What's RAID?

The purpose of this document is to explain the many forms or RAID systems, and why they are useful, and their disadvantages. RAID - Redundant Array of Inexpensive Disks - is a method of combining several hard drives into one logical unit. It can offer fault tolerance and higher throughput levels than a single hard drive or group of independent hard drives.

RAID is a mature technology that speeds up data access while at the same time protecting your data from hard disk failure. RAID is quickly becoming a necessary component in every network since data loss and downtime can prove both fatal and financially destructive. Most networks are designed to provide instant access to massive amounts of data. More and more employees have to access customer and other databases. Intranets and corporate Web sites provide access to huge databases online.

Raid Components and Concepts

First, let us define *Logical Arrays* as a split or combination of *Physical Arrays*, which in turn are one or more Physical Drives that are simply the individual hard disks that comprise these arrays. *Logical Drives* are then made of one or more *Logical Arrays*.

Mirroring refers to complete redundancy of data on identical disks. The data that is being written on one Logical Array is completely duplicated on a similar array thereby providing 100% data redundancy. The cost associated with mirroring is that the amount of available storage is reduced by 50%; writes are slightly slower albeit reads are faster in some situations.

Striping refers to a technique that allows Physical Drives in a Logical Array to be used in parallel in order to gain in performance. In this technique, data is broken down in Byte or Block levels or *stripes*, where every Byte or Block is written to a separate disk in the array. Byte level can at times be a 512-byte sector, while Block size can be selected from variety of choices. The gain in performance is similar between Reads and Writes.

In some RAID levels, striping is combined with a technique called *Parity* to enhance fault tolerance. Parity, similar to parity in memory, is simply adding a Block (Byte) of calculated parity data to several Blocks (Bytes) in such a way that any one of the Blocks (Bytes) can be reconstructed in case of loss, from the remainder of the Blocks (Bytes) and the parity Block (Byte). While Parity gains from performance of striping, its disadvantages are more complexity and loss of some disk space, which is taken up by parity information.

There are many ways to combine RAID techniques. Some standardized combinations are referred to as RAID Levels, even though ‘Level’ in this context does not denote any hierarchy or advantage. Levels are independent and different. Some RAID levels combine multiple other levels to achieve certain aims.

The RAID Advisory Board (RAB) has been active since 1992 in education and standardization of RAID technology. See http://www.raid-advisory.com/. Techniques discussed above are used in different levels. Mirroring is used in levels 1, 0+1, 10 (1+0). Striping without parity is used in level 0, 0+1, and 10. Striping with Block level parity is used in level 5 and 6. While the minimum number of drives required at each level are noted, there is no inherent maximum to number of drives in arrays other than the one imposed by controllers.

RAID-0 -- Striping

Simple *striping* is used in this level to gain in performance. This level does not offer any redundancy. Data is broken into stripes of user-defined size and written to a different drive in the array. Minimum of two disks are required. It uses 100% of the storage capacity since no redundant information is written. Recommended use for this level is when your data changes infrequently and is backed up regularly and you require high-speed access. Web servers, graphics design, audio and video editing, and online gaming are some example applications that might benefit from this level.

RAID-1 –- Mirroring

This level uses *mirroring* and data is duplicated on two drives. If either fails, the other continues to function until the failed drive is replaced. At the cost of 50% of available capacity, this level provides very high availability. Rebuild of failed drives is relatively fast. Read performance is good and write performance is fair compared to single drive read and write. A minimum of 2 drives is required. Whenever the need for high availability and vital data are involved, this level is a good candidate for use.

RAID-2, RAID-3, and RAID-4

RAID-4 interleaves stripes like RAID-0, but it requires an additional drive just to store the parity, which is used to provide redundancy. In a RAID-4 system, if any one of the disks fails, the data on the remaining disks can be used to reconstruct the data that was on the failed disk. Even if the parity disk fails, the other disks are still intact. Thus RAID-4 can survive the failure of any of its disks. RAID-2 and RAID-3 are seldom used anymore, and have mostly been made obsolete by modern disk technology. RAID-2 stores ECC information instead of parity, but since all modern disk drives incorporate ECC, RAID-2 offers little additional protection. RAID-3 is similar to RAID-4, except that it uses the smallest possible stripe size. As a result, any given read will involve all disks, making overlapping I/O requests difficult/impossible. In order to avoid delay due to rotational latency, RAID-3 requires that all disk drive spindles be synchronized. Most modern disk drives lack spindle-synchronization ability, so RAID-3 is no longer used.

RAID-5 -- Striping with Parity

One of the most popular RAID techniques, it uses Block Striping of data along with parity and writes them to all drives. In contrast to the RAID levels that write the parity information to a single drive and use the rest of the drives for data blocks, RAID-5 distributes the parity blocks amongst all drives, keeping parity separate from the data blocks generating it.

RAID-5 systems require a minimum of 3 disks. The impact on capacity is equivalent to removing one drive from the array. If any one drive fails, the array is said to be degraded, and the data blocks residing on that drive can be derived from parity and data on remainder of the drives. RAID controllers usually allow a hot spare drive to be configured that is used when the array is degraded and the array can be rebuilt in the background while normal operation continues.

RAID-5 combines good performance, good fault tolerance, with high efficiency. It is best suited for transaction processing and is often used for “general purpose” service, as well as for relational database applications, enterprise resource planning and other business systems.

RAID-6 -- Striping with dual Parity

This level is identical to level 5 except that it calculates and writes an extra parity block to all drives. While this will have the effect of reducing the useable capacity by one more drive, it reduces the window of vulnerability during the RAID-5 rebuilds considerably and can withstand the failure of a second drive during rebuilds.

The advantages of RAID-6 becomes even more pronounced as the capacity of SATA drives go up and rebuilds take longer to finish. While calculating a second parity has a negative impact on performance in software based RAID systems, the effect is very minimal when hardware RAID engines that have built in circuitry to do the parity calculations are used. RAID-6 requires a minimum of four drives to be implemented and the usable capacity is always 2 less than the number of available disk drives in the RAID set. Applications suited for this level are the same as those of level 5.

RAID-10 –- A Stripe of Mirrors

RAID-10 is an example of combining two RAID levels to achieve more targeted results. It is often confused with its brethren level 0+1 that is referred to as "*Mirrored Stripes*." While in each case drives are mirrored and blocks are striped to these drives, in RAID-10, Blocks are striped to N/2 sets of mirrored drives (N being number of drives in the array) while in level 0+1, blocks are striped to 2 mirrored sets each containing N/2 drives.

Because of RAID-10's mirroring, the storage efficiency is at 50%. This level offers excellent fault tolerance and availability. It is recommended for applications requiring high performance and high reliability that are willing to sacrifice the efficiency (twice the number of drives to achieve the capacity). These include enterprise servers and moderate size database systems.

RAID–50 -- Striping across multiple RAID--5's

Also referred to as level 5+0. It combines *Block Striping with distributed parity* with straight Block Striping of level 0. In other words it uses a Block Stripe of level 0 on Level 5 elements. The minimum number of drives is 6 and the capacity can be derived from subtracting one drive for each set of Level 5 elements. As an example, a 6-drive array would have the capacity of the five drives. Level 50 is recommended when high fault tolerance, large capacity, and random read/writes are required. It is sometimes used for large databases.

RAID-60 Striping across multiple RAID-6's

Also referred to as level 6+0, combines multiple RAID-6 sets with RAID-0 (striping). Dual parity allows the failure of two disks in each RAID-6 array. Striping helps to increase capacity and performance without adding disks to each RAID-6 array (which would decrease data availability and could impact performance in degraded mode).

Benefits of RAID

RAID provides increased storage capacities, and protects your important data from hard drive failure.

There are multiple benefits of using RAID:

- Reliability and Scalability
- Real-time data recovery with uninterrupted access when a hard drive fails
- System uptime and network availability and protection from loss
- Protection against data loss
- Multiple drives working together increase system performance

A disk system with RAID capability can protect its data and provide on-line, immediate access to its data, despite a single disk failure (some RAID storage systems can withstand two concurrent disk failures). RAID capability also provides for the on-line reconstruction of the contents of a failed disk to a replacement disk.

RAID offers faster hard drive performance and nearly complete data safety. Storage requirements are expanding as file sizes get bigger and rendering needs get more complex. If you handle very large images or work on audio and video files, faster data throughput means enhanced productivity. RAID can be backed up to tape while the system is in use.

RAID Levels

- The most common RAID levels are shown below in a tabular format, complete with pros and cons and uses for the given type of RAID.

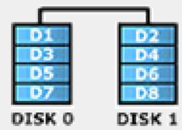
RAID - 0 (STRIPING)

RAID-0 stripes data across multiple disks without any redundant information. Data being written to the array is broken down into blocks or stripes and are distributed sequentially across the member disks of the array. This type of array provides high I/O performance at low inherent cost but provides no redundancy of Fault Tolerance. The data is not stored contiguously on a single drive, and can be accessed in parallel - that is to say the pieces of data are read back from multiple sources nearly simultaneously. Unfortunately, striping reduces the level of data availability since a disk failure will cause the entire array to be inaccessible. RAID-0 was not defined originally but has become a commonly used term.

- Minimum number of drives required: 2

Recommended Applications

- Video Production and Editing
- Image Editing
- Pre-Press Applications
- Any application requiring high bandwidth



Advantages of RAID-0

- High performance
- Very simple design. Easy to implement
- No parity overhead
- No capacity loss - all storage is usable

Disadvantages of RAID-0

- Lack of fault-tolerance
- Failure of a single drive will result in loss of all data on the array
- Should never be used in mission critical environments

RAID – 1 -- (MIRRORING / DUPLEXING)

RAID-1 provides data redundancy. Data written to one disk drive is simultaneously written to another disk drive, called the mirror. If one disk fails, the other disk can be used to run the system and reconstruct the failed disk. Since the disk is mirrored, it does not matter if one of them fails because both disks contain the same data at all times. RAID-1 provides high data availability since two complete copies of all information are maintained. In addition, read performance may be enhanced if the array controller allows simultaneous reads from both members of a mirrored pair. Higher availability will be achieved if both disks in a mirror pair are on separate I/O busses, known as duplexing.

- Minimum number of drives required: 2

Recommended Applications

- Accounting, Payroll, and Financial
- Any application requiring very high availability

Advantages of RAID-1

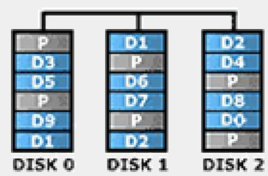
- One Write or two Reads possible per mirrored pair
- Twice the Read transaction rate of single disks, same Write transaction rate as single disks
- Fault tolerant
- Transfer rate per block is equal to that of a single disk
- Easy to recover data in case of drive failure, as no rebuild is necessary in case of a disk failure, just a copy to the replacement disk
- Easy to implement
- Simplest RAID storage subsystem design

Disadvantages of RAID-1

- Inefficient - 100% parity overhead is the highest of all RAID types.
- Becomes very costly as number of disks increase, it requires twice the desired disk space
- The RAID function is done by system software, loading the CPU/Server and degrading throughput at high activity levels. Hardware RAID recommended
- May not support hot swap of failed disk when implemented in "software"

RAID-5 (STRIPING AND PARITY)

RAID-5 stripes data and parity to generate redundancy. However, instead of requiring entirely new disk for parity storage, the parity is distributed through the stripe of the disk array. In RAID-5 both parity and data are striped across a set of separate disks. Next, the new parity is calculated. Finally, the new data and parity are written to separate disks. Data chunks are much larger than the average I/O size, but are still resizable. Disks are able to satisfy requests independently which provides high read performance in a request rate intensive environment. Since parity information is used, a RAID-5 stripe can withstand a single disk failure without losing data or access to data.



- Minimum number of drives required: 3

Recommended Applications

- File and Application servers
- Database servers
- WWW, E-mail, and News servers
- Intranet servers
- Most versatile RAID level
-

Advantages of RAID-5

- High efficiency - highest read data transaction rates, Medium Write data transaction rates
- Good aggregate transfer rate
- Cost effective - only 1 extra disk is required
- Fault tolerant
- Low ratio of ECC (Parity) disks to data disks means high efficiency
- The best choice in multi-user environments which are not write performance sensitive.

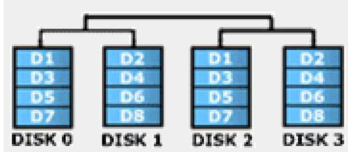
Disadvantages of RAID-5

- Disk failure has a medium impact on throughput
- Most complex controller design
- Difficult to rebuild in the event of a disk failure (as compared to RAID-1)
- Individual block data transfer rate same as single disk

RAID 0+1

RAID-01 is technically a combination of RAID-1 and RAID-0, includes both mirroring and striping, but without parity. RAID-10 is a stripe across a number of mirrored drives, and is implemented as a striped array whose segments are RAID-1 arrays. RAID-10 has the same fault tolerance as RAID-1, as well as the same overhead for fault-tolerance as mirroring alone. Advantages: Very high I/O rates are achieved by striping RAID-1 segments Excellent solution for sites that would normally use RAID-1 Great for Oracle and other databases which need high performance and fault tolerance.

- Minimum number of drives required: 4



Advantages of RAID 0+1

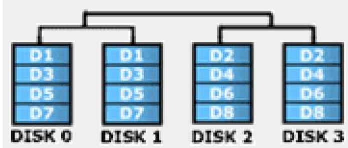
- Fault tolerant
- Very High I/O rates

Disadvantages of RAID 0+1

- Very expensive - Expensive to maintain As with Raid-1 total capacity is equal to half of the total capacity of all disk in the array
- High overhead
- Very limited scalability

RAID-10 A Stripe of Mirrors

RAID-10 is not RAID 0+1. RAID-10 uses RAID-1 mirroring and RAID-0 striping, and has both security and sequential performance. RAID-10 is a striped RAID-0 array whose segments are mirrored RAID-1. It is similar in performance to RAID 0+1, but with better fault tolerance and rebuild performance. It has the same fault tolerance as RAID-1 with the same overhead for fault tolerance as mirroring alone. Typically four plus 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 appropriate for redundant storage of large files, and because parity is not calculated, write operations are very fast.



- Minimum number of drives required: 4

Recommended Applications

- Database server requiring high performance and fault tolerance

Advantages of RAID-10

- High fault tolerance
- High I/O rates achieved by striping RAID-

1 segments
- Faster rebuild performance than RAID 0+1
- Under certain circumstances, RAID-10 array can sustain multiple simultaneous drive failures

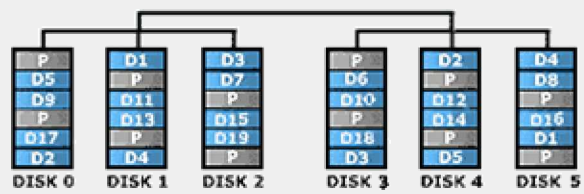
Excellent solution for sites that would have otherwise gone with RAID-1 but need some additional performance boost

Disadvantages of RAID-10

- Very expensive
- High overhead
- All drives must move in parallel to proper track lowering sustained performance
- Very limited scalability at a very high inherent cost

RAID-50 – A Stripe Across a RAID-5 array

RAID-50 is a striped RAID-0 array which is striped across a RAID-5 array. Performance is improved compared to RAID-5 because of the addition of the striped array. Fault tolerance is also improved.



- Minimum number of drives required: 6

Advantages of RAID-50

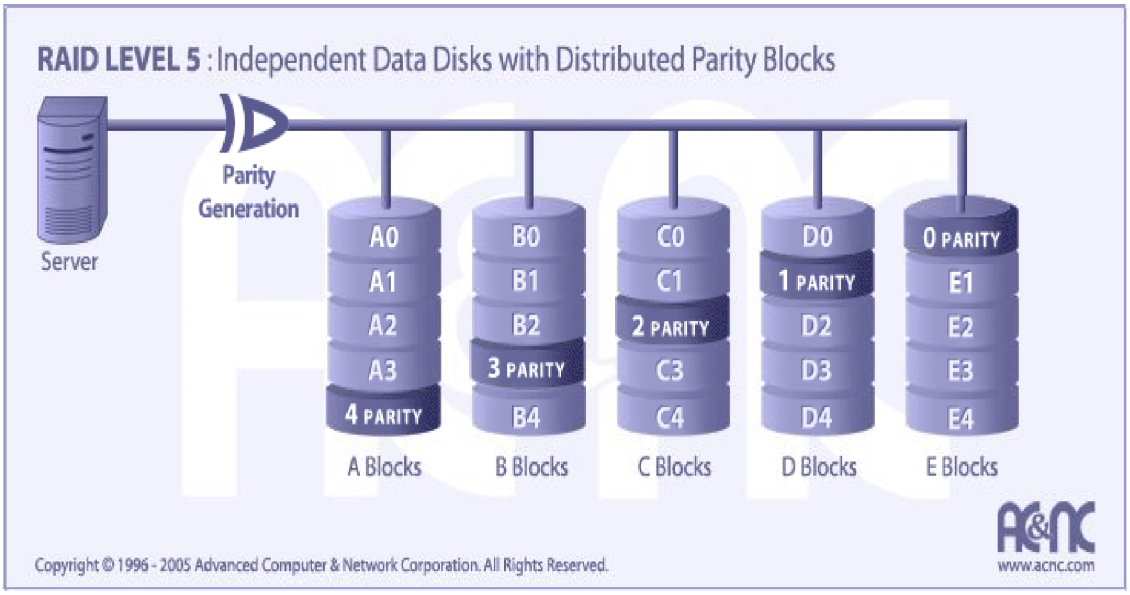
- Higher fault tolerance than RAID-5
- Higher efficiency than RAID-10
- Higher I/O rates

Disadvantages of RAID-50

- Very complex and expensive to implement

More on RAID-5

Each entire data block is written on a data disk; parity for blocks in the same rank is generated on Writes, recorded in a distributed location and checked on Reads.



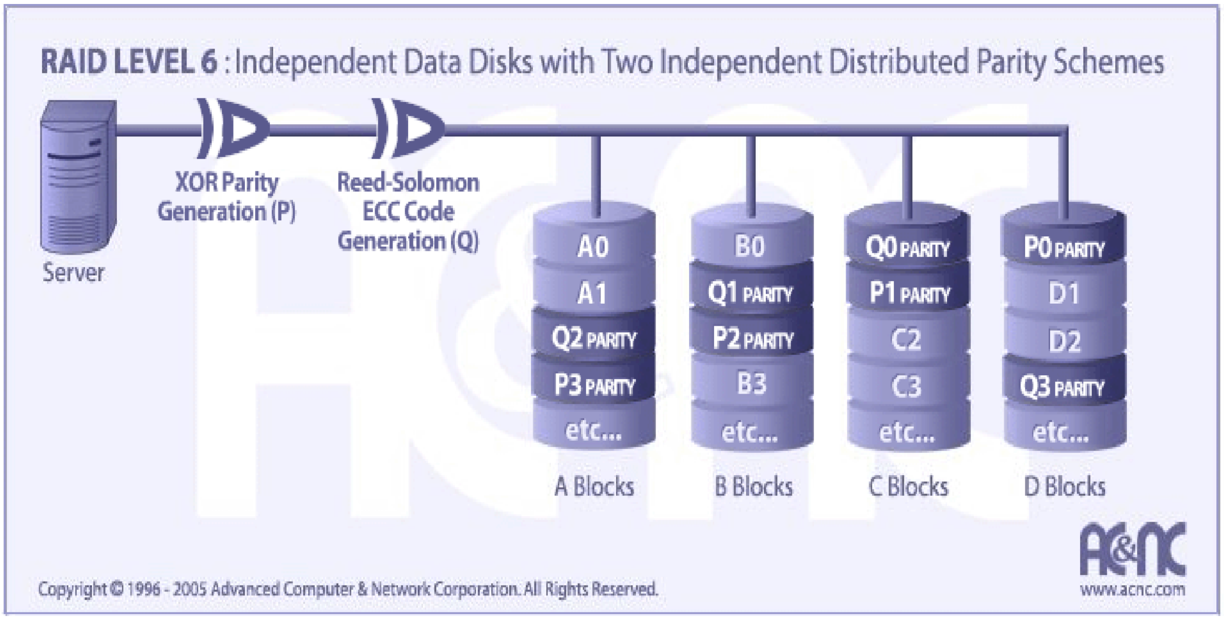
The following table lists advantages and disadvantages of RAID-5.

RAID-5 Advantages	RAID-5 Disadvantages
Highest Read data transaction rate	Disk failure has a medium impact on throughput
Medium Write data transaction rate	Most complex controller design
Low ratio of ECC (Parity) disks to data disks means high efficiency	Difficult to rebuild in the event of a disk failure (as compared to RAID-1
Good aggregate transfer rate	Individual block data transfer rate same as single disk

RAID-6: Dual Parity Stripes

Two independent parity computations must be used in order to provide protection against double disk failure. Two different algorithms are employed to achieve this purpose.

RAID-6 requires a minimum of 4 drives to implement



RAID-6 Characteristics and Advantages

RAID-6 is essentially an extension of RAID-5 which allows for additional fault tolerance by using a second independent distributed parity scheme (dual parity).

Data is striped on a block level across a set of drives, as in RAID-5.

A second set of parity is calculated and written across all the drives.

RAID-6 provides for an extremely high data fault tolerance and can sustain multiple simultaneous drive failures.

RAID-6 protects against multiple bad block failures while non-degraded.

RAID-6 protects against a single bad block failure while operating in a degraded mode.

Raid-6 Disadvantages

More complex controller design.

Controller overhead to compute parity addresses is extremely high.

Write performance can be brought on par with RAID-5 by using a custom ASIC for computing Reed-Solomon parity.

Requires N+2 drives to implement because of dual parity scheme.

Is Raid-5 Going Away and being replaced by RAID-6?

Has RAID-5's time finally come? Is it dead, and if so, why? When you consider RAID-5 against RAID-10, you might wonder why RAID-5 ever won out. Both protect against drive failure and data loss, but RAID-10 is much more straightforward to implement and has much higher performance. RAID-5 requires less drive overhead for basically the same level of data protection, as each RAID-5 requires just one drive's worth of storage to protect all the other drives in the array. Thus, if you had a five drive RAID-5, it would be 80% efficient, whereas an 8-drive RAID-10 is only 50% efficient.

As drives get bigger and cheaper, one would think that companies can afford to just throw away half their storage for the benefits of RAID-10. This might be true for PATA and SATA directly attached drives. But in an enterprise, do we ever have "too much" network storage in the office? In fact, the IT crew is always trying to get more budget for more drives, switch ports, etc. But more components always means more hardware that will break and a bigger facility just to hold all "our storage stuff."

Have you ever been to a computer center that was not crowded, with things stashed in every available place? So dropping from RAID-5's 80-90% efficiency to RAID-10's 50% just won't fly. Bigger drives just mean that we will find more ways to fill them. So it looks like the requirement for RAID-5 efficiency is here to stay.

And along comes RAID-6

RAID-6 is like RAID-5, but it uses two different types of parity stripes to support two concurrent drive failures rather than just one. Would you ever have two drives failing at once? Mean Time Between Failures on currently shipping drives are up over a million hours – that's 114 years! What's the chance of two drives failing at once? Probably slim, but

Murphy's Law and its insidious Corollary still apply. Murphy's corollary assures that not only will a drive break, but it will break at the worst possible time.

For RAID arrays, if a non-failed drive breaks during a rebuild, your storage is truly vulnerable. Again, the chance of this happening is slim, but today's drives are pretty large and during the time it takes to rebuild the array you're vulnerable to data loss.

Why did that drive fail to begin with? Maybe it wasn't just a random drive failure. Maybe it was system related, such as a fan failing and temperatures rising, or noise on the power cables, or flakey cables. When taking environmental failures into account it's common to reduce the second drive's MTBF to 1/10th the value of the first drive.

Now take into account all the systems you've installed or shipped. What's the chance of just one of those systems experiencing a two drive failure? The chance of failure for each individual installation is still relatively low, but the chance that at least one of those installations will lose data can be pretty high.

A second way to get a two drive failure is purely human error. When a drive in a RAID-5 fails, a well-designed system will light a fault LED next to the failed drive. Assuming that the system is in use 24/7, the administrator will remove that failed drive from the live system in order to replace it with a new drive. Hopefully he or she is able to do that without (a) removing the wrong drive, or (b) yanking hard enough on the drive carrier to dislodge an adjacent drive. Of course neither should ever happen, but accidents do happen.

The single biggest reason for using RAID-6 is based on the chance of drive errors during an array rebuild after just a single drive failure. Rebuilding the data on a failed drive requires that all the other data on the other drives be pristine and error free. If there is a single error in a single sector, then the data for the corresponding sector on the replacement drive cannot be reconstructed. Data is lost.

In the drive industry, the measurement of how often this occurs is called the Bit Error Rate (BER). Simple calculations will show that the chance of data loss due to BER is much greater than all the other reasons combined. PATA and SATA drives have historically had much greater bit errors per drive than SCSI and SAS drives, causing some vendors to recommend RAID-6 for SATA drives if they're used for mission critical data.

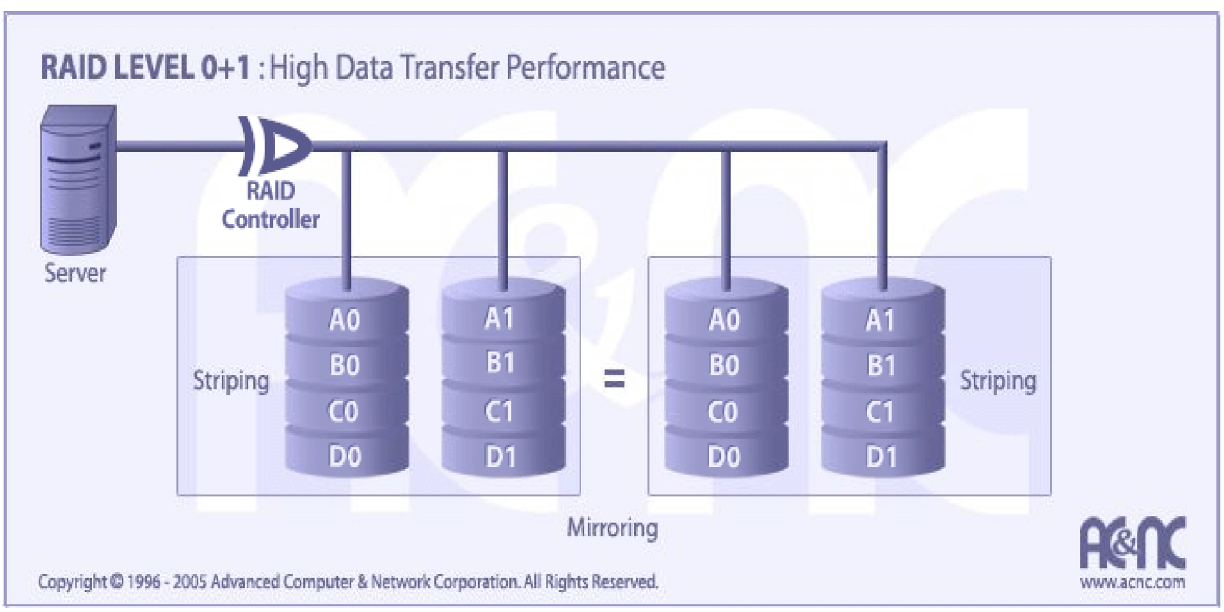
A wise man once said, "*If it sounds too good to be true, it is too good to be true!*" What's RAID-6's downside? In read operations the performance is basically identical to RAID-5 since there is no need to read or manipulate the parity data, assuming that the array contains no failed drives. And on long sequential write operations the overhead of calculating the additional parity is not significant compared to all the other data that is being written. A well designed RAID-6 controller should give 90% of the performance of a similar RAID-5 controller.

Significant degradation may occur on short random writes, which are typical in transaction database updates. Most database administrators needing excellent performance choose to run their databases on RAID-10 arrays. The bottom-line is that in all the access patterns that matter, RAID-6 performance is close enough to RAID-5 performance to make the issue moot.

Most major RAID vendors, including EMC, have started shipping products incorporating RAID-6. Although different vendors use different algorithms, the results are the same – they can stay up, even with two drive failures. Eventually, all the major vendors will support hardware RAID-6. Once everyone supports RAID-6, there really is no need for RAID-5.

The moral of the story? Make sure you use RAID-6.

Raid 0+1 -- A Mirror of Stripe Sets



RAID 0 + 1 Characteristics and Advantages

RAID 0+1 is implemented as a mirrored array whose segments are RAID-0 arrays. RAID 0+1 has the same fault tolerance as RAID-5, and has the same overhead for fault-tolerance as RAID-1, mirroring alone.

RAID 0+1 has high I/O rates are achieved thanks to multiple stripe segments.

RAID 0 + 1 is an excellent solution for sites that need high performance but are not concerned with achieving maximum reliability.

RAID 0 + 1 Disadvantages

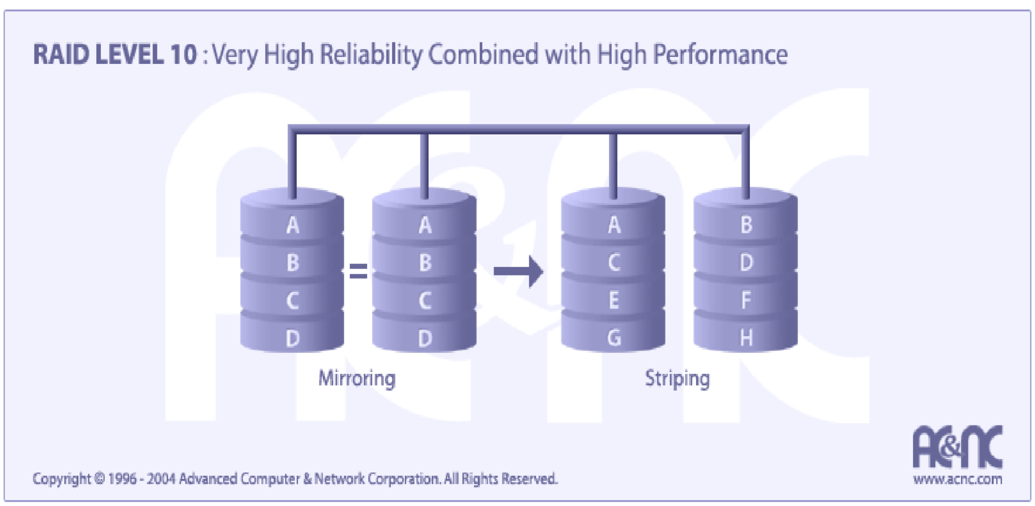
RAID 0+1 is NOT to be confused with RAID-10. A single drive failure will cause the whole array to become, in essence, a RAID-0 array

RAID 0 + 1 is very expensive / with high overhead

All the RAID 0 + 1 drives must move in parallel to properly track, which can lower sustained performance

RAID 0 + 1 has very limited scalability at a very high inherent cost.

Raid-10 Stripe Sets of Mirrored Drives Raid 1 + 0



RAID-10 Characteristics and Advantages

RAID-10 is implemented as a striped array whose segments are RAID-1 arrays and RAID-10 has the same fault tolerance as RAID-1.

RAID-10 has the same overhead for fault-tolerance as mirroring alone.

High I/O rates are achieved by striping RAID-1 segments.

Under certain circumstances, RAID-10 array can sustain multiple simultaneous drive failures.

RAID-10 provides an excellent solution for sites that would have otherwise gone with RAID-1 but need some additional performance boost.

RAID-10 Disadvantages

Very expensive / High overhead

All drives must move in parallel to properly track lowering sustained performance.

Very limited scalability at a very high inherent cost.

RAID-50 (5+0) Block Striping with Distributed Parity

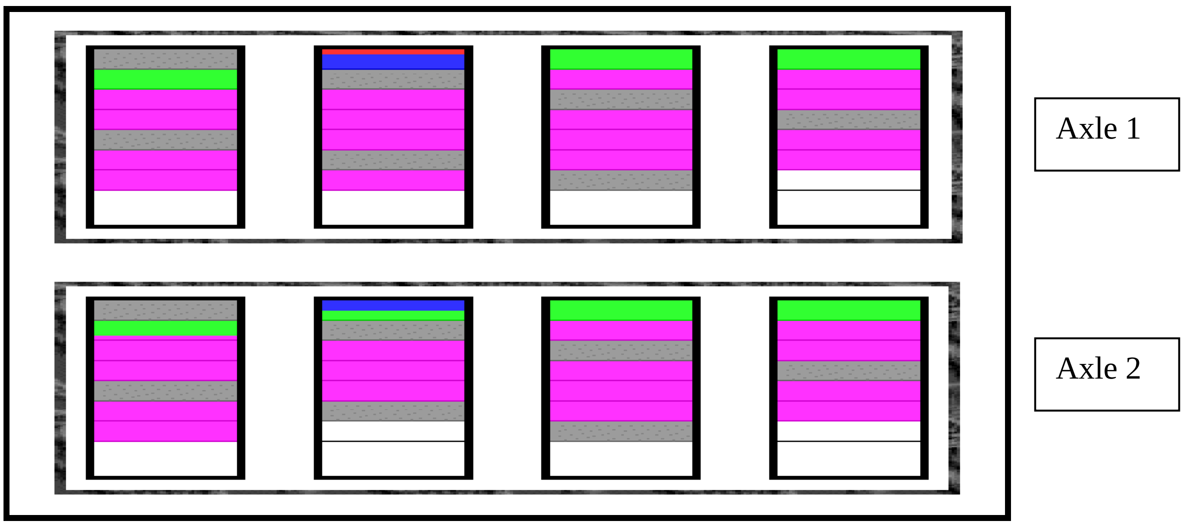
Simply stated, a RAID-50 array is a RAID-0 array on top of a RAID-5 array. Thus, RAID-50 forms large arrays by combining the block striping and parity of RAID-5 with the straight block striping of RAID-0. RAID-50 improves upon the performance of RAID-5 through the addition of RAID-0, particularly during writes. RAID-50 also provides better fault tolerance than the single RAID level does. Because of the improved speed and fault tolerance, RAID-50 is excellent for transactional environments. RAID-50 systems require a high-end hardware controller.

RAID-50 Axles

When you create a RAID-50, you must specify the number of axles. An axle refers to a single RAID-5 array that is striped with other RAID-5 arrays to make RAID-50. The number of drives in the RAID-50 array must be factorable into two integers, one of which must be 2 or higher and the other 3 or higher. We can easily deduce the minimum number of drives in a RAID-50 array to be 2x3 or 6. With 12 drives, you could have a either 2x6 or a 3x4 array. With 16 drives, you could have either an 8x2 or a 4x4 array. There are limitations on the number of axles and drives controllers can support. Some drive configurations and enclosures might yield an unbalanced RAID-50.

In a RAID-5 array, you can keep going with one failed drive. In a RAID-50 array, you can keep going with one failed drive per axle! The diagram below shows a RAID-50 2x4 array with two 4-drive axles. Our diagram will help us understand how a RAID-50 system would store files in its array.

There are four files to be stored in our RAID-50 array, each of varying color and size. First off, our array will use a grey 16KB stripe, so files will be stored in 16KB chunks. The tiny red file is 4 KB in size; the blue file is 20 KB; the green file is 100 KB; and the magenta file is 500 KB. The data will be evenly striped between these two RAID-5 arrays using RAID-0. Then within each RAID-5 array, the data is stored using striping with the 16 KB parity blocks.



How we store the four files: We assume a top/bottom expansion, and that the 4KB red file, the 20KB blue file, 100KB green file, and 500KB magenta file are to be stored on the array, in order, with the small red file first. Since we are using a 16KB stripe, all of the 4KB red file, and 12KB of the second blue file, were sent to Axle 1; the remaining 8KB of the blue file and the first 8KB of the green 100KB file went to Axle 2. Then 16KB of the green file went to Axle 1, the next 16KB went to Axle 2, and so forth, until all the magenta file is stored.

RAID-50 Array Capacity and Storage Efficiency

To discuss capacity and efficiency, we will use the abbreviation NDA as the number of drives in a RAID-50 axle.

The capacity formula is

(Smallest Drive Size) * (NDA - 1) * (# of Axles).

To illustrate, suppose we purchase a 12 bay SAS enclosure for holding 15K RPM, 300 GB, SAS drives. There are two possible RAID-50 configurations with 12 drives, the 2x6 and the 3x4 arrays. For the 2x6 array, the capacity is 300GB * (6-1) * 2 or 300 * 10 or 3.0 TB. For the 3x4 array, the capacity is 300GB * (4-1) * 3 or 2.7TB.

The RAID-50 efficiency formula is (NDA – 1)/ NDA.

The efficiencies for our two 12 drive array configurations are (6–1) / 6 or 84% for the 2x6 array and (4 – 1) / 4 or 75% for the 3x4 array.

Since both the capacity and efficiency are better for the 2x6 array, why would we choose the 3x4 array? The only answer is that the more drives are on an axle, the longer it takes to computer the required parities.

If we purchased a 16 bay SAS enclosure, to hold the same 300 GB, 15 K SAS Drives, we would have two choices, a 2x8 array or a 4x4 array. For the 2x8 array, our capacity would be 300 * 7 * 2, or 4,200 GB or 4.2 TB, and the efficiency would be (8-1)/8 = 7/8 or 87.5%. The 4x4 array would have a capacity of 300 * 3 * 4 = 3.6KB and an efficiency of (4-1)/4 or 75%.

As a final RAID-50 example, suppose we purchased a 24 bay enclosure to hold 24 1GB 7,200 RPM SATA drives. Of course, these spin much slower than the SAS drives, but are much, much less expensive. In November, 2007, a median Web price for the 300 GB, 15K SAS was $800, and $400 for the 1 TB, 7,200 RPM SATA. For a 24 drive array, we could have 3 arrays, a 2x12, a 3x8, or a 4x6 array. For the 2x12, the capacity and efficiencies would be 1 TB * (2-1) * 12 or 12TB and (2-1)/2 = 50%. For the 3x8 array, the capacity is 1 TB * (8-1) * 3 = 21 TB and the efficiency is (8 – 1) / 8 or 87.5%. For the 4x6 array, the capacity is 1TB * 5 * 4 or 20TB, and the efficiency is 80%.

Curiously, the drive, enclosure, and controller costs would be the about the same for either the 12 SAS drives or the 24 SATA drives. With 15K RPM SAS drives, the maximum storage we get is 3.0 TB, but with the 7.2K RPM SATA drives, we get 20 TB for 7.2K SATA.

Looking at these numbers, you might ask, “Why go with SAS drives?” SAS drives sell because of their extra speed and inherent drive reliability, and dual-ported connectors, which give extra drive redundancy. It is well known that hardware costs are only a small portion of the storage dollar.

Estimates on the annual costs to maintain such storage range from 4 to 7 times the initial hardware cost for smaller systems to 30 – 50% of the initial costs for larger systems.

See the two articles

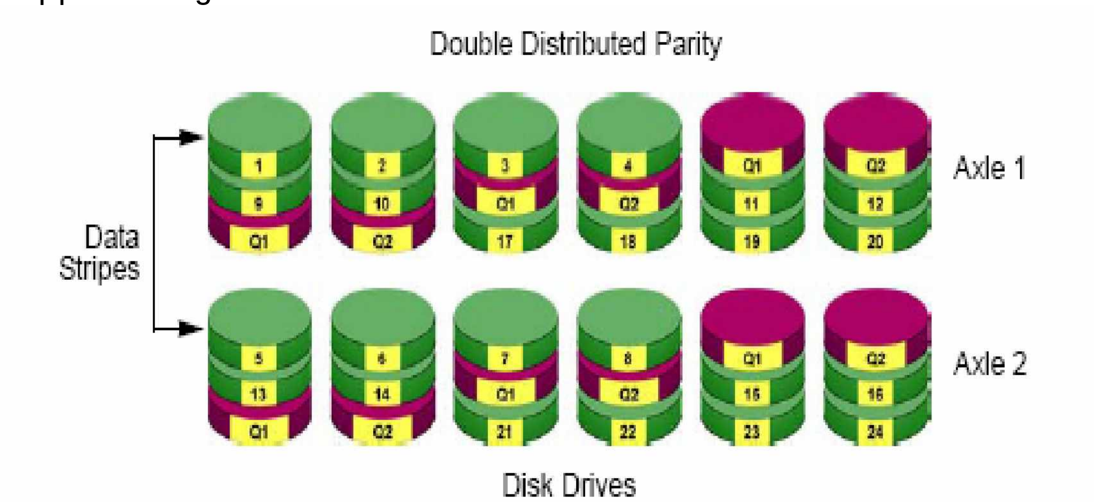
http://tjscott.net/storage/sata.sas.overview.pdf and

http://tjscott.net/storage/tco.analysis.scsi.iscsi.pdf

for more information on these issues.

RAID-60 (6+0) Striping of Double Parity

RAID-60 combines both RAID-6 and RAID-0 features. Data is striped across disks as in RAID-0, and it uses double distributed parity as in RAID-6. RAID-60 provides data reliability, good overall performance and supports larger volume sizes.



RAID-60 Striping of Double Distributed Parity disk arrays

RAID-60 also provides very high reliability because data is still available even if multiple disk drives fail. *When you create a RAID- 60, you must specify the number of axles.* A RAID-60 axle refers to a single RAID-6 array that is striped with other RAID-6 arrays to make RAID-60. An axle can have from four or more physical drives. The greater the number of axles, the greater the number of disk drives that can fail without the RAID-60 array going offline. RAID-60 arrays consist of eight or more physical drives.

Possible RAID-60 Axles

No. of Drives in RAID 60 Array	No. of Axles in RAID 60 Array	No. of Drives per Axle
8	2	4
10	2	5
12	2	6
	3	4
15	3	5
16	2	8
	4	4

RAID-60 Advantages	Raid-60 disadvantages
High Read data transaction rate	High disk overhead – equivalent of two drives used for parity per axle
Medium Write data transaction rate	Slightly lower performance than RAID-5
Good aggregate transfer rate	More complex to setup than RAID-5
Safest RAID Level	

RAID-60 Capacity and Efficiency

The RAID-60 Capacity formula is (Minimum Drive Size in array) * (NDA – 2) * NDA and the Efficiency formula is (NDA-2)/NDA. The "-2" is used instead of the "-1" in RAID-50 formulas to account for the extra parity drive.

As an example, for a 2x6 array of 15K 300GB SAS drives, the capacity would be 300GB * (6-2)/6 * 2 = 2.4TB and the efficiency would be 67%. Both of these are lower than with RAID-50, but being able to lose two drives per axle and keep working is a tremendous extra benefit.

Conclusions on RAID-50 and RAID-60

RAID-50 or RAID-60 is best suited for high capacity volume arrays with greater than 16 disk drive single RAID configurations. Using RAID-50 in these types of configurations allows for one drive failure per one disk drive per axle. RAID-60 allows you to lose up to two disks per axle thus making RAID-50 and RAID-60 the array of choice for safer, more redundant, high capacity Single RAID volumes with greater than 16 disk drives.

As demand for disk storage increases, we will ultimately need bigger arrays.

The following table, from a www.cerberos.com RAID whitepaper, summarizes strengths and weaknesses of all the standard RAID levels.

RAID Level	0	1	5	6	10	50	60
Efficiency	100%	50%	Good (1 drives per set)	Good (2 drives per set)	50%	Reasonably Good	Fair
Fault Tolerance	None	Excellent	Good	Very Good	Excellent	Good	Very Good
Availability	Low	Excellent	Good	Very Good	Excellent	Good	Very Good
Random Read	Good	Fair	Fair-Good	Fair-Good	Fair-Good	Very Good	Very Good
Random Write	Good	Fair	Fair	Fair	Good	Good	Good
Sequential Read	Very Good-Exc.	Fair	Very Good	Very Good	Very Good	Very Good	Very Good