Microsoft SharePoint Server 2010 on Cisco Unified Computing System

Medium Farm Solution-Performance and Scalability Study

White Paper

June 2011

Contents

Executive Summary

A Microsoft SharePoint Server 2010 Environment comprising of various servers which collectively hosts the core applications and provide services, termed as SharePoint Farm is responsible for providing various functions to its users. A three-tier architectural topology in which the SharePoint tiers (Web, Application, and Database) are deployed using independent servers responsible for each tier is among the most used SharePoint 2010 farm topology. To size each tier of a SharePoint farm demands detailed study of the workload requirement and its usage patterns along with the performance capabilities of each hardware component deployed in the system.

This paper illustrates performance of a medium SharePoint farm, built using Cisco UCS servers implementing three tier architecture. A load generation framework to perform a load test was developed by the SharePoint engineering team at Cisco, to measure the performance metrics while keeping the much important response time less than 1 second. The paper shares the test results which provide accurate guidance, better understanding of the performance impact of different SharePoint workloads, assist our customers in sizing and designing the best farm architecture to support different workloads, while recommending the best infrastructure elements for the optimal SharePoint implementation.

This study details information on how the recommended farm architecture could support 50000 users with 10% of total users concurrently working and show how to achieve sub second response time. Also, highlights the performance benefits of new Cisco Servers used for this study.

Introduction

Turning users into participants, allowing users to easily create, share and connect with information, applications and people is what Microsoft SharePoint offers. SharePoint 2010 evolves from an earlier version of SharePoint Office 2007. SharePoint 2010 provides all the good features present in the earlier version like collaboration, information sharing, document management along with several new features and important architectural changes to improve the product.

Essentials for any SharePoint capacity planning would be to understand requirements like Total number of users, Number of concurrent Users, Number of service applications etc. And how would the requirement scale in the future. Factors like custom developed application which can be a part of an implementation can also influence the sizing of storage and servers.

Objective

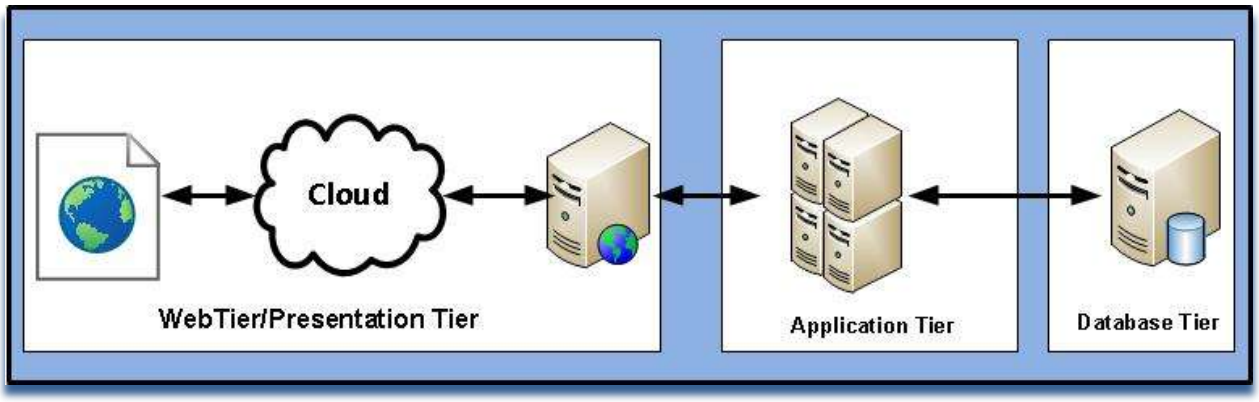
The main objective of this paper is to study the capacity and performance characteristics of a medium SharePoint 2010 farm using Microsoft Visual Studio 2008 Team System to generate the workload which consists of SharePoint Collaboration, Search and design based on the most adopted three-tier architectural topology built using Cisco UCS Blade Server B200 M2, Cisco UCS Blade Server B250 M2.

SharePoint 2010 Three Tier Architecture

We at Cisco have made sure to design a SharePoint Medium Farm Topology which typically depicts the real world incorporation or depicts the most realistic implementation with best of technologies developed by Cisco . A three tier architecture was incorporated, to build a SharePoint 2010 medium farm to achieve its objective.

Figure 1 details the workflow of three tier web application into five simplistic events.

Figure 1. Three Tier Web Application



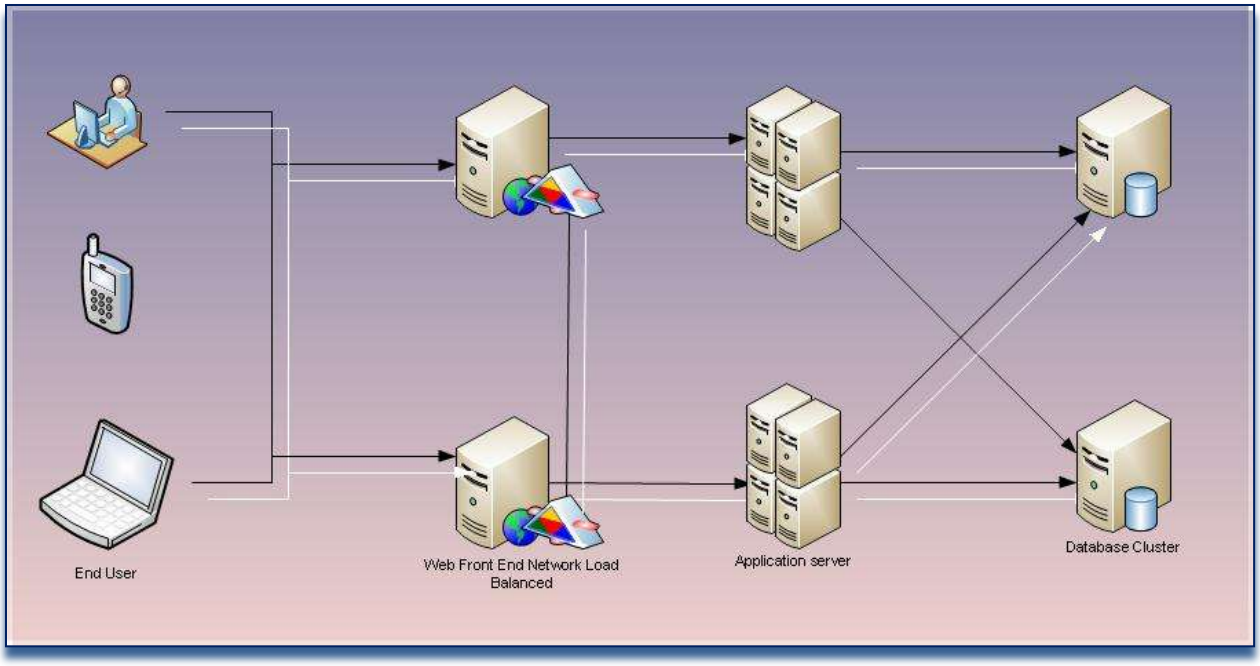
Why a Three Tier Architecture?

Among many major benefits three-tier architecture provides, a few are illustrated below:

- Maintainability: Three-tier architecture follows a modular approach; each tier is independent of each other. Thus each tier can be updated without affecting application as a whole.
- Scalability: Scalability is a major benefit to incorporate three-tier architecture, we can scale each tier as and when required without dependence on the other tiers. I.e. independently scaling each tier. For instance in scaling of web servers, provisioning of servers at multiple geographical locations enables faster end user response time, avoiding high network latency. Another aspect is scaling of application servers which require high computing resources, thus a farm of clustered application servers can provide on demand application performance.
- Availability and Reliability: Applications using three tier approach can exploit three-tier modular architecture to scale components and servers at each tier , thus providing redundancy , avoiding single point of failure and in-turn improving availability of overall system.

Figure 2 depicts the aspect of Deployment for Reliability and Availability.

Figure 2. Deployment for Reliability and Availability



Cisco Unified Computing

Leadership from Cisco

Cisco is the undisputed leader in providing network connectivity in enterprise data centers. With the introduction of the Cisco Unified Computing System, Cisco is now equipped to provide the entire clustered infrastructure for a three-tier web deployment. The Cisco Unified Computing System provides compute, network, virtualization, and storage access resources that are centrally controlled and managed as a single cohesive system. With the capability to scale to up to 320 servers and incorporate both blade and rack-mount servers in a single system, the Cisco Unified Computing System provides an ideal foundation to cater ever increasing demand for improved performance and scalability of deployments.

Introducing the Cisco Unified Computing System

The Cisco Unified Computing System is a next-generation data center platform that unites compute, network, storage access, and virtualization into a cohesive system designed to reduce total cost of ownership (TCO) and increase business agility. The Cisco Unified Computing System server portfolio consists of the Blade Server platform, B-Series and the C-Series Rack Mount platform. We chose the Cisco UCS B-Series Blade Server platform for this study. The system integrates a low-latency, lossless Infrastructure Components.

Figure 3. Cisco Unified Computing System

Cisco UCS Manager

Cisco Unified Computing System (UCS) Manager provides unified, embedded management of all software and hardware components of the Cisco UCS. It controls multiple chassis and thousands of virtual machines.

Cisco UCS Manager provides unified management domain with centralized management capabilities. Cisco UCS Manager is embedded device management software that manages the system from end to end as a single logical entity through an intuitive GUI, a command-line interface (CLI), or an XML API.

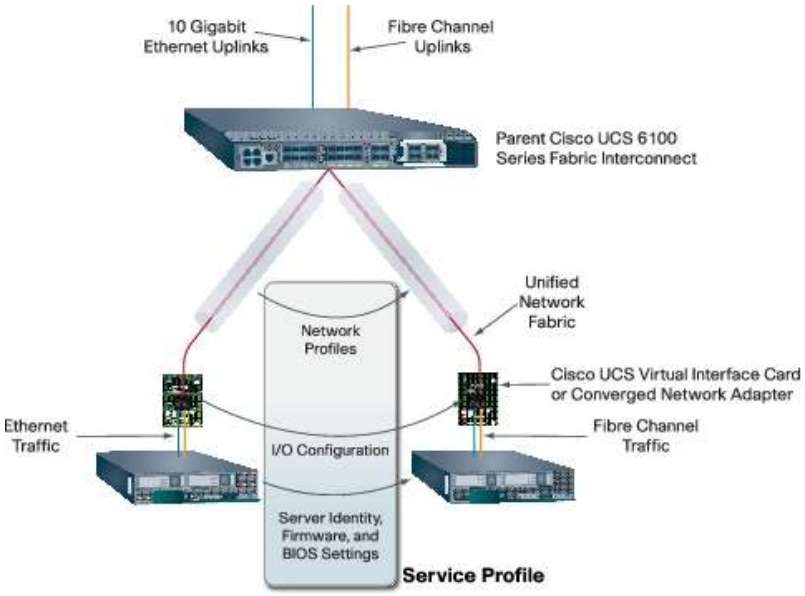
Service Profiles and Templates

Service profiles are the central concept of Cisco UCS. Each service profile serves a specific purpose: ensuring that the associated server hardware has the configuration required to support the applications that need be hosted.

The service profile maintains configuration information about the server hardware, interfaces, fabric connectivity, and server and network identity. Cisco UCS manager can deploy the service profile on any physical server at any time. All service profiles are centrally managed and stored in a database on the fabric interconnects.

The real power of the service profile becomes evident in templates. A service profile template parameterizes the UIDs that differentiate one instance of an otherwise identical server from another.

Figure 4. Service Profile



An Overview of Cisco Unified Computing System

10 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. The system is an integrated, scalable, multi-chassis platform in which all resources participate in a unified management domain.

The main system components include:

Compute—the system is based on an entirely new class of computing system that incorporates blade servers based on Intel Xeon 5500 Series Processors. The Cisco UCS blade servers offer patented Cisco Extended Memory Technology to support applications with large datasets and allow more virtual machines per server.

Network—the system is integrated onto a low-latency, lossless, 10-Gbps unified network fabric. This network foundation consolidates what today are three separate networks: LANs, SANs, and high-performance computing networks. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing power and cooling requirements.

Virtualization—the system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.

Storage access—the system provides consolidated access to both SAN storage and Network Attached Storage (NAS) over the unified fabric. Unifying storage access means that the Cisco Unified Computing System can access storage over Ethernet, Fibre Channel, Fibre Channel over Ethernet (FCoE), and iSCSI, providing customers with choice and investment protection. In addition, administrators can pre-assign storage-access policies for system connectivity to storage resources, simplifying storage connectivity and management while helping increase productivity.

Management—the system uniquely integrates all the system components, enabling the entire solution to be managed as a single entity through the Cisco UCS Manager software. The Cisco UCS Manager provides an intuitive graphical user interface (GUI), a command-line interface (CLI), and a robust application programming interface (API) to manage all system configuration and operations. The Cisco

UCS Manager helps increase IT staff productivity, enabling storage, network, and server administrators to collaborate on defining service profiles for applications. Service profiles are logical representations of desired physical configurations and infrastructure policies. They help automate provisioning and increase business agility, allowing data center managers to provision resources in minutes instead of days.

Working as a single, cohesive system, these components unify technology in the data center. They represent a radical simplification in comparison to traditional systems, helping simplify data center operations while reducing power and cooling requirements. The system amplifies IT agility for improved business outcomes. The Cisco Unified Computing System components illustrated in Figure 1 include, from left to right, fabric interconnects, blade server chassis, blade servers, and in the foreground, fabric extenders and network adapters.

Fabric Interconnect

The Cisco UCS 6100 Series Fabric Interconnects are a core part of the Cisco Unified Computing System, providing both network connectivity and management capabilities for the system (Figure 2). The Cisco UCS 6100 Series offers line-rate, low-latency, lossless 10 Gigabit Ethernet and FCoE functions.

The Cisco UCS 6100 Series provides the management and communication backbone for the Cisco UCS B-Series Blade Servers and Cisco UCS 5100 Series Blade Server Chassis. All chassis, and therefore all blades, attached to the Cisco UCS 6100 Series Fabric Interconnects become part of a single, highly available management domain. In addition, by supporting unified fabric, the Cisco UCS 6100 Series provides both the LAN and SAN connectivity for all blades within its domain.

From a networking perspective, the Cisco UCS 6100 Series uses a cut-through architecture, supporting deterministic, low-latency, line-rate 10 Gigabit Ethernet on all ports, independent of packet size and enabled services. The product family supports Cisco low-latency, lossless 10 Gigabit Ethernet unified network fabric capabilities, which increase the reliability, efficiency, and scalability of Ethernet networks. The fabric interconnect supports multiple traffic classes over a lossless Ethernet fabric from the blade through the interconnect. Significant TCO savings come from an FCoE-optimized server design in which network interface cards (NICs), host bus adapters (HBAs), cables, and switches can be consolidated.

The Cisco UCS 6100 Series is also built to consolidate LAN and SAN traffic onto a single unified fabric, saving the capital and operating expenses associated with multiple parallel networks, different types of adapter cards, switching infrastructure, and cabling within racks. Fibre Channel expansion modules in the interconnect support direct connections from the Cisco Unified Computing System to existing native Fibre Channel SANs. The capability to connect FCoE to native Fibre Channel protects existing storage system investments while dramatically simplifying in-rack cabling.

Figure 5. Cisco UCS 6120XP 20-Port Fabric Interconnect (Top) and Cisco UCS 6140XP 40-Port Fabric Interconnect.

The Cisco UCS 6100 Series is equipped to support the following module options:

- Ethernet module that provides 6 ports of 10 Gigabit Ethernet using the SFP+ interface.
- Fibre Channel plus Ethernet module that provides 4 ports of 10 Gigabit Ethernet using the SFP+

interface; and 4 ports of 1/2/4-Gbps native Fibre Channel connectivity using the SFP interface .

- Fibre Channel module that provides 8 ports of 1/2/4-Gbps native Fibre Channel using the SFP interface for transparent connectivity with existing Fibre Channel networks.
- Fibre Channel module that provides 6 ports of 1/2/4/8-Gbps native Fibre Channel using the SFP or SFP+ interface for transparent connectivity with existing Fibre Channel networks.

Figure 6. From left to right: 8-Port 1/2/4-Gbps Native Fibre Channel Expansion Module; 4-Port Fibre Channel plus 4-Port 10

Cisco UCS 2100 Series Fabric Extenders

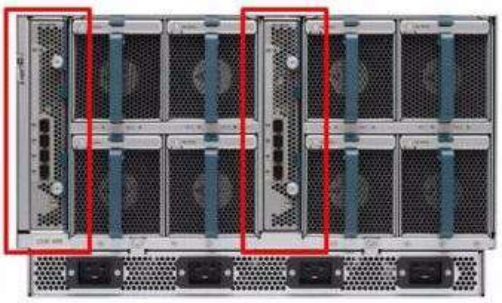
The Cisco UCS 2100 Series Fabric Extenders bring the unified fabric into the blade server enclosure, providing 10 Gigabit Ethernet connections between blade servers and the fabric interconnect, simplifying diagnostics, cabling, and management.

The Cisco UCS 2100 Series extends the I/O fabric between the Cisco UCS 6100 Series Fabric Interconnects and the Cisco UCS 5100 Series Blade Server Chassis, enabling a lossless and deterministic FCoE fabric to connect all blades and chassis together. Since the fabric extender is similar to a distributed line card, it does not do any switching and is managed as an extension of the fabric interconnects. This approach removes switching from the chassis, reducing overall infrastructure complexity and enabling the Cisco Unified Computing System to scale to many chassis without multiplying the number of switches needed, reducing TCO and allowing all chassis to be managed as a single, highly available management domain.

The Cisco 2100 Series also manages the chassis environment (the power supply and fans as well as the blades) in conjunction with the fabric interconnect. Therefore, separate chassis management modules are not required.

The Cisco UCS 2100 Series Fabric Extenders fit into the back of the Cisco UCS 5100 Series chassis. Each Cisco UCS 5100 Series chassis can support up to two fabric extenders, enabling increased capacity as well as redundancy.

Figure 7. Rear view of Cisco UCS 5108 Blade Server Chassis with two Cisco UCS 2104XP Fabric Extenders



Cisco UCS Chassis

The Cisco UCS 5100 Series Blade Server Chassis is a crucial building block of the Cisco Unified Computing System, delivering a scalable and flexible blade server chassis for today's and tomorrow's data center while helping reduce TCO.

Cisco's first blade server chassis offering, the Cisco UCS 5108 Blade Server Chassis, is six rack units (6RU) high and can mount in an industry-standard 19-inch rack. A chassis can house up to eight half-width Cisco UCS B-Series Blade Servers and can accommodate both half- and full-width blade form factors.

Four single-phase, hot-swappable power supplies are accessible from the front of the chassis. These power supplies are 92 percent efficient and can be configured to support non-redundant, N+ 1 redundant and grid-redundant configuration. The rear of the chassis contains eight hot-swappable fans, four power connectors (one per power supply), and two I/O bays for Cisco UCS 2104XP Fabric Extenders.

A passive mid-plane provides up to 20 Gbps of I/O bandwidth per server slot and up to 40 Gbps of I/O bandwidth for two slots. The chassis is capable of supporting future 40 Gigabit Ethernet standards.

Figure 8. Cisco Blade Server Chassis (front and back view)

Cisco UCS for SharePoint 2010

The Cisco Unified Computing System addresses many of the challenges faced during three-tier application deployment and maintenance. For instance, non-disruptive scaling of servers by addition of new servers to manage ever increasing demand of higher throughput.

Comprehensive Management

The system uses an embedded, end-to-end management system that uses a high-availability active-standby configuration. Cisco UCS Manager uses role and policy-based management that allows IT departments to continue to use subject-matter experts to define server, network, and storage access policy. After a server and its identity, firmware, configuration, and connectivity are defined, the server, or a number of servers like it, can be deployed in minutes, rather than the hours or days that it typically takes to move a server from the loading dock to production use. This capability ensures scaling-out of three-tier application without the tedious, manual assembly of individual hardware components and makes it a straightforward process.

Radical Simplification

The Cisco Unified Computing System represents a radical simplification compared to the way that servers and networks are deployed today. It reduces network access-layer fragmentation by eliminating switching inside the blade server chassis. It integrates compute resources on a unified I/O fabric that supports standard IP protocols as well as Fibre Channel through FCoE encapsulation. The system eliminates the limitations of fixed I/O configurations with an I/O architecture that can be changed through software on a per-server basis to provide needed connectivity using a just-in-time deployment model. The result of this radical simplification is fewer switches, cables, adapters, and management points, helping reduce cost, complexity, power needs, and cooling overhead.

High Performance

The system's blade servers are based on the Intel Xeon 5600 and 7500 series processors. These processors adapt performance characteristics based on application demands by increasing/decreasing the clock rate on specific processor cores as workload and thermal conditions permit. These processors, combined with patented Cisco Extended Memory Technology, deliver the best system performance along with the memory footprint needed to support large in-server caches. The system is integrated within a 10 Gigabit Ethernet–based unified fabric that delivers the throughput and low-latency characteristics needed to support the demands of the cluster's public network, storage traffic, and high-volume cluster messaging traffic.

Ready for the Future

The system gives room to scale while anticipating future technology investments. The blade server chassis, power supplies, and midplane are capable of handling future servers with even greater processing capacity. Likewise, the chassis is built to support future 40 Gigabit Ethernet standard when it becomes available.

Servers Used in Designing the SharePoint 2010 Farm Under Test

Cisco UCS B200 M2 Blade Server

The Cisco UCS B200 M2 Blade Server is a half-width, two-socket blade server. The system uses two Intel Xeon 5600 Series Processors, up to 96 GB of DDR3 memory, two optional hot-swappable small form factor (SFF) serial attached SCSI (SAS) disk drives, and a single mezzanine connector for up to 20 Gbps of I/O throughput. The server balances simplicity, performance, and density for production-level virtualization and other mainstream data center workloads.

Figure 9. Cisco UCS B200 M2 Blade Server

With the scope of this performance study UCS B200 M2 blade servers are used for the Web front end/ Application server tier and memory is configured with 32 GB. Blade B200 M2 Servers can support up to 96 GB of memory, which can accommodate future growth of increasing user's workloads and hosting more of SharePoint 2010 services and have room for applications demands in production environment.

Cisco UCS B250 M2 Extended Memory Blade Server

The Cisco UCS B250 M2 Extended Memory Blade Server is a full-width, two-socket blade server featuring Cisco Extended Memory Technology. The system supports two Intel Xeon 5600 Series Processors, up to 384 GB of DDR3 memory, two optional SFF SAS disk drives, and two mezzanine connections for up to 40 Gbps of I/O throughput. The server increases performance and capacity for demanding virtualization and large-data-set workloads with greater memory capacity and throughput.

Figure 10. Cisco UCS B250 M2 Extended Memory Blade Server

With the scope of this performance study UCS B250 M2 blade servers are used for the database tier and memory is configured with 40 GB. Blade B250 M2 Servers can support up to 384 GB of memory, which can accommodate multiple SharePoint 2010 farm, multiple instances and have room for application demands in production environment.

SharePoint 2010 Medium Farm Topology

Demystifying SharePoint Farm

In context with SharePoint, the term 'farm' is used to describe a collection of one or more SharePoint servers and one or more SQL servers that together provide a set of basic SharePoint services bound together by a single Configuration Database in SQL.

Farm marks the highest level of SharePoint administrative boundary. Everything that happens inside SharePoint happens inside this administrative boundary called Farm.

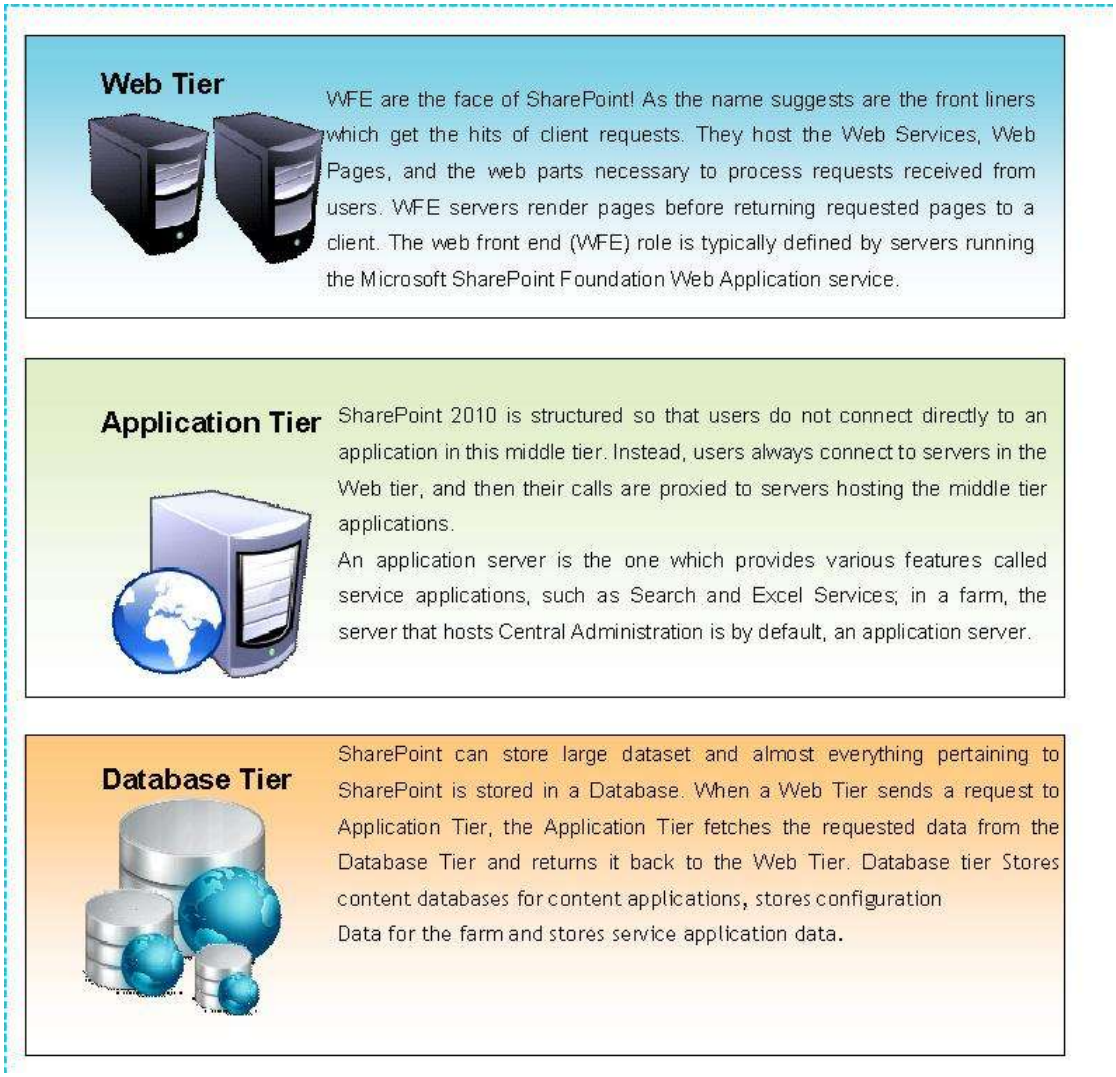
SharePoint 2010 can be configured as a small, medium, or large farm deployment. Remember, the topology service provides you with an almost limitless amount of flexibility, so you can tailor the topology of your farm to meet your specific needs.

Some of the common questions are How many servers would it require to create a Farm? Or what would a size of the farm be? What capacity a farm should posses? Etc. Proper sizing can be only gauged by analyzing the demand characteristics that a given solution is expected to handle i.e. understanding both the workload characteristics such as number of users, concurrent users at peak time, most frequently used operations, dataset characteristics such as content size and distribution. Based on current Enterprise needs a three tier SharePoint 2010 Medium farm was created for this performance study.. The farm under test have multiple servers serving various tiers to fulfill the most desired realistic enterprise needs apart from being flexible, scalable, maintainable and to be proofed from failures, making it a reliable farm.

A SharePoint Farm basically comprises of three tiers:

1. Web Tier
2. Application Tier
3. Database Tier

The following section briefly introduces each tier and its roles which together make a SharePoint farm.



WFE servers form the SharePoint connection point for clients that request content or services. This means that all client requests results with some load on the WFE servers.WFE servers render pages before returning requested pages to a browser. WFE servers do not require large quantities of disk storage, but rely heavily on processor and memory for performance.

The following table describes the processor and memory load characteristics for WFE servers.

Service application	CPU load	Memory load
SharePoint Foundation Service	High	High
Timer Service	Medium	Medium
Logging Service	Medium	Medium
Sandboxed Solutions	Low	Low
Workflow Capabilities	High	High

With the scope of this performance study UCS B200 M2 blade servers was used for the Web front end tier and memory was configured with 32 GB for the test.

Application Server

Different service applications have different workload profiles, but one can dedicate specific servers to specific service applications. Scale out can be achieved by assigning multiple servers for a specific service application.

Different service applications have different workload profiles, but most application services do not require local storage on the application server. The main hardware requirements for application servers are processor and memory.

The following table describes the processor and memory load characteristics for application servers.

Service application	CPU load	Memory load
SharePoint Foundation Service	None	None
Timer Service	Medium	Medium
Logging Service	None	None
User Profile Service	Medium	Medium
Word Viewing Service	High	Medium

With the scope of this performance study UCS B200 M2 blade servers was used for the Application tier and memory is configured with 32 GB for the test.

Database

All the data including content, configuration and metadata are stored in the SQL server. Not all service applications affect database servers, because some service applications do not require databases. However, storage access times and storage capacity are a key requirement for this role.

Service application	CPU load	I/O load	Storage
SharePoint Foundation Service	Medium	High	High
Timer Service	None	None	None
Logging Service	Medium	High	High
User Profile Service	High	High	Medium
Word Viewing Service	None	None	None
Sandboxed Solutions	None	None	None
Workflow Capabilities	None	None	None

With the scope of this performance study UCS B250 M2 blade servers was used for the Database Tier and memory is configured with 40 GB for the test.

SharePoint 2010 Search

SharePoint 2010 Search service is typically a significant benefits for users, but it places a large workload burden on the farm. When considering farm performance, one must often consider search performance specifically in the context of the farm.

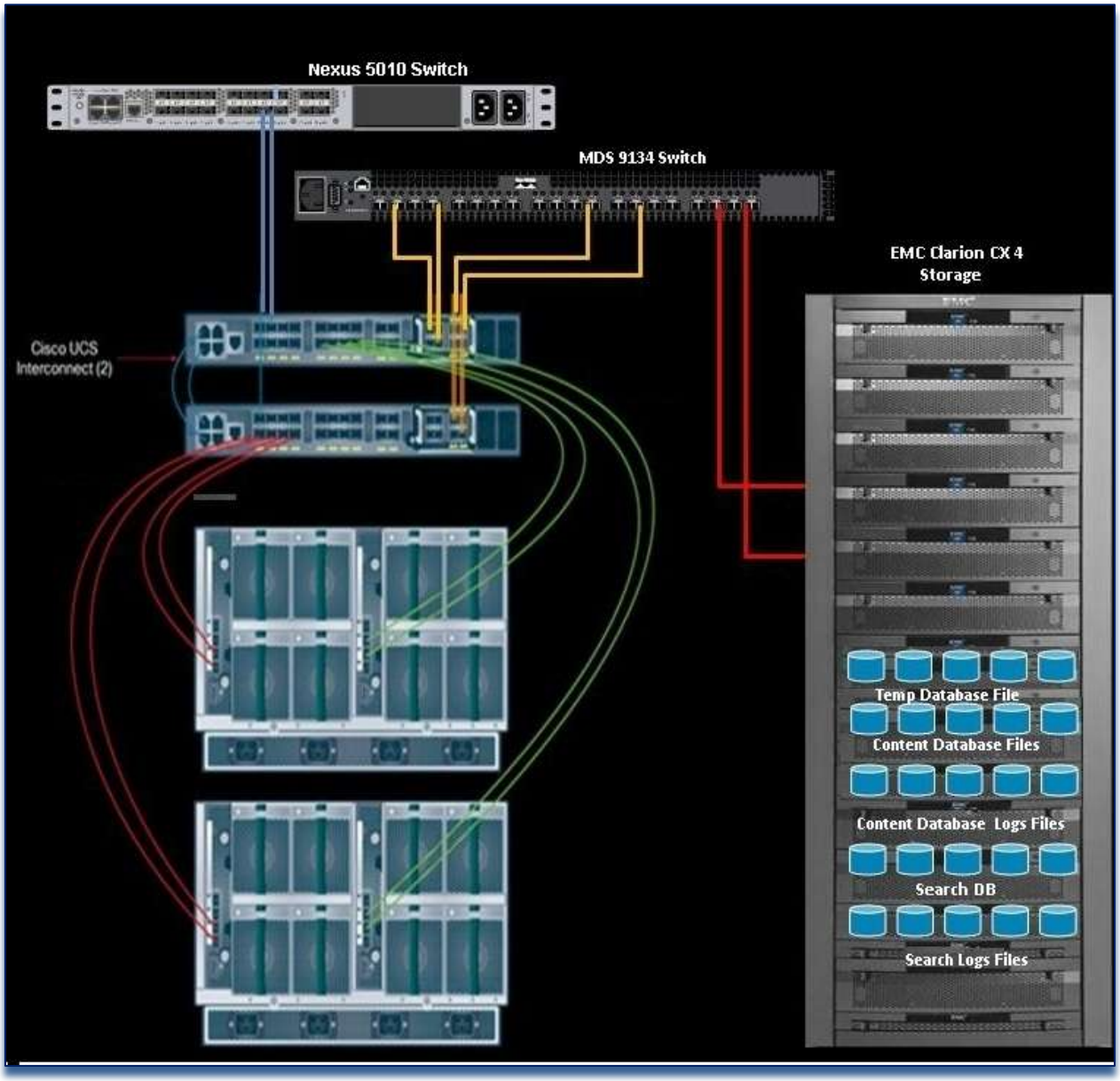
Search servers have two components:

- The crawl component crawls and indexes content primarily in the SharePoint content databases, although it can also index other types of storage Repository. The crawl role builds the index and submits index updates to the search query role. Crawl components aggressively use CPU bandwidth. Memory is not as critical for the crawl component.
- The query component responds to user search requests. When users enter a search in a SharePoint site, SharePoint 2010 submits the query to a server that hosts the query role to return a result set. All servers that host the query role have a copy of the index that the crawl role

generates. Web Front end server hosts the other additional service -Search Query services discussed in this paper

The below figure (Figure 12) depicts the SharePoint farm setup with various hardware component which occupies different tiers of the medium farm under test.

Figure 11. Configuration of the SharePoint 2010 Farm



The following table lists the various hardware and software components which occupies different tiers of the medium farm under test.

Vendor	Name	Version	Description
Cisco	Cisco 6120 XP	UCSM 1.3(1o)	Cisco UCS 6100 Series Fabric Interconnects
Cisco	Nexus 5010 Switch	NX-OS	Nexus Switch 5000 series

Vendor	Name	Version	Description
Cisco	MDS 9134 Switch	NX -OS	Fiber channel switch
Microsoft	Microsoft SharePoint server	SharePoint 2010	Web Front End
Microsoft	Microsoft SharePoint server	SharePoint 2010	Web Front End
Microsoft	Microsoft SharePoint Server	SharePoint 2010	Application Server
Microsoft	Microsoft SharePoint server	SharePoint 2010	Application Server
Microsoft	SQL Server	2008 R2	Database Server
Microsoft	SQL Server	2008 R2	Database Server
Microsoft	VSTS 2008	SP1	Load test Controller
Microsoft	VSTS 2008	SP1	Load test Agents

The following tables lists the details of all the hardware components used in the deployment.

Servers	Role	Processor	Memory Max Supported	Memory Configured	Storage	Network	Protocol
Cisco UCS Blade Server B200 M2	WFE Presentation Tier	Intel Xeon 5680	96 GB	32 GB	Internal Drive SAS 2x2.5”	10 GB	FCoE
Cisco UCS Blade Server B200 M2	CA ,Search APPLICATION Tier	Intel Xeon 5680	96 GB	32 GB	Internal Drive SAS 2x2.5””	10 GB	FCoE
Cisco UCS Blade Server B250 M2	SQL Server Database Tier	Intel Xeon 5680	384 GB	40 GB	EMC CLARION	10 GB	FC
Cisco UCS Blade Server B200 M2	VSTS CONTROLLER	Intel Xeon 5680	96 GB	32 GB	Internal Drive SAS 2x2.5”	10 GB	FCoE
Cisco UCS Blade Server B200 M2	VSTS AGENT	Intel Xeon 5680	96 GB	32 GB	Internal Drive SAS 2x2.5”	10 GB	FCoE

Network

Network Components	Quantity
Cisco Fabric interconnect 6120	2
Cisco MDS 9134 Fiber channel Switch	2
Nexus 5000 Series Switch	2

Storage

Storage Subsystem	EMC Clarion CX 4
Protocol	FC

Cisco UCS Statelessness

Cisco UCS brings in the concept of statelessness which makes the personality settings of one sever to be applied to another automatically at failover. This aids SharePoint administrators in executing change management activities like hardware upgrades.

It makes use of Cisco UCS Service Profiles' which capture the configuration settings for servers and LAN/SAN network access they require and all those low-level device configuration tasks that are needed to have the system up and running making it highly efficient and time saving at failover. Service Profile, in turn has a link to the storage parameters like host, storage group for the server in question.

Handling Hardware Crash

Recovering from a crash has now being made simple with Cisco UCS. In case of SharePoint server hardware crash at any tier, it's just a matter of associating the service profile to a new blade. With this the entire set of components, end-to-end is configured automatically without any further follow-up tasks or post processing.

SharePoint 2010 High Availability

Medium farm topology discussed in this paper, is architected with SharePoint High Availability. High Availability for a three tier SharePoint Farm under test was implemented in the fashion described.

To implement high availability for Web Front End servers, we chose two Web Front End servers that host the same Web application, with Microsoft Load balancing (NLB), a feature in Windows Server 2008.

NLB is available in Windows Server 2008 Web, Standard, Enterprise, and Datacenter editions.

- To implement high availability for service applications, we chose two application servers that host the same services.
- To implement high availability for databases, we chose two machines to run SQL Server. You can configure these for database mirroring (which requires duplicate storage), or you can configure them as part of a SQL Server failover cluster (which requires shared storage).

For the purpose for this paper. Database tier was implemented with SQL mirroring along with witness server. SQL Mirroring is integrated with SharePoint 2010 Application.

Farm administrator can configure the SQL environment to mirror a SharePoint content database first. When a new database is created, the SharePoint administrator will be prompted to enter the failover server. SharePoint then communicates with the witness server, and if issues are discovered, SharePoint will fail over to the hot backup of the environment.

Note: SharePoint 2010 server does not support all the service application databases for SQL mirroring.

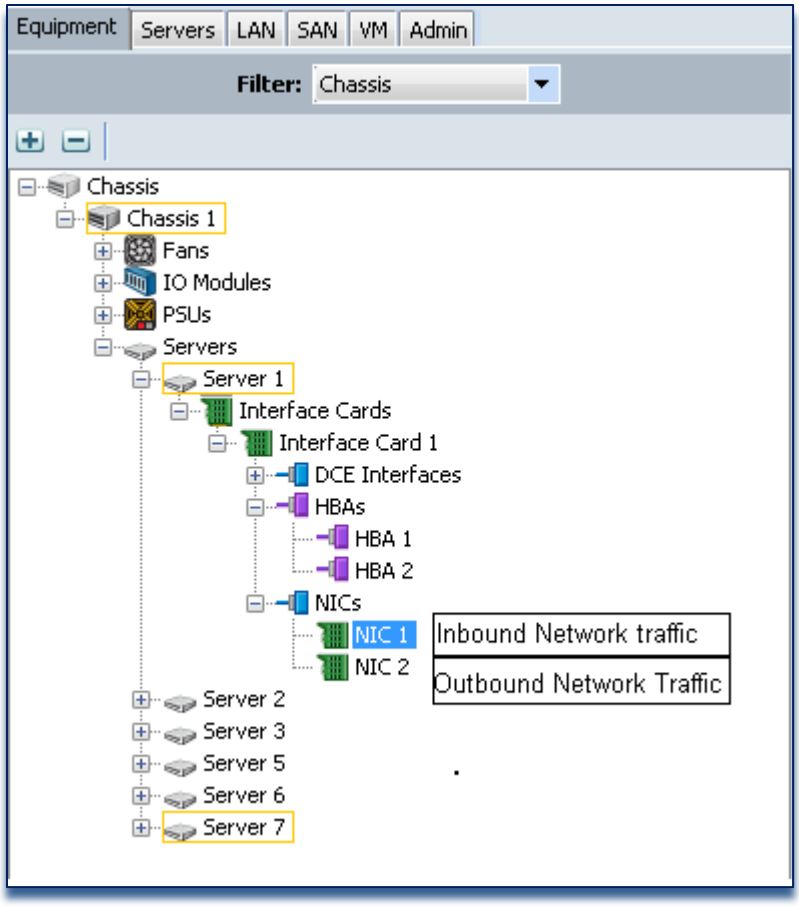
Network Design Considerations

In SharePoint environment, there is the possibility of high network traffic between clients and WFE's, and between WFE's and the database.WFE and Application servers.

Addition of new features in SharePoint 2010 such as Office Web Apps, digital asset storage and playback introduce additional network traffic requirements over previous versions of Office SharePoint.

Best practice is to separate the Client-WFE HTTP traffic from the WFE-database traffic and thus improve network efficiency.

To fulfill the high network demand of Microsoft SharePoint 2010 environment Cisco UCS system which is integrated onto a low-latency, lossless, 10-Gbps unified network fabric were employed. This network foundation consolidates what today are three separate networks: LANs, SAN's, and high-performance computing networks. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing power and cooling requirements making it efficient and a cost-effective network.

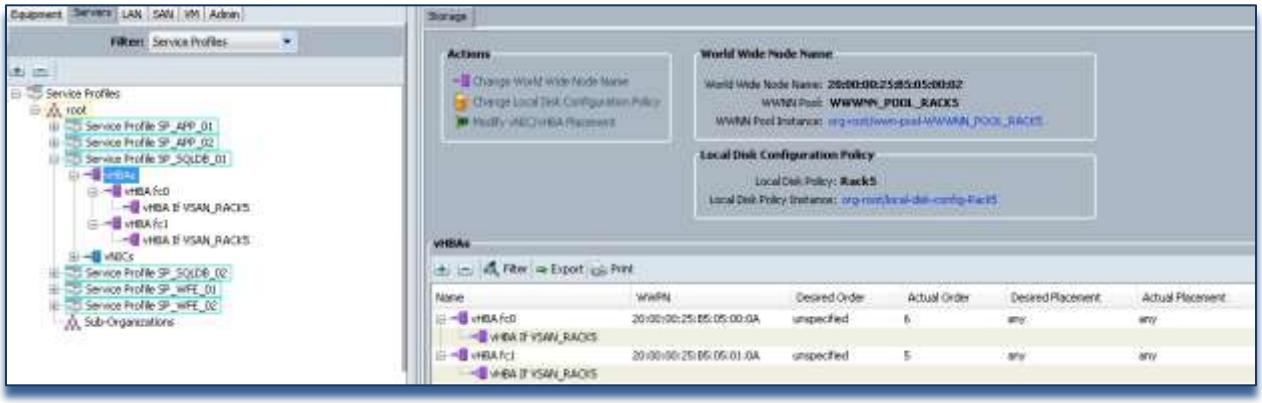


The access layer LAN configuration consists of a pair of Cisco Nexus 5010, a family of low-latency, line-rate, 10 Gigabit Ethernet and FCoE switches for our SharePoint 2010 deployment. Four 10 Gigabit Ethernet uplink ports are configured on each of the Cisco UCS fabric interconnects, and they are connected to the Cisco Nexus 5010.

Storage Considerations

When planning for content storage on SharePoint 2010, one must choose suitable storage architecture. SharePoint 2010 content storage has a significant dependency on the underlying database; therefore, database and SQL Server requirements will drive your storage choices.

In our scenario we have taken Storage area network (SAN), which uses a network infrastructure to connect the server running SQL Server to separate disk storage volumes.



A pair of Cisco MDS 9134 Multilayer Fabric Switches were used in the configuration to connect to the Fibre Channel port of the Cisco UCS fabric interconnects .Fibre Channel expansion module ports to a separate EMC storage .Fibre Channel ports. Cisco MDS 9000 Family single initiator zone was used to connect to the EMC Fibre Channel ports.

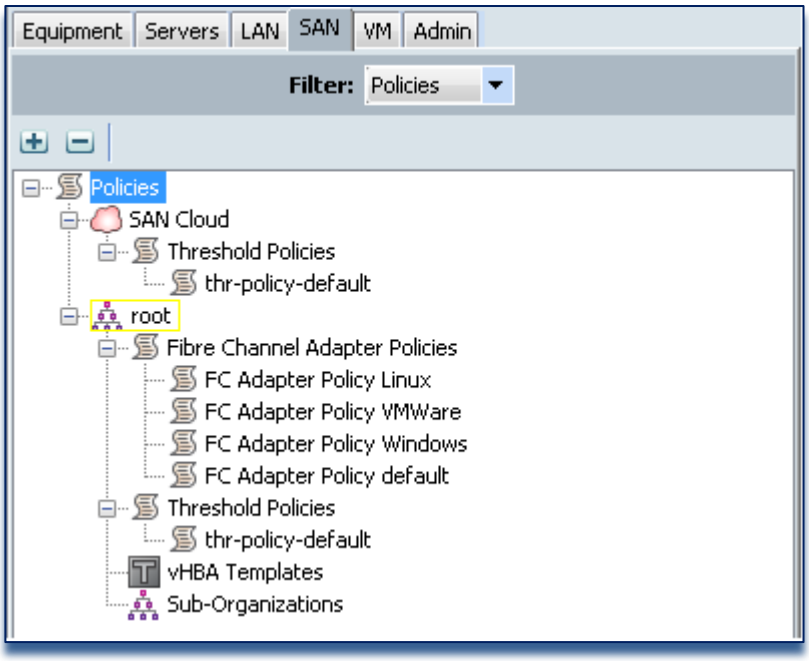
Storage Provisioning

SharePoint database storage is provisioned on separate LUN's for databases and logs.

Disks are configured with Raid 5 and Raid 10. Databases (.mdf) files hosted on Raid 5 and (.ldf) on Raid 10.

FC is used as the transport protocol to storage subsystem.

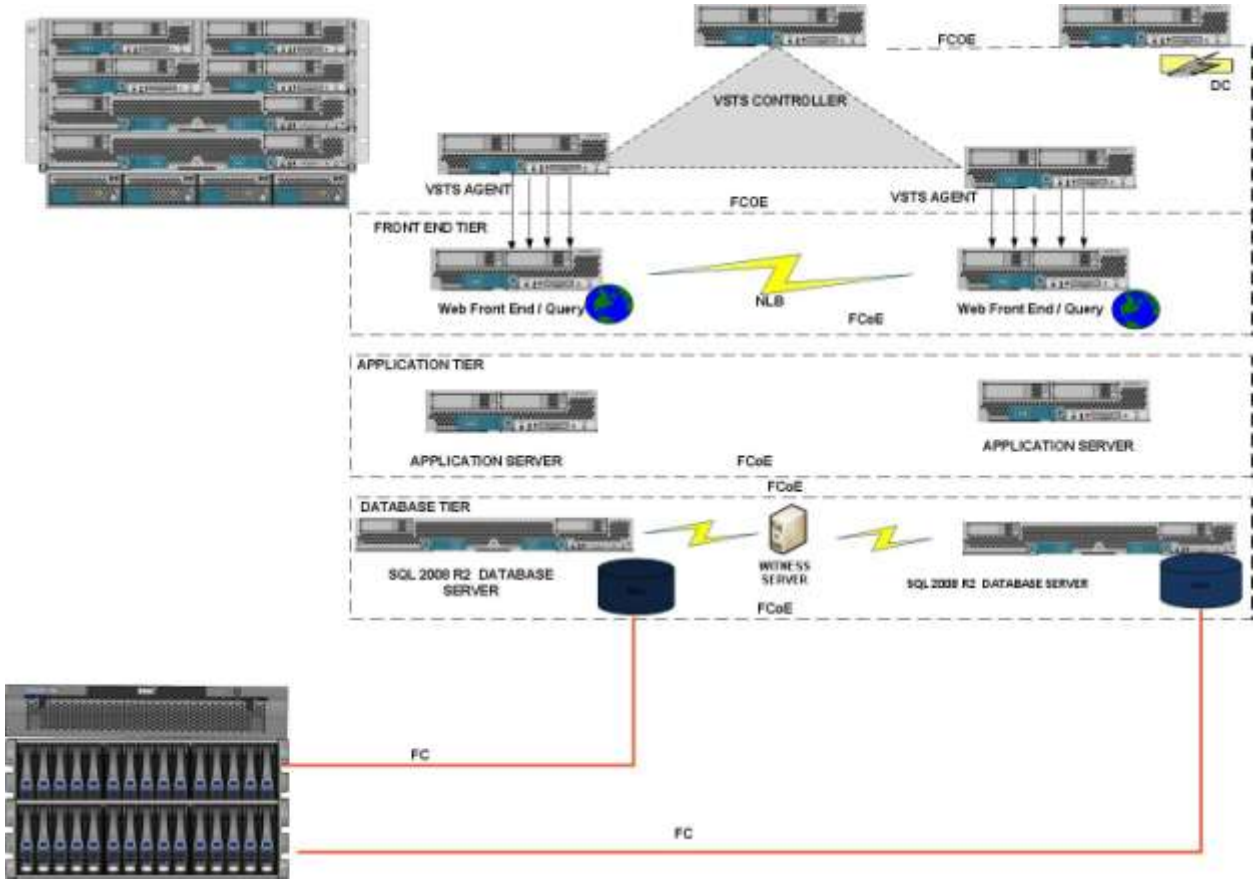
Cisco UCS pre-assign storage-access policies for system connectivity to storage resources, simplifies storage connectivity and management while helping to increase productivity.



SharePoint Farm Under Test

Figure 12 shows the specific architecture.

Figure 12. Architecture of the SharePoint Farm Under Test



The physical architecture of the test farm consists of number of servers and the network infrastructure, which together create the SharePoint Environment for this paper and is tailored with respect to the size and topology explained below.

Modeling a SharePoint Server 2010 environment begins with analyzing current requirements and estimating the expected future demand and targets for the deployment. Decisions are made on the key solution that an environment must support, and establish all important metrics and parameters. We at Cisco have modeled the SharePoint Server 2010 environment under test by taking into account the enterprise workload characteristics like number of users and the most frequently used operations and dataset i.e. content size and content distribution demands and tailored in accordance with Microsoft recommendations.

Workload Characteristics

- Workload Characteristics considered to size the farm under test

  Sizing a SharePoint environment workload is among the key factor. The system under test shall sustain the described work load demands, user base and usage characteristics. The following table provides some key requirements that are part of the workload.

Workload Characteristics	Value
Number Of Users (of unique users in a 24 hour period)	**50000**
Concurrent users at Peak Time (Distinct users generating requests in a given time frame)	**5000**

Dataset capacity

- Defining Dataset capacity of the farm under test

  A dataset which will hold the SharePoint content of the defined workload.

- Provides a flexibility of distributing of the content in the data store was created.

The following table provides few key metrics which were used to determine the capacity of the dataset for the test.

Dataset Characteristics	Value
Content Database size each	200
Number of sites	4000
Search index size (number of items)	4197282
Total number of documents	5759000
Number of content databases	2
Total number of databases	21
Number of site collections	400
Number of Web applications	5

Performance Test Framework

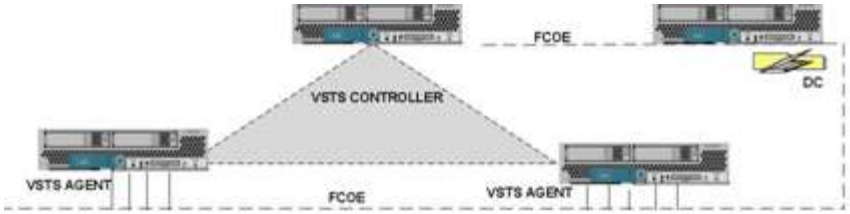
This performance test intends to determine the responsiveness, throughput, reliability, and/or scalability of a SharePoint 2010 farm under a given workload. The results from this performance test and analysis can help our customers estimate the hardware configuration required for a SharePoint 2010 farm to support up to 50000 users with 10% concurrency under production operation.

VSTS Test Rig

A group of servers are used to generate a simulated load for testing. A single controller machine was used run tests remotely and simultaneously on several servers with the help of one or more agents, collectively called a as rig. The rig is employed to generate more load than a single computer alone can generate. The controller is used to coordinate the agents i.e. send the load test to the agents and the agents perform the test received from the controller. The controller is also responsible to collect the test results.

Figure 13 depicts the rig created for this performance test. The rig consists of one controller and multiple agents.

Figure 13. Controller and Test Agents



Note: The agent takes as input a set of tests and a set of simulation parameters. A key concept in Test Edition is that tests are independent of the computer on which they are run.

Test Methodology

For this performance test the load on the farm was applied using a Microsoft LTK (Load Testing Kit). The kit was modified to make it more flexible and enhanced to generate the desired load. The Load Testing Kit (LTK) generates a Visual Studio Team System 2008 (VSTS) load test based on Windows SharePoint Services 3.0 Internet Information Services logs. The content database of size 180 GB was created which has sites and site collections and other important things which constitute the dataset. To get more details about the created dataset please refer to the “Defining Dataset capacity of the farm under test”

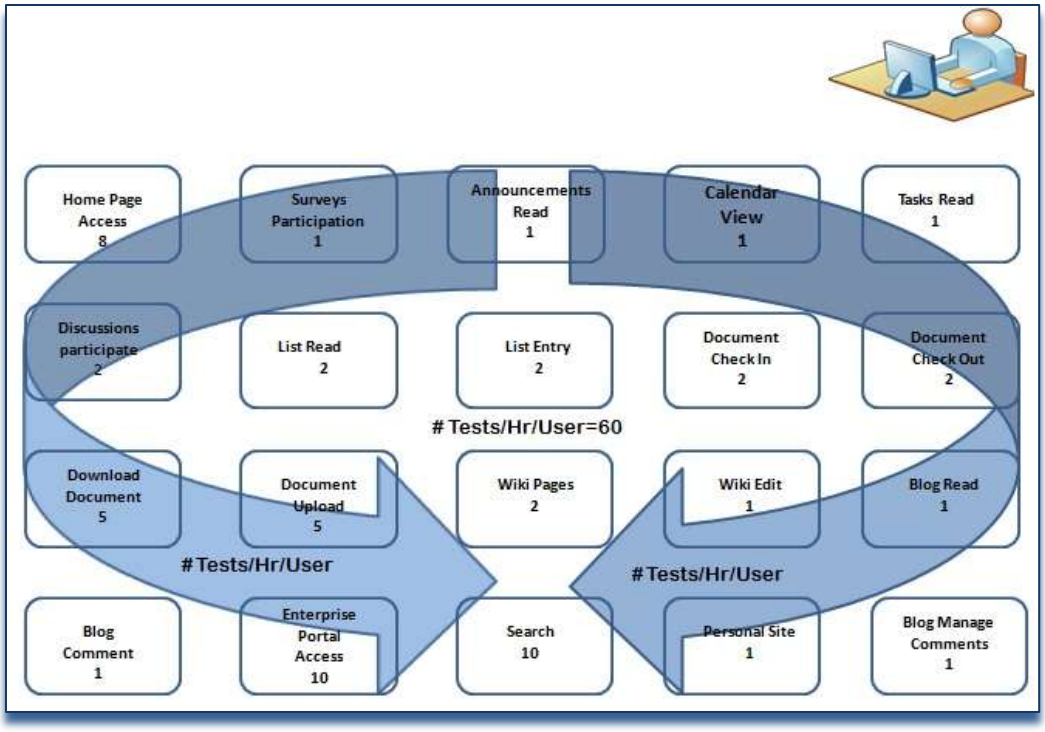
Workload Mix (60 RPH)

Requirement on the farm includes the number of users, and their usage characteristics. This performance test considers a heavy profile, were a single active user requests 60 different requests per hour to the SharePoint farm under test i.e. 60 requests/hour/ user.

User activities or usage characteristics are modeled based on an enterprise business environment needs such as organizations like marketing and engineering. In general the environment hosts central team sites, publishing portals for internal teams as well as enterprise collaboration of organizations, teams and projects. Sites created in these environments are used as communication portals host’s applications for business solutions and provides channel for general collaboration. Searching, editing and uploading documents, participate in discussions, blog post, comment on blogs, etc. are among must do sort of daily activities. Considering these activities to be part of a typical enterprise user, following set of activities where included in the workload used for the performance test.

Figure 14 displays various requests which one user shall make over a period of one hour depicting a workload generated during peak hour. These workloads were loaded onto the SharePoint farm with a brief warm up time.

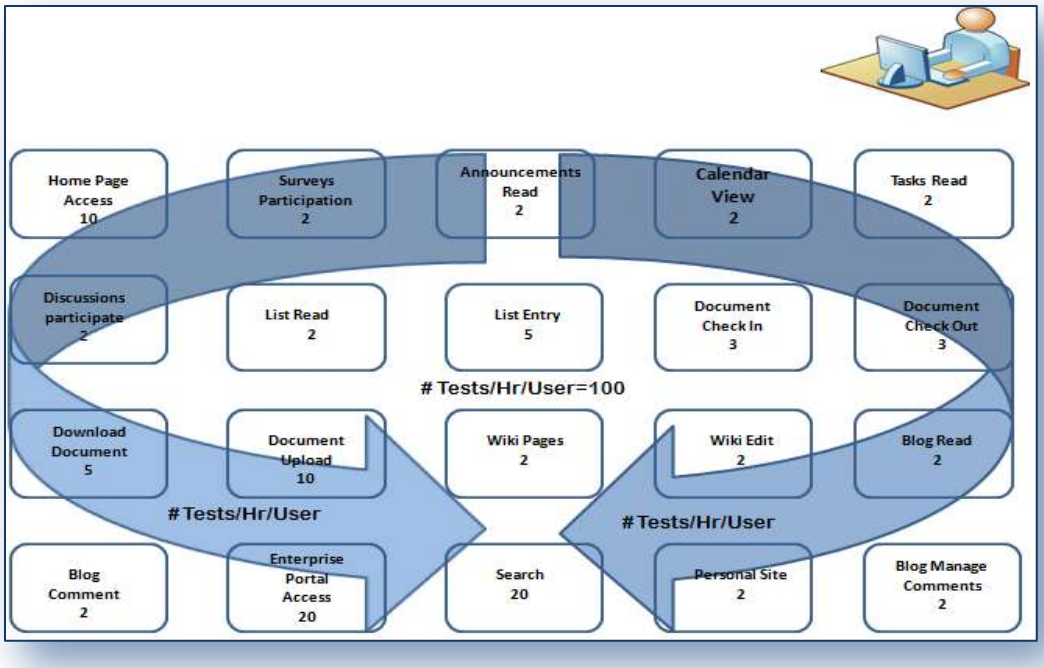
Figure 14. Requests Over One Hour



Workload Mix 2 (100 RPH)

This performance test considers an extreme heavy profile, were a single active user requests 100 different requests per hour to the SharePoint farm under test, for example 100 requests/hour/user.

Figure 15. Extreme Heavy Profile



Performance Tuning

Caching Of SharePoint 2010 Farm

SharePoint 2010 has several methods of caching data and objects to help improve the Performance for the end user. When a client requests a page, a copy of the page is stored temporarily in the output cache. Although the duration of the cache is typically small (the default is 60 seconds), this can greatly assist the performance of WFE servers and reduce latency.

Cache profiles are available so that site administrators can control the output cache behavior. Administrators can also determine whether different users, such as content editors can receive cached pages. One can adjust the output cache at the site, site collection, or Web application level.

SharePoint 2010 uses the object cache to temporarily store objects such as lists, libraries, or page layouts on the WFE server. Caching enables the WFE server to render pages more quickly, reducing the amount of data that is required from the SQL Server databases.

SharePoint 2010 has a BLOB cache that temporarily stores digital assets, such as image or media files, but you can use it with any file type. Using the BLOB cache in conjunction with the Bit Rate Throttling feature in IIS 7.0 also enables the progressive download feature for digital assets. Progressive download enables the download of media files in chunks of data, and playback can start after the client downloads the first chunk rather than the whole file. One can enable and control the size of the BLOB cache at the Web application level. The default size is 10 GB and the default setting is disabled.

In the performance test we have enabled the SharePoint 2010 cache to improve the overall response time of the farm.

Environment Configuration Tuning

The following table enumerates settings which were made on the environment to enhance its performance and capacity.

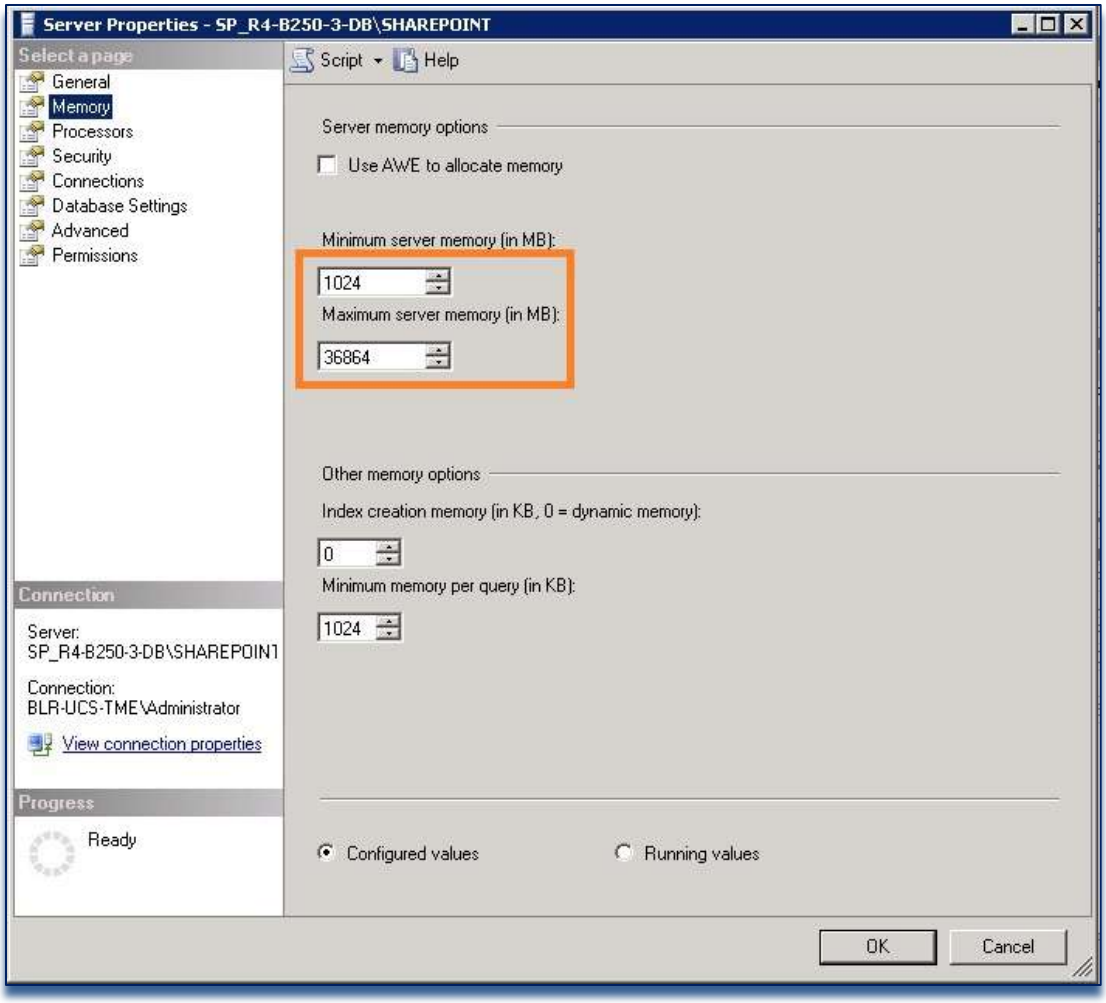
Settings	Value	Notes
Site collection: Object Caching (On | Off) Anonymous Cache Profile (select) Object Cache (Off | n MB) Cross List Query Cache Changes (Every Time | Every n seconds)	 ON Enabled ON – 10GB 60 seconds	Enabling the output cache improves server efficiency by reducing calls to the database for data that is frequently requested.
Site collection cache profile (select)	Intranet (Collaboration Site)	“Allow writers to view cached content” is checked, bypassing the ordinary behavior of not letting people with edit permissions to have their pages cached.

Settings	Value	Notes
Object Cache (Off | n MB)	ON – 1GB	The default is 100 MB. Increasing this setting enables additional data to be stored in the front-end Web server memory.
Usage Service: Trace Log – days to store log files (default: 14 days)	5 days	The default is 14 days. Lowering this setting can save disk space on the server where the log files are stored.
Database Server – Default Instance: Max degree of parallelism	1	The default is 0. To ensure optimal performance, Microsoft strongly recommend that you set *max degree of parallelism* to 1 for database servers that host SharePoint Server 2010 databases. For more information about how to set *max degree of parallelism*, see max degree of parallelism Option (http://go.microsoft.com/fwlink/?LinkId=189030).

SQL Server Memory Configuration Tuning

SQL Server is by default set up to use all the memory available on Server. You can establish upper and lower limits to the amount of memory (buffer pool) used by the SQL Server database engine with the min server memory and max server memory configuration options.

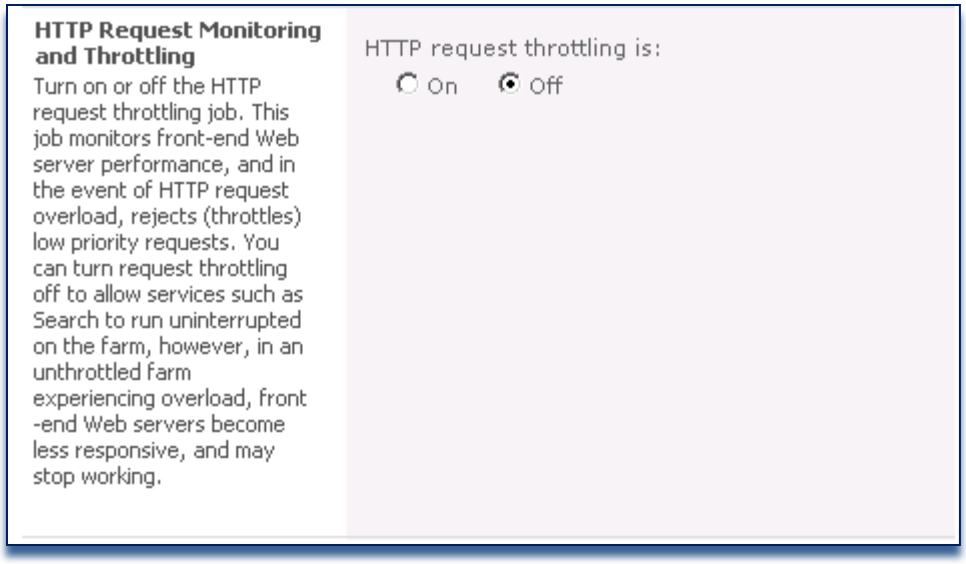
Provisioning memory 384 GB had caused significantly unused, For the purpose of the performance test SQL Server memory was limited to maximum 36 GB.



HTTP Throttling

Http throttling is a new feature in SharePoint 2010 that allows the server to discard of serving requests when it is too busy. Every five seconds a job will run that will check the server resources compared to the levels configured. By default the Server CPU, Memory, Request in Queue and Request wait time are being monitored. After three HTTP GET unsuccessful checks, the server will enter a throttling period and will remain in this state until a successful check is completed. Requests generated prior to the server entering into the throttling mode will be completed. Any new HTTP GET and Search Robot request will generate a 503 error message and will be logged in the event viewer. Also while the server is in a throttling period no new timer jobs will be started.

For the purpose of the performance test HTTP throttling was turned off.



Performance Results and Analysis

SharePoint 2010 server performance in general varies from environment to environment, depending on the complexity of the deployment and the components involved in the architecture.

Performance of the SharePoint architecture is determined by the user experience.

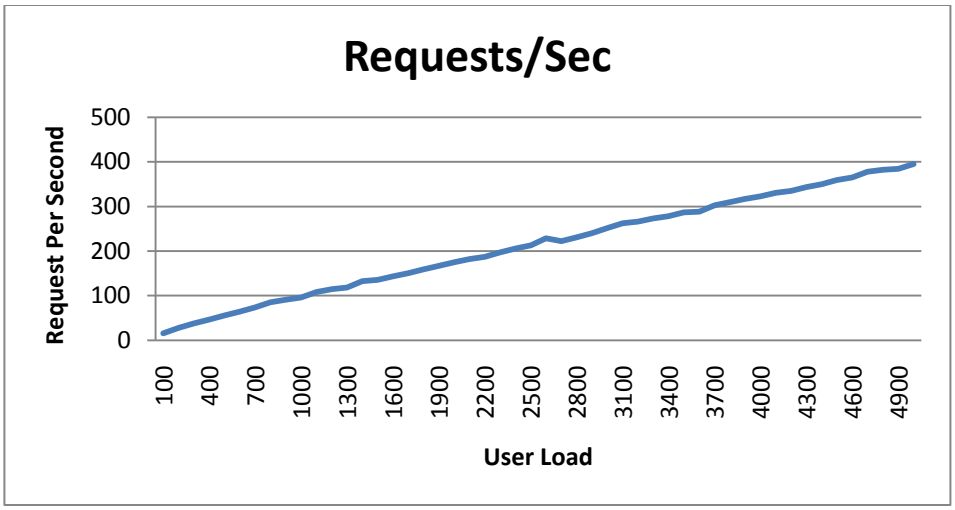
The following are the major performance counters that measure user experience.

- Requests per second: Number of requests per second taken by the SharePoint Server
- Average response time : Amount of time the SharePoint Server takes to return the results of a request to the user
- Average page response time

The following graphs illustrates the results as we applied the described workload (60 RPH) on the created SharePoint Farm.

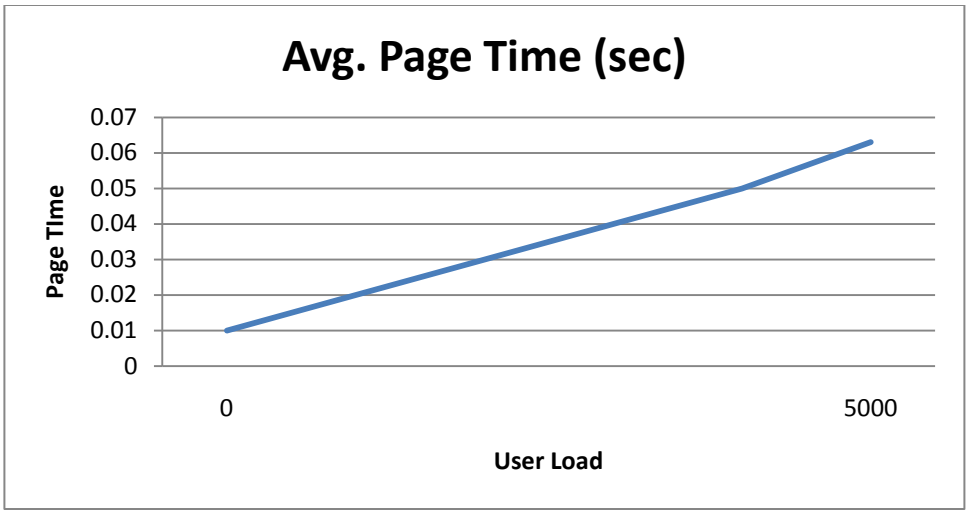
Request Per Seconds

The following graph depicts highest achieved request per second for several user loads (60 RPH). The graph attests the smooth performance of the Cisco UCS servers B200 M2 and B250 M2 as the request per second scale linearly with the user load without causing any or much stress on the server. The following graph also signifies the opportunity to further scale up the user load with an ensured stable server performance.



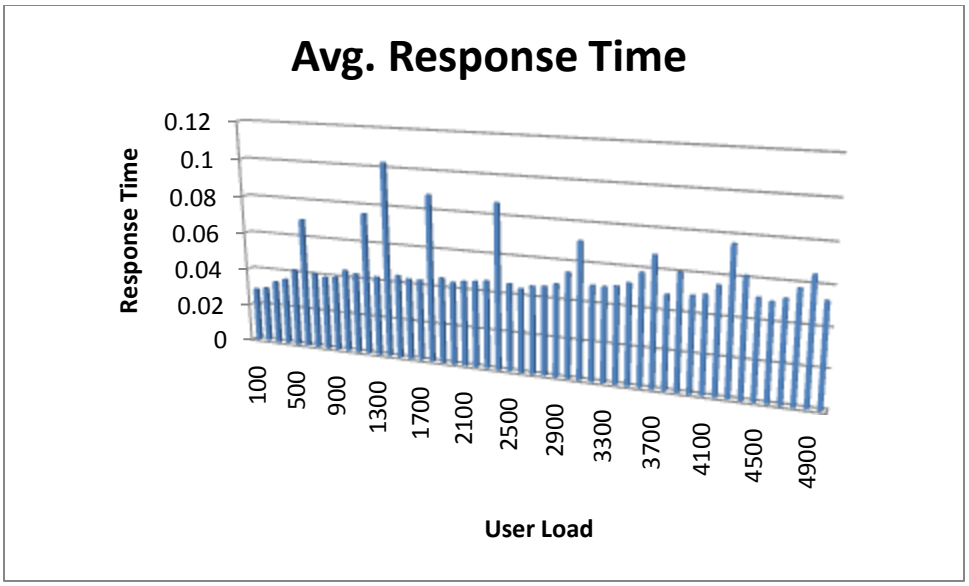
Average Page Time

The graph below depicts the SharePoint farm average page time well below 1 sec. the SharePoint achieved sub-second response time in the performance test for a concurrent user load 5000 user, so the response time varies with increased load on the SharePoint 2010.



Average Response Time

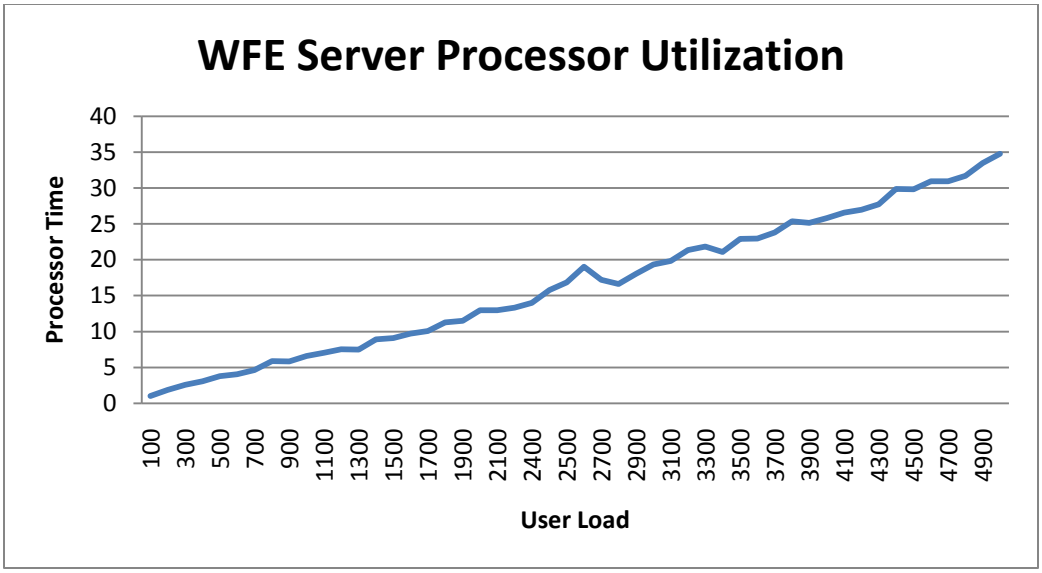
The graph below depicts average response time metrics for several user loads (60 RPH) of the SharePoint farm built using Cisco UCS B200 M2 and B250 M2. The designed SharePoint Farm can support more than 50000 users and achieve sub one second response time which remains the primary objective of this performance study. The noticeable spike in the graph are the result of the web front end cache flush. The duration of the cache is typically small (the default is 60 seconds), but this can greatly assist the performance of WFE servers and reduce latency. The graph shows the average response time to be well below One second proving the efficiency and potential of Cisco UCS servers and making it the smartest of choice.

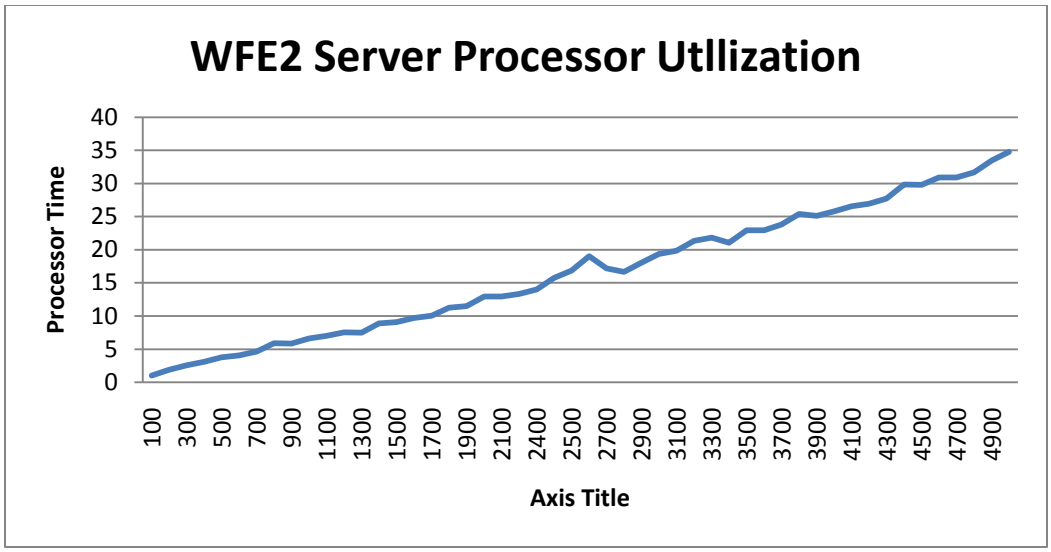


Processor Utilization

Web Front End Server

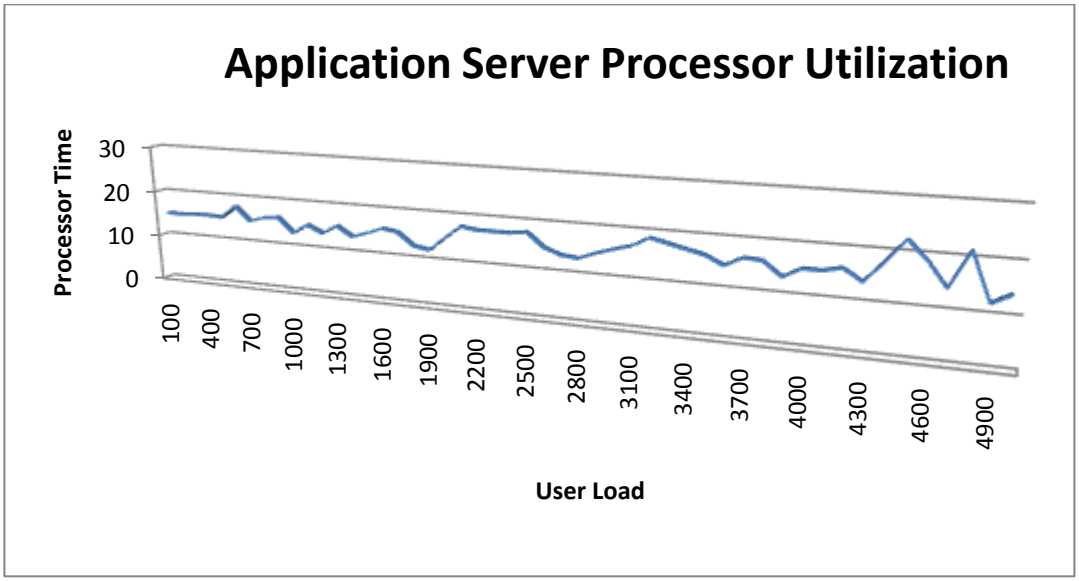
The graphs below depicts the CPU utilization of the SharePoint Web Front End servers. Under heavy workloads of 50000 users and with NLB to balance the load at the web front end tier, Cisco UCS B200 M2 servers CPU utilization remained around 35%. The graph shows the linear growth in the CPU utilization for increase in the user load and their workloads. The graph showcases Cisco UCS B200 M2 server's capacity to accommodate much more workloads with ease without causing any stress.

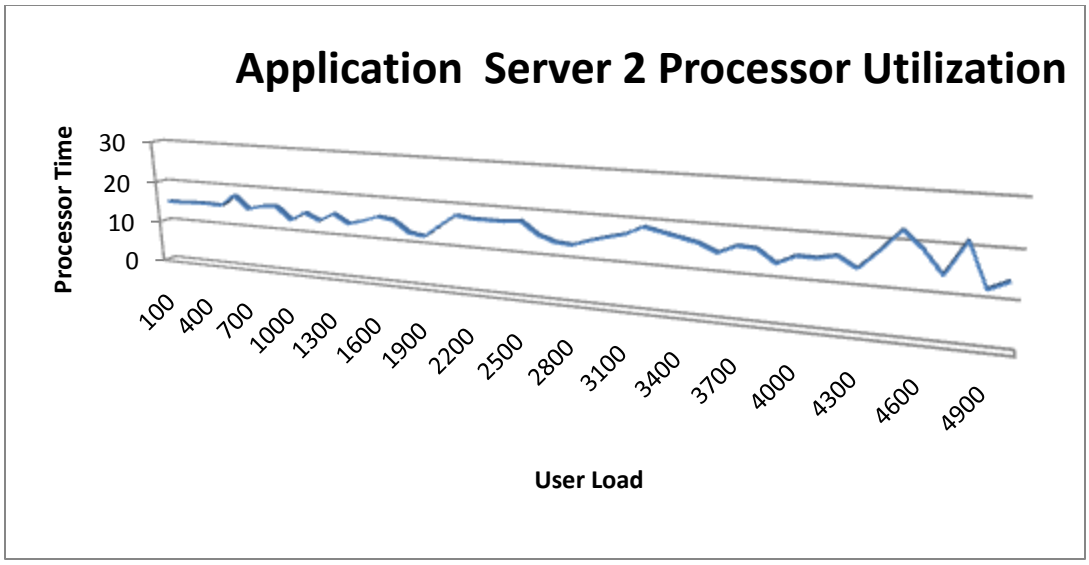




Application Server

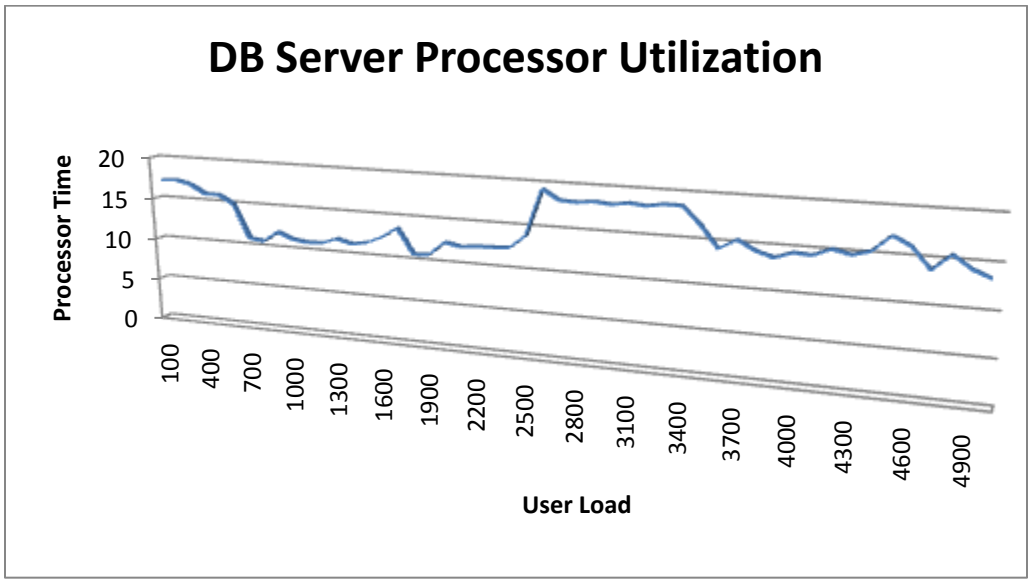
The graphs below depict the Application Server CPU Utilization. The Application server hosted the Central Administration and search services. Noticeable spikes are due to the Incremental search crawl at the time of the performance test. At Application Tier Cisco UCS B200 M2 servers remained quite stable with CPU utilization around 30%. Again providing opportunity to scale and accommodate more services.





Database Server

The graph below depicts SharePoint database server CPU utilization .There are noticeable CPU spike; as mentioned in the previous graph incremental search crawl services update the search databases; these search crawls has resulted in slight CPU spikes. On an average overall CPU utilization at the database tier with Cisco UCS B250 M2 server remained around 20 percent. Again Cisco UCS servers B250 M2 providing the provision to accommodate more without causing stress on its performance.

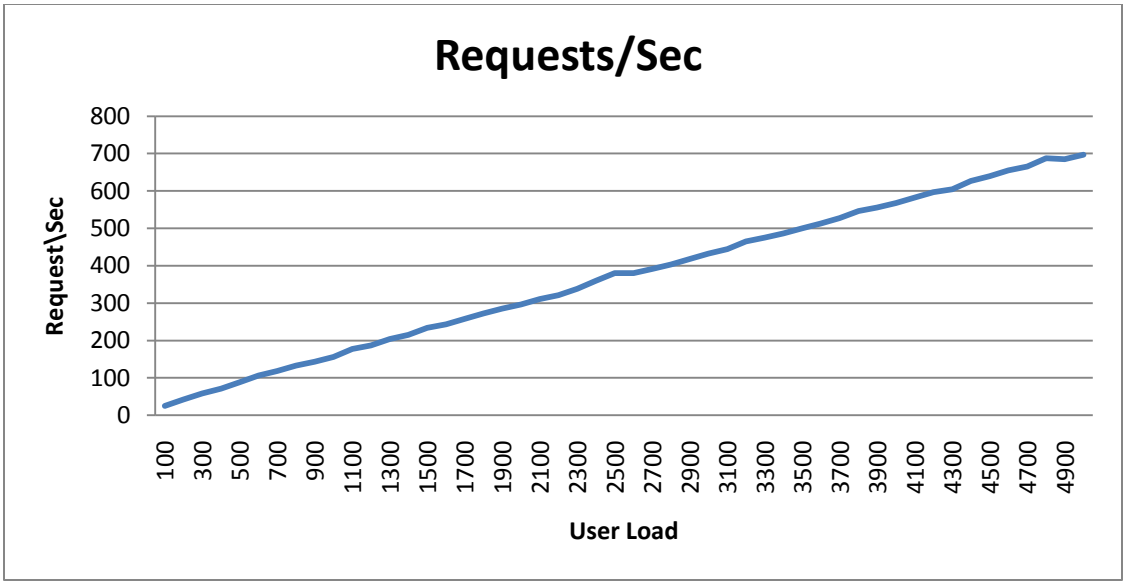


Performance Results for Workload mix 2 (100 RPH)

Request Per Seconds

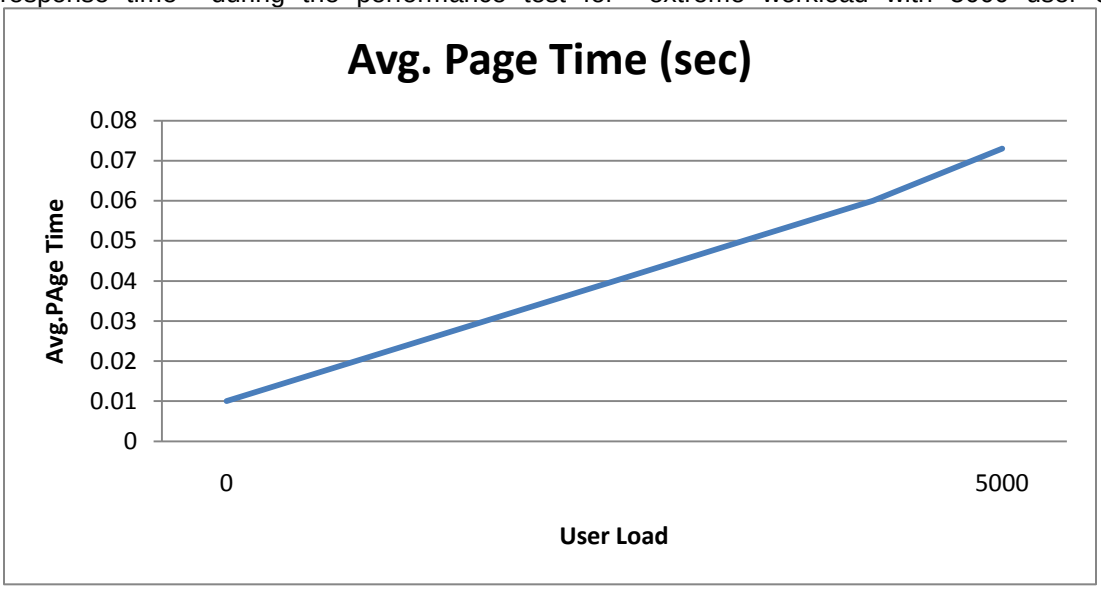
The following graph depicts highest achieved request per second for extreme workloads of (100 RPH). An extreme workload condition where one user requests 100 requests an hour and with 5000 users concurrency each requesting 100 different request in an hour. A significant increase in the request per second was obtained, due to the increase in workloads as compared to the previous workload of (60 RPH). The graph attests the smooth performance of the Cisco UCS servers B200 M2 and B250 M2 as the request per second scale linearly with the applied extreme workload without causing any or much stress on the server. The following graph also signifies the scalability provided by the powerful Cisco UCS

servers to further scale up the user load with an ensured stable server performance.



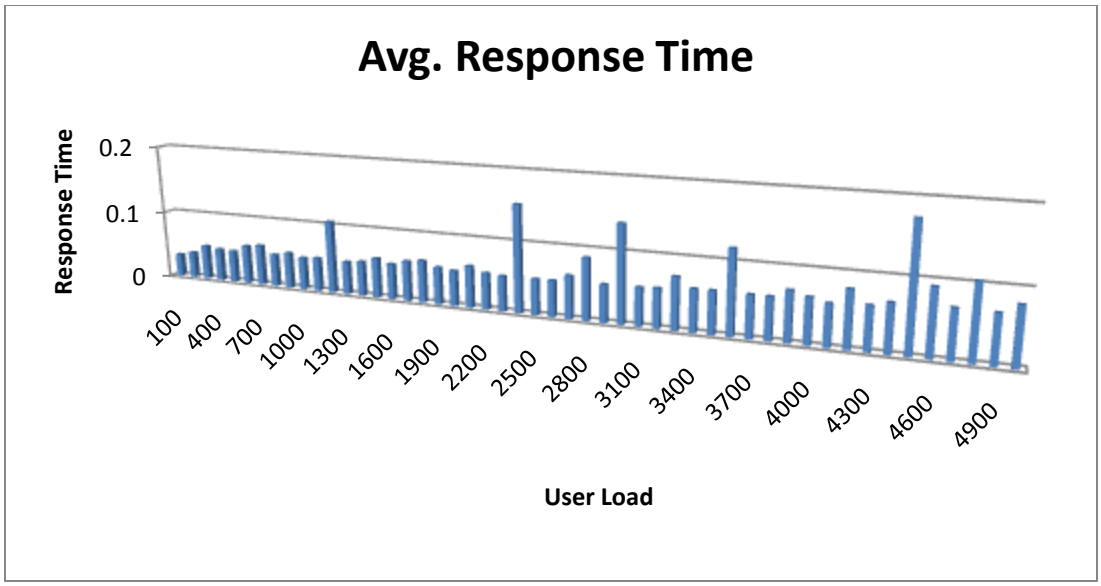
Average Page Time

From below graph you can see that the SharePoint farm average page time well below 1 sec even for an extreme workloads of 100 RPH. The SharePoint 2010 on Cisco UCS server achieved sub-second response time during the performance test for extreme workload with 5000 user concurrency.



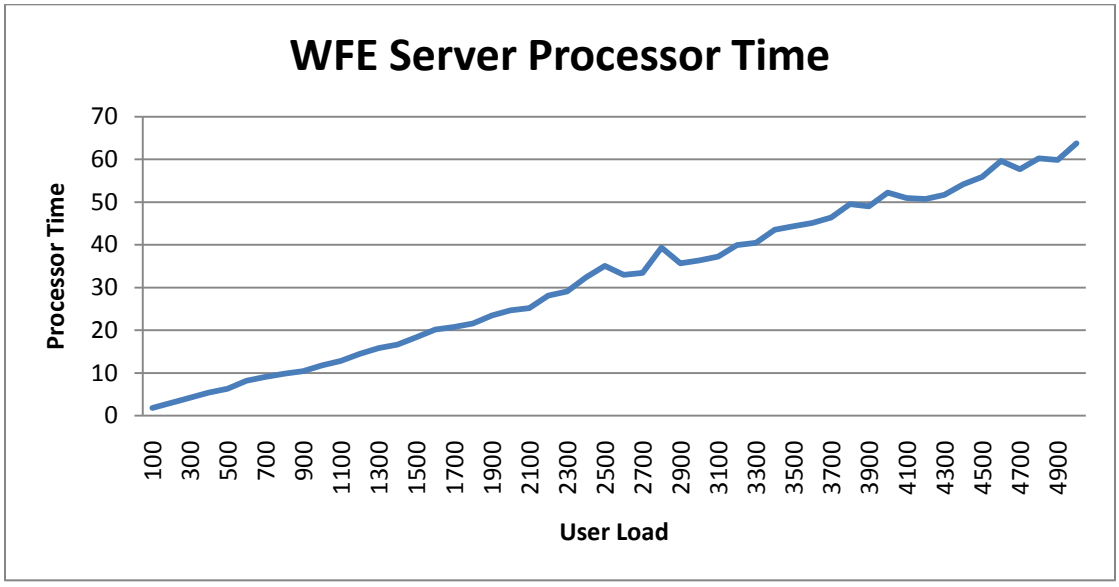
Average Response Time

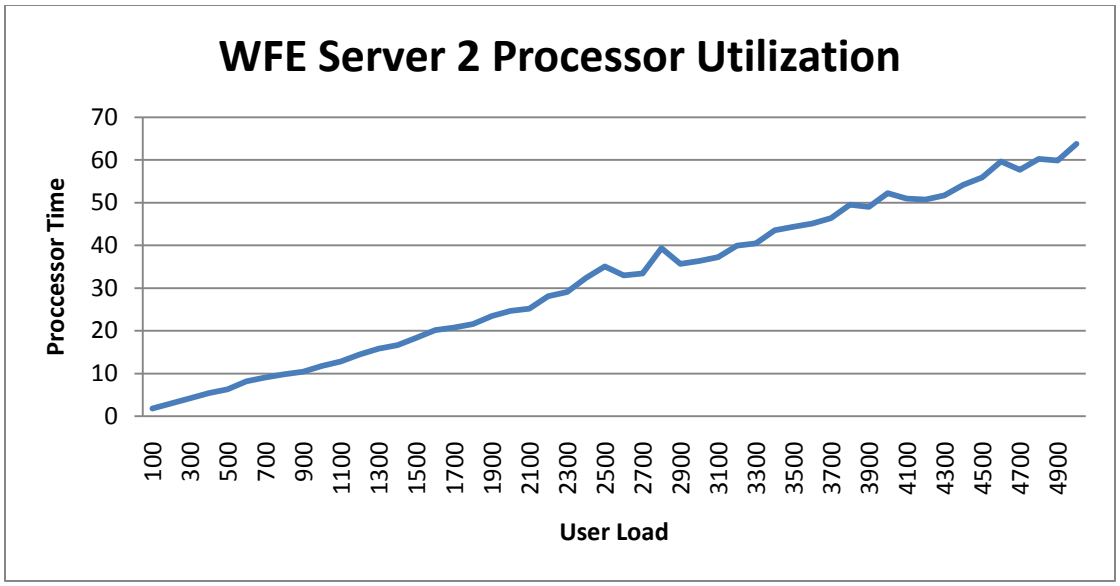
The graph below depicts yielded average response time metrics of the performance test for extreme user loads on the SharePoint farm built using Cisco UCS B200 M2 and B250 M2. The designed SharePoint Farm supported 50000 users and achieved sub one second response time even for the extreme user loads of 100 RPH; which remains the primary objective of this performance study. The noticeable spikes in the graph are the result of the web front end cache flush. The duration of the cache is typically small (the default is 60 seconds), but this can greatly assist the performance of WFE servers and reduce latency. The graph shows the average response time to be well below One second proving the efficiency and potential of Cisco UCS servers and making it the smartest of choice.



Web Front-End Server

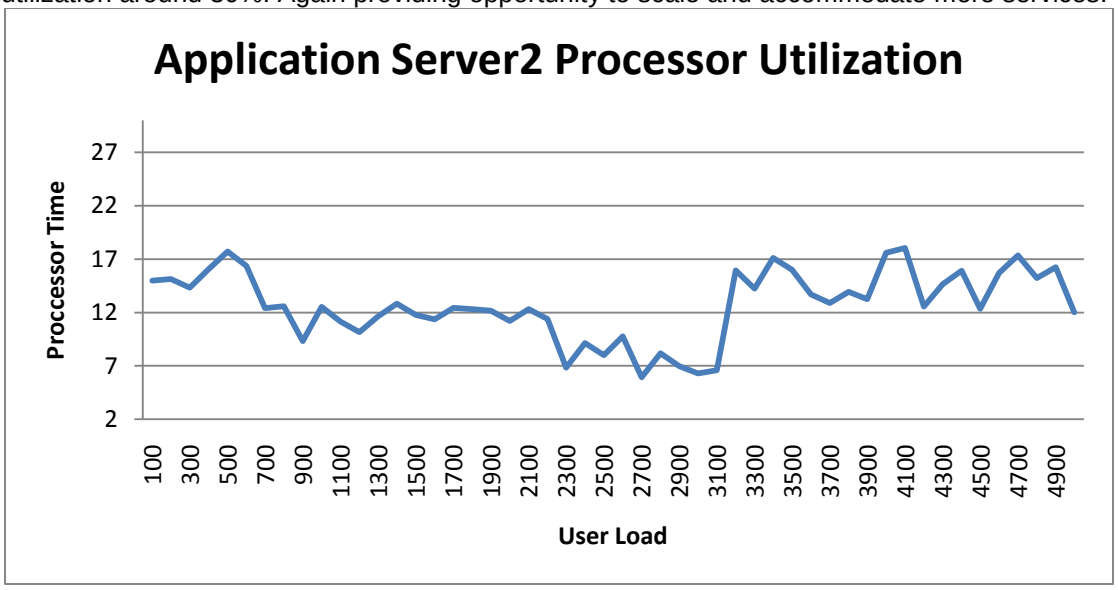
The graph depicts the CPU utilization of the SharePoint Web Front End servers. Under Extreme workloads, user loads and with NLB to balance the load at the web front end tier, Cisco UCS B200 M2 servers CPU utilization remained around 60%. The graph shows the linear growth in the CPU utilization for increase in the user load and their workloads.

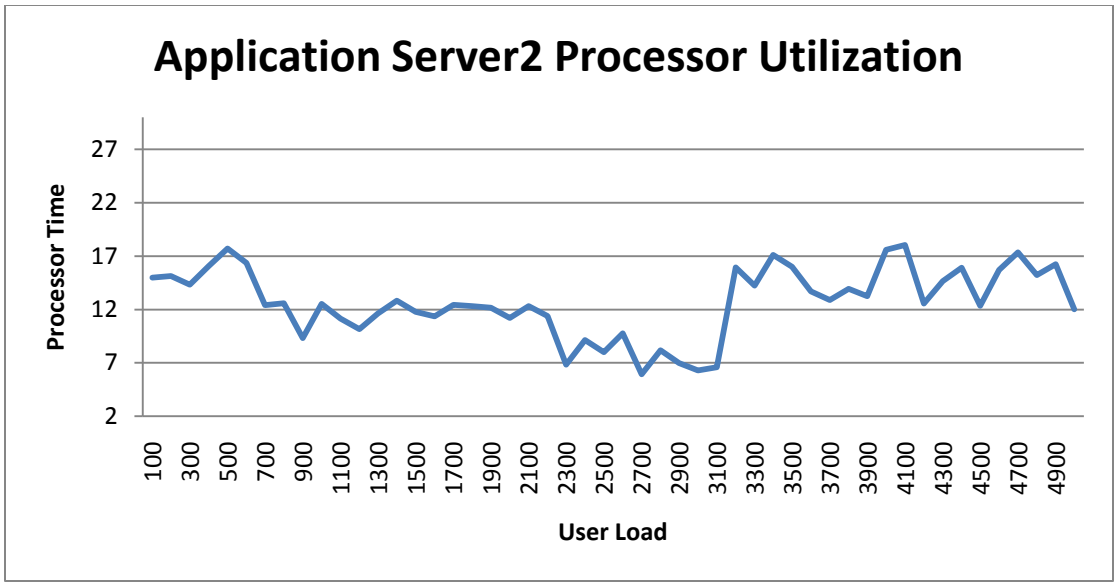




Application Server

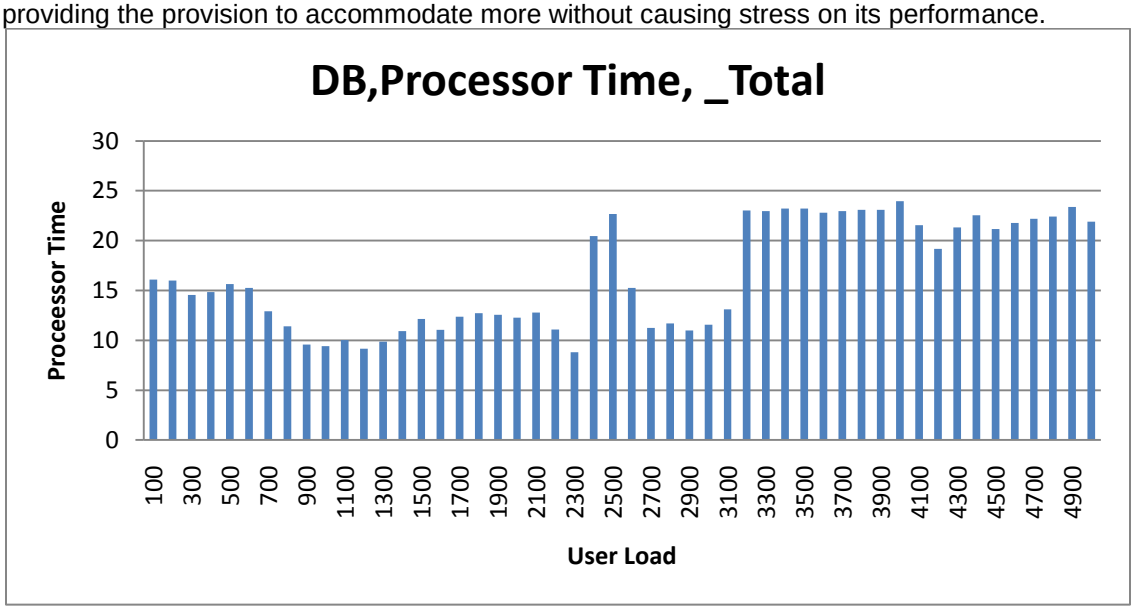
The graph depicts the Application Server CPU Utilization. The Application server hosted the Central Administration and search services. Noticeable spikes are due to the Incremental search crawl at the time of the performance test. At Application Tier Cisco UCS B200 M2 servers remained quite stable with CPU utilization around 30%. Again providing opportunity to scale and accommodate more services.





Database Server

The below graph depicts SharePoint database server CPU utilization .There are noticeable CPU spike; as mentioned in the previous graph incremental search crawl services update the search databases; these search crawls has resulted in slight CPU spikes. On an average overall CPU utilization at the database tier with Cisco UCS B250 M2 server remained around 30 percent. Again Cisco UCS servers B250 M2 providing the provision to accommodate more without causing stress on its performance.



Summary of the Performance Results and Analysis

The testing was functionally successful meeting the criteria set to achieve a 50000 user workload with approximately 10% concurrency. The following table provides the summary of the performance results and also results with respect to the most vital realistic enterprise concerns.

Important Concerns	Performance Results
Response Time (User Concern)	Less than One Second
Throughput (Business Concern)	60 Tests Per User Per Hour

Important Concerns	Performance Results
Request Per Second	400 RPS
Throughput (Business Concern)	100 Tests Per User Per Hour
Request Per Second	700 RPS
Resource Utilization (System Concern)	Under 50% CPU utilization Except Workload mix 2, where WFE server processor utilization reached 60%

This white paper provides a brief introduction to SharePoint 2010 and makes one familiar to the key terminology and concepts which aids in understanding of SharePoint reference architecture.

Three-tier architecture provisions an ideal SharePoint topology. Several servers at individual tier renders various SharePoint components; together to make up a SharePoint 2010 medium farm. Servers at web tier renders web and search query functions, servers on the application tier are responsible for search indexing and various service application functions and server at the database tier hosts SQL Server databases for the farm.

SharePoint 2010 finds its implementation to fulfill various enterprise demands. Implementation of SharePoint requires an in depth study of requirements, requirements which demands implementation. Demand Characteristics such as expected user load, concurrent users at peak time, and request per second and content size define the size of a SharePoint implementation. This performance study paper was intended to understand the performance capacity of SharePoint medium farm configurations built using Cisco UCS B200 M2 and B250 M2. Cisco UCS servers are high performing servers providing improved application performance and operational efficiency.

The performance study showed that SharePoint medium farm could easily support 50,000 users and with a minimum 10% concurrency. SharePoint farm comprising of Cisco UCS B200 M2 and B250 M2 servers and Network access of 10 Gbps connectivity between the tiers provided an average response time well below 1 second. UCS servers with their various innovative technological performance benefits (Unified Fabric, Extended Memory, etc.) essentially yielded the targeted performance results.

For More Information
For more information about the Cisco and Microsoft alliance, visit www.Cisco.com/ go/microsoftalliance or e-mail Cisco_microsoft_info@Cisco.com.

Acknowledgments

This document was made possible through the efforts of Abrar SY, with thanks to Vadiraja Bhatt.

References

Cisco Unified Computing System Architecture Overview

http://wwwin.Cisco.com/marketing/unifiedcomputing/solutions/launches/unifiedcomputing/files/cucs_solution_overview.pdf

Cisco UCS B-Series Blade Servers

http://www.Cisco.com/en/US/products/ps10280/tsd_products_support_series_home.html

Cisco UCS Manager

http://www.Cisco.com/en/US/products/ps10281/tsd_products_support_series_home.html

Cisco UCS 6100 Series Fabric Interconnects

http://www.Cisco.com/en/US/products/ps10276/tsd_products_support_series_home.html

Cisco Nexus 5000 Series Switch CLI Software Configuration Guide

http://www.Cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/configuration/guide/cli_rel_4_0_1a/CLIConfigurationGuide.html

Microsoft SharePoint 2010 on FlexPod with VMware

http://www.Cisco.com/en/US/docs/solutions/Enterprise/Data_Center/App_Networking/SharePoint_FlexPod.html

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.**

Printed in USA

Document number: UCS-TR1000xx 06/11