HP Smart Array and HP StorageWorks 60 Modular Smart Array (MSA60) performance brief for Microsoft Exchange Server 2010

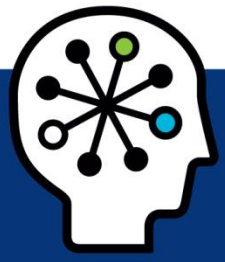


Table of contents

Executive summary

Microsoft® Exchange Server 2010 (Exchange 2010) introduces many new design features and thus many new configuration options to consider. Most importantly for this paper, Exchange 2010 completely re-architects the database engine for better performance. This paper will look at how the new database engine performs and how HP Smart Array and HP StorageWorks products can be configured to use different classes and types of disk storage.

The performance of an Exchange Server 2010 solution depends on the implementation of a suitable storage solution. The storage subsystem must not only provide suitable disk space to store mailbox content but also be able to handle the transactional database performance requirements, measured as input/output operations per second (IOPS). If not properly configured, storage subsystems are commonly a bottleneck or support issue for Exchange solution deployments.

This paper describes testing performed by HP Alliances, Performance and Solutions (APS) engineering to determine best practices for the Exchange 2010 storage configuration with HP Smart Array controllers and the HP StorageWorks 60 Modular Smart Array (MSA60). This paper describes the workloads used, the results and analysis, the resultant strategy and best practice recommendations to deploy this type of storage for the new Exchange application.

Target audience: This paper is intended for people who will be proposing solutions, providing installation services or consulting, or who may be assisting in deploying Microsoft Exchange Server 2010 on HP ProLiant systems and HP StorageWorks storage technologies. It will also be of interest to IT professionals who may be deploying and/or managing Exchange Server 2010 solutions. The paper focuses primarily on guidelines developed during testing performed in the Alliances, Performance and Solutions (APS) Engineering labs regarding deployment and performance on HP ProLiant servers and HP StorageWorks technologies.

This white paper describes testing performed in October 2009.

Exchange Server 2010

The new version of Microsoft Exchange Server 2010 includes a new high availability (HA) model that can use multiple copies of databases. Exchange 2010 introduces the Database Availability Group (DAG) concept or feature. The DAG model supports switchovers or failover (unplanned) at the database level, which is different than Exchange 2007 CCR (Cluster Continuous Replication) that required the entire server to failover.

With this type of design and deployment, if we have 3 or more copies we can consider the copies as sufficient to protect our data, thus we may not need RAID protection offered by the controller. We can then consider using RAID-less "Just a Bunch Of Disks" (JBOD) storage, which means that all the disks in this type of configuration are being independently addressed and have no collective properties or parity protection as RAID would offer. The controller treats each disk separately and each physical disk maps to a logical disk. This configuration does not provide any data redundancy at the controller level.

Note
In Smart Array terminology JBOD is referred to as RAID0 in the HP Array Configuration Utility.

If only 2 copies of databases are planned to be configured then RAID storage configuration should be used.

To summarize:

- Standard RAID configuration – Two (or more) database copies
- RAID-less (JBOD) configuration – Only if three or more database copies

Exchange Server 2010 database changes

This new version of Exchange Server 2010 has significantly improved database disk I/O performance. The key improvements in the Extensible Storage Engine (ESE) database include more sequential database access, and moving from smaller data I/O patterns in Exchange Server 2007 to larger less frequent data I/O transfer in Exchange Server 2010. The page size of databases has been increased to improve the database cache read performance with more sequential database writes. The new database I/O changes in Exchange Server 2010 were designed to allow new storage configurations that were not feasible in Exchange Server 2007 such as larger Midline SATA or SAS drives that do not have the high performance or reliability of Enterprise SAS drives. These new storage configurations are targeted to reduce the overall storage cost without affecting the appropriate availability and service level options.

Another consideration for Exchange 2010 is whether or not to honor database and log isolation which ensures that transaction logs and databases from the same database reside on different physical spindles. This is no longer a requirement for Exchange data protection if high availability using database availability groups (DAG) are deployed. The performance consideration is that all disk activity, including transaction logs writes and database reads and writes impact the same disk.

Jetstress 2010

In order to simulate the new Exchange Server 2010 database I/O patterns and performance, the Microsoft Jetstress 2010 utility was used in this testing. This tool performs input/output (I/O) operations against storage subsystems to simulate Exchange Server 2010 and can determine the maximum I/O operations per second (IOPS) performance within a defined disk latency threshold. Successful performance criteria are measured in terms of highest IOPS to the database drive(s) with sub-20ms read latencies. A sub-20ms read and 10ms write latency threshold is considered important for Exchange 2010 database disk latency, because above this threshold end users may experience delays due to slower responding disks.

HP Smart Array and StorageWorks

The following HP Smart Array and StorageWorks products were used in this testing.

HP Smart Array P411 Controller

The HP Smart Array P411 controller is a 2.0 serial attached SCSI (SAS) PCI-Express (PCIe) RAID controller. The P411 controller is capable of supporting up to 100 SAS or Serial ATA (SATA) hard drives. The P411 controller supports all the RAID levels such as RAID10, 5. It is available with a 256 MB or 512MB (upgrade option) cache module, and battery-backed write cache (BBWC). This controller is capable of supporting up to 64 logical drives. See http://h18000.www1.hp.com/products/servers/proliantstorage/arraycontrollers/smartarrayp411/index.html for more information.

HP Smart Array P800 Controller

The HP Smart Array P800 controller is a 16-port serial attached SCSI (SAS) PCI-Express (PCIe) RAID controller. The P800 controller is capable of supporting up to 108 SAS or Serial ATA (SATA) hard drives. The P800 supports 512MB of battery-backed write cache (BBWC) with two battery packs that can store data for up to four days during a power outage. This controller is capable of supporting up to 32 logical drives. The P800 adds dual domain for additional fault tolerance and resiliency. Addition of second I/O Module adds Dual Domain capability providing higher level of redundancy and reliability. See http://h18000.www1.hp.com/products/servers/proliantstorage/arraycontrollers/smartarrayp800/index.html for more information.

HP StorageWorks 60 Modular Smart Array (MSA60)

The HP StorageWorks 60 Modular Smart Array (MSA60) is a 2U Serial Attach SCSI (SAS) disk drive storage enclosure supporting 3.5 inch SAS, Serial ATA (SATA) and SAS/SATA Midline (MDL) drives. The MSA60 supports the cascading of shelves in a 1+3 configuration to allow a maximum of 48 drives in an 8U configuration behind each of the two external ports on the HP Smart Array P800, the HP Smart Array E500 or the HP Smart Array P411 Controllers for a total of 96 drives in 8 enclosures (this single controller port incorporates four lanes for a total max throughput of 12Gb/s for SAS). On HP ProLiant servers Dual Domain support is available for the HP Smart Array P800 attached to a MSA60 with the addition of the HP StorageWorks Dual Domain I/O Module Option. See http://h18006.www1.hp.com/storage/disk_storage/msa_diskarrays/drive_enclosures/msa60/index.html for more information.

SATA and SAS drives

Midline (MDL) SATA and MDL SAS drives were used in this testing in addition to Enterprise SAS drives. Note that MDL SATA and MDL SAS drives do not have the high performance or reliability of Enterprise SAS drives and were not recommended for Exchange databases in the past. See http://h18000.www1.hp.com/products/servers/proliantstorage/serial/sata/entry/index.html for more information.

Test configuration

The following hardware and software configuration was used for performance testing.

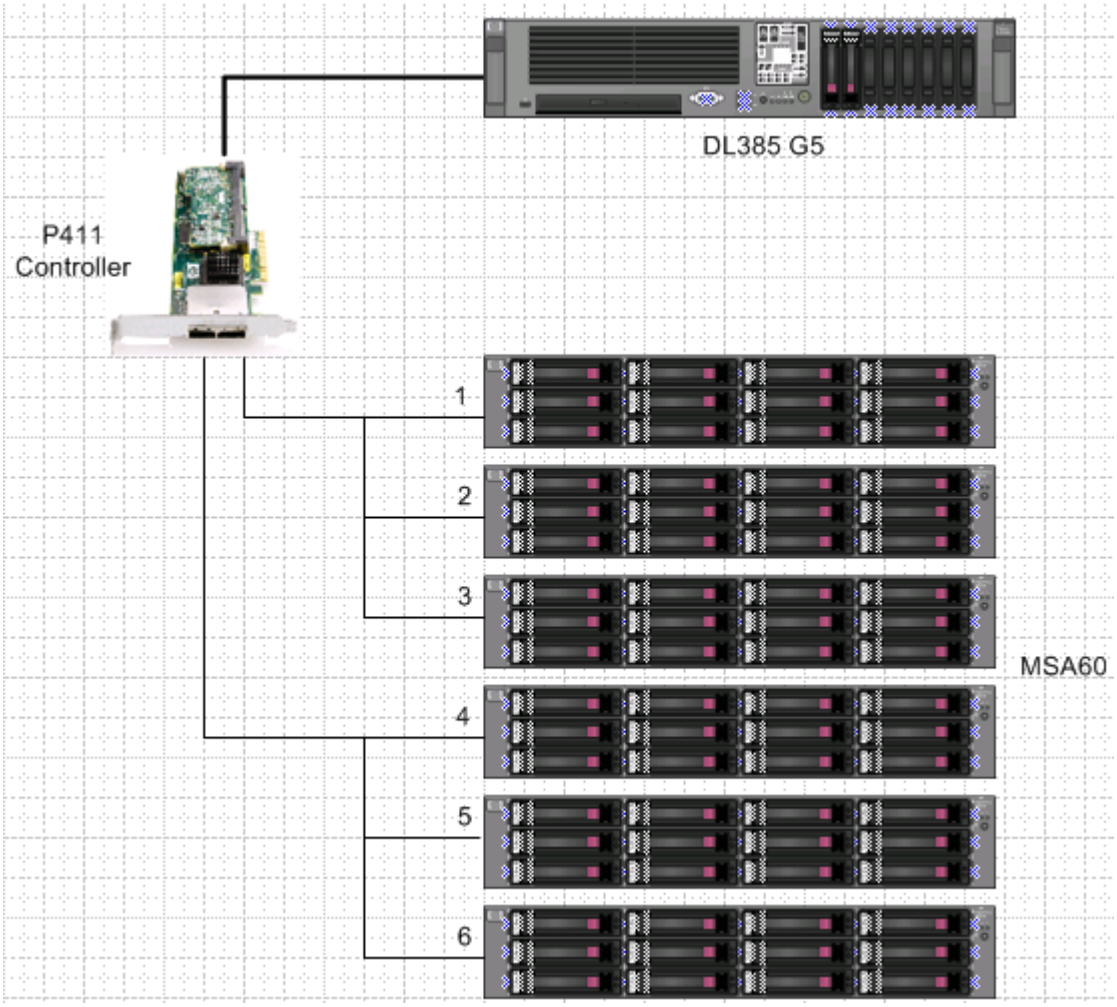
Server hardware and software

- Server: HP ProLiant DL385 G5, 2 x2.3 GHz CPU, Memory: 16GB
- Microsoft Windows® Server 2008 R2 Enterprise (x64 Edition)
- HP Smart Array P411 controller (Firmware Version: 2.5)
- HP Smart Array P800 controller (Firmware Version: 6.86)
- HP StorageWorks MSA60 drive enclosures with up to 12 disks per enclosure, with a total of six enclosures available for testing
- A range of drive types from 300GB 15K Enterprise SAS to 500 GB and 1 TB MDL SATA and SAS drives.

Storage design

The storage design was built with six MSA60 enclosures, for a total of 72 disks connected to a single P411 controller as shown in Figure 1.

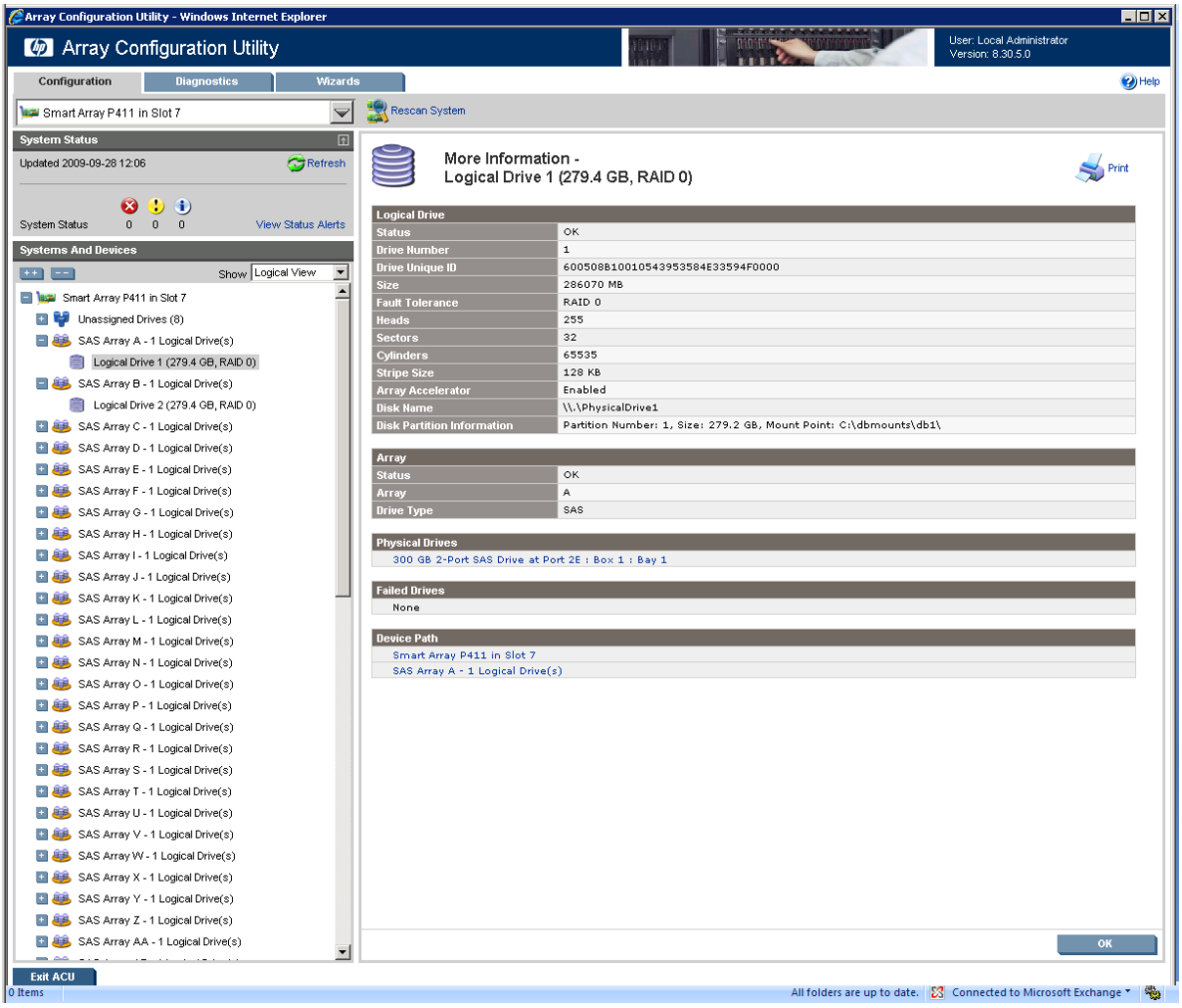
Figure 1: Architecture diagram



Array design

Figure 2 shows an example of the HP Array Configuration Utility (ACU) (version 8.30.5.0) as used to create the arrays and logical drives necessary for each test.

Figure 2. Array Configuration Utility (ACU)



RAID protection levels

RAID provides different levels of data protection using technologies such as mirroring and striping. The considerations for choosing one over the other include balancing the capacity and performance needs of the application data placed on disk.

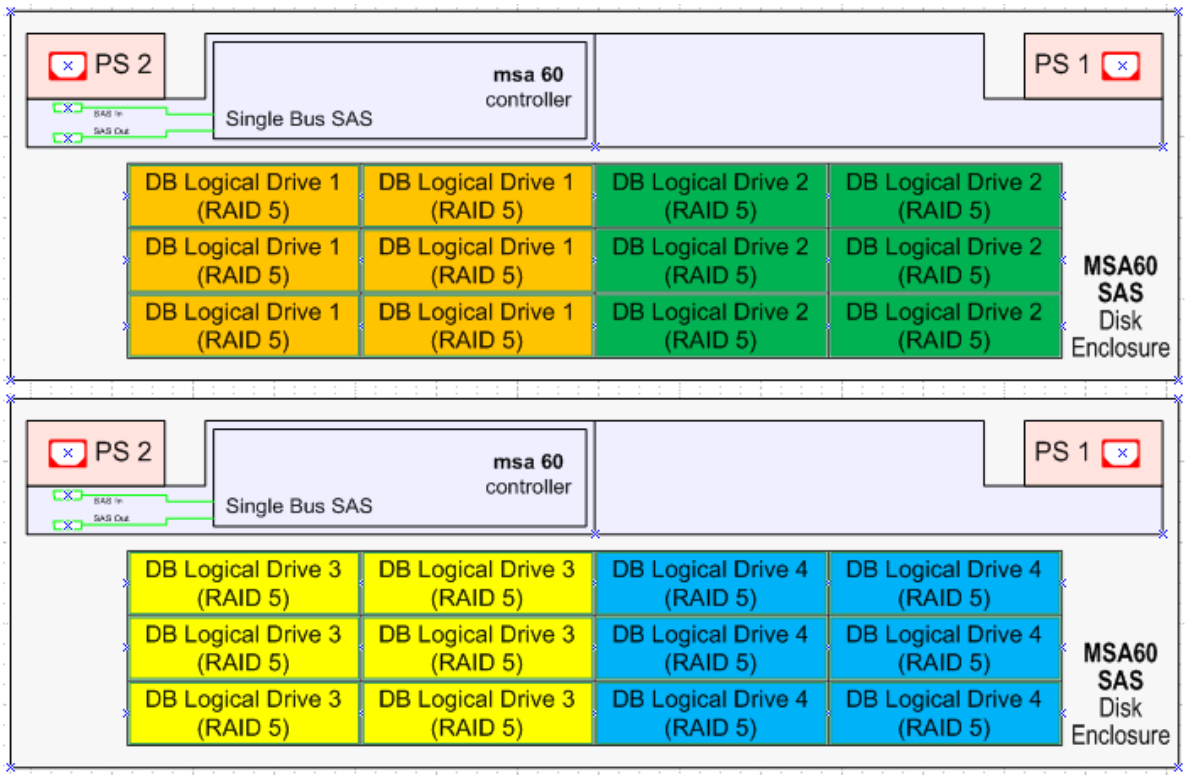
RAID10

RAID10 uses striping of mirrored data for the best performance and data protection but at a high cost of capacity by using data mirroring.

RAID5

RAID5 provides additional storage capacity with the same number of disks, but it also incurs additional write overhead, lowering the overall disk I/O performance of write-intensive applications such as Exchange databases. In addition, RAID5 is more at risk if a disk fails and the array is under rebuild conditions – if another disk fails during this rebuild interval, the array data will be lost. Array performance is also degraded during the array rebuild condition, however the HP Smart Array configuration utility (ACU) offers some control over the impact by allowing you to set the rebuild priority to high, medium or low.

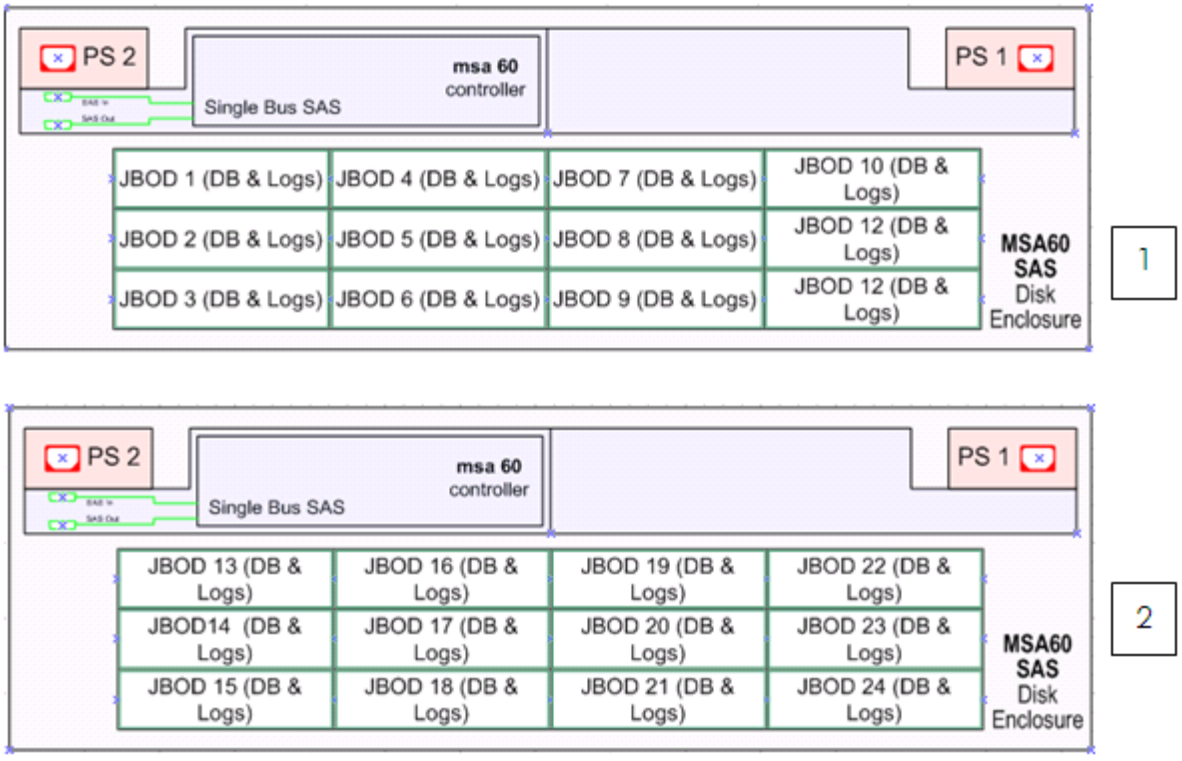
Figure 3. Physical storage layout (24 disk RAID5 configuration)



JBOD

JBOD is not a type of RAID protection, but stands for "Just a Bunch of Disks", which means that all the disks in this type of configuration are being independently addressed with no collective properties. The controller treats each disk separately as mapped to its own logical disk. This configuration does not provide any data redundancy. Figure 4 shows the physical storage layout of a 24 Disk JBOD configuration.

Figure 4. Physical storage layout (24 disk JBOD configuration)



Array configuration

The following section is designed to give you an example of how easy it is to configure the Smart Array and logical disk volumes using automated scripting. It is not at all necessary to use other tools such as traditional point and click interfaces.

Example scripts

Configuration using scripting can be divided into 3 parts:

- First, create the arrays and logical drives step using the HP Array Configuration Utility Command Line Interface (HPACUCLI)
- Next, create the logical disk volumes using Diskpart
- Finally, format (or quick format) the disks

The HP Array Configuration Utility Command Line Interface (HPACUCLI) can be run as follows. (See Figure 4 for an example of how to identify the drive enclosure, port and target numbering).

Note that the numbering in these example scripts will need to be adjusted for your specific configuration. Also, each HPACUCLI command may take several seconds to complete and this is normal.

HPACUCLI

```
cd "C:\Program Files (x86)\Compaq\Hpacucli\Bin"
hpacucli ctrl slot=3 show config
REM Get the list of unassigned drives and use in next steps…
for /L %a in (1,1,12) do hpacucli ctrl slot=3 create type=LD drives=1E:1:%a
for /L %a in (1,1,12) do hpacucli ctrl slot=3 create type=LD drives=1E:2:%a
…repeat for additional ports and enclosures
for /L %a in (1,1,12) do hpacucli ctrl slot=3 create type=LD drives=2E:2:%a
```

Create the mount point folders if you will not be using drive letters:

Mount points

```
md c:\dbmounts
cd c:\dbmounts
for /L %a in (1,1,64) do md db%a
```

Next, create the logical disk volumes using Diskpart and assign them to the mount points created above. Note that the disk numbering should be verified in Server Manager or diskmgmt.msc before proceeding. At this point all of the disks can be initialized using the disk manager as GPT disks.

Diskpart

```
for /L %a in (1,1,64) do call dpx.cmd %a
```

DPX.cmd file

```
echo create vol sim disk=%1 > c:\dpx.txt
echo assign mount=c:\dbmounts\db%1 >> c:\dpx.txt
diskpart /s c:\dpx.txt
```

Finally, format the disks (quick format is preferred):

Format

```
for /L %a in (1,1,64) do format c:\dbmounts\db%a /v:db%a /fs:ntfs /A:64K /q
```

The figure below shows example output of HPACUCLI and how to identify the drive enclosure, port and target numbering for your environment.

Figure 5. Array Configuration Utility (HPACUCLI) example output

Partition Alignment

For this configuration logical volumes can be created using the Windows logical disk manager (server manager or diskmgmt.msc) or the Windows Server 2008 diskpart utility. Note that Windows Server 2008 no longer *requires* the use of diskpart for partition creation and sector boundary alignment as Window Server 2003 did. In Windows Server 2008 the disk partitions are created with a 1 MB offset (except when the disk capacity is less than 4 GB).

Test procedures

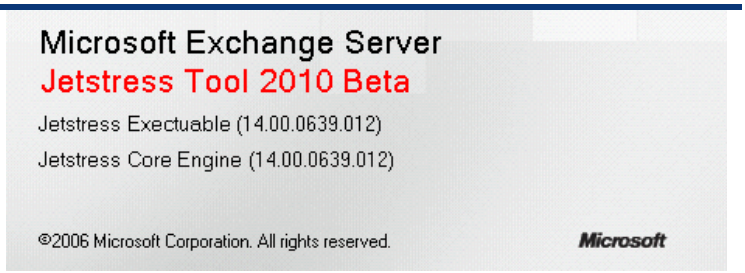
Microsoft Jetstress 2010 beta version 14.00.0639.012 was used to simulate Exchange disk I/O load. There are many different ways to configure Jetstress to perform tests against the configuration. In order to compare the results obtained from different tests, it is important to use the same workload for comparison. When these tests were conducted, the latest available version for Jetstress 2010 beta was used.

The first time Jetstress 2010 is run, the appropriate database must be created for the Jetstress testing. For more information on Jetstress, the tool and documentation can be downloaded from http://www.microsoft.com/downloads/details.aspx?displaylang=en&FamilyID=13267027-8120-48ed-931b-29eb0aa52aa6

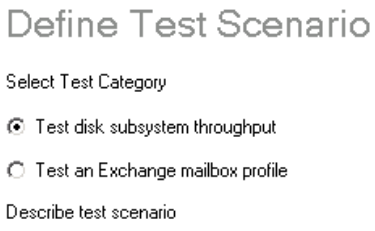
Microsoft Jetstress 2010 requires only some Exchange ESE files to drive the Exchange 2010 I/O against the storage subsystem and not a full Exchange Server installation. For this testing, the Exchange 2010 (v14.0.639.19) ESE files were used.

The parameters that were defined for all tests in the Jetstress utility are the following:

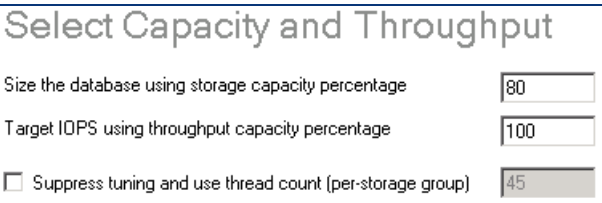
1. First, verify the Jetstress version



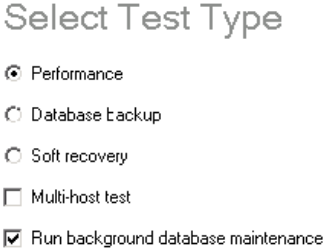
2. Define the test scenario as throughput test



3. In this step, select the capacity and throughput percentages as shown below.



4. Select "Performance" as a test type.



Test results

The following section details the test results using Jetstress 2010 to find the maximum IOPS of which each configuration was capable, keeping the disk latency counters below the previously explained limits (such as sub-20ms read latencies).

The key findings are:

- The Smart Array P411 controller supports more drives than the Smart Array P800 controller in a JBOD configuration (64 as compared to 32) and is capable of supporting a higher workload (e.g. approximately 6500 IOPS for Enterprise SAS) in this configuration. As mentioned earlier, the Smart Array P800 controller supports dual domain for additional fault resilience, but the P411 does not.
- Enterprise SAS drives provide substantially more IOPS per disk than MDL SAS or SATA. Note that the actual IOPS per disk will depend on the total number of disks in the configuration and whether RAID or JBOD is used. For example, Enterprise SAS disks produced over 200 IOPS while MDL SAS and SATA produced around 70 IOPS per disk. However, Enterprise SAS drives are only available at lower capacities than MDL SAS or SATA drives, thus if the largest capacities are desired then sizing (mailboxes per disk) must be reduced to fit. A later section will cover the balance between IOPS and mailbox sizing.
- RAID5 introduces a performance penalty over JBOD reducing the overall IOPS by around 40%. However, this can be considered the cost of data protection. Note that this reduction in performance is substantial enough for MDL SAS and SATA that RAID5 is not recommended for these types of drives. The RAID5 reduction was less on the Smart Array P800 controller when using Enterprise SAS drives and large RAID5 configurations could actually outperform large JBOD configurations.

JBOD configurations

This set of tests was run to determine the performance for each disk type using the JBOD configuration. A total of 24 disks of each type were used for this configuration connected to single Smart Array P411 or P800 controller. There were a total of 24 JBOD disks presented to the server – each containing both Exchange database and logs.

RAID5 configurations

This set of tests was run to determine the performance for each disk type using the RAID5 configuration. A total of 24 disks were used for this configuration connected to single Smart Array P411 or P800 controller. RAID5 was configured with six disk RAID sets with one disk of capacity given up for parity, so the 24 physical disks were configured as four (4) logical disk volumes presented to the server – each containing both Exchange database and logs.

Performance results

Table 1 shows the Total IOPS performance results of 12 spindles for MDL SAS/ SATA in JBOD and RAID5 configurations for the P411 controller.

Table 1. Total IOPS comparisons with 12 spindles

Disk Type	JBOD	RAID5
MDL SAS /MDL SATA	1,254	Not recommended
Enterprise	3,915	2,334

Note that in the above comparison, RAID5 uses six disk RAID sets with one disk of capacity given up for parity, so it does have an impact on potential mailbox sizes.

Table 2 shows the Total IOPS performance results of 24 spindles for MDL SAS/ SATA in JBOD and RAID5 configurations for the P411 controller.

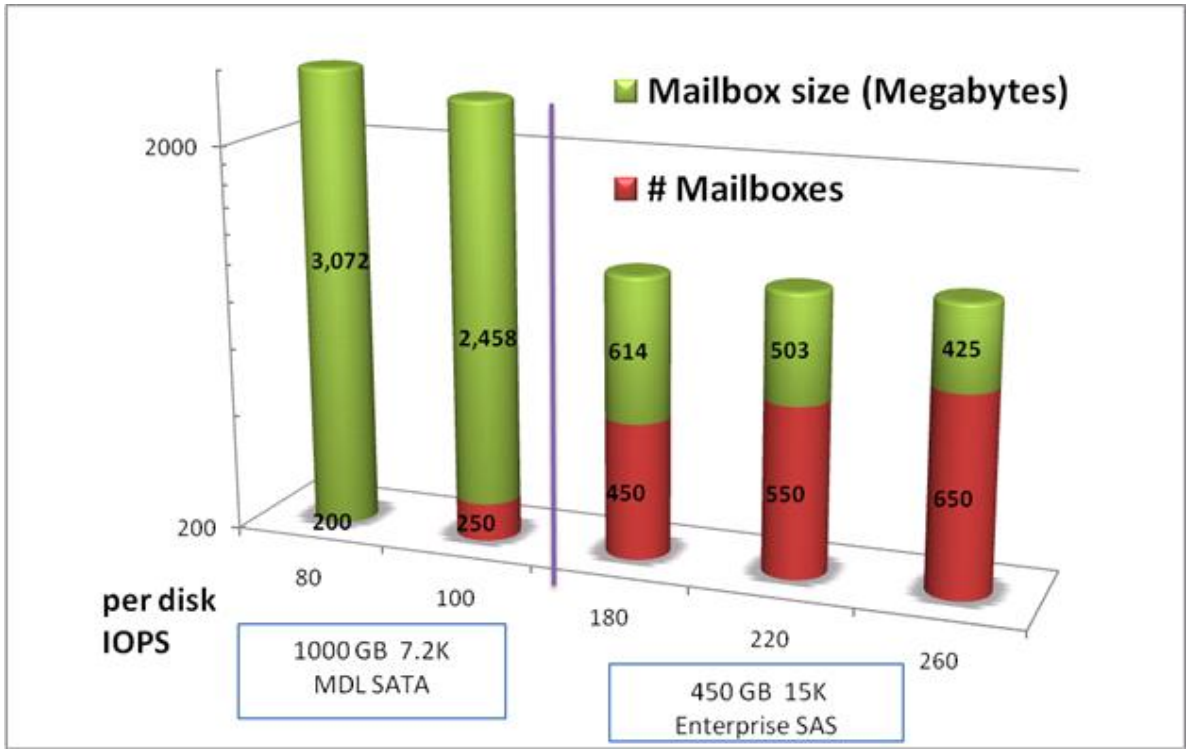
Table 2. Total IOPS comparisons with 24 spindles

Disk Type	JBOD	RAID5
MDL SAS /MDL SATA	2,449	Not recommended
Enterprise	6,667	4,124

IOPS and mailbox sizing

Figure 6 below compares the number of mailboxes that can be supported for MDL SAS/SATA versus Enterprise SAS based on the estimated IOPS per disk and a 0.4 per-user I/O profile. The resulting size for the mailbox is then illustrated, ranging from 3GB per user for 200 users on a 1TB MDL SAS/SATA disk to around 500MB for up to 500 users per Enterprise SAS disk. These numbers are based on a 60% database fill ratio or footprint on each disk, and RAID protection can be added, which will reduce performance, as was detailed earlier in the test results. The additional overhead of RAID can be accommodated by adding additional disks for data protection as in RAID5 for Enterprise SAS or RAID10 and MDL SAS/SATA disks.

Figure 6. Number of mailboxes and resulting size for MDL SAS/SATA versus Enterprise SAS disks



Summary

Microsoft Exchange Server 2010 completely re-architects the database engine for better performance and introduces many new design features and thus many new configuration options to consider. This paper shows how the new database engine performs and how HP Smart Array and StorageWorks products can be configured to use different classes and types of disk storage.

The tests that were performed produced the following conclusions:

- The Smart Array P411 controller supports more drives than the Smart Array P800 controller in a JBOD configuration (64 as compared to 32) and is capable of supporting a higher workload (e.g. approximately 6500 IOPS for Enterprise SAS) in this configuration. As mentioned earlier, the Smart Array P800 controller supports dual domain for additional fault resilience, but the P411 does not.
- Enterprise SAS drives provide substantially more IOPS per disk than MDL SAS or SATA. Note that the actual IOPS per disk will depend on the total number of disks in the configuration and whether RAID or JBOD is used. For example, Enterprise SAS disks produced over 200 IOPS while MDL SAS and SATA produced around 70 IOPS per disk. However, Enterprise SAS drives are only available at lower capacities than MDL SAS or SATA drives, thus if the largest capacities are desired then sizing (mailboxes per disk) must be reduced to fit.
- RAID5 introduces a performance penalty over JBOD reducing the overall IOPS by around 40%. However, this can be considered the cost of data protection. Note that this reduction in performance is substantial enough for MDL SAS and SATA that RAID5 is not recommended for these types of drives. The RAID5 reduction was less on the Smart Array P800 controller when using Enterprise SAS drives and large RAID5 configurations could actually outperform large JBOD configurations.

Implementing a proof-of-concept

As a matter of best practice for all deployments, HP recommends implementing a proof-of-concept using a test environment that matches as closely as possible to the planned production environment. In this way, appropriate performance and scalability characterizations can be obtained. For help with a proof-of-concept, contact an HP Services representative http://www.hp.com/hps/contacts/index.html.

In addition to Jetstress there are other tools such as Microsoft Exchange LoadGen, which can be used to validate the server and storage design of the deployment.

For more information

For further information on HP Exchange solutions including best practices and additional testing of HP StorageWorks or ProLiant products with Exchange, please see http://www.hp.com/solutions/exchange

For additional information on HP ProLiant products including the ProLiant DL380, please see http://www.hp.com/go/proliant.

The HP ProLiant DL385 G5 Server Maintenance and Service Guide is available at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00291791/c00291791.pdf

For general information on Exchange sizing and best practices, there are several white papers and planning tools that are free to download from, http://www.hp.com/solutions/activeanswers

To help us improve our documents, please provide feedback at http://h20219.www2.hp.com/ActiveAnswers/us/en/solutions/technical_tools_feedback.html.

Technology for better business outcomes

4AA2-1181ENW, Revision 2, February 2010