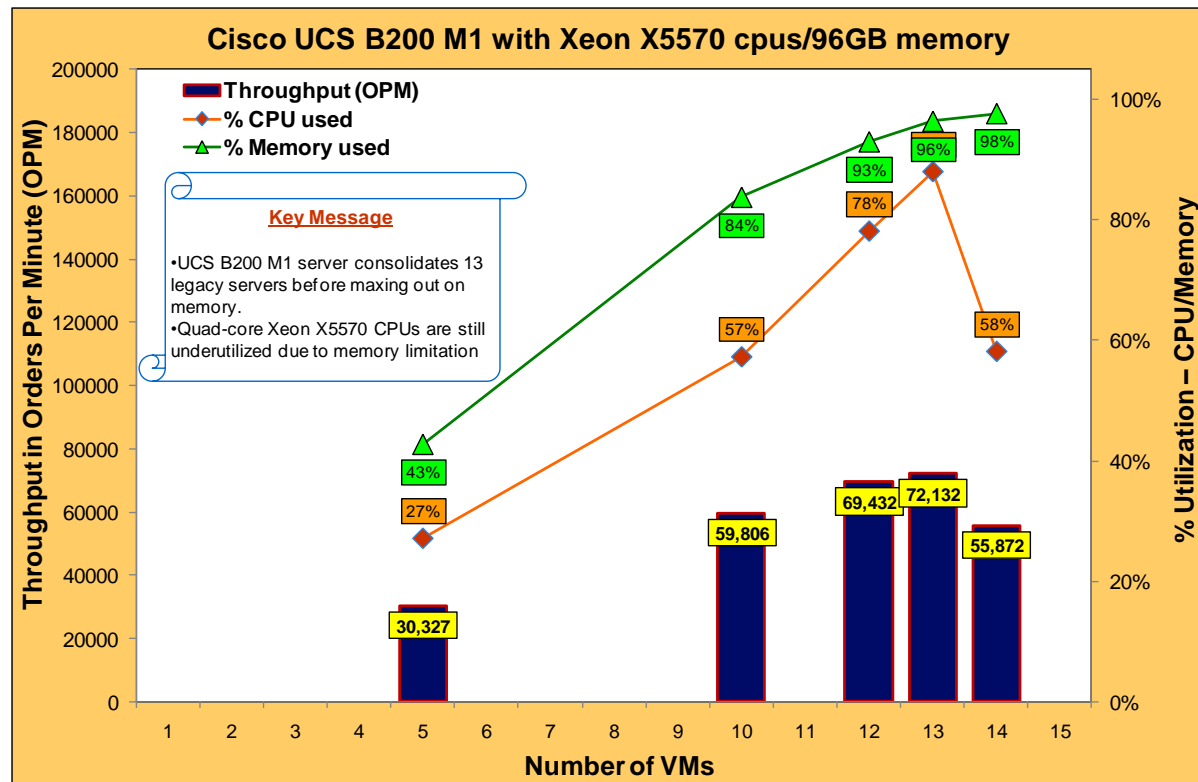


Figure 9 shows consolidation of the older servers onto the Cisco UCS B200 M1 platform. Each virtual machine is assigned one vCPU and 8 GB of memory. Virtual machines each with a performance level similar to that of baseline platform are added, and steady-state performance is reached. CPU, memory, throughput, and response times are monitored as virtual machines are added. As can be seen in the figure, the performance increase is linear for up to 10 virtual machines. But at higher virtual machine levels, memory oversubscription starts to occur,

and hence throughput starts decreasing, and CPU utilization increases more dramatically. After 13 virtual machines have been added, the overall throughput drops, showing severe contention issue on memory.

Figure 10. Cisco UCS B250 M1 Blade Server Replacing 20 Legacy Servers

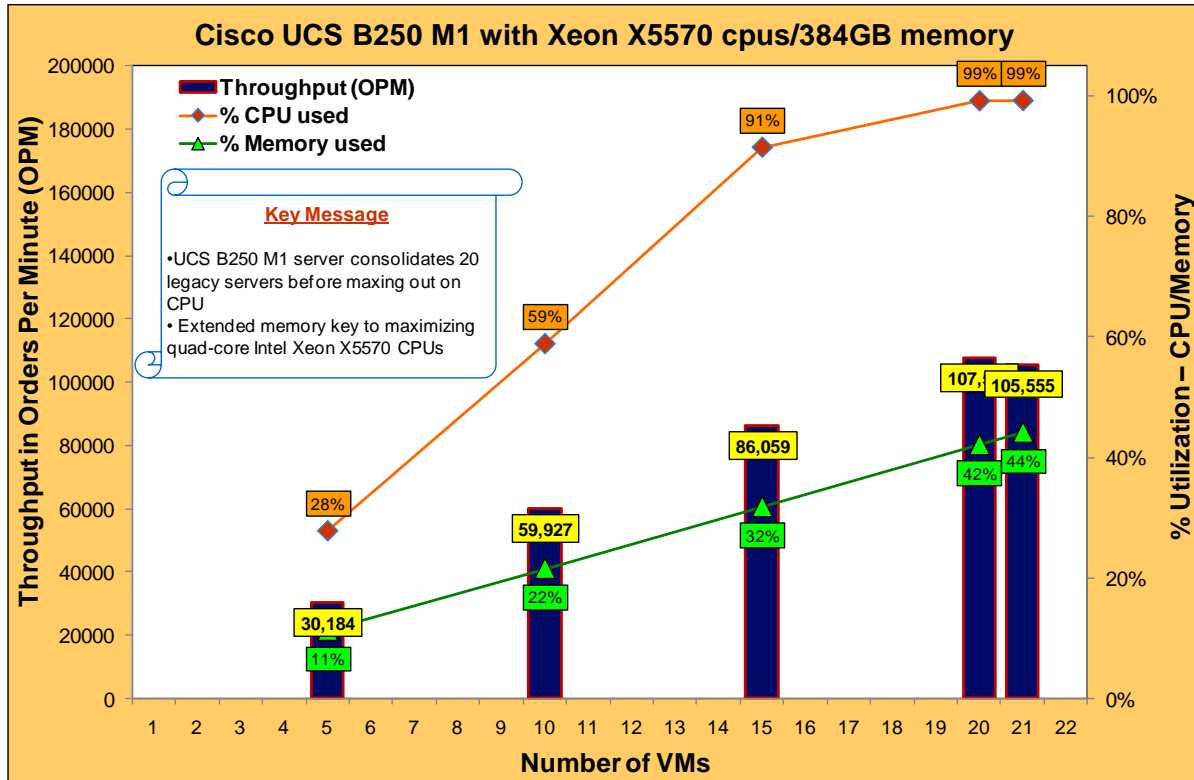


Figure 10 shows consolidation of older servers onto the Cisco UCS B250 M1 platform. Note that this platform supports more memory and is configured with 384 GB. Each virtual machine is assigned one vCPU and 8 GB of memory. Virtual machines each with a performance level similar to that of the baseline platform are added, and steady-state performance is reached. CPU, memory, throughput, and response times are monitored as virtual machines are added. As can be seen in the figure, the performance increase is linear for up to a much higher virtual machine count than on the Cisco UCS B200 M1 platform. But at a virtual machine count of about 20, the platform CPU has reached its maximum, while memory capacity is still available. Beyond 20 virtual machines, the overall throughput drops due to CPU contention. Given that memory utilization was below 50 percent, comparable performance could be attained if the server were configured with 192 GB using the 4-GB DIMMs. Thus, the Cisco UCS B250 M1 was able to optimize the use of the Intel Xeon X5570 CPUs.

Figure 11. Cisco UCS B200 M2 Blade Server Replacing 13 Legacy Servers

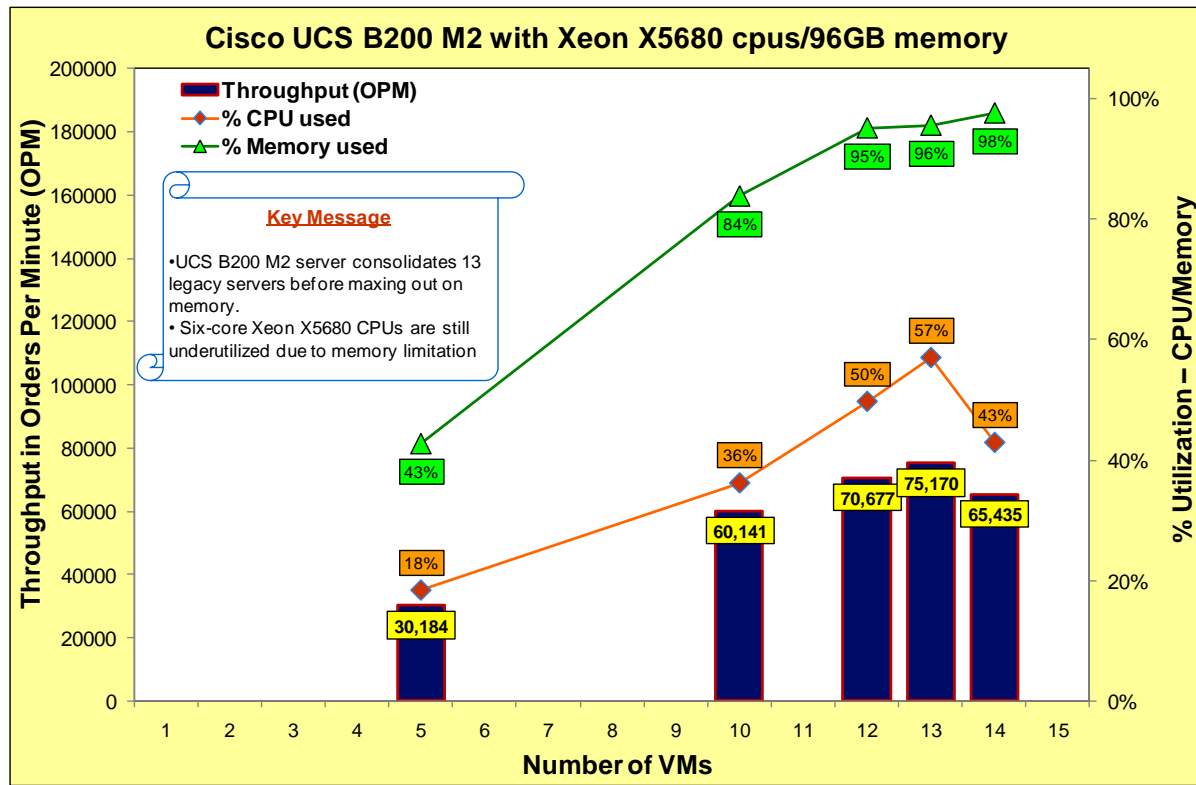


Figure 11 shows consolidation of older servers onto the Cisco UCS B200 M2 platform using Intel Xeon X5680 CPUs. Each virtual machine is assigned one vCPU and 8 GB of memory. Virtual machines each with a performance level similar to that of the baseline platform are added, and steady-state performance is reached. CPU, memory, throughput, and response times are monitored as virtual machines are added. As can be seen in the figure, the performance increase is linear for up to 10 virtual machines. But at higher virtual machine levels, memory oversubscription starts to occur, and hence the throughput starts decreasing, and CPU utilization increases more dramatically. After the 13 virtual machines have been added, the overall throughput drops; showing severe contention for memory.

Note that this server delivers virtual machine capacity similar to that of the Cisco UCS B200 M1 server, but with significantly more idle CPU cycles, because both were limited by the server memory capacity of 96 GB.

Figure 12. Cisco UCS B250 Blade Server Replacing 29 Legacy Servers

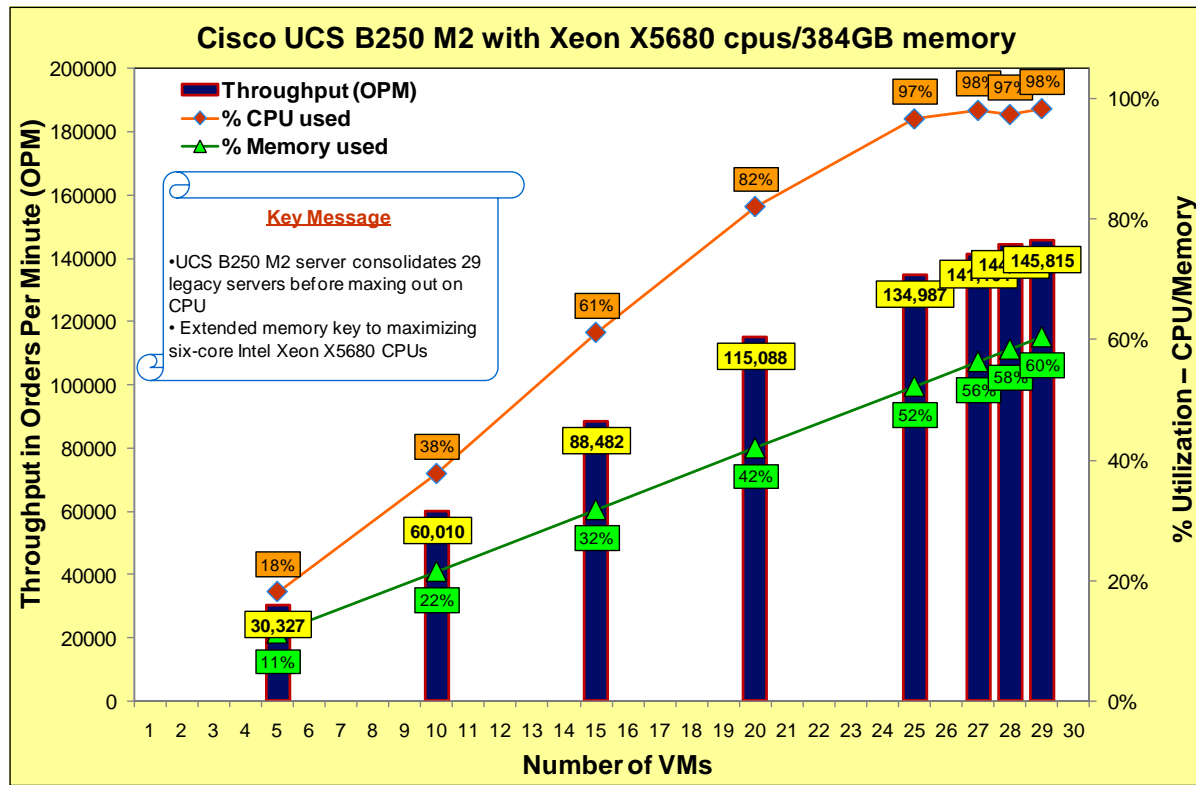


Figure 12 shows consolidation of older servers onto the Cisco UCS B250 M2 platform using Intel Xeon X5680 CPUs. Note that this platform supports more memory and is configured with 384 GB. Each virtual machine is assigned one vCPU and 8 GB of memory. Virtual machines each with a performance level similar to that of the baseline platform are added, and steady-state performance is reached. CPU, memory, throughput, and response times are monitored as virtual machines are added. As can be seen in the figure, the performance increase is linear for up to a much higher virtual machine count than on the Cisco B200 M2 platform. But at a virtual machine count of about 29, the platform CPU has reached its maximum, while memory capacity is still available. This demonstrates a consolidation ratio that is twice in overall throughput, but at an aggregate level and virtual machine count. Thus, the Cisco UCS B250 M2 was able to optimize the use of the Intel Xeon X5680 CPUs.

6 Summary Comparison

Figure 13 provides a summary comparison of all the results. It is clear that Cisco Extended Memory Technology delivers benefits for both for Cisco UCS B-Series M1 and M2 platforms.

Figure 13. Summary of Results

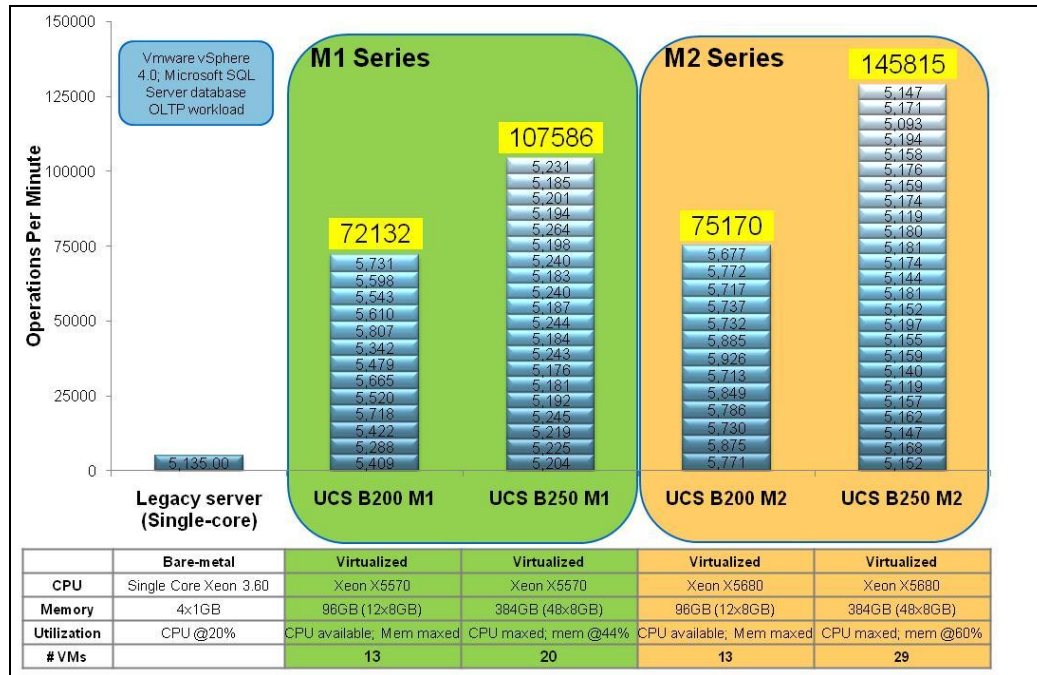
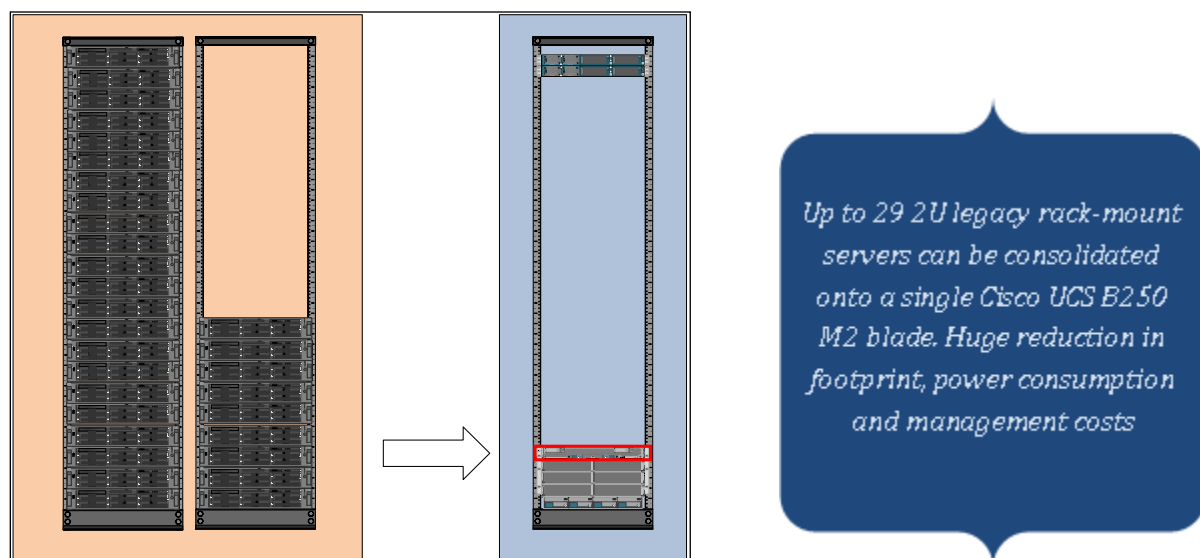


Figure 14 shows the physical racks in the data center to illustrate the effect of server consolidation at a 29:1 ratio. Such massive reduction in the server footprint obviously saves space, power, and management costs.

Figure 14. Illustration of the Physical Racks in the Data Center



7 Conclusion

Server consolidation using Virtualization technology has tremendous benefits for optimizing data center resources. The Cisco Unified Computing System with EMC CLARiiON storage provides an ideal platform for server consolidation. Cisco's unique technologies such as Cisco Extended Memory Technology enable an exceptional level of consolidation that stands out in the industry for this class of platform, thereby resulting in significantly improved ROI and lower TCO for IT managers.

8 Appendix

Table 3 and Table 4 shows the peak number of virtual machines that the Cisco UCS B-Series Blade Servers could handle during testing before dipping below the acceptable threshold and the minimum client score from the median score.

Table 3. Cisco UCS B-Series Blade Server Test Results

	Cisco UCS B250 M2 Extended Memory Blade Server with Intel Xeon Processor X5680	Cisco UCS B250 M1 Extended Memory Blade Server with Intel Xeon Processor X5570	Cisco UCS B200 M2 Blade Server with Intel Xeon Processor X5680	Cisco UCS B200 M1 Blade Server with Intel Xeon Processor X5570
Virtual machine 1	5,152	5,204	5,771	5,409
Virtual machine 2	5,168	5,225	5,875	5,288
Virtual machine 3	5,147	5,219	5,730	5,422
Virtual machine 4	5,162	5,245	5,786	5,718
Virtual machine 5	5,157	5,192	5,849	5,520
Virtual machine 6	5,119	5,181	5,713	5,665
Virtual machine 7	5,140	5,176	5,926	5,479
Virtual machine 8	5,159	5,243	5,885	5,342
Virtual machine 9	5,155	5,184	5,732	5,807
Virtual machine 10	5,197	5,244	5,737	5,610
Virtual machine 11	5,152	5,187	5,717	5,543
Virtual machine 12	5,181	5,240	5,772	5,598
Virtual machine 13	5,144	5,183	5,677	5,731
Virtual machine 14	5,174	5,240		
Virtual machine 15	5,181	5,198		
Virtual machine 16	5,180	5,264		
Virtual machine 17	5,119	5,194		
Virtual machine 18	5,174	5,201		
Virtual machine 19	5,159	5,185		
Virtual machine 20	5,176	5,231		
Virtual machine 21	5,158			
Virtual machine 22	5,194			
Virtual machine 23	5,093			
Virtual machine 24	5,171			
Virtual machine 25	5,147			
Virtual machine 26	5,169			
Virtual machine 27	5,142			
Virtual machine 28	5,036			
Total	144,306	104,236	75,170	72,132

Table 4. OPM, CPU Utilization, and Memory Utilization for the Blade Servers

Cisco UCS B250 M2 Extended Memory Blade Server with the Intel Xeon Processor X5680 and 384 GB of RAM			
Number of VMs	OPM	CPU utilization	Memory utilization
5	30,327	18.07%	11.24%
10	60,010	37.69%	21.50%
15	88,482	61.15%	31.77%
20	115,088	82.07%	42.02%
25	134,987	96.64%	52.27%
27	141,164	98.04%	56.38%
28	144,306	97.37%	58.44%
29	145,815	98.34%	60.49%
Cisco UCS B250 M1 Extended Memory Blade Server with the Intel Xeon Processor X5570 and 384 GB of RAM			
Number of VMs	OPM	CPU utilization	Memory utilization
5	30,312	27.73%	11.24%
10	59,927	58.83%	21.51%
15	86,059	91.46%	31.76%
20	107,586	99.17%	42.04%
21	105,555	99.22%	44.08%
Cisco UCS B200 M2 Blade Server with the Intel Xeon Processor X5680 and 96 GB of RAM			
Number of VMs	OPM	CPU utilization	Memory utilization
5	30,184	18.44%	42.82%
10	60,141	36.21%	83.84%
12	70,677	49.72%	95.13%
13	75,170	56.99%	95.58%
14	65,435	42.92%	97.60%
Cisco UCS B200 M1 Blade Server with the Intel Xeon Processor X5570 and 96 GB of RAM			
Number of VMs	OPM	CPU utilization	Memory utilization
5	30,288	27.16%	42.79%
10	59,806	57.28%	83.81%
12	69,432	78.14%	93.02%
13	72,132	88.01%	96.43%
14	55,872	58.18%	97.71%



Table 5 shows the read, write, and total input/output operations per second (IOPS) for the median run of the four test servers. Higher numbers are better, as they show that the system is handling more requests.

Table 5. IOPS Comparison

IOPS	Cisco UCS B250 M2 Extended Memory Blade Server with Intel Xeon Processor X5680	Cisco UCS B250 M1 Extended Memory Blade Server with Intel Xeon Processor X5570	Cisco UCS B200 M2 Blade Server with Intel Xeon Processor X5680	Cisco UCS B200 M1 Blade Server with Intel Xeon Processor X5570
Read	3,534	2,794	1,287	1,507
Write	7,814	6,769	4,523	4,381
Total	11,348	9,563	5,810	5,888

Study conducted by Principled Technologies and posted at:

<http://www.principledtechnologies.com/clients/reports/Cisco/UCSB250B200.pdf>



Cisco Systems, Inc.

170 West Tasman Drive

San Jose, CA 95134-1706

USA

www.cisco.com

Tel: 408 526-4000

800 553-NETS (6387)

Fax: 408 527-0883

© 2010 Cisco Systems, Inc. All rights reserved. Cisco, the Cisco logo, and Cisco Systems are registered trademarks or trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries. All other trademarks mentioned in this document are the property of their respective owners. (0805R)

Document number: UCS-TR100002