ATTO Technology, Inc.
Corporate Headquarters
155 Crosspoint Parkway
Amherst, NY 14068

Phone: 716-691-1999
Fax: 716-691-9353
www.attotech.com
sales@attotech.com

RAID Overview:
Identifying What RAID Levels Best Meet Customer Needs

Diamond Series RAID Storage Array

I. Introduction

RAID is an acronym for "Redundant Array of Independent *(or Inexpensive)* Disks." It refers to a set of methods and algorithms for combining multiple disk drives as a group in which the attributes of the multiple drives are better than those of the individual disk drives. RAID can be used to improve *data integrity* (risk of losing data due to a defective or failing disk drive), *cost*, or *performance*. The various RAID implementations that are available today offer different tradeoffs between these three factors.

II. Origins of RAID

The concept of RAID was first defined in 1988, when a group of computer scientists at the University of California Berkeley, (David Patterson, Garth Gibson, and Randy Katz) published a paper titled "A Case for Redundant Arrays of Inexpensive Disks (RAID)."

The group observed that computer CPU speed and memory size were growing exponentially, while I/O performance was increasing at a much slower rate. Unless I/O performance could be significantly improved, computer systems would not be able to take full advantage of the rapidly increasing CPU and memory performance gains.

At the time, hard drive manufacturers addressed this issue by designing and building *Single Large Expensive Disks* (SLED). While storage capacities of these disk drives were sufficient for the times, I/O performance was still not keeping up as the inherent mechanical limitations of the hard drives were significantly slower when compared to electronic circuitry.

To overcome these limitations, the UC Berkley scientists proposed that, instead of storing all data on one disk drive (with only one spindle), several small inexpensive disks (with many spindles) be combined to *stripe* the data (split the data across multiple drives), such that Reads or Writes could be done in parallel. To simplify the I/O management, a dedicated controller would be used to facilitate the striping and present these multiple drives to the host computer as one large *logical* drive. They estimated the performance improvements would be an order of magnitude greater than using SLEDs.

The problem with this approach was that the small inexpensive PC disk drives of the time were less reliable than the SLEDs. An artifact of striping data over multiple drives is that if one drive fails, all data on the other drives is rendered unusable. It would be analogous to deleting every 3rd or 4th sentence out of a book, then not knowing what sequence the sentences were written in. To compound this problem, by combining several drives together, the probability of one drive failing increases dramatically.

To overcome this pitfall, the scientists proposed adding extra drives to the RAID group to store redundant information. The thought was that; if one drive failed, another drive within the group would contain the missing information, which could then be used to regenerate the lost information. Since all the information was still available, the end user would never be impacted with down time and the rebuild could be done in the background. If users requested information that had not already been rebuilt, the data could be reconstructed on the fly and the end user still would not know about it.

III. Original RAID Levels

The group outlined six RAID architectures (levels) ranging from "Level 0 RAID" to "Level 5 RAID." These levels provided alternative ways of achieving storage fault tolerance, increased I/O performance and true scalability.

They used three main building blocks in their architectures:

1. **Data Striping** - Data from the host computer are broken up into smaller chunks and distributed to multiple drives within a RAID array. Each drive's storage space is partitioned into stripes. The stripes are interleaved such that the logical storage unit is made up of alternating stripes from each drive. Major benefits are I/O performance gains and the ability to create large logical volumes. Used in RAID 0.
2. **Mirroring** - Data from the host computer are duplicated on a block-to-block basis across two disks. If one disk drive fails, the data remain available on the other disk. Used in RAID levels 1 and 1+ 0.
3. **Parity** - Data from the host computer are written to multiple drives. One or more drives are assigned to store parity information. In the event of a disk failure, parity information is combined with the remaining data to regenerate the missing information. Used in RAID levels 3, 4 and 5.

RAID 0 plus the five original RAID levels developed by the Berkley scientists (along with the RAID group's performance, reliability and cost assumptions) are listed as follows:

RAID Level 0
- Striped Disks
  - Performance: Very Good
  - Reliability: Poor – Less than a single disk drive (Non-Redundant)
  - Cost: Low

RAID Level 1
- Mirrored Disks
  - Performance: Slightly better than a single drive
  - Reliability: Excellent.
  - Cost: High (must purchase 2X disks).

RAID Level 2
- Uses bit interleaving and ECC. (This feature is built into most modern disk drives now)
  - Performance: Same as RAID level 1 for large I/Os
On small I/Os it is very bad; have to read all disks; no parallelism
  - Reliability: Good
  - Cost: Cost is better than mirroring with 20%-to-40% cost overhead

RAID Level 3
- Uses byte interleaving with parity instead of ECC. Parity data are stored on a dedicated drive
  - Performance: Same as RAID level 0 for reads, Writes are slightly slower
  - Reliability: Good
  - Cost: Cost is one additional disk per RAID group

RAID Level 4
- Same as RAID Level 3, but uses sector interleaving instead of bit interleaving
  - Performance: Same as RAID level 0 for Reads, Writes are slightly slower
  - Reliability: Good
  - Cost: Cost is one additional disk per RAID group

RAID: Level 5
- Same as RAID Level 4, but distributes stripe parity across all disks
  - Performance: Writes are slightly faster than RAID 3 and 4, but Reads tend to be considerably slower
  - Reliability: Good
  - Cost: Cost is one additional disk per RAID group

IV. RAID Today

The most widely used RAID levels in use are:

RAID 0: Striping
RAID 1: Mirroring
RAID 10: Mirroring and Striping
RAID (0+1): Striping and Mirroring
RAID 3: Striping with Dedicated Parity Disk
RAID 5: Striping with Distributed Parity

As noted earlier, the different RAID levels offer a variety of performance, data availability, and data integrity levels depending on the specific I/O environment; however, it is important to remember that:

- ***RAID levels are not progressive -*** In other words, increasing the RAID level from 0 to 1 to 2 to 3 etc. does not give progressively better data integrity, performance, or cost. Each RAID level is independent and the numbering is arbitrary (use of the term RAID *level* creates some confusion).

- ***Not all RAID levels are redundant -*** RAID 0 provides no data redundancy; in fact, it is more prone to data loss than individual disk drives, because if any drive fails in a RAID 0 group, all data are lost.

- ***There are no standards for RAID -*** Each vendor has its own implementations and may use different terminology. Some vendors have invented their own RAID terminology (e.g., EMC's RAID-S and Storage Computer's RAID 7). Vendors who claim to implement RAID 3 are actually implementing a modified RAID 3. Combinations such as 10, 0+1, and 53 are all vendor defined. Storage users must closely examine exact RAID implementations from specific vendors.

- ***RAID can be implemented in various places in the computer system -*** The storage devices, the Host Bus Adapter (HBA) and the host operating system can all implement RAID. It is possible to use a combination of these; for example, RAID 0 (striping) in the storage array combined with RAID 1 (mirroring) in the operating system. Each location has benefits and shortcomings, which need to be understood by computer system architects.

- ***Physical vs. Logical drive numbering -*** Physical numbering refers to the physical components in the storage array. Logical (or virtual) numbering refers to the "disks" or "volumes" that the host operating system "sees" in the storage device. These two can be very confusing to new storage users.

- ***Logical disks do not always map 1-to-1 with physical disks -*** In RAID, several physical disk drives (or portions of several physical drives) can be grouped into a logical disk or Logical Unit (*LUN*). Each LUN can be broken into logical blocks of 512 bytes each, numbered 0 through "n" (the Logical Block Number or *LBN*). For example, a 100GB LUN has approx. 200,000,000 logical blocks.

- ***Logical volumes are very similar to logical drives -*** A logical volume is composed of one or several logical drives. The member logical drives can be the same RAID level or different RAID

levels. The logical volume can be divided into partitions. During operation, the host sees a non-partitioned logical volume or a part of a partitioned logical volume as one single physical drive.

V. RAID Summary

RAID can be a powerful tool in a storage environment. Using a RAID storage subsystem has the following advantages:

- Provides fault-tolerance by mirroring or parity operation
- Increases disk access speed by breaking data into several blocks when Reading/Writing to several drives in parallel
- Simplifies management by weaving multiple drives together to form a large volume or groups of volumes

Today, the major RAID levels available offer the following characteristics:

A. RAID 0: Striped Disk Array

Raid 0 is not a fault tolerant RAID solution, if one drive fails, all data within the entire array is lost. It is used where raw speed is the only (or major) objective. It provides the highest storage efficiency of all array types.

Pros

- Improved I/O performance
- Most capacity-efficient RAID level
- Ability to create large logical volumes

Cons

- RAID 0 does not utilize disk space for redundancy
- If one disk fails, all data within the stripe set is lost

Configuration

- RAID 0 arrays are made by grouping two or more physical disks together to create a virtual disk and making this virtual disk appear as one physical disk to the host. Each physical drive's storage space is partitioned into stripes. The stripes are then interleaved so that the virtual disk is made up of alternating *stripes* from each drive.
- To increase performance, RAID 0 writes block level data across all available stripes in the RAID 0 group, enabling parallel disk I/O, that optimizes I/O performance.
- Ideally, the size of the stripe is large enough to fit one record. The record is broken into smaller sizes and evenly distributed across all drives in the stripe group.

Uses of RAID 0: RAID 0 should be used with applications that require the highest level of performance and use non-critical or temporary data, such as:

- Full-motion video editing applications
- Prepress editing applications
- Scratch files for CAD
- Any application where the original content is backed up and can be easily restored. The time saved in doing normal data processing work with RAID 0 more than makes up for the time lost in infrequent disk crash events

B. RAID 1: Mirrored Disk Array

RAID 1 provides complete protection and is used in applications containing mission-critical data. It uses paired disks, where one physical disk is partnered with a second physical disk. Each physical disk contains the same exact data to form a single virtual drive.

Complete data protection is achieved by simultaneously Writing two exact block-level copies of data to each disk in a mirrored pair. There is no striping. Read performance is improved because either disk can be read at the same time. Write performance is the same as for single-disk storage. RAID-1 provides the best performance and the best fault-tolerance in a multi-user system.

With RAID 1, the host will see what it believes to be a single physical disk of a specific size. (The host does not know or care about the mirrored pair) The RAID controller manages where the data are written and read. This allows one disk to fail without the host ever knowing, providing time for service personnel to replace the failed drive and initiate a rebuild.

Pros:

- Highest level of protection - Mirroring provides 100% duplication of data
- Read performance is faster than a single disk (if the array controller is capable of performing simultaneous reads from both devices of a mirrored pair)
- Delivers the best performance of any redundant array type during a rebuild
    - No re-construction of data is needed. If a disk fails, copying on a block-by-block basis to a new disk is all that is required.
    - No performance hit when a disk fails; storage appears to function normally to outside world.
- The only choice for fault tolerance if only two drives are used

Cons

- RAID 1 writes the information twice because of this, there is a minor performance penalty when compared to writing to a single disk
- I/O performance in a mixed Read-Write environment is essentially no better than the performance of a single-disk storage system
- Requires two disks for 100% redundancy, doubling the cost

Uses of RAID 1:
RAID 1 provides the most complete protection; however, it also requires duplication of physical disks. In the past, this RAID level was used exclusively in smaller mission-critical networks to keep costs down. As the cost of storage arrays declines, many system architects are reconsidering the use of RAID 1 in larger applications. Typically, these applications involve mostly Read-only operations or light Read-Write operations. An example of a typical RAID 1 implementation is a data entry network. It is recommended for applications where:

- Data availability is very important
- Speed of Read access is very important
- Read activity is heavy
- Logging or record keeping are needing

C. RAID 10: Mirroring and Striping

RAID 10 consists of multiple sets of mirrored drives. These mirrored drives are then striped together to create the final virtual drive. The result is an extremely scalable mirror array, capable of performing reads and Writes significantly faster (since the disk operations are spread over more drive heads).

Pros:

- Very high reliability
    - Because there are multiple mirror sets, this configuration can actually handle multiple disk failures and still survive (*with one exception)
- Provides highest performance with data protection
- By striping multiple mirror sets, RAID 10 can create larger virtual drives. The host computer will see what it believes to be a single physical disk of a specific size
- Can be tuned for either a request-rate- intensive or transfer-rate-intensive environment

*Disk failures occurring within the same mirror set are the exception, which is extremely rare

Cons:

- Like RAID 1, RAID 10 Writes the information twice because of this, there is a minor performance penalty when compared to writing to a single disk
- I/O performance in a mixed Read-Write environment is essentially no better than the performance of a single-disk storage system
- Requires an additional disk to make up each mirror set

Uses of RAID 10:
Applications where high performance and reliability are paramount are ideal for RAID 10. Examples would be **on-line transaction processing environments** and **financial transaction processing environments**. It is recommended for applications where:

- Data availability is critically important
- Overall performance is very important

D. RAID (0+1): Striping and Mirroring

Not to be confused with RAID 10 (they are very different). Raid 0+1 flips the order of RAID 10. Drives are first striped, and then these drives are mirrored. Typically, two or more disks are striped to create one segment and an equal number of drives are striped to form an additional segment. These two striped segments are then mirrored to create the final virtual drive.

Pros

- High I/O performance
- Ability to create large logical volumes

Cons

- Reliability is less than RAID 1 and 10. If one disk fails, you essentially have a RAID 0 configuration. Due to the multiple disks that make up the RAID 0 segment, the probability of a disk failure is greater
- Requires duplicate drives. Capacity of physical drive is half

Uses of RAID 0+1:
Applications that require high performance, but are not overly concerned with achieving maximum reliability.

E. RAID 3: Striping with Dedicated Parity Disk

RAID 3 is a fault-tolerant version of RAID 1 (Striping). Fault tolerance is achieved by adding an extra disk to the array and dedicating it to storing parity information. Parity information is generated and written during Write operations and checked on Reads. It requires a minimum of three drives and provides data protection.

In the event of a disk failure, data recovery is accomplished by calculating the exclusive OR (XOR) of the information recorded on the other drives. Since an I/O operation addresses all drives at the same time, RAID 3 cannot overlap I/O. For this reason, RAID 3 is best for single-user systems with long record applications.

Pros

- Good data protection
- Good Write performance
- Good Read performance
- The amount of useable space is the number of physical drives in the array minus 1

Cons

- A single disk failure reduces the array to RAID 0
- Performance is impacted when degraded
- Poor performance with small data transfers
- Limited to single-user environments

Uses of RAID 3:
This version of RAID is best suited for:

- Single-user, single-tasking environments with large data transfers
- Heavy Write applications
- Environments where large volumes of data are stored

F. RAID 5: Striping and Parity

Raid 5 is similar to RAID 3, but the parity is not stored on one dedicated drive. Instead, parity information is interspersed across the drive array. RAID 5 requires a minimum of 3 drives. One drive can fail without affecting the availability of data. In the event of a failure, the controller regenerates the lost data of the failed drive from the other surviving drives.

By distributing parity across the array's member disks, RAID Level 5 reduces (but does not eliminate) the Write bottleneck. The result is asymmetrical performance, with Reads substantially outperforming Writes. To reduce or eliminate this intrinsic asymmetry, RAID level 5 is often augmented with techniques such as caching and parallel multiprocessors.

Pros:

- Best suited for heavy Read applications
- The amount of useable space is the number of physical drives in the virtual drive minus 1

Cons

- A single disk failure reduces the array to RAID 0
- Performance is slower than RAID 1 when rebuilding
- Write performance is slower than Read (Write penalty)
- Block transfer rate is equal to single-disk rate

Uses of RAID 5:

RAID 5 is a general-purpose RAID storage solution. It is recommended for applications where:

- Data availability is important
- Large volumes of data are stored
- Multi tasking applications are using I/O transfers of different sizes
- Good Read and moderate Write performance is important

G. Comparing RAID configurations

Each of the described RAID levels offers different cost and performance characteristics. The following table compares the cost, data availability, and I/O performance of the commonly known RAID levels. I/O performance is shown both in terms of large I/O requests, or relative ability to move data, and random I/O request rate, or relative ability to satisfy I/O requests, because each RAID level has inherently different performance characteristics relative to these two metrics.

RAID Level	Name	Description	Disk Costs	Data Availability	Large Write Data -transfer Speed	Large Read Data-transfer Speed[1]	Random Write Request Rate	Random Read Request Rate
0	**Data Striping**	Data distributed across the disks in the array. No parity data	*Lowest 1x*	Lower than single-disk	Very high	Very high	Very high	Very high
1	**Mirroring**	Data is duplicated on 2 separate disks. Backup data is on duplicate disk	*High 2x*	Higher than RAID Level 3 and 5	Slightly lower than single disk	Higher than single disk (up to 2x)	Similar to single disk	Up to 2x single disk
3	**RAID 3, Parallel Transfer Disks with Parity**	Each user data block distributed across all data disks. Parity check data stored on one disk.	*x+ 1*	Much higher than single disk; comparable to RAID 5	Highest of all listed types	Highest of all listed types	Approximately 2x single disk	Approximately 2x single disk
5	**RAID 5**	Independent disks. User data distributed as with striping; Parity check data distributed across disks.	*x+ 1*	Much higher than single disk; comparable to RAID 3	Lower than striping	Moderately higher than striping	Much lower than single disk	Moderately higher than striping
10	**Mirroring Stripes**	User data are mirrored across separate pairs of striped disks.	*2x*	Highest	Higher than single disk	Very high	Higher than single Disk	Very high
0+ 1	**Striped Mirrors**	Data striped across separate pairs of mirrored disks.	*2x*	Higher than RAID Level 3 and 5	Higher than single disk	Very high	Higher than single disk	Very high

[1] The data transfer capacity and I/ O request rate columns reflect only I/ O performance inherent to the RAID model, and do not include the effect of other features, such as caching.

H. RAID Set-up Considerations

Setting up a fault tolerant RAID array involves trading off economy for MTDL (Mean Time to Data Loss). MTDL is the probable time to failure for any component that makes data inaccessible.

If your data is backed up and performance and cost are your primary concern, then RAID 0 is the logical choice.

If you determine that you need data protection, then you have two choices to protect your data: Parity or Mirrored arrays.

The “cost” of data protection for a Parity RAID array is the equivalent of one disk per RAID group. At first glance, to select Parity RAID over Mirrored RAID would seem the obvious choice. However, you must take into consideration:

- When a drive fails in a parity RAID array (RAID 3 or 5), the array becomes a RAID 0-stripe group. If a second drive failure occurs before the array completes a rebuild, then you loose all data within the array.

If you are comfortable with this, then Parity RAID is probably the most economical solution for you.

It would also seem to be very cost effective to build parity arrays with several disks. However, consider the following when configuring your parity array:

- More disks in a parity RAID array affects write performance adversely
- More disks in any RAID array increases the probability of a drive failure
- By creating a parity array with several disks, the capacity of the array skyrockets, dramatically increasing resynchronization time after a disk failure. This has a major impact on array performance and forces the array to run "unprotected" for an extended period

If you determine that the performance and/or protection limitations of parity RAID are too great, then a Mirrored (RAID 1) array or a dual-level array (such as RAID 10) should be your choice.

VI. Summary

Selecting the proper RAID level and setup for a disk storage array is the key to properly balancing system costs and performance needs. This primer and its overview are an excellent starting point to properly select your RAID storage activities. Should you have any further questions, the professionals at ATTO Technology are ready to assist you in selecting a RAID system and setup to best meet your needs and goals.

About ATTO Technology, Inc.:

ATTO Technology, Inc., Amherst, New York, is a global leader in storage and storage-infrastructure solutions for direct- , network- and fabric-attached environments. Our lines of Host Adapters, Fibre Channel Bridges and Hubs, iSCSI Bridges SAN software, and RAID Storage Arrays improve data availability, productivity and total cost-of-ownership.

The technology behind the Diamond Storage Array stems from over 15 years of research and design by ATTO. The ATTO Diamond Storage Array provides up to 7.2 TB of high-performance, Enterprise-class RAID-protected disk storage in a 3U 19" rack form factor at a very aggressive price point. The Diamond line consists of four models designed to meet the needs of Digital Video, Backup, Near-line and Imaging/Fixed Content markets.