Chose the best RAID solution for Exchange 2003

Written by Marc Grote - mailto:grotem@it-training-grote.de

Abstract

In this article I will show how to find the best RAID solution to store your Exchange Server databases and transaction logfiles.

Let's begin

Storage Disk subsystems bottlenecks can cause performance problems, sometimes more than server side CPU or RAM bottlenecks. Slow disk subsystems are often the cause for many performance bottlenecks.

As a first indicator your disk subsystem is slow when:

- Average read and write latencies over 20 ms.
- Latency spikes over 50 ms for more than a few seconds.

As a general tip you should invest in high performance hard disks with 15.000 rpm. It is better to have smaller capacity disks that utilize each spindles performance than to use fewer spindles with large capacity. Fast storage with a sufficient amount of spindles is one of the most important investments you can make in your messaging infrastructure.

What is impacting disk I/O in Exchange

- Jet Database (Exchange Database files (*.EDB))
- Streaming media files (*.STM)
- Exchange Transaction logfiles (*.LOG)

Other Activities that Impact I/O

- Configuration of "Zero out deleted database pages" in Exchange System Manager for the Exchange Server. With this setting enabled, Exchange will overwrite deleted pages with zeros.
- SMTP Mail transmission can also additionally load your hard disks with more I/O.
- If you enable Content Indexing on your Exchange databases you will estimate excessive paging and more disk I/O on your harddisks.

What is RAID

In computing, a redundant array of independent disks, also known as redundant array of inexpensive disks (commonly abbreviated RAID) is a system of using multiple hard drives for sharing or replicating data among the drives. Depending on

the version chosen, the benefit of RAID is one or more of increased data integrity, fault-tolerance, throughput or capacity compared to single drives. In its original implementations (in which it was an abbreviation for "Redundant Array of Inexpensive Disks"), its key advantage was the ability to combine multiple low-cost devices using older technology into an array that offered greater capacity, reliability, or speed, or a combination of these things, than was affordably available in a single device using the newest technology

RAID Level

I will give you an overview of the commonly used RAID levels. The description of the RAID levels is copied from www.wikipedia.com

RAID Level 0

A RAID 0 (also known as a striped set) splits data evenly across two or more disks with no parity information for redundancy. It is important to note that RAID 0 was not one of the original RAID levels, and is not redundant. RAID 0 is normally used to increase performance, although it can also be used as a way to create a small number of large virtual disks out of a large number of small physical ones. A RAID 0 can be created with disks of differing sizes, but the storage space added to the array by each disk is limited to the size of the smallest disk—for example, if a 120 GB disk is striped together with a 100 GB disk, the size of the array will be 200 GB.
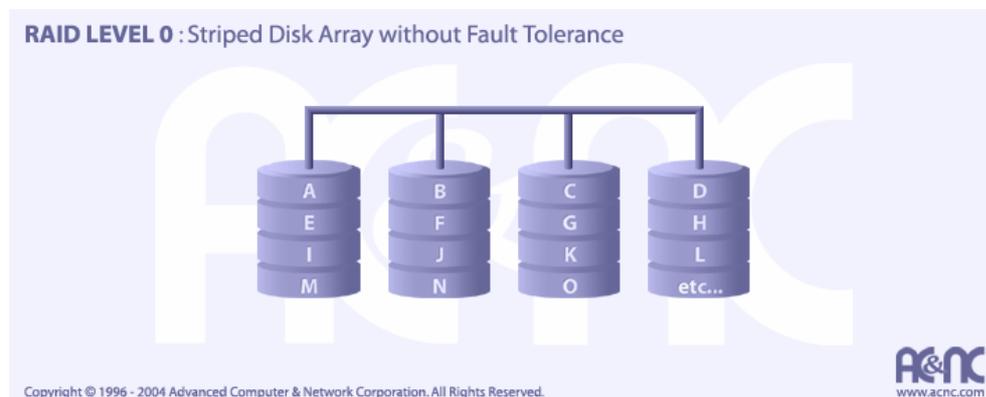

Figure 1: RAID Level 0

Please note: Raid-0 is never recommended for storing Exchange databases

RAID Level 1

A RAID 1 creates an exact copy (or mirror) of a set of data on two or more disks. This is useful when write performance is more important than minimizing the storage capacity used for redundancy. The array can only be as big as the smallest member disk, however. A classic RAID 1 mirrored pair contains two disks, which increases reliability by a factor of two over a single disk, but it is possible to have many more than two copies. Since each member can be addressed independently if the other fails, reliability is a linear multiple of the number of members.
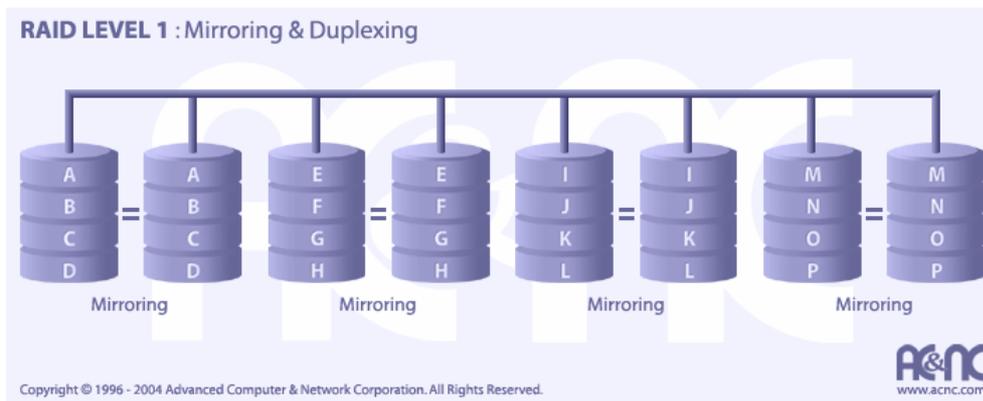
Figure 2: RAID Level 1

RAID Level 5

A RAID 5 uses block-level striping with parity data distributed across all member disks. RAID 5 has achieved popularity due to its low cost of redundancy. Generally RAID-5 is implemented with hardware support for parity calculations
Every time a data "block" (sometimes called a "chunk") is written on a disk in an array, a parity block is generated within the same stripe.
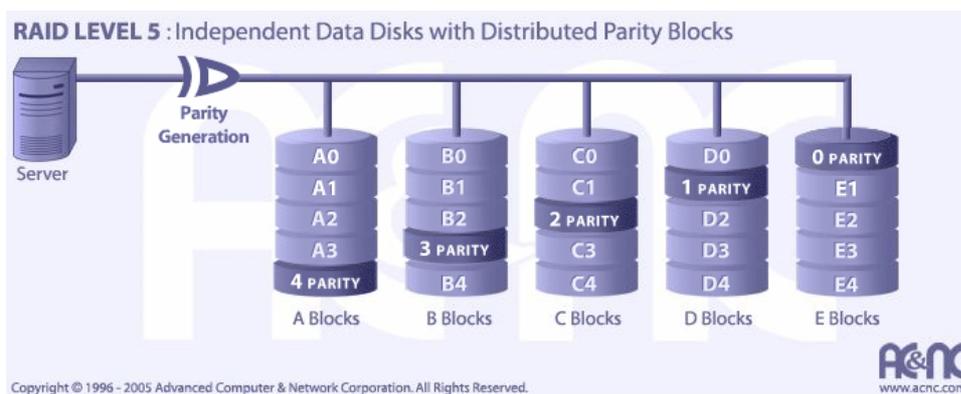

Figure 3: RAID Level 5

RAID Level 10

A RAID 10, sometimes called RAID 1+0, or RAID 1&0, is similar to a RAID 0+1 with exception that the RAID levels used are reversed—RAID 10 is a stripe of mirrors. One drive from each RAID 1 set could fail without damaging the data. However, if the failed drive is not replaced, the single working hard drive in the set then becomes a single point of failure for the entire array. If that single hard drive then fails, all data stored in the entire array is lost
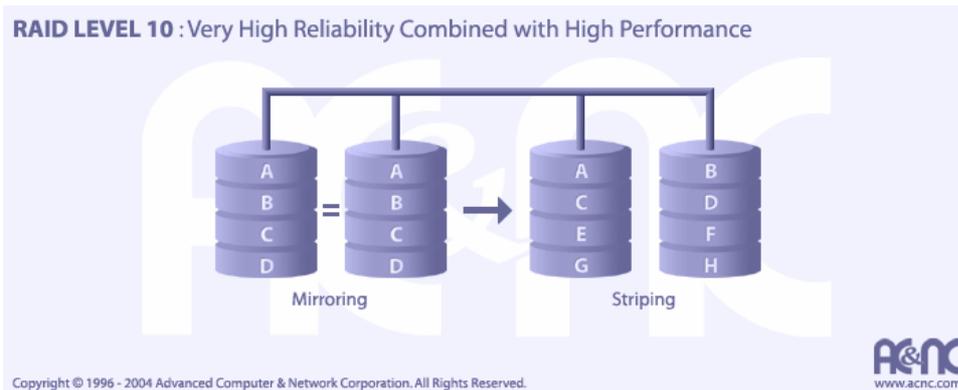
Figure 4: RAID Level 10

One example of a recommended configuration

RAID 1 =System volume, Operating System, Exchange binaries
RAID 1 = Pagefile
RAID 0+1 = SMTP and MTA queues
RAID 1 = Logfiles from one Exchange Storage Group
RAID 0+1 = Exchange Databases from Storage Group

In general, Raid-5 does not provide the best reliability, availability and performance.

Some other tips

- Place database files from Exchange Storage Groups on separate hard disks.
- Never place the Content indexing files on the same disk as the pagefile
- Place the Transaction log files for each Storage Group to a dedicated RAID-1 or RAID-0+1 array.
- Place the SMTP queue on a separate disk subsystem.
- Place the page file on separate disks on RAID1 because a lost of a pagefile could cause a stop error.

Diskpar

You can use DiskPar (part of the Windows Server 2003 Resource Kit tools) to verify that your disk tracks are sector-aligned. You can use DiskPar to create aligned partitions (as compared with non-aligned partitions that are created with Disk Manager from Windows Server 2003), you can increase disk performance up to 20 percent (as stated from Microsoft).

Jetstress

You can use Jetstress (part of the free Exchange downloads from the Microsoft Exchange website) to verify the performance and stability of a disk subsystem prior to putting your Exchange servers into production. Jetstress helps you to verify disk performance by simulating Exchange disk I/O load. Jetstress simulates Exchange database load from the number of users you specified. Jetstress is in parts similar to the Exchange 2003 tool Loadsim but more specialised in database I/O tests.

Conclusion

A high performance RAID solution for Exchange is not so difficult to implement. The recommended RAID solution depends highly on you Exchange I/O usage and the number of mailboxes on your Exchange Servers. You should analyze your requirements before you buy Server Hardware and disk subsystems.

Related Links

Disk Geometry
http://www.msexchange.org/tutorials/Disk-Geometry.html
Optimizing Storage for Exchange Server 2003
http://www.microsoft.com/technet/prodtechnol/exchange/2003/library/optimizestorage.mspx
Disk Subsystem Performance Analysis for Windows
http://www.microsoft.com/whdc/device/storage/subsys_perf.mspx
Jetstress
http://www.microsoft.com/downloads/details.aspx?familyid=94b9810b-670e-433a-b5ef-b47054595e9c&displaylang=en