SQL Server Consolidation Using Cisco Unified Computing System C460 and Microsoft Hyper-V

White Paper

October 2011
Contents

Executive Summary..................................................................................................................3
Introduction.................................................................................................................................3
Audience and Scope..................................................................................................................4
Today's Challenges.......................................................................................................................4
Considerations for Consolidation..............................................................................................5
Cisco UCS C-Series Rack-Mount Servers....................................................................................6
Cisco UCS P81E Virtual Interface Cards....................................................................................9
  Product Overview....................................................................................................................9
  Features and Benefits.............................................................................................................9
Cisco UCS P81E Architecture....................................................................................................10
Tested Configurations and Methodology...................................................................................11
  Test Bed Configuration........................................................................................................13
  Workload............................................................................................................................15
Test Results................................................................................................................................16
Conclusion.................................................................................................................................19
Acknowledgements..................................................................................................................19
Executive Summary

The purpose of this document is to provide a study of the Microsoft SQL Server consolidation on Cisco Unified Compute System (UCS). This study will prove that consolidated SQL Server implementations using the Cisco Unified Computing System can meet the scalability, availability and performance requirements mandated by today’s high-volume database implementations.

The goal of this white paper is to provide a framework for choosing among virtualization and multi-database consolidation strategies for SQL Server Database Engine supporting OLTP applications by highlighting consolidation benchmark numbers based on technical analysis.

This white paper also provides a scaling and consolidation study to help customers design a consolidation and virtualization strategy for deploying Microsoft SQL Server software on the Cisco Unified Computing System.

This experiment will prove that the Cisco Unified Computing System can help companies realize a reduction in both capital and operational expenditures along with the ability to achieve the level of infrastructure agility required to meet the challenges of today’s fast-paced and ever-changing business requirements.

The tests detailed in this white paper reveal best consolidated benchmark numbers on a virtualized environment and recommends server and virtual machine configurations as well as workload sizes supported by the server.

Introduction

In today’s economic climate, enterprises are taking a closer look within their IT organizations to identify potential areas in which cost-saving strategies can be implemented to help reduce operating expenses. One of the challenges that IT organizations face today is how to develop an infrastructure that allows flexibility, redundancy, high-availability, ease of management, security, and access control while at the same time reducing costs, hardware footprints and complexity.

The growth in IT industry has realized tremendous growth in hardware computing capacity, database applications and physical server sprawl has resulted in costly and complex computing environments containing many over-provisioned and under-utilized database servers. Many of these servers implement a single instance of SQL Server realizing ten to fifteen percent CPU utilization on average which is not an optimal use of server resources. Additionally, due to the complexities introduced by all of this growth, many database administrators today are overburdened with redundant management and administrative tasks that must be implemented on each of the many servers they are responsible for. Also, during a catastrophic, physical database server failure, there is a significant impact on the server administrator as they typically are the resources that are required to provision another physical server into the environment and prepare it for the DBA.

Database server consolidation is an area where companies can realize considerable cost savings with regards to both op-ex and cap-ex. Database server consolidation can also help companies to achieve the infrastructure agility they are seeking to stay competitive and provide the fastest time to market for their solutions.

Consolidation, in general terms, is the combining of various units into more efficient and stable larger units. When applied to an IT department, consolidation specifically translates into improved cost efficiency from higher utilization of resources, standardization and improved manageability of the IT environment, and (more recently) a focus on a “green” IT environment through reduced energy consumption. One of the important components in the IT environment is the database. Databases tend to be very widespread
across the enterprise, because they are very efficient for handling relational storage of data and are designed to be platforms for a wide variety of applications.

Because databases form the foundation of so many business systems, IT departments can easily lose control of the number of databases that need to be maintained, because each group may simply create their own database to solve a specific problem they may be experiencing. This leads to a proliferation of databases and machines running database instances also known as database sprawl. Thus databases are one of the prime candidates for consolidation.

**Audience and Scope**

This white paper is intended for customers, partners, solution architects, storage administrators, and database administrators who are evaluating, planning a consolidation and/or virtualization strategy. It provides an overview of various considerations and reference architectures for consolidated and virtualized SQL Server 2008 R2 deployments.

**Today’s Challenges**

IT organizations face an enormous challenge to keep operations running around the clock with increasing demand and growing complexity, to the point that it becomes difficult for employees to request resources to fulfill their tasks. This has led to circumventing the IT standards and procedures to get the job done faster, which in turn has led to server sprawl especially at the database tier. Databases are the engines that provide data to applications in a typical three- or four-tier environment; web-tier, application-tier, database-tier, and an optional storage-tier. With the advent of standardized hardware, a tiered architecture allows compartmentalization of applications and separation of resources. Hardware at each tier provides a specific function and requires a specific set of tools to manage and maintain. All of the components are connected through a network, in most cases Ethernet. This separation of resources provides an easy way for employees to bypass and deploy database servers for test and development without the IT department’s oversight and management.

The apparent benefit of ease of deployment and development can easily be overshadowed by potential challenges such as incompatibility issues that may arise during deployment to production, thus delaying time to market. Security vulnerabilities may also be introduced to the organization due to un-patched systems. Over time, however, servers that were deployed for a specific project that is no longer active still remain in the organization just in case a need arises to continue testing and quality assurance (QA). In the meantime, CPU cycles are lost to idle systems that consume valuable wattage, adding to the bottom line of the business costs.

It very quickly becomes apparent that unmanaged server-sprawl can potentially grow very complex, adding to the data center footprint and the costs for management, security, power, and cooling.
Figure 1. SQL Server consolidation on virtualized environment

Considerations for Consolidation

Consolidation projects are typically started to achieve specific goals such as creating room for new servers or reducing operating expenditure. These goals can be broadly grouped into the following categories:

- Lack of space in the data center
- Reducing costs and improving efficiency
- Standardization and centralization
- IT agility
- Green IT
This white paper focuses on consolidation strategies for online transaction processing (OLTP) applications storing data in the SQL Server Database Engine. OLTP applications typically focus on fast response times with smaller but more frequent queries and updates dealing with relatively small amounts of data. And the consolidation strategy applies primarily to the consolidation of the SQL Server instance supporting this application.

Some general traits that make an application a good candidate for consolidation are low machine resource utilization, moderate performance requirements, little active development, and low maintenance costs. Another factor to consider is the impact on the application’s network and I/O latency, because both the network and storage resources become shared as part of consolidation.

While consolidating database applications we considered the following strategy and testing was carried out on a stand-alone hardware/software configurations.

**Testing Effort**
- SQL Server Consolidation and scale test
- Single Server Scalability test

**Strategy**
- Single server scalability testing on Cisco UCS C460 M1 large (100GB) Database deployed on a single virtual machine
- Hosting multiple virtual machine’s with the exact same configuration on a single physical server with each virtual machine running single SQL Server Database

**Cisco UCS C-Series Rack-Mount Servers**

Cisco UCS C-Series Rack-Mount Servers (Figure 2) based on Intel Xeon 5500, 5600 and 7500 series processors, extend Cisco Unified Computing System innovations to an industry standard rack-mount form factor, including a standards-based unified network fabric, Cisco VN-Link virtualization support, and Cisco Extended Memory Technology.

Designed to operate both in standalone environments and as part of the Cisco Unified Computing System, these servers enable organizations to deploy systems incrementally—using as many or as few servers as needed—on a schedule that best meets the organization’s timing and budget. Cisco UCS C-Series servers offer investment protection through the capability to deploy them either as standalone servers or as part of the Cisco Unified Computing System. Current generation of Cisco UCS C-Series Rack-Mount Servers is depicted in Figure 2 and a high level comparison of features is in Table 1.

One compelling reason that many organizations prefer rack-mount servers is the wide range of I/O options available in the form of PCI Express (PCIe) adapters. Cisco UCS C-Series servers supports spectrum of I/O options, which includes interfaces supported by Cisco as well as adapters from third parties.
When deployed as standalone servers in a heterogeneous environment, Cisco UCS C-Series servers can be managed just like any other x86-architecture servers. Popular enterprise management tools using OS-resident host agents work without modification. Cisco UCS Integrated Management Controller (CIMC) gives administrators the tools they need to manually control server functions, including remote keyboard, video, and mouse (KVM); power on and off; and standard Simple Network Management Protocol (SNMP) traps for system monitoring. Cisco UCS C series rack mounts can also be managed by the Cisco UCS Manager.

Table 1. Comparison of Cisco UCS C – Series Rack-Mounts Server features

<table>
<thead>
<tr>
<th></th>
<th>Cisco UCS C200 M1 &amp; M2</th>
<th>Cisco UCS C210 M1 &amp; M2</th>
<th>Cisco UCS C250 M1 &amp; M2</th>
<th>Cisco UCS C260 M1 &amp; M2</th>
<th>Cisco UCS C460 M1 &amp; M2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multicore Processors</strong></td>
<td>Up to 2 x Intel® Xeon® processor 5600 series</td>
<td>Up to 2 x Intel® Xeon® processor 5600 series</td>
<td>Up to 2 x Intel® Xeon® processor 5600 series</td>
<td>Up to 4 x Intel® Xeon® processor E7-2800 product family</td>
<td>Up to 4 x Intel® Xeon® processor E7-4800 product family</td>
</tr>
<tr>
<td><strong>Form Factor</strong></td>
<td>1 RU</td>
<td>2 RU</td>
<td>2 RU</td>
<td>2 RU</td>
<td>4 RU</td>
</tr>
<tr>
<td><strong>Maximum memory</strong></td>
<td>192 GB</td>
<td>192 GB</td>
<td>384 GB</td>
<td>1 TB</td>
<td>1 TB</td>
</tr>
<tr>
<td><strong>Internal disk drive</strong></td>
<td>Up to 8</td>
<td>Up to 16</td>
<td>Up to 8</td>
<td>Up to 16</td>
<td>Up to 12</td>
</tr>
<tr>
<td><strong>Maximum Storage Capacity</strong></td>
<td>Up to 8 TB</td>
<td>Up to 16 TB</td>
<td>Up to 8 TB</td>
<td>Up to 16 TB</td>
<td>Up to 12 TB</td>
</tr>
<tr>
<td><strong>Built-In RAID</strong></td>
<td>0 and 1 (SATA only)</td>
<td>0 and 1 (5 SATA drives only)</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td><strong>Optional RAID</strong></td>
<td>0, 1, 5, 6, 10, 50, and 60</td>
<td>0, 1, 5, 6, 10, 50, and 60</td>
<td>0, 1, 5, 6, 10, 50, and 60</td>
<td>0, 1, 5, 6, 10, 50, and 60</td>
<td>0, 1, 5, 6, 10, 50, and 60</td>
</tr>
<tr>
<td><strong>Integrated networking</strong></td>
<td>2x integrated Gb Ethernet; 10 Gb unified fabric</td>
<td>2X integrated Gb Ethernet; 10 Gb unified fabric</td>
<td>4X integrated Gb Ethernet; 10 Gb unified fabric</td>
<td>2X GbE LOM ports; 2X 10 GbE ports</td>
<td>2X 10 Gigabit Ethernet LAN-on-motherboard</td>
</tr>
</tbody>
</table>
We used UCS C460 M1 in our testing. The Cisco UCS C460 M1 has the processor, memory, and local storage needed to handle the most demanding applications and server consolidation workloads. It expands the capabilities of unified computing to the most demanding and enterprise-critical applications and is the only Intel Xeon 7500 Series based rack-mount system with a built-in migration path to unified computing.

By extending unified computing to a wider range of workloads, IT teams can liberate resources and respond more quickly to business demands. The scalable performance and reliability features of Intel Xeon 7500 series processors helps increase virtualization and unify performance intensive applications in a standardized, simplified infrastructure.

The Cisco UCS C460 M1 server features:

- 4 rack unit (RU) rack-mount chassis
- 64 dual in-line memory module (DIMM) slots and up to 1 TB based on Samsung's 40 nm class DDR3 technology
- Up to 12 small form-factor (SFF) optional SAS or SATA hot plug hard drives 10 x PCIExpress (PCIe) slots
- Two 1/10 Gb Ethernet LAN on Motherboard (LOM) ports, two 10/100/1000 LOM ports plus baseboard management controller (BMC)

<table>
<thead>
<tr>
<th></th>
<th>Cisco UCS C200 M1 &amp; M2</th>
<th>Cisco UCS C210 M1 &amp; M2</th>
<th>Cisco UCS C250 M1 &amp; M2</th>
<th>Cisco UCS C260 M1 &amp; M2</th>
<th>Cisco UCS C460 M1 &amp; M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/O through PCIe</td>
<td>Two, half-length, PCIe x8 slots: one full height and one low profile slot</td>
<td>Five, full-height, PCIe, x8 slots: two full length and three half-length slots</td>
<td>Five PCIe slots: Three low-profile, half-length x8 slots; 2 full-height, half-length, x16 slots</td>
<td>Three low-profile, half-length x8 slots; 2 full-height, half-length x16 slots; 1 low-profile, half-length x4</td>
<td>Ten, full-height, PCIe slots: 4 half-length slots, 6 three quarter length slots; 2 Gen 1 slots, 8 Gen 2 slots</td>
</tr>
</tbody>
</table>
Figure 3. Cisco UCS C-Series are designed to operate in a wide range of Data Center Environments shown in Figure 3, including those using Cisco Unified Computing System, Cisco Nexus Family Products, and discrete Ethernet and fiber channel switches from Cisco and third parties.

Cisco UCS P81E Virtual Interface Cards

Product Overview

A Cisco innovation, the Cisco UCS P81E Virtual Interface Card is a virtualization-optimized Fibre Channel over Ethernet (FCoE) PCI Express (PCIe) 2.0 x8 10-Gbps adapter designed for use with Cisco UCS C-Series Rack-Mount Servers. The virtual interface card is a dual-port 10 Gigabit Ethernet PCIe adapter that can support up to 128 PCIe standards-compliant virtual interfaces, which can be dynamically configured so that both their interface type (network interface card [NIC] or host bus adapter [HBA]) and identity (MAC address and worldwide name [WWN]) are established using just-in-time provisioning. In addition, the Cisco UCS P81E can support network interface virtualization and Cisco VN-Link technology.

Features and Benefits

Table 2. Summary of the features and benefits of Cisco UCS P81E

<table>
<thead>
<tr>
<th>Feature</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen 2x8 PCIe interface</td>
<td>Allows high throughput and is PCI 2.0 compliant</td>
</tr>
<tr>
<td>22 watts (W) typical power demand</td>
<td>Reduces power and cooling requirements compared to separate Ethernet and Fibre Channel adapters</td>
</tr>
<tr>
<td>Unified I/O</td>
<td>Helps reduce TCO by reducing the overall number of NICs, HBAs, cables, and switches because LAN and SAN traffic is consolidated on the same adapter and fabric</td>
</tr>
<tr>
<td>Up to 128 dynamic virtual adapters</td>
<td>Allows up to 2 Fibre Channel and 16 Ethernet adapters; will allow up to 128 virtual Fibre Channel and/or Ethernet adapters (future) Helps create a highly flexible I/O environment needing only one card for all I/O</td>
</tr>
</tbody>
</table>
### Feature | Benefit
--- | ---
configurations, including virtualized environments
**Cisco VN-Link technology** | Provides network visibility to virtual machines; configurations and policy follow the virtual machine during migration.
**Management** | Local management of adapter; managed by Cisco UCS Manager (future)
**Support Lossless Ethernet** | Uses priority flow control (PFC) and IEEE 802.3x Pause to enable FCoE as part of the Cisco unified fabric
**Pass through switching** | Reduces virtual machine latency and increases performance through remotely configurable virtual NICs (future)
**Broad OS and hypervisor support** | Supports customer requirements for Microsoft Windows Server 2008, Red Hat Enterprise Linux 5.4, and VMware vSphere 4 Update 1 and 5.0

### Cisco UCS P81E Architecture
The Cisco UCS P81E brings adapter consolidation to physical environments. The adapter can be defined as multiple different NICs and HBAs. For example, one adapter can replace two quad-port NICs and two single-port HBAs, resulting in fewer NICs, HBAs, switches, and cables.
Figure 4. P81-E Architecture

Tested Configurations and Methodology
Tests were conducted on UCS C460 consists of four Intel Xeon CPU X7560 (@ 2.27 GHz), 512GB of memory (64 number of DIMM each of 8GB) and one Cisco UCS P81E (Monterey park cards) with 10GB bandwidth. Microsoft windows 2008 R2 64Bit was installed on the physical server and enabled Hyper-V role. Multiple virtual machines were created where each virtual machine was assigned 48GB of memory and Windows Server 2008 R2 64 bit was installed on the virtual machine.
Table 3. Hardware and software used during testing

**Physical Server**
C460 M1  
512 GB Memory  
Intel Xeon CPU X7560 @ 2.27 GHZ  
OS – Windows Server 2008 R2 64 Bit

**Virtual Machine configuration – 10 Numbers**
4 vCPU's  
48 GB of memory  
Windows 2008 R2 enterprise version  
SQL Server 2008 R2 enterprise version

**Storage Configuration**
RAID 5 (4+1) Disks x 5 – LUN for SQL Data  
RAID 1/0 (8+8) Disks – LUN for SQL Log  
RAID 5 (4+1) Disks – LUN for VHD files

Table 4. Storage LUN layout

<table>
<thead>
<tr>
<th>Raid Type</th>
<th>Raid Group</th>
<th>Disk Used</th>
<th>Lun</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>RG: 1 (4+1)</td>
<td>5 Disks</td>
<td>VM1 - DATA</td>
<td>SP-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VM2 - DATA</td>
<td>SP-B</td>
</tr>
<tr>
<td>5</td>
<td>RG: 2 (4+1)</td>
<td>5 Disk</td>
<td>VM3 - DATA</td>
<td>SP-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VM4 - DATA</td>
<td>SP-B</td>
</tr>
<tr>
<td>5</td>
<td>RG: 3 (4+1)</td>
<td>5 Disk</td>
<td>VM5 - DATA</td>
<td>SP-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VM6 - DATA</td>
<td>SP-B</td>
</tr>
<tr>
<td>5</td>
<td>RG: 3 (4+1)</td>
<td>5 Disk</td>
<td>VM7 - DATA</td>
<td>SP-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VM8 - DATA</td>
<td>SP-B</td>
</tr>
<tr>
<td>5</td>
<td>RG: 3 (4+1)</td>
<td>5 Disk</td>
<td>VM9 - DATA</td>
<td>SP-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VM10 - DATA</td>
<td>SP-B</td>
</tr>
<tr>
<td>1/0</td>
<td>RG: 4 (8+8)</td>
<td>16 Disk</td>
<td>LOG LUN (VM1, VM2 ----- VM10)</td>
<td>Alternate SP Owner's</td>
</tr>
<tr>
<td>5</td>
<td>RG: 5 (4+1)</td>
<td>5 Disk</td>
<td>VHD OS ( VM1, VM2 ----- VM6)</td>
<td>Alternate SP Owner's</td>
</tr>
</tbody>
</table>
Figure 5. LUN layout from Storage

Figure 6. Connectivity Diagram
We used single Monterey Park (P81-E, Palo) cards (CNA – FCoE card) in our testing as FCoE card. We created 2 vHBA’s and 11 vETH’s shown in figure 7. vHBA’s were used as initiators ports and each port was connected to a Nexus Switch (FCoE Switch). Each initiator port was connected and zoned with two front end FC ports of EMC – CLARiiON CX4 (used as a target) for dual redundancy/high availability. Out eleven vETH’s one vETH was configured as Ethernet/Management port for physical server and rest were mapped with 10 virtual machines. This configuration helped to ensure that each virtual machine can be managed individually by an IP address.

**Figure 7.** P81E configuration – Two vHBA’s and 11 vNIC’s were created from a single Palo adapter

A recognized open source Online Transaction Processing (OLTP) benchmark application was used in our tests to generate OLTP workload. This workload was used to evaluate performance and scalability of the server.

Generated OLTP workload (100GB Database) was loaded to each virtual machine running one SQL Server Instance and hosting one large database. We used the following parameters from OLTP client machine to push the heavy workload to each database.

**Table 5.** OLTP workload client parameter

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Concurrent users (Varies based on Workload &amp; db_Size)</td>
<td>18</td>
</tr>
<tr>
<td>Startup rate (users/sec)</td>
<td>10</td>
</tr>
<tr>
<td>Warm-up time before statistics are kept (min)</td>
<td>15</td>
</tr>
<tr>
<td>Runtime during which statistics are kept (min)</td>
<td>30</td>
</tr>
</tbody>
</table>
In our test case processor and memory utilization was captured along with OPM (Order per Min). Each test was run for 30 min with 15 min of warm up time and utilization of each virtual machine and physical server was captured using the Performance Monitor.

### Table 6. Physical and virtual machine configuration

<table>
<thead>
<tr>
<th>PHYSICAL MACHINE</th>
<th>VIRTUAL MACHINE</th>
<th>DATABASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>C460 M1</td>
<td>Each Hyper-V VM Configuration</td>
<td>One Database in one Virtual Machine with single database of size 100GB.</td>
</tr>
<tr>
<td>4 Core Intel Xeon CPU X7560 @ 2.27 GHz, 512GB Memory of 1033 MHz speed of each DIMM</td>
<td>vCPU – 4</td>
<td>Min SQL Server Memory – 40 GB</td>
</tr>
<tr>
<td></td>
<td>VM Memory – 48 GB</td>
<td>Max SQL Server Memory – 45 GB</td>
</tr>
<tr>
<td></td>
<td>One Network Adapter per VM</td>
<td>Max Degree of Query Parallelism - 1</td>
</tr>
<tr>
<td></td>
<td>Each VM was assigned with one data and log VHD</td>
<td></td>
</tr>
</tbody>
</table>

### Workload

An open source OLTP Simulation application was used to generate OLTP workload. The driver of this tool was included to stress the database component.

Cisco UCS C460 was used for the testing running Microsoft Windows 2008 R2 Datacenter Edition with Hyper-V enabled. During this consolidation we created 10 Virtual Machine with 4 vCPU and 48 GB memory assigned to each virtual machine. Microsoft SQL Server 2008 datacenter version was installed on each virtual machine. We created database of 100GB in one SQL Server Instance running on each virtual machine. Each virtual machine was assigned a VHD (virtual hard drive) file to install an operating system. We created ten Raid 5 LUNS for data and ten Raid 1/0 LUNS for log (refer to table 4). Data and log LUNs were attached as disks to each virtual machine through hyper V’s SCSI controller. This configuration made sure that DATA and LOG files do not lie on the same raid disk.

All data were stored on storage system (EMC-CX4) and were accessed through Fiber Channel. Client machine was running Windows 2003 64 bit to initiate the workload using OLTP workload client for each database.

We started the test with one database in single SQL Server Instance on one virtual machine and captured the benchmark results and resource utilization. We started adding virtual machine one at a time to the same physical server and started the workload all virtual machines simultaneously. In this way we were able to scale 10 virtual machines with 48GB of virtual machine memory on the same physical server. During this iteration, benchmark scores (OPM, TPM and IOPS) and throughput scaled linearly till 8 virtual machine’s. After eight virtual machine’s OPM dropped down by 8% as physical server ran out of memory. However the validation results were expected while running benchmark iteration on 10 virtual machines holding single SQL server instance because memory was 98% utilized on the physical server.

Client machine ran a single instance of OLTP workload driver with 18 threads which simulated a OLTP workload. The load generating client ran with .001 think time which queued requests quickly and the C460 was able to handle the load properly. We ran this workload on all virtual machines simultaneously
for 30 min with 15 warm up time. We considered Cisco UCS rack mount C460 M1 was able to deliver good performance as total OPM of the server did not decrease.

Test Results

The following graph in Figure 8 describes number of average OPM recorded from each virtual machine. The tests was started with one database of 100GB size in SQL server instance running on virtual machine and test results were captured at the end of the test. We kept on adding one virtual machine at a time with same memory and database configuration. From every virtual machine count we captured benchmark metrics of the OLTP workload. We were able to get best utilization of CPU and memory of each virtual machine until 10 virtual machines were deployed on the same physical server.

Figure 8. Average OPM per Virtual Machine

![Average OPM per VM](image)

Figure 8 shows we were able to scale up to 10 virtual machines with 48 GB of memory per virtual machine out of 512GB total physical memory. Beyond 8 virtual machine’s we saw OPM dropped by 10 percent due to memory constraint. It was observed that running Hyper-V virtual machines on the single server the most significant factor was the physical memory constraint. This is because each virtual machine must reside in non-paged-memory and the non-paged-memory cannot be paged back to disk. The physical server that hosts virtual machine should have available physical memory equal to the sum of memory allocated for each virtual machine. Hyper-V manager was used to allocate 48 GB of memory in hyper-v manager, the actual physical memory used when running a virtual machine was approximately 49492 MB with overhead of hypervisor layer. As 10 virtual machine were enabled the memory consumed by Hyper-V virtual machine would be approximately 492220 MB.
Figure 9. Average transaction per minute per virtual machine

Figure 9 shows average transaction per min. Beyond 8 virtual machines we saw a drop in TPM because there was memory constraint and more number of transactions were happening.

Figure 10. Total data IOPS from SQL Server

Figure 10 shows total platform I/O throughput obtained in each iteration of the test. The IOPS increased with the increasing number of virtual machines. Since each iteration activates an additional instance of SQL Server running the same workload on each virtual machine, the regular increase in workload was expected.
Figure 11. Average disk response time from SQL Data Disks

Figure 11 shows the average physical disk response time. Storage latency plays a critical role in the ability of an application such as SQL Server to maintain the level of performance for the response times required by OLTP applications. Given that the host platform has no bottlenecks, both the storage array and the storage network must respond quickly to keep response times low. We could see with the increasing number of VM’s response time increased linearly. Beyond 8 VM’s as there was a memory constraint there were more physical reads than logical reads hence response time increases with increasing number of VM’s which was expected.

Figure 12. Physical server CPU utilization

Figure 12 summarizes physical server utilization in concurrently running VM’s from 1 through 10 with concurrent SQL server instances attached to each VM. Physical server utilization was measured from hypervisor counter “Hyper-V Hypervisor Virtual Processor Hypervisor Run Time”. From the figure we see
that CPU utilization increased linearly after enabling one virtual machine at a time. Virtual machine CPU utilization is very much consistent which is expected because every SQL Server instance is performing the same work.

Conclusion

The following test study provided an overview of consideration to develop a strategy to maximize hardware utilization and reducing the cost associated with database sprawl. It describes the strategies and reference architectures as a starting point to consolidate SQL Server on the Cisco Blade server using a building block approach to design, configure, and deploy using best practice recommendations to simplify IT. To simplify the design and deployment on a virtualization infrastructure – Cisco offers Solution Architecture bundles for Rack Mounts, Microsoft Hyper-V. This bundle provides configuration and best practices to achieve full redundancy with no single point of failure, scalability and ease of management.

The tests performed in Cisco labs showed that significant gains can be achieved by developing a strategy to maximize hardware utilization, reducing sprawl, power and cooling costs by consolidating and virtualizing SQL Server on the latest Cisco rack mount servers, while providing the performance to meet the most demanding customers’ workloads.

Acknowledgements

This paper was made possible through the efforts of Sheshgiri I and Vadi Bhatt with special thanks to Frank Cicalese.