

Performance report for Microsoft Office Communications Server Consolidated Configuration on HP BladeSystem



Introduction	2
Test methodology	4
Test topology	5
SQL Server storage subsystem	7
Test results	8
Conclusions	13
For more information	14

Introduction

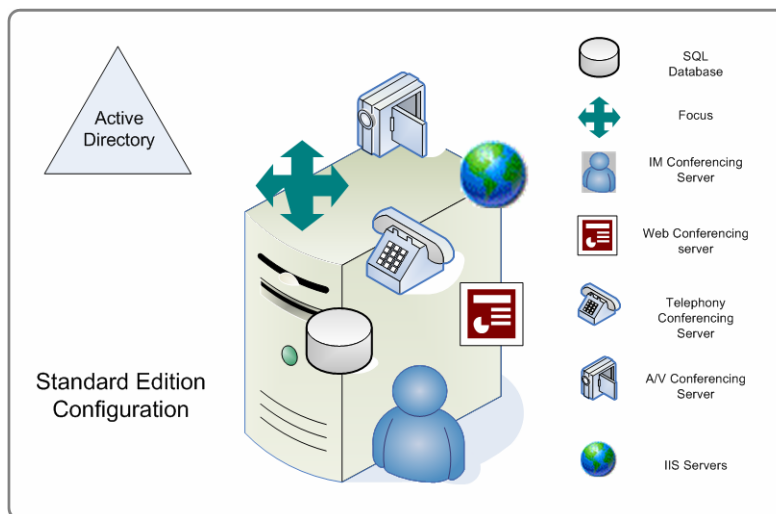
Microsoft® Office Communications Server (OCS) 2007 is an extensive update to Live Communications Server (LCS) 2005. In addition to the instant messaging and presence features of LCS, OCS provides Enterprise Voice with integrated voice over IP (VoIP) and telephony integration, audio/video conferencing, on premise web conferencing and federation enhancements.

OCS servers can host a core set of services including: Instant Messaging (IM) Conferencing Server, Web Conferencing Server, Telephony Conferencing Server, Audio/Video Conferencing server, Internet Information Services (IIS), and the Focus, which helps manage meeting participants. Database services, provided by Microsoft SQL Server, are also required.

There is also a set of OCS server roles in addition to those core services to provide incremental functionality such as access for users over the Internet, archiving and call detail reporting, telephony integration, public IM connectivity, and access through a browser based client. These server roles are typically deployed on servers separate from the core services described above.

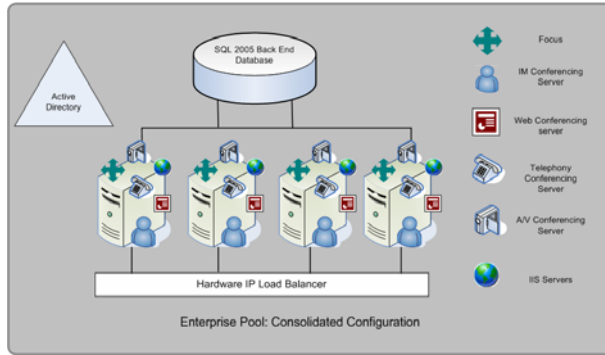
OCS can be deployed in three primary configurations. The base configuration is OCS Standard Edition (SE) which includes all core services and SQL Server 2005 Express Edition on a single server. With all functionality on a single server, Standard Edition provides no fault tolerance and is not recommended for deployments over 5000 users. Figure 1 depicts a Standard Edition deployment.

Figure 1. Microsoft Office Communications Server Standard Edition



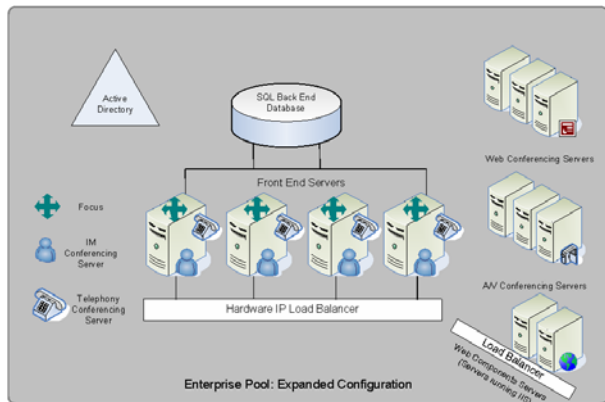
The second method of deployment is OCS Enterprise Edition (EE) in a Consolidated Configuration. In this configuration a pool of servers is created where all core services are on each of the servers and a hardware load balancer distributes traffic among the members of the pool. Microsoft SQL Server 2005 is on a separate server for back end database services. Figure 2 depicts an Enterprise Edition Consolidated Configuration. Some advantages of a Consolidated Configuration are that the SQL Server can be scaled for performance independently of other roles and can also be clustered for increased availability and maintenance. Another benefit is that the number of OCS servers can be scaled to meet each customer's needs. Multiple OCS servers behind a hardware load balancer provide redundancy for maintenance and high availability. The Consolidated Configuration is the simpler of the two Enterprise Edition configurations and is generally recommended for up to 30,000 users.

Figure 2. Microsoft Office Communications Server Enterprise Edition Consolidated Configuration



The third method of deployment is OCS Enterprise Edition in an Expanded Configuration. This configuration breaks out IIS, web conferencing and audio/video conferencing on to separate pools of servers. In this configuration those functions can be scaled separately to meet specific needs. Figure 3 depicts an Enterprise Edition Expanded Configuration. As with a Consolidated Configuration, the SQL Server is separate and can be scaled and clustered to address those needs. By breaking out the server functionality in to separate server pools, each type of functionality can be scaled independently as needed. The Enterprise Edition Expanded Configuration is recommended for up to 125,000 users.

Figure 3. Microsoft Office Communications Enterprise Edition Expanded Configuration



Figures 1, 2 and 3 are from “Microsoft Office Communications Server 2007 Technical Overview”. For more information about these topologies please visit the Microsoft Office Communications Server Technical Library located here: <http://technet.microsoft.com/en-us/library/bb676082.aspx>

Each of the topologies shown above do not include the additional servers to provide services outside of the core services such as edge servers for Internet access, mediation servers for telephony integration, archiving and call detail record services, etc.

The focus of this report is performance testing of the core services of an OCS Enterprise Edition Consolidated Configuration as shown in Figure 2 above and as deployed on an HP BladeSystem and HP StorageWorks solution. As noted above, one requirement for any solution deploying more than one server in a pool is a hardware load balancer. In this solution a F5 BIG-IP Local Traffic Manager is used.

Test methodology

One key element in testing any solution is the user profile. OCS can be utilized in many different ways, from presence to IM, audio & video conferencing to web conferencing with a mix of internal and external users, with authenticated and anonymous users. It is impossible to model a user profile to satisfy every organization's planned or actual usage.

This document uses a single profile for each of the users of a certain type, as described in the list below for the 30,000 users under test. While this will not meet every organization's needs, it provides a reference point from which to plan an OCS deployment. This profile distribution assumes that an organization of this size will not deploy all services at the same time to all users, but will take an incremental approach, where presence and IM are rolled out to most users and VoIP and Audio & Video conferencing are rolled out to fewer users.

The users are categorized as follows:

- 25,500 (85%) – Presence and IM
- 3,000 (10%) – Presence and VoIP
- 1,500 (5%) – Presence and Audio & Video (AV) conferencing

The presence profile is defined to include the number of times a user's presence status is changed per hour. In the profile under test, this is defined as 6 times per hour. The VoIP profile includes conversations per hour and duration. In the test profile, these are defined as 6 calls per hour at 3 minutes each. The IM and AV conferencing profiles include conferences per user per day and maximum length, which are defined as 48 per day at 5 minutes each for IM and 2 conferences per day at 15 minutes each for AV conferencing.

In this test environment 6 load drivers are used to simulate the user activity. Four load drivers are used for presence and IM activity, and 1 each is used for VoIP and AV conferencing. The load drivers are configured with a user per second logon rate to vary the rate at which users are logged in to the environment.

With a static profile, one of the parameters to vary is the user logon rate. This is important because user productivity and help desk resources can be negatively impacted if the OCS environment cannot handle peak logon activity. In these tests, the logon rate is varied from 1 user per second per load driver, up to 6 users per second for each. This means that the cumulative logon rate varies from 6 users per second to 36 users per second.

One important point to note is that since the VoIP and AV Conferencing load drivers have different numbers of users, they will complete the user logon process before the 4 that are running the presence and IM load. While the initial logon rate may be up to 36 users per second, that rate will drop as load generators complete the logon process at different times.

A key element in SQL Server performance is the database cache. The memory in the SQL server should be sized so that the entire database can be held in RAM which reduces response time since RAM is much faster than physical disk.

To analyze this aspect of the solution, tests were performed in 2 stages. Before the first stage, the SQL and OCS servers were rebooted to reset any cache elements. Then a test of each user logon rate was performed. At the conclusion of that test, the load generators were stopped and the database was idle, but not rebooted. For the second stage of the test, the load generators were started, requiring a new logon for each user, but the database was primed from the previous test. This methodology shows the difference between a cold and a warm database cache.

The simulation performed in this document includes the general topology as show previously in Figure 2. It should be noted that this testing does not include access from the Internet, telephony integration, or other additional services above those provided by the pool of servers under test.

As with any business critical application, OCS should be tested in a lab or proof of concept environment before being deployed throughout an organization. The purpose of this report is to summarize testing of a 30,000 user OCS Enterprise Edition Consolidated Configuration as deployed on HP BladeSystem and StorageWorks. The main areas of focus are SQL Server and OCS Server performance primarily concerning CPU, memory, and disk resources.

Important

As with any laboratory testing, the performance metrics quoted in this paper are idealized. In a production environment, these metrics may be impacted by a variety of factors.

HP recommends proof-of-concept testing in a non-production environment using the actual target deployment and user profile as a matter of best practice for all application deployments. Testing the actual target application in a test/staging environment identical to, but isolated from, the production environment is the most effective way to estimate systems behavior.

Test topology

The document “Microsoft Office Communications Server 2007 Planning Guide”, from the Microsoft Technical Library noted above, states that a Consolidated Configuration pool can support up to 30,000 users with four servers in the pool. The specifications from that guide for the OCS servers and the SQL server are noted below.

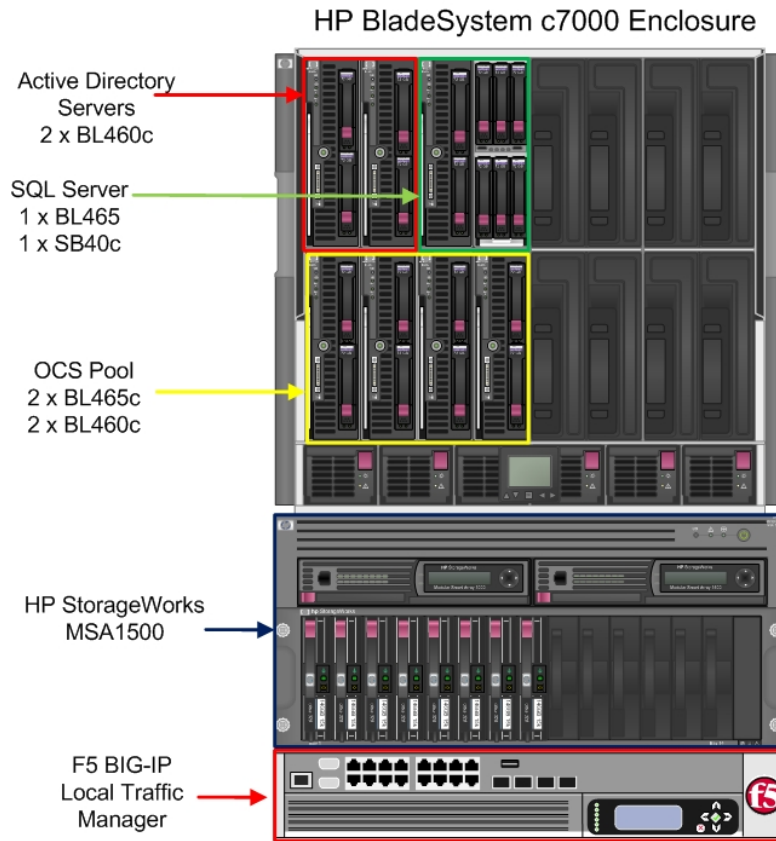
	OCS Consolidated Configuration pool member	SQL Server (up to 50,000 users)
CPU	2 x Dual-Core 2.6GHz+	2 x Dual-Core 2.6GHz+
CPU Cache	1MB L2	2MB L2
Memory	4GB	8GB
Network	1Gb	1Gb
Disks	2 x 18GB – boot/system	2 x 18GB – boot/system 1 x 36GB 15K – database log 1 x 36GB 15K – database log 8 x 36GB 15K – database files

For this test, the pool of 4 servers was divided with two servers having AMD™ processors, and two servers having Intel® processors

Figure 4 shows the architecture of the OCS EE Consolidated Configuration under test. For configuration details for the F5 BIG-IP Local Traffic Manager (LTM), please visit the F5 website at: <http://www.f5.com/pdf/deployment-guides/microsoft-ocs-ltm94-dg.pdf>

For OCS Enterprise Edition installation and configuration details please refer to “Office Communications Server 2007 Enterprise Edition Deployment Guide” at the OCS Technical Library previously mentioned.

Figure 4. OCS Enterprise Edition Consolidated Configuration architecture under test



HP ProLiant BL460c and BL465c servers were used for the OCS pool. These servers each have a total of 4 processor cores – two Dual-Core Intel Xeon® 5160 3.0GHz processors or two Dual-Core AMD Opteron™ 2220 2.8GHz processors, respectively. All four servers used 4GB of RAM and a single gigabit Ethernet interface.

The SQL Server was a BL465c also with two dual-core 2.8GHz processors and 8GB of RAM. Each server used a RAID1 mirror pair for the operating system. The details of the SQL Server storage subsystem will be described later in this document.

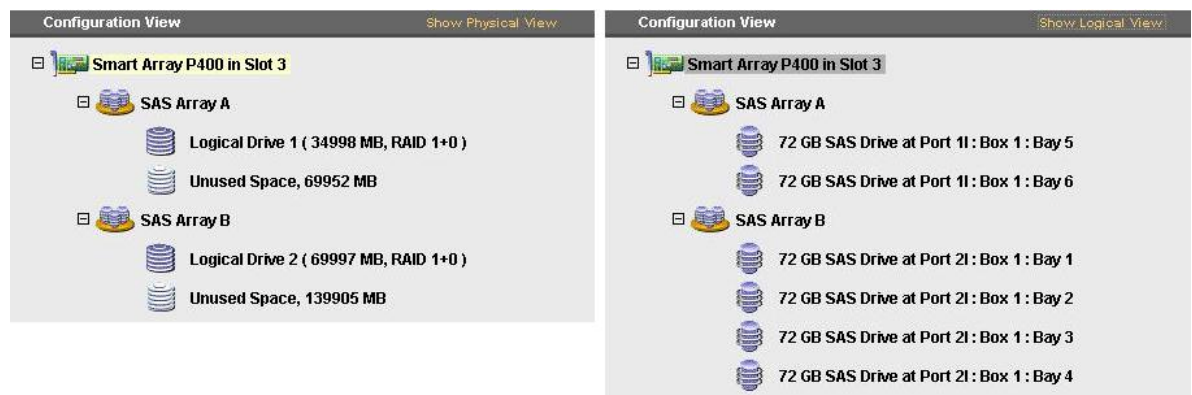
The HP BladeSystem c7000 Enclosure was fully configured with six power supplies and 10 fans. For Ethernet and Fibre Channel connectivity, two HP GbE2c Ethernet Blade Switch for c-Class BladeSystem and two Brocade 4Gb SAN Switch for HP c-Class were used.

For more information about deploying OCS on HP BladeSystem, please read this document “Designing a Microsoft Office Communication Server 2007 Architecture with HP BladeSystem” located here: <http://www.hp.com/solutions/microsoft/uc>

SQL Server storage subsystem

The SQL Server storage subsystem is critical to the overall performance of the OCS solution. In this report two different storage subsystems were tested. The BL465c SQL Server was first paired with an HP StorageWorks SB40c storage blade to provide back end database services. The SB40c uses the HP Smart Array P400 serial attached SCSI (SAS) storage controller for the six 72GB 15K RPM Small Form Factor (SFF) SAS disks in the storage blade. The P400 has 2 channels with 2 disks on one channel and the other 4 disks on the other channel. The SB40c was configured so that the 2 disks for the database logs were on one channel, and the other 4 disks supported the database files on the other channel. Figure 5 shows that configuration in the Array Configuration Utility (ACU). Note the 'Port 1I' and 'Port 2I' notation which indicates through which of the 2 internal ports those disks are controlled. Logical drives from each array are presented to Windows® as a RAID1+0 volume as shown in Figure 5.

Figure 5. SB40c Storage Blade configuration, logical and physical view



In this series of tests, firmware version 4.12 and driver version 6.8.0.64 were used for the Smart Array P400 controller. The Microsoft storport.sys driver was updated to version 5.2.3790.4021 with KB932755.

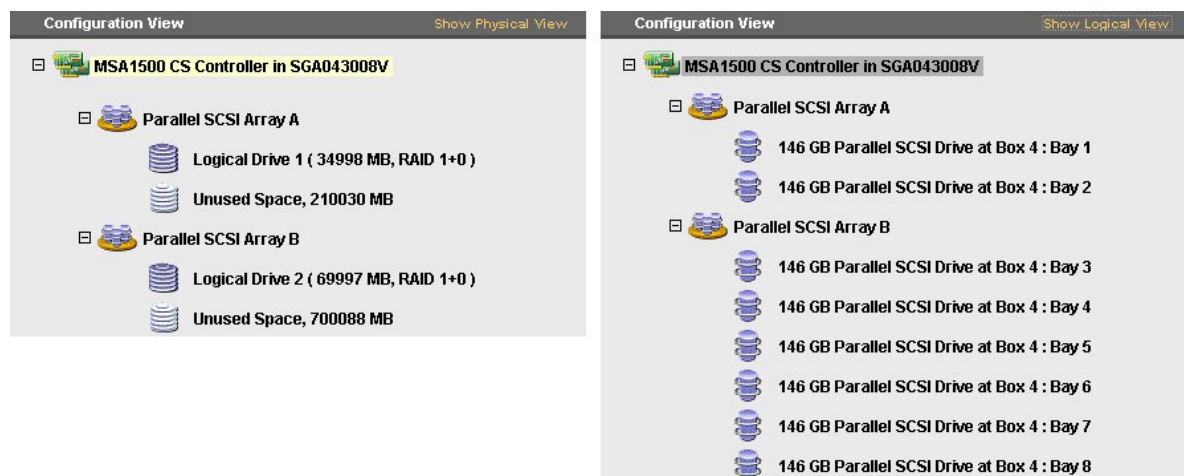
For another series of tests, the SQL Server was paired with an HP StorageWorks 1500cs Modular Smart Array (MSA1500). This Fibre Channel array offers the option of clustering the SQL Server for high availability and also offers a larger number of disk drives for increased performance. We'll compare the results of these two storage solutions later in this document.

The MSA1500 array was configured with eight 146GB 15K parallel SCSI disk with 2 disk drives for log files and 6 for database files. These disks were configured as shown in Figure 6. The MSA1500 was configured with firmware version 7.00 which supports active/active controllers. For more information on implementing active/active MSA1500 controllers, please refer to this document: "HP StorageWorks 1000/1500 Modular Smart Array application note Migrating to active/active controllers in Windows environments" at

<http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00841968/c00841968.pdf>.

The physical and logical configurations of the array are outlined in Figure 6.

Figure 6. MSA1500cs configuration, logical and physical view



The Emulex LPe1105-HP 4Gb FC HBA for HP c-Class BladeSystem was used to connect the BL465c server to the Fibre Channel SAN with driver version 7-2.00a12 and the default HBA parameters.

Test results

For each test run the performance can be viewed in two parts, the ramp up and the steady state. The ramp up is defined as the period during which the load generators are logging in users and performing the first several minutes of the user profile. Steady State is defined as the period when the system is in relative equilibrium with limited variation in performance elements such as CPU utilization, disk IO, etc.

At the OCS Team Blog site, Microsoft posted an article “How can I tell if my server is healthy in less than 10 counters?” <http://communicationsserverteam.com/archive/2007/09/10/9.aspx> This article outlines the following counters, the applicable server and the range to watch.

Performance Counter	Server Role	Range
Processor\% Processor Time	OCS Server	less than 80%
LC:USrv – 00 – DBStore\Usrv – 002 – Queue Latency (msec)	OCS Server	less than 100msec at steady state
LC:USrv – 00 – DBStore\Usrv – 004– Sproc Latency (msec)	OCS Server	less than 100msec at steady state
LC:SIP - 07 - Load Management\SIP - 000 - Average Holding Time For Incoming Messages	OCS Server	Less than 6 seconds
LC:SIP – 01 – Peers\Flow Controlled Connections counter	OCS Server	zero during steady state
LC:SIP – 04 – Responses \ SIP – 051 – Local 503 Responses/sec	OCS Server	zero during steady state
LC:SIP – 04 – Responses \SIP – 053 – Local 504 Responses/sec	OCS Server	zero during steady state
LC:SIP – 01 – Peers\SIP – 017 - Sends Outstanding	OCS Server	zero during steady state

Performance Counter	Server Role	Range
MSSQL Buffer Manager\Page life expectancy (seconds)	SQL Server	greater than 3600
Processor\% Processor Time	SQL Server	less than 80%

The blog article also mentions SQL Server disk latency as a critical factor, but it does not name it as one of the 10 counters and does not outline the counter thresholds. We'll look more at SQL Server disk performance later in this section.

During the logon process, one of the key factors that drives SQL Server load is the user logon rate, measured in users per second. In this report the following user logon rates are examined: 6, 18, 30, and 36. Keep in mind that in this test, no users are logged in at the start of the test, and all are logged in at the end of the logon period. Production environments will not likely see this pattern. In many organizations users may stay logged in and mark their presence as 'Away' or 'Busy' until they return for their next work day or they may remain logged in on their mobile device.

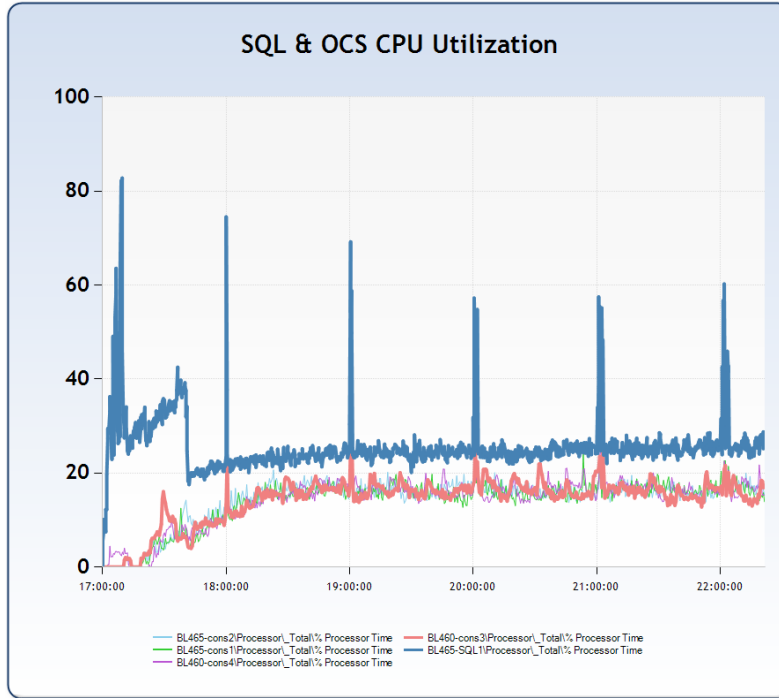
The table below provides a high level summary of the test results for the user logon rates mentioned above and for the steady state of each test run.

Logon rate users per second	SQL Server CPU% ramp up/ steady state	OCS Servers CPU% ramp up/ steady state	Storage array used	Storage cache setting Read/Write Ratio	Exceptions
6	20/30	5/20	SB40c	50/50	None
18	40/30	5/20	SB40c	50/50	None
30	55/30	7/20	SB40c	50/50	None
36	65/30	7/20	SB40c	50/50	None
36	70/30	7/20	MSA1500	50/50	None

The exception column represents any exceptions against the counters from the Microsoft Blog article. In each of these test runs there were no exceptions against those counters.

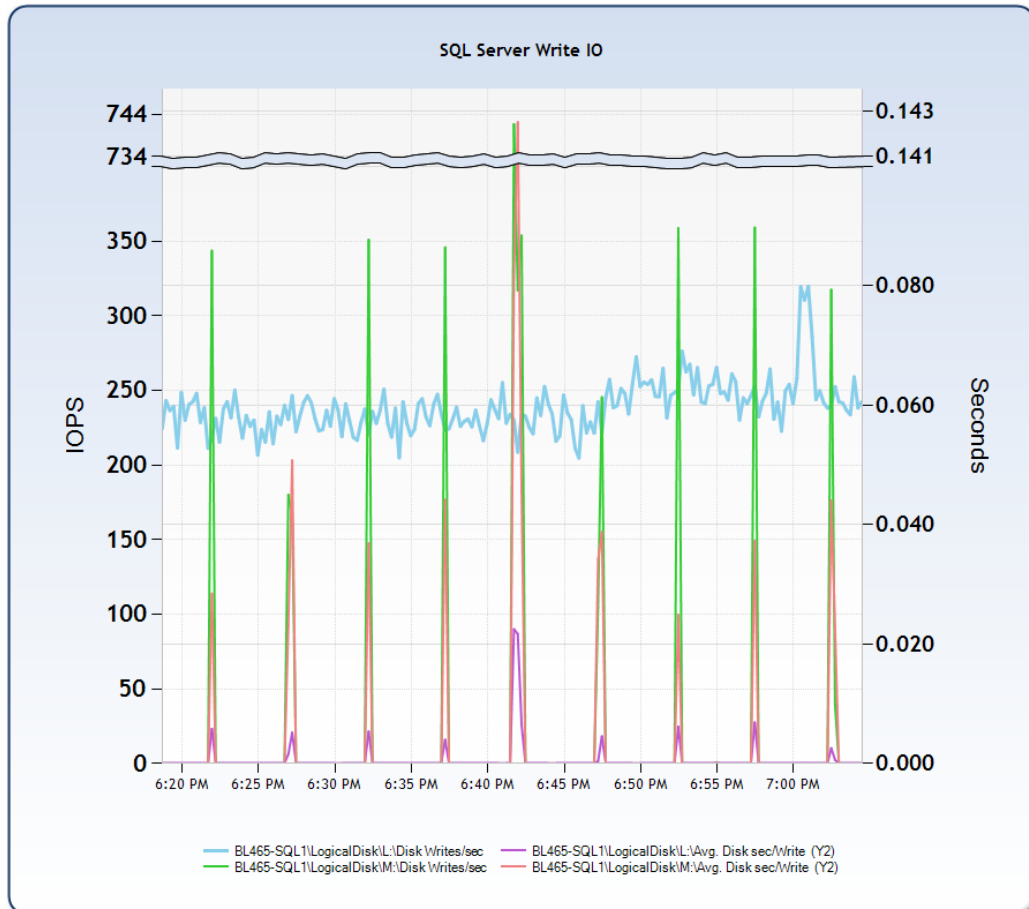
In the test runs with 18 or more users per second, there were SQL Server CPU spikes of an additional 20-40%, but these spikes were brief and did not impact any of the key performance counters on the SQL or OCS servers. An example SQL server ramp up and transition to steady state is shown in Figure 7 below. The hourly CPU spikes are caused by the load generators updating their calendar status every hour.

Figure 7. SQL and OCS server ramp up and steady state CPU utilization



Every 5 minutes the SQL Server flushes the transaction logs to the database resulting in a spike in writes/second to the database disk. If the storage subsystem cannot handle these spikes, then the queue and sproc latencies can be impacted such that communications between the OCS servers and clients can be throttled. Since there is a spike every 5 minutes, analyzing averages over a long period of time will not be useful in finding bottlenecks. Figure 8 is an example of one set of such disk IO on the SQL Server.

Figure 8. SQL Server write IO

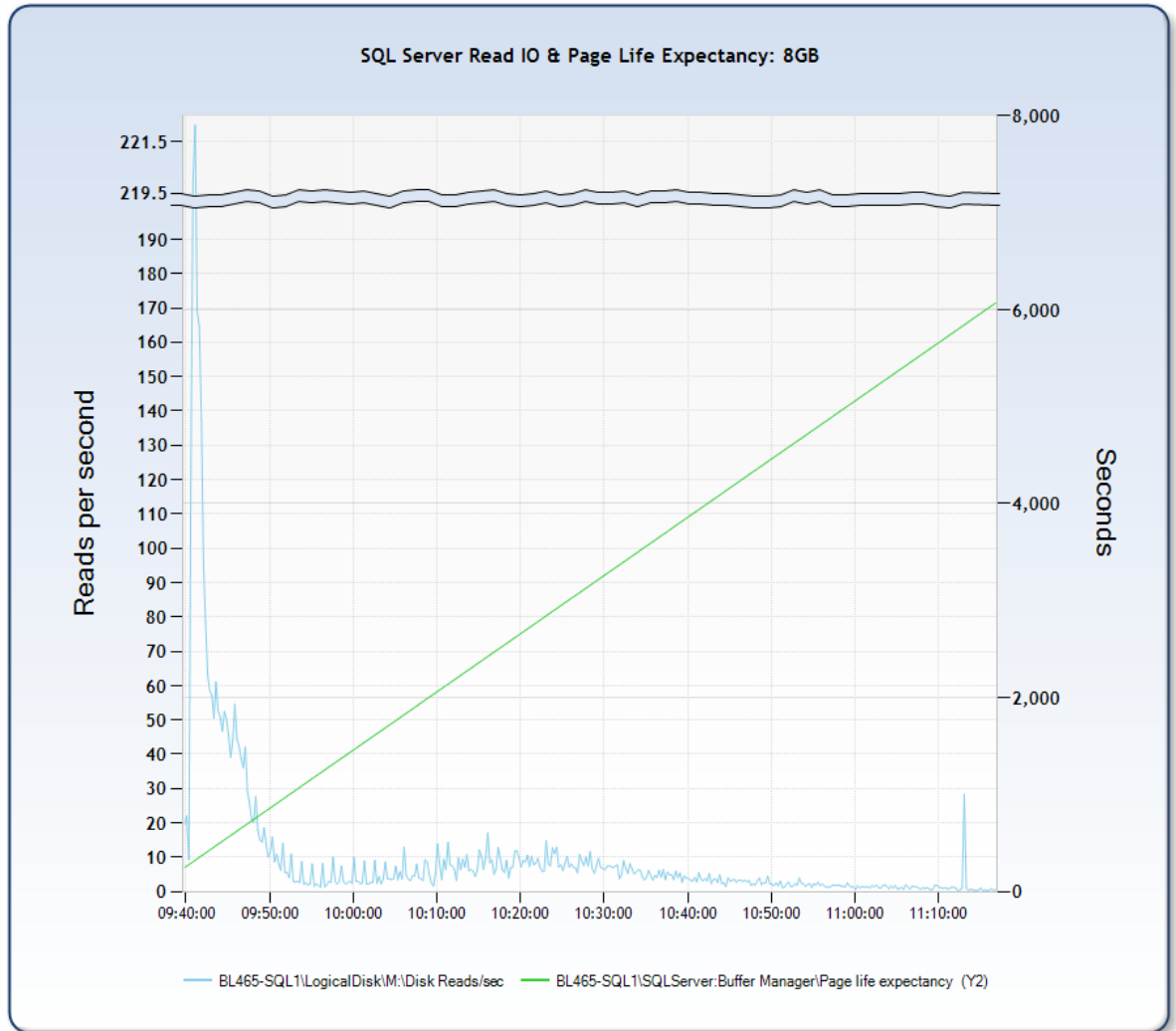


The blue horizontal line is the disk write IO to the log drive and is fairly constant during this sample and during steady state of a test run. The green line represents the disk write IO to the database disk. Each of the spikes are at 5 minutes intervals and occur when the logs are flushed to the database disk. Both the green and blue lines use the scale on the left of the graph. The red and the pink lines represent disk write latency and use the scale on the right of the graph. Log disk write latency peaks at about 21ms and database disk latency peaks at 142 ms, but the latencies are near zero for much of the time.

In looking at another of the test runs, of the 2331 data samples for database disk writes per second, only 190 of the samples had values above zero with a maximum of 1230 writes/second. Of the samples that were non zero, the average was 462, compared to an average of 32 writes/second for the entire sample set. Similarly with write latency for that sample set, the average of the entire set is 2.5ms, but the average of the non-zero samples is 30ms with a maximum of 180ms.

With properly sized memory for the SQL Server, the database should reside completely in memory during steady state which reduces the number of reads from the database disk and improves response time since memory access is much faster than disk access. Figure 9 shows this on a newly rebooted SQL server such that the database cache has not been primed, or read in to memory.

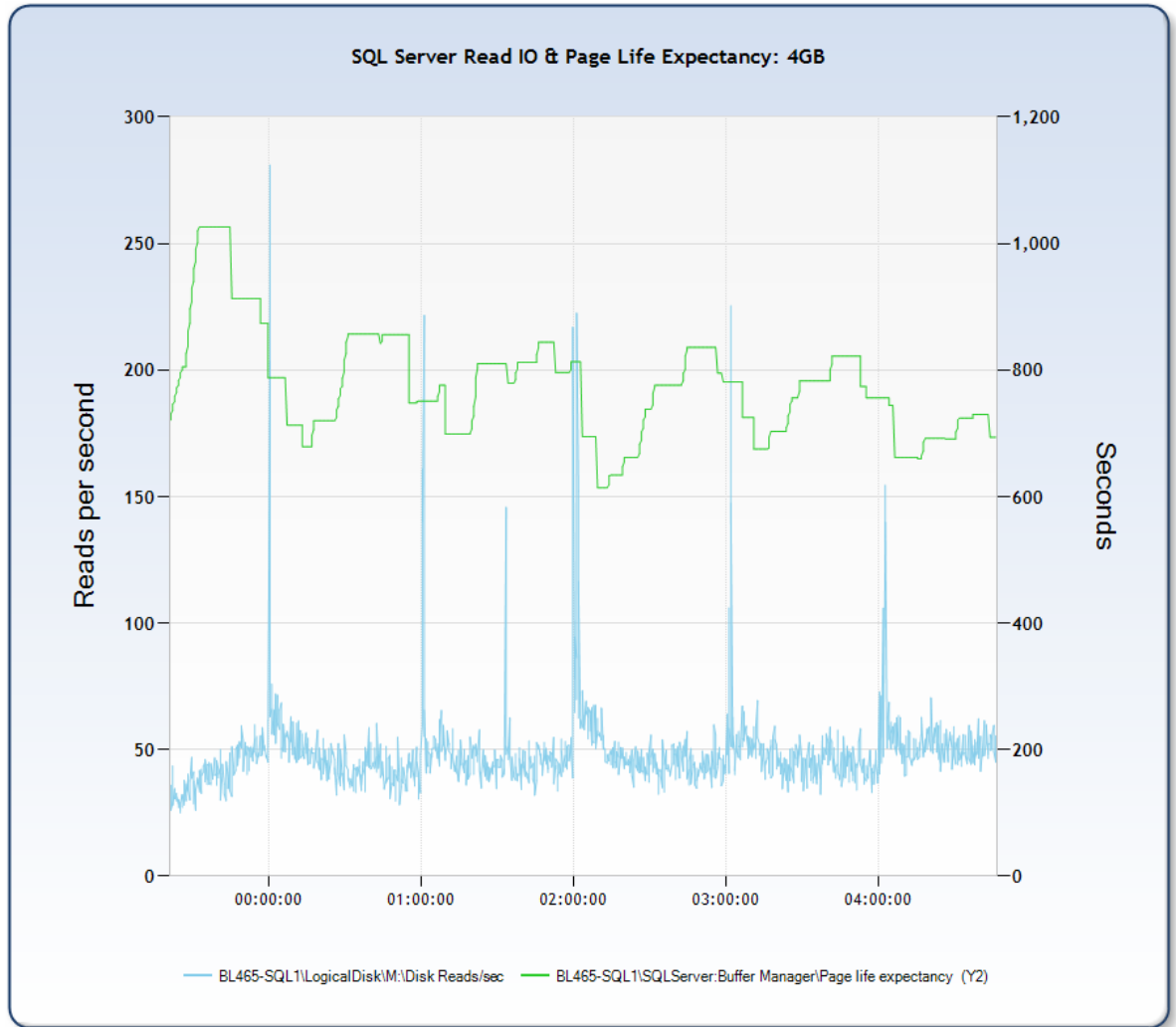
Figure 9. SQL Server database read IO with 8GB memory



This sample includes the ramp up and transition to steady state. The disk read IO peaks when the test starts, levels off a bit during ramp up, and increases in the 10:10:00 time frame as the database activity shifts from user logon to user activity. The chart also shows a steadily increasing database page life expectancy, which means that database pages are expected to be held in memory longer and reducing access time to that data.

Figure 10 shows the same configuration, but with 4GB of memory in the SQL Server. Rather than tapering off to near zero, the database reads hover near 50 per second and peak at over 250. This figure also shows that the database page life expectancy never goes over 1100, which should not be below 3600 during steady state. This is an example of a SQL Server with not enough RAM.

Figure 10. SQL Server database read IO with 4GB memory



As the solution was designed and deployed, these test results show that none of the disk, CPU or memory subsystems were stressed to the point of impacting OCS Server functionality for this user profile. Reducing the memory on the SQL server to 4GB shows stress that could lead to a poor user experience, as could reducing other system resources.

Conclusions

This testing shows that deploying an OCS Consolidated Configuration as defined on HP BladeSystem and StorageWorks for 30,000 users with this profile should provide a very positive user experience. This is a solution that businesses can deploy and grow in to. There are server and CPU resources to spare to allow for growth as users become more accustomed to the features and expand the way they use the system.

For more information

For more information on the HP BladeSystem c-Class, please visit www.hp.com/go/bladesystem

For additional information on Microsoft Office Communications Server 2007, please visit www.hp.com/solutions/microsoft/uc

For additional information on Microsoft Unified Communications, please visit www.hp.com/services/sprb/uc

For additional Exchange and HP BladeSystem information, please visit www.hp.com/go/bladesolutions/exchange

For additional Microsoft solutions information, please visit www.hp.com/solutions/activeanswers

To help us improve our documents, please provide feedback at www.hp.com/solutions/feedback

© 2008 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. Intel and Xeon are trademarks of Intel Corporation in the U.S. and other countries. AMD and AMD Opteron are trademarks of Advanced Micro Devices, Inc.

4AA1-7558ENW, January 2008

