About this RAID Primer
This RAID Primer is intended as a guide to help the reader determine which PCI-SCSI disk array controller and which RAID level are appropriate for a given system configuration. Various configurations are presented, which will give the reader a wide variety of options. By increasing the reader’s understanding of RAID configurations, and the many enhancements to system efficiency that RAID provides, we are removing the mystery that surrounds RAID.

If there are questions or further information is needed, please contact:
ICP vortex Computersysteme GmbH,
Phone: +49-7132-9620-800 or 900,
Fax: +49-7132-9620-200
Introduction - What is RAID?

Part I - RAID Levels
- How are the Drives Organized?
  RAID 0 or Data Striping
  RAID 1 or Drive Mirroring / Drive Duplexing
  RAID 2 or Hamming System
  RAID 3 or Byte Striping with Parity Drive
  RAID 4 or Block Striping with Parity Drive
  RAID 5 or Block Striping with Striped Parity
  RAID 10 or Mirrored Striping Array

Part II - Typical Configuration Examples
  1. NT or Unix* Workstation
  2. Fast Workstation Working with Large Files
  3. Small NetWare* Server (36 GB)
  4. Small Windows* NT or Unix* Server (36 GB)
     Add more Capacity
  5. Mid size NetWare* File Server (72 GB)
     Add Capacity to the Mid size NetWare* File Server
     Adding Speed and Redundancy to the Server
  6. Mid size Windows* NT or Unix* Application Server (72 GB)
     Adding more Capacity to the 3 Channel System
  7. Large Database Server
     Adding even More Capacity to the Database Server

Part III - Index
Introduction - What is RAID?

With RAID (Redundant Array of Independent Disks) several independent hard drives (the gray pots in the picture) are combined together to form one large logical array (dashed pot). Not only data is stored on this array, but “redundancy information” is also added. The redundancy information may be either the data itself (mirroring), or parity information calculated out of several data blocks (RAID 3, 4, or 5).

The operating system (like Windows®, NetWare®, or Unix®) thus no longer deals with individual drives, but instead with the array. Improving accessibility is the major objective of RAID. RAID prevents downtime in the event of a hard disk failure, but it will not recover data that has been deleted by the user or destroyed by a major event like theft or a fire. Because of this you still need a backup to secure your system from these problems even if you install a RAID system.
Part I - RAID Levels - How are the Drives Organized?

A research group at UC Berkeley has defined six RAID levels. Each level spreads the data across the drives of the array in a different way and is optimized for specific situations.

RAID 0 or Data Striping

This RAID system combines two or more hard drives such that the data (ABCD...) coming from the user is cut into smaller blocks (4-128kB, represented by one character) and these blocks are stored in turns on the different drives of the RAID 0 array. By doing this, two or more hard drives are combined and the read/write performance, especially for sequential access, can be improved. However, no redundancy information is stored in a RAID 0 array, which means that if one hard drive fails, all data is lost. This lack of redundancy is also stated in the number 0, which indicates no redundancy. RAID 0 is thus usually not used in servers where data protection is a concern.

RAID 1 or Drive Mirroring / Drive Duplexing

In a RAID 1 system, identical data is stored on two hard drives (100% redundancy). If either hard drive fails, the system continues working with the remaining drive. For smaller systems, RAID 1 is usually the best solution for high data availability. However, this RAID level becomes quite expensive for large capacities because twice the used disk capacity is needed in order to provide complete redundancy.
RAID 2 or Hamming System

A RAID 2 system cuts the data into single bytes and writes them onto the data disks of the array. Error correction code is calculated using the Hamming algorithm and stored on additional hard disks. This system was used during the early days of RAID, before hard disks had their own built-in correction code. Today’s built-in ECC makes the ability of RAID 2 to find and correct single byte errors no longer necessary, so this RAID level is no longer used.

RAID 3 or Byte Striping with Parity Drive

In a RAID 3 configuration, the data is cut into single bytes, which are written in turns onto the data drives (typically 2-4 drives) of the array. A parity byte is added for each “data row” and is stored on an additional drive – the parity drive. (This is very similar to the parity used for RAM on the motherboard, where 8 data bits are secured by one parity bit). If a hard disk fails in a RAID 3 array, the user data can be recalculated by the RAID system from the remaining data and the stored parity information. Since modern motherboards and operating systems no longer deal with single bytes, RAID 3 is not of much interest any more.
RAID 4 or Block Striping with Parity Drive

RAID 4 is very similar to RAID 3, with the difference that the data is cut into larger blocks (16, 32, 64 or even 128 Kbytes) rather than into single bytes. These blocks are stored on the data drives (similar to RAID 0). For each row, a parity block (similar to the “sum” of the data in that row) is calculated and stored on the parity drive. With this parity information, it is possible to recalculate the lost data in case of a hard drive failure. The performance of a RAID 4 system is very good for sequential reads and writes (storing very large files). Random writes, however, require an update of the corresponding parity block on the parity drive for each user I/O (in effect, the “sum” must be recalculated every time the new data is added). This means that the random user I/Os have to wait for the parity information to be written to the parity drive, resulting in a rather low random write performance.
RAID 5 or Block Striping with Striped Parity

The difference between RAID 4 and this array is that in a RAID 5 system, the parity is distributed over all hard drives in the array. This speeds up random write I/Os, since there is no longer a dedicated parity drive that may become a bottleneck. The read performance is typically similar to a RAID 4 system. This is the standard RAID level for high capacity servers.

RAID 10 or Mirrored Striping Array

This new RAID level is a combination of RAID 1 (mirroring) and RAID 0 (striping), having both features of these arrays – security and sequential performance. Sometimes it is also referred to as RAID 0+1. Typically four hard drives are used, because RAID 10 creates two pairs of mirrored arrays and combines these arrays to form one RAID 0 array. RAID 10 is especially appropriate for redundant storage of large files, and because parity is not calculated, write operations are very fast.
Part II - Typical Configuration Examples

1. NT or Unix* Workstation

These operating systems work very intensively with swap files. In high load situations (many user I/Os on the hard disk), the swap file must also be accessed very often, which can easily result in a bottleneck on the hard disk. The use of a caching controller can significantly improve the hard drive performance in such situations. Typically you will install 64 to 128MB of RAM, and you may want to consider using two independent hard drives – one for the boot drive with the swap file and one for the user data. Doing this will completely separate the swap and user I/Os and thus improve the performance even more.

Since this system has no redundancy, the hard drives need not be mounted in hot-plug shuttles.

To terminate the SCSI bus, use the terminators on the controller and a separate LVDS or multimode terminator connected to the other end of the SCSI bus.

If a CD-ROM is also attached to the controller, it should be not used for termination, i.e. do not put it at the end of the SCSI cable. Since standard CD-ROMs are relatively slow SCSI devices, they usually incorporate only passive SCSI termination and therefore cannot be used to terminate Ultra SCSI systems.
2. Fast Workstation Working with Large Files

For video or DTP workstations where large files are stored on a hard drive, performance is more important than data protection. Storing large files means sequential data transfer, which can be improved using a striping configuration (RAID 0).

A very powerful solution would be to store the large data files on an 72 GB RAID 0 system (two 36 GB hard drives striped together) and having an additional 36 GB hard drive as a boot drive and also for the swap and program files. This keeps the system I/Os separate from the data I/Os. Since RAID 0 is not failure tolerant and does not support any kind of Hot Plug, the hard drives can simply be mounted inside the workstation.

3. Small NetWare* Server (36 GB)

A server requires security. For smaller systems, a mirrored solution (RAID 1) fits best. Whenever the necessary storage capacity can be handled by one hard disk, a mirrored array should be considered. A single channel mirroring controller provides a very cost-sensitive solution, with the two mirrored hard disks connected to the same SCSI channel.
A two-channel mirroring controller provides even more performance and redundancy, because the mirrored hard disks can be connected to separate SCSI channels. The two hard disks can then be accessed simultaneously by the controller, increasing performance, and even if a complete SCSI channel should fail (through cable problems, terminator failure, channel blocked by a SCSI device, etc.), the other SCSI channel is still up and running, with all data intact.

The big advantage of hardware mirroring compared to software mirroring (e.g. the built-in mirroring function on some server operating systems) is that hardware mirroring copies every byte from the first hard disk to the second, providing 100% mirroring. The NetWare* partition, the MS-DOS* partition and even the master boot record are all mirrored. Hardware mirroring also supports Hot Fix drives (spare drives used in case of a disk failure), Hot Plug, and even Auto Hot Plug to exchange a failed hard drive during operation without shutting down the operating system.
Finally, hardware mirroring puts no additional load on the motherboard, even during a mirror update or repair process.

The hard drives are usually mounted in removable shuttles inside the servers to support Hot Plug. To make the exchange of a defective hard drive very easy, separate internal active terminators can be used (in which case, the termination of a new hard drive, if available, should always be deactivated before installation). Tapes and CD-ROMs also may be connected to the RAID controller or to a separate low-cost SCSI adapter. Connecting these slower SCSI devices to a separate SCSI adapter will free the RAID controller’s SCSI buses from the slow data transfer of these devices. Therefore this is the preferable solution.

4. Small Windows* NT or Unix* Server (36 GB)

Operating systems like Windows* NT or Unix* work very intensively with swap files to virtually increase the RAM installed in the computer. Lots of RAM is needed during heavy loading, which means that the swap files are used very intensively.
These swap I/Os can interfere with the user I/Os if both are located on the same drive (whether that drive is a single hard drive or a RAID system). In this situation, the read/write head of the drive(s) would permanently move back and forth between the swap file and the user data, during which the hard drives are not doing any data transfer. This problem can be avoided if the swap file and the user data are kept separate. One Array can be configured for the boot/swap drive (e.g. two mirrored 36 GB drives) and another for the user data (e.g. two mirrored 72 GB drives). Four hard disks are installed in the system, but the operating system will deal with 2 logical drives (i.e. the two mirrored arrays).

A one-channel RAID 1 controller allows for a very inexpensive setup, or a two-channel RAID 1 card will provide more performance and redundant SCSI channels.

Add more Capacity

If the installed capacity becomes too small later on, simply buy two new hard drives, connect them to the controller and configure them as a new mirroring array. The operating system will then see just one additional drive.
Since six hard disks are already installed in this setup, it is time to think about a Hot Fix drive. A Hot Fix drive is an installed spare available to immediately replace a failed drive. The RAID controller will start the repair of the non-redundant array right away, so the system remains in a non-redundant state for only a very short time.

5. Mid-size NetWare* File Server (72 GB)

When building a new server with 72 GB or more, it makes sense to consider a RAID 5 system. Combining three 36 GB hard drives provides a user capacity of 72 GB (due to the way RAID 5 calculates parity). NetWare* does not usually require a separate boot drive, because the system files are typically loaded once during boot and do not need to be reloaded very often during normal operation.
The three hard drives usually fit in a mid sized server enclosure along with a tape and a CD-ROM. Since RAID 5 allows the exchange of a failed hard drive during operation, a shuttle system with a separate active SCSI terminator should be used for the hard drives. For best performance and security tapes and CD-ROMs should be connected to a separate low-cost SCSI host adapter.

Add more Capacity

Adding more user capacity to the file server can be done online (NetWare* keeps running, the users are still logged in and even continue reads and writes) with an online expansion of the RAID system. To accomplish this, the mechanics (an additional empty shuttle for the new hard disk) must already be installed in the system. The online expansion of the RAID system will include the new hard disk in the existing disk array, reorganize the data and parity information and make the additional storage capacity available to the operating system and the user. But online expansion should not be overdone, i.e. not too many hard disks should be linked together in one big disk array. Better performance can usually be obtained by configuring additional RAID arrays instead – even on the same controller. With about five hard disks in one array you typically get the best performance.
Adding Speed and Redundancy to the Server

This setup can be further improved by using a four channel controller and attaching each one of the four hard drives of the RAID 5 system to a separate SCSI channel. By doing this the controller can access all hard drives at the same time, making the system even faster.

An additional advantage is that SCSI channels are now also redundant, which means that, even if a complete SCSI channel fails (due to cable-, connector-, or termination problems), we lose only one hard drive of the redundant RAID 5 array.

6. Mid-size Windows* NT or Unix* Application Server (216 GB)

Since these operating systems work very intensively with swap files it is advisable to separate the boot drive and swap file from the normal user I/Os (see discussion above). A good solution is two 36 GB hard drives for the mirrored boot drive and four 72 GB hard drives in a RAID 5 configuration for the user data (adding up to 216 GB user capacity).
More security and performance can be added to the server by connecting the hard drives to a 4-channel controller. In this setup the hard drives of each redundant array are connected to independent SCSI channels. In case of a cable failure you may lose two hard drives but since they belong to different RAID arrays, each array remains active.

To add even more security to the system, add an additional hard disk and define it as a Hot Fix drive, which is an installed spare available to immediately replace a failed drive. The RAID controller will start the repair of the non-redundant array right away, so the system remains in a non-redundant state only for a very short time. If this 72 GB spare
drive is defined as a Pool Hot Fix drive, it can be used either for a failed drive in the RAID 5 array or for failure of one of the mirrored 36 GB hard drives. If the latter is the case, the 72 GB may be replaced later by a 36 GB hard drive and be configured again as a Hot Fix drive.

For such a setup, a four-channel RAID controller is the perfect choice. Since only seven hard drives are used in this configuration, it should be most possible to mount them directly in the server enclosure. If this poses a problem, an external enclosure may be used (for the four hard drives forming the RAID 5 array or even for all the hard drives). In this case, the hard drives may still be striped over the SCSI channels to achieve the highest level of redundancy and performance. In both setups, it is advisable to use separate active terminators for the SCSI bus.

Adding more Capacity to the Four-Channel System

If more capacity is needed later on, just add another four hard disks and define them as a new RAID 5 system. The operating system will see three logical drives: the boot drive (RAID 1), the data drive (RAID 5) and another data drive (RAID 5).
If there is not enough room in the server, an additional external enclosure must be used. If this becomes necessary, the boot drive and Hot Fix drive may still remain inside the server. In this setup the controller will no longer be at the end of the SCSI bus and therefore the termination of the controller has to be disabled.

7. Large Database Server

The issue of separating different I/Os on mechanically independent arrays becomes even more important in a database server. Besides the swap I/Os of the operating system, there are additional I/Os which are performed in parallel – index and data read/write operations.
Any access to the database must access both the index file and the data file. If both files are located on the same array in a heavily loaded database server, the hard drives must perform many time-consuming movements of the read/write heads (moving the heads from the index file to the data file back and forth). Because of this, database management system companies recommend “load balancing” when laying out data on the hard drives. This means using independent hard drives for the different database files. A disk array automatically balances the load, but one can achieve even more performance by installing independent arrays for these independent I/Os.

The nine hard drives in our example forming the two RAID 1 arrays (boot and index drive), the RAID 5 array (data files) and the Hot Fix drive may be mounted with the motherboard itself in a 19-inch rack. Everything is close together in such a system, which avoids SCSI cables that are too long. Again, separate active terminators should be used in such a setup to avoid problems when replacing defective hard drives. Additionally, the power supply and the cooling system must be able to handle the requirements of such a setup.

Adding even More Capacity to the Database Server

If more capacity is needed in such a setup, just add another bunch of hard disks and configure them as a new RAID system. This may either be a RAID 1 or a RAID 5 array depending on the needed capacity. If you are only dealing with large files and the array is usually used by only one process at a time, a RAID 4 system may give you a little more performance compared to the RAID 5 in some situations.

If data availability is not very critical, but performance must be as high as possible, a RAID 0 system may be the array you need. If both security and performance are important the best solution is to combine a RAID 1 (security) and RAID 0 (performance) into a RAID 10 system.
Part III - Index

Adding More Capacity  17, 19, 22, 24
Adding Speed and Redundancy  20
Backup  8
Fast Workstation Working with Large Files  14
Hot-Fix Drive  18
Hot-Plug Shuttles  13
Index File  23
Large Database Server  23
Load Balancing  23
Low-Cost SCSI Adapter  16
Master Boot Record  15
Mid size NetWare* File Server  19
Mid size Windows* NT or Unix* Application Server  20
NT or Unix* Workstation  13
Online Expansion  19
RAID  - Redundant Array of Independent Disks  8
RAID 0 - Data Striping  9
RAID 1 - Drive Mirroring / Drive Duplexing  9
RAID 10 - Mirrored Striping Array  12
RAID 2 - Hamming System  10
RAID 3 - Byte Striping with Parity Drive  10
RAID 4 - Block Striping with Parity Drive  11
RAID 5 - Block Striping with Striped Parity  12
RAID Levels  9
Random Write Performance  11
Random Writes  11
Redundancy Information  8
SCSI Channel Complete Failure  15
Sequential Data Transfer  14
Small NetWare* Server  14
Small Windows* or Server  16
Stripe Size  11
Swap File  13
UC Berkeley  9