Optimizing Storage Solutions for Microsoft® Exchange Server 2007

A Dell Technical White Paper

By Farrukh Noman and Casey Birch
Dell | Product Group - Enterprise
INTRODUCTION

The need for storage solutions is ever-increasing based on requirements faced by IT departments due to the amount and nature of information being stored, the length of time that data remains relevant, and the manner in which that data is being retrieved. Messaging services like Exchange is one of the major applications driven by large storage and bandwidth requirements. Furthermore, the heavy dependence on email in businesses has made it essential to have a highly available and reliable storage solution ensuring business continuity and providing fast data recovery for failures. With these demanding criteria, many organizations are rethinking their storage strategies to meet the optimal performance requirements of Exchange environment.

Different storage technologies and their implementation designs have created too many options for the customers to choose from. While achieving higher performance within their budgetary constraints, many customers are struggling through these criteria to define a logical choice that best fits into their environment. This white paper defines some of the features considered important for storage selection and provide Dell-optimized storage strategies suitable for a high-performing, scalable, flexible, and reliable Microsoft™ Exchange 2007 environment.

Storage Topology Selection

Exchange is the most widely used messaging application and requires careful planning for its storage selection and deployment. The storage solution may vary from one company to other based on their organizational needs. Selecting a proper storage solution for Exchange deployment depends on a few fundamental questions:

- Is local storage or remote storage required?
- What kinds of storage features are required?
- What level of performance is needed?
- What are the scalability requirements?
- What are the availability or fault tolerance requirements?

Based on the answers to these questions, system administrators can choose a storage technology and implementation suited to meet their needs.

Direct Attached Storage

Traditionally (with versions prior to Exchange 2007), Direct Attached Storage (DAS) topology has been widely used for small Exchange deployments, mainly because it is less expensive and satisfied small businesses requirements. SCSI-based storage systems were the most dominant technology in the DAS space. The performance was limited though, due to the shared parallel nature of the SCSI bus and the number of storage systems that can be served on SCSI connectors. Storage Area Networks (SANs) were typically employed for enterprise-level Exchange deployments to overcome the restrictions of DAS, because many SAN arrays offered advanced software-based features for managing storage, taking backups, and replicating data. But with the advent of Serial attached SCSI (SAS) technology, DAS limitations improved both in terms of performance and connectivity. SAS storage, with its point-to-point technology and higher throughput for I/O intensive applications, combined with the new replication and high availability features in Exchange 2007, attracts organizations running Exchange 2007 of all sizes.
Figure 1: Example of Direct Attached Storage connectivity to Exchange mailbox server

Dell offers high-capacity SAS-based storage systems like Dell PowerVault™ MD1000, MD1120 and MD3000 to make low-cost, highly reliable, and very large mailboxes a reality. The MD1000 and MD3000 storage enclosures can house up to 15 3.5-inch hard drives in 3U of rack space; the MD1120 accommodates up to 24 2.5-inch hard drives in 2U rack space. These PowerVault storage systems connect directly to the mailbox servers through high speed SAS controllers and can be daisy-chained to multiple enclosures. Figure-1 shows multiple PowerVault MD1120 enclosures daisy chained in an Exchange environment.

Fibre Channel Storage Area Network

Although DAS may be adequate for businesses in terms of performance, it may not be ideal for enabling storage consolidation and central administration. DAS storage systems are isolated and require planned down time for most routine maintenance activities. Storage Area Network (SAN) solutions have enabled consolidation in the Enterprise storage space and give administrators a central point to administer all storage for multiple applications. Traditional SANs use a Fibre Channel SCSI-based serial protocol that can accommodate high-performance transfer of block data while overcoming the maintenance, backup, and restore limitations of a DAS solution.

Choosing a traditional Fibre Channel-based SAN solution addresses the limitations posed by a DAS solution, but are far more expensive since they require investments in the form of additional storage infrastructure. In large corporate industries, Fibre Channel SANs have been in use for quite a long time due to their robustness and high performance. These businesses have invested in the FC infrastructure from the very beginning as it was the only dominant SAN technology available at that time. FC SAN can be considered a good option for large customers who already have the infrastructure and expertise to support it.

Figure 2: Example of Fibre Channel SAN connectivity to Exchange mailbox servers

Dell offers high performance Dell / EMC® CX4 series FC storage arrays with UltraFlex™ technology providing modular FC design and 4 Gbps interfaces. Dell/EMC storage systems are highly scalable and suitable for very large Exchange deployments. Figure-2 shows a CX4-series array connected to Exchange mailbox servers in FC SAN environment.

iSCSI Storage Area Network

Internet SCSI (iSCSI) technology is another alternative and overcomes the restrictions provided by complex and expensive Fibre Channel solution. iSCSI technology has matured over the past few years and, from an Exchange perspective, is on par with FC SAN in terms of performance and reliability. It delivers SAN-like SCSI block traffic over an existing LAN-based TCP/IP infrastructure. It is recommended to dedicate an IP network for an iSCSI SAN to segregate iSCSI and typical LAN traffic; however this network does not require additional expertise or equipment to implement, unlike a FC SAN. Business continuity features like remote site replication can take advantage of the existing IP network and easily make it work on Wide Area Networks (WANs). iSCSI SANs are becoming more popular with IT organizations due to their ease of management and their ability to leverage the existing IP networking expertise.
Dell offers high-performance Dell EqualLogic™ PS Series storage arrays in the iSCSI SAN space. The PS Series iSCSI solution is an intelligent and virtualized storage array meeting the scalable needs of a growing environment. It removes many of the complexities associated with a traditional SAN by automating tasks internally with minimal human intervention. Figure-3 shows EqualLogic arrays connected to Exchange mailbox servers in an iSCSI SAN with remote replication over the WAN.

**Figure 3: Example of iSCSI SAN connectivity to Exchange mailbox servers**

---

**Disk Type Selection**

The disk drive is a key component in a storage system and plays an important part in determining the overall performance of a solution. In an Exchange solution, besides the implemented RAID technology, the type and speed of drives define the I/O performance that can be sustained by a storage subsystem. The quality and reliability of the physical drives are also of interest. This section describes some criteria that should be taken into consideration when selecting a storage disk type.

**Disk Technology – Fibre-Channel, SAS & SATA**

Fibre Channel drives were designed for greater reliability and performance required by enterprise applications. The FC architecture with arbitrated loop mechanism allows linking the drives within enclosures over high-speed backplanes and connecting with the initiators without protocol conversion. The FC technology has been in the Enterprise space for quite a while with fibre transfer rates of 4 Gbps (soon to increase to 8 Gbps). But these drives are slightly more expensive due to complex controller designs. They are structurally designed with more-platters, more heads and higher rpm motors with sophisticated electronics using dual-processors, multi-host, and a vibration control mechanism.

The SAS technology has already proved itself better than parallel SCSI in performance and connectivity. Its architecture is based on full-duplex, dual-port design providing additional data accessibility. A SCSI bus controller is limited to 15 storage devices whereas a SAS expander module can attach up to 128 devices. With constant improvement in technology, SAS speed has doubled from 3 to 6 Gbps, essentially offering the same performance and reliability as FC drives at a lower cost. With these features and cost, SAS is becoming more attractive and gaining popularity at the expense of FC.

SATA (Serial ATA) technology, following the same path as SAS, is the successor to the parallel IDE/ATA protocol. SAS and SATA support the same underlying protocol and therefore both types of drives can work on the same interconnected network. The parallel IDE or ATA drives are focused for the embedded desktop applications, but SATA II, with its higher speed of 600MBps and native command queuing (NCQ) feature, is now actually competing for the Enterprise market. Due to its lower MTBF (Mean Time Between Failure) and higher latency compared to FC and SAS drives, SATA is positioned in the Enterprise space where cost is more important than reliability.

Choosing between FC, SAS, and SATA drives is based on some of the factors mentioned above as well the as the supported architecture in deployment. A customer satisfied with existing FC infrastructure would likely continue to use FC drives even with a slightly higher differential price. SAS – besides its lower cost – has another compelling feature that allows interchanging and mixing of SAS and SATA drives in enclosures, which would be of special interest for tiered storage designs. SATA drives, with its increasingly high capacities and a lower cost, is becoming a value proposition for many customers. With trade-offs in certain scenarios, the low reliability factor of SATA drives can be improved by employing RAID techniques.
Disk Drive Speed

Disk drive speed directly affects the disk positioning and data transfer rates and thus is an important factor in determining the performance of a drive. Enterprise class drives for FC and SAS come in both 10K and 15K rpm speeds whereas SATA drives are available in 5400 and 7200 rpm speeds.

The speed factor also determines the power consumption in a disk drive because the spindle motor is the primary consumer of power among all drive components. A higher speed drive consumes more power, generates more heat, and makes more noise while providing higher IOPS (I/O per second). For the same reason, higher speed spindles in an enclosure demand more cooling. SATA drives are usually favored for applications where I/O performance requirements are low and power/cooling requirements are also limited.

Disk Drive Form-factor

Traditionally 3.5-inch form factor FC/SAS drives have been widely used in the Enterprise storage systems, but SAS, in its emerging space, is leaping one step forward by providing small form-factor (SFF) 2.5-inch SAS drives. These drives occupy less space because of their smaller platter size. SFF drives have less power consumption and are also less noisy due to their smaller components. Since more SFF drives can fit into a unit enclosure space with less power consumption compared to 3.5-inch drives, the rack density and power utilization can be greatly improved within large deployments. But there can be a trade-off with the capacity requirements since SFF SAS drives currently come in smaller capacities compared to 3.5-inch drives. Therefore some deployments with a 3.5-inch drives solution might be a better architecture. SFF SAS are currently offered in capacities of 146 GB with 10K rpm and 73 GB with 15K rpm whereas 3.5-inch SAS drives are available in higher capacities of up to ~450 GB.

Dell offers both 2.5-inch and 3.5-inch SAS drives based storage architectures. Dell PowerVault MD1000 supports 15 3.5-inch disk drives in a 3U SAS enclosure whereas PowerVault MD1120 supports 24 2.5-inch drives in a 2U SAS enclosure. Dell / EMC CX-series and EqualLogic both support 3.5-inch form factor drives.

RAID Policy Selection

In an Exchange environment, special consideration is required when allocating RAID type and disks for database/log volumes. Selecting a particular RAID level for Exchange depends on the workload, I/O performance requirement, and disk failure rate.

Impact on I/O Performance and Capacity

The overall IOPS (I/O per second) that can be sustained by a particular storage system depends on the implemented RAID policy and speed of drives. As mentioned in the previous section, higher speed disks can drive higher data transfer rates resulting in more IOPS. RAID policy defines the protection and availability at disk level, which in turn, impacts the way the read and write I/Os are handled in the overall RAID system. There are various RAID policies providing different levels of protection at the expense of performance and capacity. In general, RAID 10 is considered more reliable and provides higher IOPS but also requires more spindles. The data is mirrored and striped across multiple RAID 1 sets and therefore only half of the drives’ space is usable. On the other hand, RAID 5 and RAID 50 comparatively provide less performance but are more cost-effective in terms of providing higher capacity. The more capacity available in RAID 5 is due to the fact that all drives are usable for data except one, which is reserved for parity. The lower performance is a result of a write penalty, where each write is equivalent to four IOPS at the disk level.

The data and log LUNs are usually recommended on separate RAID sets to avoid mixing of random and sequential I/Os. While most of the storage systems strictly follow this recommendation, the Dell EqualLogic array, with its intelligent handling of random and sequential I/Os, allows mixing of data and logs on the same RAID type without impacting the availability and performance.

Degraded Mode and Rebuild Times

The performance is also impacted with disk failure rate and rebuild time in a particular RAID type. When a disk fails in RAID 10, the read performance only decreases on the RAID 1 set with the failed disk, so the overall impact is minimal. The RAID 10 rebuild also involves only two disks – the active disk in the RAID 1 set and the spare disk. In RAID 5 or RAID 50, one disk failure impacts the reads and writes on all the disks. The write I/Os are more affected than reads due to the parity calculations. RAID 50 is comparatively more robust than RAID 5 due to higher availability with more sets and capacity to handle more disk failures.

It is extremely important to reserve the recommended number of hot-spare for the selected RAID type. During a disk failure without a hot-spare, the data continues to be available in degraded mode with some penalty on the performance. With the availability of a hot-spare, the degraded RAID is rebuilt to a healthy point to handle the next failure.

Choosing RAID types in an Exchange deployment are largely influenced by factors like IOPS, capacity and RAID rebuild time requirements and degraded mode performance. RAID 10 is considered for higher performance and reliability at the expense of more spindles. RAID 5 is considered for deployments where capacity is a greater concern over performance. It cannot handle multiple disk failures and therefore should be avoided if there are possibilities of higher disk failure rates. RAID 50 provides slight improvement over RAID 5 by sustaining more disk failures.
High Availability Considerations
High availability for the mailbox server application and mailbox data is a crucial requirement of every messaging system. Choosing a type of availability may vary from one organization to other depending on the down-time that can be sustained during a primary server or site failure. Regardless of the down-time factor and service availability features, every organization should also have an updated backup of data to handle disaster scenarios. Exchange Server 2007 offers various high-availability options that provide data or application availability or both depending on the feature chosen and the topology being deployed.

Cluster Continuous Replication
Cluster Continuous Replication (CCR) is based on the Microsoft Cluster Services (MSCS) Majority Node Set (MNS) cluster model. The active and passive node in this cluster model each maintains its own copy of the database, and the replication takes place in the form of asynchronous transaction log copy and replay. The passive database copy in CCR can be utilized for offloading required database backup activity and to reduce backup frequency. During a primary node failure, the incomplete messages are queued up in a Hub Transport server and resubmitted to the passive node, which becomes primary/active after an automatic failover.

CCR clustering model ensures the availability of both the mailbox application and the database storage. It enables highly available Exchange deployments without requiring complex shared storage systems. Host-based RAID controllers with JBODs (Just a bunch of ordinary disks) make up a robust and inexpensive CCR solution. If deploying Exchange 2007 on Dell MD1120 or MD1000 storage arrays, Dell strongly recommends using CCR.

Single Copy Cluster
Single Copy Cluster (SCC) is based on the shared storage Microsoft Cluster Services (MSCS) clustering model wherein the active and passive nodes are connected to a single copy of a database residing on a shared external storage. It is recommended that each active mailbox server in the cluster have a corresponding passive node so that when multiple active node failures occur, there are sufficient passive nodes to handle the failover, though many organizations deploy between three and seven active nodes with only one or two passive nodes. Unlike CCR, SCC ensures availability of only the mailbox application.

Though SCC was the only form of clustering in previous versions of Exchange, CCR is recommended above SCC for the following reasons: 1) CCR protects both the application and data level, while SCC only protects the application level; and 2) SCC requires a level of complexity in that two servers share the same data. SCC should only be deployed in an environment where some form of backup methodology already exists or the storage is capable of handling highly available cluster requirements. Large deployments with FC or iSCSI SAN can be considered good candidates for Single Copy clustering because the storage systems themselves are highly redundant and provide internal mechanisms for protecting the data. If deploying SCC in a DAS environment, Dell PowerVault MD3000 is designed to meet the clustering requirements.

Standby Continuous Replication
Standby Continuous Replication (SCR) is the new replication feature provided in Exchange 2007 SP1 which supports multiple replication targets per storage group. It copies the log data from an active node to multiple target locations and provides the flexibility of replaying the logs at a later time. The replay lag time is used to prevent logical corruption on the target node in situations where the production data is already corrupted. The target node is manually promoted to the production environment after a failure. SCR is a recommended disaster recovery solution that should be deployed in addition to some local high-availability mechanism.

Local Continuous Replication
Local Continuous Replication (LCR) is a single-server solution that provides availability by creating and maintaining a copy of the Exchange storage groups’ logs and database on a second set of disk volumes connected to the same mailbox server. In case of an active copy failure, the mailbox server can be pointed manually to start using the passive copy as the production version. The passive database copies can be used to offload the required backup activities from the active databases with minimal impact to the end user response time. LCR does not provide application availability or site resiliency, and therefore should not be deployed as the only availability mechanism in the environment. Some other form of application availability method is strongly recommended in an LCR deployment.

Table 1 provides a high-level comparison of all the mailbox high availability features. Specific deployment rules apply for deploying these features. The LCR, CCR, and SCC features cannot be deployed to coexist on the same mailbox server configuration, and are mutually exclusive; SCR can be deployed in combination with these features to increase redundancy and site resiliency.

Table 1: Mailbox Availability Feature Comparison
<table>
<thead>
<tr>
<th>Feature</th>
<th>SCC</th>
<th>LCR</th>
<th>CCR</th>
<th>SCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability Level</td>
<td>Application</td>
<td>Data</td>
<td>Application and Data</td>
<td>Application and Data</td>
</tr>
<tr>
<td>Automatic Failover</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Native Data Replication</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Site Resilience for Disaster Recovery</td>
<td>No</td>
<td>No</td>
<td>Yes*</td>
<td>Yes</td>
</tr>
<tr>
<td>Replication Targets (passive nodes) per Storage Group</td>
<td>None</td>
<td>Single</td>
<td>Single</td>
<td>Multiple</td>
</tr>
<tr>
<td>Replay Delay Option</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Microsoft Windows Server Catalog Listing for Cluster Solution Hardware</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*- It is generally not recommended to deploy CCR in a geo-cluster scenario, but to use SCR for site resiliency.

Putting It All Together
The storage design for mailbox servers is a critical component of any Exchange 2007 architecture. Advances in storage technologies and new features in Exchange 2007 have multiplied the options to consider when designing storage for an Exchange deployment. Given that both DAS and SAN can provide the performance and scalability that Exchange 2007 requires in organizations of any size, this decision must be made based on other considerations such as strategy, budget, etc.

Dell Optimized DAS Solutions for Exchange
For DAS deployments, the Dell PowerVault MD1120 will typically provide the most compelling solution. The MD1120 is based on 2.5" or SFF SAS drives, and therefore provides benefits related to high density storage solutions. The MD1120 packs more spindles per U than 3.5" SAS arrays, providing more IOPS and a lot of capacity in a smaller footprint while saving power costs, since SFF drives use roughly half the power of full size SAS drive. At the current time, the 10K RPM drives provide the best balance of cost, performance and capacity for the MD1120.

Transaction log disks should be configured as RAID 1 or RAID 10, according to Microsoft’s recommendations. Database drives in RAID 10 are typically preferred for performance bound configurations, whereas RAID 5 might be considered from a cost point of view in a capacity bound configuration. Keep in mind the degraded performance of RAID 5 in the case of a disk failure, and possibly plan additional spindles to accommodate.

One thing that has enabled the credibility of lower cost DAS solutions for Exchange 2007 is the implementation of replication combined with clustering features. As detailed above, CCR gives the most in high availability by providing it for both the server and storage layers. It is not recommended to implement DAS solutions for Exchange without deploying CCR. Figure-4 shows a CCR enabled DAS implementation of Exchange with Dell PowerVault MD1120 storage systems.
In large Exchange deployments on DAS, sometimes the concern is raised about manageability of islands of storage. To minimize the overhead of managing dispersed storage arrays it is recommended to design a storage building block that remains consistent across many or even all mailbox servers. The building block consists of a predetermined disk layout optimized for Exchange that is repeatable from one server to the next. Ideally, logical disks for databases and logs are split between two MD1120 arrays in a redundant design as shown in Figure-5. When properly designed for the end state of the mailbox server, the configuration needs only minimal management or change.

Dell Optimized SAN Solutions for Exchange

For SAN deployments, organizations must choose between Fibre Channel and iSCSI. Either topology supports Exchange equally well. Dell EqualLogic PS 5000 series arrays remove many of the complexities associated with SAN. They utilize existing infrastructure of and expertise in IP networks to provide a high performing and redundant storage fabric. Additionally, the built-in
intelligence and tools dramatically reduce the time and effort in deploying and managing a SAN. Those with existing expertise in Fibre Channel and Fibre Channel infrastructure may choose Dell EMC Clarion CX4 series arrays. Often, organizations with an FC SAN already have disciplined practices for deployment, configuration, and management of the SAN.

Deciding which PS5000 array to deploy Exchange on will depend on a number of factors unique to each organization like number of mailboxes, mailbox profile, and mailbox size. The various PS5000 arrays are distinguished by the type and speed of drives whether SATA, 10K RPM SAS, or 15K RPM SAS. Dell recommends either RAID 10 or RAID 50 configurations on PS 5000 arrays used for Exchange.

Multiple high-availability options exist for Exchange setup in a SAN environment. As with DAS solutions, CCR is considered the recommended high-availability (HA) model, because it provides availability at both the server and storage levels. However, unlike most DAS solutions, both the PS 5000 and CX4 arrays are composed of redundant and highly available components so that there are virtually no single points of failure. Therefore, CCR may not be as critical for SAN as it is for DAS, but is still recommended if the storage internal replication mechanisms are not deployed. Figure-6 shows a SAN implementation of Exchange using Dell EqualLogic storage arrays in a CCR based environment.

![Figure 6: CCR based SAN solution for Exchange using Dell EqualLogic storage arrays](image)

Figure 6: CCR based SAN solution for Exchange using Dell EqualLogic storage arrays
Conclusion
When designing storage for Exchange 2007, there are many options to consider and choices to make in regards to storage topology, drives, RAID types, and HA configuration. The enterprise solutions group at Dell has undertaken the task of recommending optimized storage solutions for Exchange based on Microsoft recommendations, internal testing, and the capabilities of various Dell DAS and SAN storage arrays.

Dell-optimized storage solutions for Exchange provide the following benefits:

<table>
<thead>
<tr>
<th></th>
<th>Dell SAN Solutions for Exchange</th>
<th>Dell DAS Solutions for Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reducing Complexity</strong></td>
<td>Dell EqualLogic arrays require no additional infrastructure expertise beyond IP networking to implement. Additionally, the intelligence and management features in the array greatly reduce the time and effort in deploying SAN.</td>
<td>Deploying and managing PowerVault DAS solutions require virtually no additional expertise beyond managing internal server storage.</td>
</tr>
<tr>
<td><strong>Reducing Costs</strong></td>
<td>The PS 5000 series arrays remove much of the additional costs associated with deploying SAN in that an expensive fibre channel infrastructure is not needed to support it.</td>
<td>DAS solutions offer the highest value in terms of dollars per GB and dollars per IOPS. Additionally the SFF drives of the MD1120 only require roughly half the power of 3.5&quot; drives, reducing the cost even more.</td>
</tr>
<tr>
<td><strong>Reducing Data Center Footprint</strong></td>
<td>Dell EqualLogic arrays have a significant performance advantage over other storage arrays as evidenced in comparing ESRP results among multiple vendors. For performance bound implementations, the PS 5000 allows you to do more with less, reducing the number of arrays needed to support large environments.</td>
<td>Using SFF drives allows PowerVault MD1120 to be one of the highest density enterprise storage solutions in the market allowing you to deploy more mailboxes per U than most arrays using 3.5&quot; drives.</td>
</tr>
</tbody>
</table>

More information about Dell-optimized storage solutions for Exchange can be obtained at [www.dell.com/exchange](http://www.dell.com/exchange). Dell services include assessment, design, and implementation tailored for messaging deployments. Dell also offers end-to-end Exchange messaging solutions that include partner offerings for security, archiving, backup, and recovery. More information can be obtained at [www.dell.com/secureexchange](http://www.dell.com/secureexchange).