

Save

An Introduction to RAID 6

ULTAMUS™ RAID

Overland
STORAGE®
Simply Protected

The highly attractive cost per GB of SATA storage capacity is causing RAID products based on the technology to grow in popularity. SATA RAID is now being used to provide storage capacity for a broad range of applications. 750 GB SATA drives in RAID configurations are increasingly being deployed in business-critical applications.

SATA RAID delivers on a sound value proposition: inexpensive storage capacity with "reasonable" performance coupled with RAID data protection. While SATA drives can be used in any RAID level, the most typically encountered configuration for general purpose storage is RAID 5. RAID 5 provides a reasonable balance among the important RAID product dimensions of performance, cost, capacity and redundancy.

But RAID 5 using SATA drives has an increasingly noticeable "Achilles' heel" that becomes evident when a catastrophic disk drive failure occurs. A catastrophic disk drive failure is typically caused by a motor failure or a head crash. When a drive suffers a catastrophic failure the disk is rendered physically inoperable.

How Does RAID 5 Handle a Drive Failure?

Drive failures occur occasionally and the mechanism to deal with this event is well known and well understood. Although RAID 5 can protect against the catastrophic failure of one disk drive as shown in Figure 1, it cannot protect against a second catastrophic failure. While the RAID 5 set loses a single drive, it loses fault tolerance and operates in a "degraded" state until the RAID set is rebuilt. A replacement drive or the use of a hot spare allows the array to begin "rebuilding."

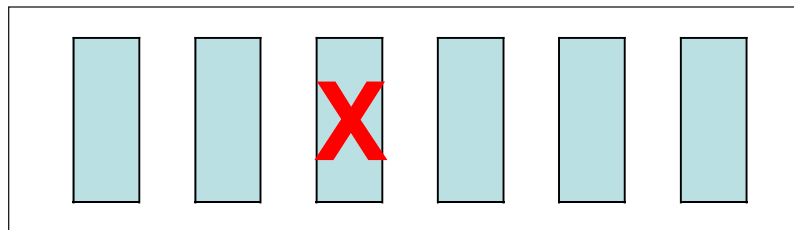


Figure 1. RAID 5 array with failed disk drive. No longer fault tolerant.

The problem with RAID 5 is that if a second disk drive fails while the array is in either a degraded or rebuilding state as shown in Figure 2, the array fails and there is data loss (i.e. the system is no longer capable of recovering the data).

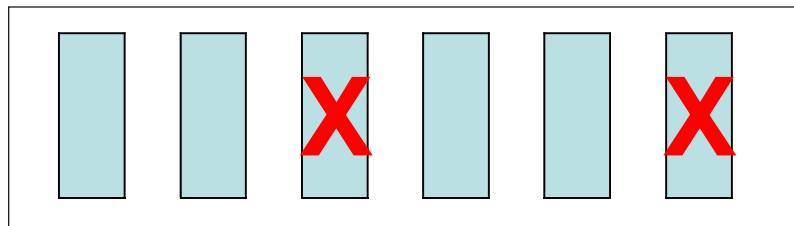


Figure 2. RAID 5 array with two failed disk drives. Data is lost and must be recovered from backups.

As SATA drives increase in capacity, the RAID 5 rebuild time—and hence the period of data vulnerability—increases in length. As an example, 12 750GB drives typically take several hours to rebuild. To overcome the possibility of dual drive failures, storage architects are turning to RAID 6 solutions that can protect critical data even with the loss of two disk drives.

What is RAID 6?

RAID 6, first described by RAID inventors Gibson, Patterson, Katz, Chan and Lee in 1993 extended the capabilities of RAID 5 into tolerance for dual-drive faults. Titled "P+Q Redundancy Using Reed-Solomon" the paper described RAID 6 as an improvement of the fault tolerance of RAID 5.

It is RAID 6's ability to handle multiple drive failures that makes it the conservative IT architect's RAID level of choice. The extra parity block is what enables RAID 6 to protect against catastrophic data loss because of its ability to handle two simultaneous drive failures as shown in Figure 3. And as arrays continue to expand, with more and higher capacity disk drives participating in a RAID set, the need for RAID 6 becomes increasingly apparent.

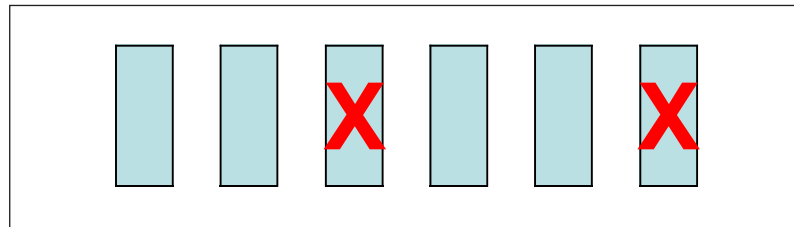


Figure 3. RAID 6 array with two failed disk drives. There is no data loss.

A Review of RAID Levels

RAID 6 extends the original RAID levels, giving it a familiar feel. But the RAID model can be a confusing one. RAID 0 stripes data across all the drives, but does not offer redundancy. RAID 1 mirrors data onto another disk drive, so one disk's data is identical to the others. RAID 2 is obsolete because it's based on Error Correcting Codes, which are now found on all disk drives. RAID 3 introduces the parity block, a checksum for previously written blocks. RAID 3 stripes at the byte level and writes the parity blocks to a specific disk in the array. RAID 4 is similar to RAID 3, but parity is written at the block level instead of the byte level. RAID 5 borrows from RAID 3 and the concept of parity blocks, but spreads the parity blocks around each drive in the array.

Some IT architects argue that because of the reliability of SATA drives, the need for dual drive failure protection is diminished. After all, if the disk drives in a RAID set each have MTBFs in excess of 1,000,000 hours, then aren't the odds of two drives failing at nearly the same time pretty low?

Actual experience, however, debunks the myth of independent disk drive MTBFs. What we observe is that multiple SATA disk drive failures are frequently linked to some common event like a power supply spike or a "hot spot" in the enclosure—sometimes caused by something as simple as a temporary loss of airflow. In addition, many dual drive failures are not really the result of a failing component, but are the result of the accidental removal of the wrong drive from an array that has lost fault tolerance.

RAID 6 is a close relative of RAID 5, with an additional parity block inserted. The extra parity block is different from the first in the way it's generated and calculated. Generating parity blocks can slow down data transfer speeds, and because RAID 6 performs this process twice, it may write data more slowly than RAID 5.

RAID 6 not only writes more data by writing the stored data along with the parity blocks, but it does so twice. Reading from the disk isn't so much a problem because the parity blocks are skipped, meaning data can be delivered quickly.

In all cases except RAID 0, RAID reduces effective or "usable" disk capacity. In RAID 1, if 10 drives of 100GB capacity are grouped, the true capacity of the array will be 500GB. That's because all data is duplicated, essentially reducing the available capacity by half. Similarly with RAID 5 and 6, 10 to 20 percent of the total array is used for parity information, depending on the number of disks in the array.

Although RAID storage is purchased for capacity and performance reasons, redundancy is the most important RAID product characteristic. Redundancy is the reason why most RAID products are purchased. RAID 6 extends redundancy. As arrays grow, the possibility of concurrent drive failure—along with concerns over lost or corrupted data—grows as well.

Cost Considerations

Because the similarities between RAID 5 and 6 are so strong, cost is an important differentiator. RAID 6 requires an additional disk drive. But are the prices of disks really a concern? Certainly for smaller systems the answer is no, but for arrays containing hundreds of drives, the costs add up. Thus RAID 6 makes the most sense in applications that can justify the additional cost. Examples include business and mission-critical, and financial- and health care-related application storage.

Pros and Cons of RAID 6

RAID 6 is conceptually very much like RAID 5 with one key feature addition: the ability to tolerate more than one drive failure without data access interruption or even worse—data loss. This is its chief advantage over RAID 5, and it's a big one.

There are two downsides to this extra protection however. For example, with an equal number of disks in each array, a RAID 6 array will cost more per usable gigabyte than a RAID 5 array because the RAID 6 array will use two of the disks for parity data, where the RAID 5 array will use only one disk. When a RAID 5 set and a RAID 6 set both use the same number of physical disks, RAID 6 usable capacity will be reduced by the capacity of one disk, thus increasing the cost per gigabyte for usable storage.

A second RAID 6 downside is lower random write performance due to the number and frequency of dual-parity calculations required.

Although RAID 5 allows one array drive to fail without impairing data integrity, as soon as a drive does fail, the array is without protection. Stored data is at risk until the array is rebuilt—and a second drive could potentially fail with catastrophic results—at any time during the rebuild.

RAID 6 Advantages

Data is striped on a block level across a set of drives in a fashion similar to RAID 5 but a second set of parity is also calculated and written across all the drives. When a single disk suffers a total failure, the array can still tolerate either another complete failure or a bad block before entering the same state of vulnerability from which a degraded RAID 5 array would suffer. This is possible because there are two redundant disks as opposed to just one. So with the loss of one disk, the array can proceed through a rebuild to a spare disk, and still be protected against data loss by the second redundant disk.

RAID 6 Disadvantages

A fast and sophisticated RAID controller is needed to implement and compensate for the overhead of dual Reed-Solomon parity computations. RAID 6 also requires N+2 drives to implement because of the two-dimensional parity scheme. Finally, RAID 6 rebuilds can affect performance.

Recommended RAID 6 Applications

RAID 6 is best suited when multiple users need a protected storage solution for mission critical applications, including:

- File and application servers
- Database servers
- E-mail, news and web servers
- Intranet servers

Real-Life RAID reliability

In the Storage Advisers Weblog (<http://storageadvisors.adaptec.com/2005/11/02/actual-reliability-calculations-for-raid/>) Tom Treadway, Chief Technology Officer (CTO) for the Adaptec storage software group presents a set of calculations concluding that RAID 6 can deliver a greater than 100X improvement in RAID reliability over RAID 5.

During Intel Developer Forum Spring 2005, Intel presented a study showing the probability of failure on a RAID 5 system in three different configurations, supporting the conclusion that the probability of data loss during a data recovery operation increases with both the number of disk drives in the system and the capacity of each of those disk drives. The article is available here: <http://www.hardwaresecrets.com/article/314/2>

Sample configurations displaying RAID 5 failure probability:

- Configuration 1: RAID 5 with five enterprise-class 30 GB (120 GB total). Frequency on which data recovery operations is necessary: one every 23 years. Probability of a system failure during the data recovery operation: 0.12% (i.e. one error every 834 data recovery operations).
- Configuration 2: RAID 5 with five desktop-class 300 GB hard disk drives (1.2 TB total). Frequency on which data recovery operations is necessary: one every 2.3 years. Probability of a system failure during the data recovery operation: 11% (i.e. one error every nine data recovery operations).
- Configuration 3: RAID 5 with 50 desktop-class 300 GB hard disk drives (15 TB total). Frequency on which data recovery operations is necessary: one every 3 months. Probability of a system failure during the data recovery operation: 70% (i.e. one error every two data recovery operations).

RAID 6 Overview

Common Name: RAID 6.

Technique Used: Block-level striping with dual distributed parity.

Description: RAID 6 stripes blocks of data and parity across an array of drives like RAID 5, except that it calculates two sets of parity information for each parcel of data. The goal of this duplication is to improve fault tolerance. RAID 6 can handle the failure of any two drives in the array while other single RAID levels can handle at most one fault. The write performance of RAID 6 is typically less than that of RAID 5 due to the added overhead of more parity calculations, but may be slightly faster in random reads because data is spread over an additional disk. As with other RAID levels, performance can be adjusted through the use of different stripe sizes.

Disk Drive Requirements: A minimum of four disk drives is required. The maximum number of drives is dictated by the RAID controller. Disk drives should be of identical size and type.

Array Capacity Calculation: (Size of Smallest Drive) * (Number of Drives - 2).

Storage Efficiency Calculation: If all drives are the same size, $(\text{Number of Drives} - 2) / \text{Number of Drives}$.

Fault Tolerance: Very good to excellent. RAID 6 can tolerate the simultaneous loss of any two drives in the array.

Availability: Excellent.

Degradation and Rebuilding: Due to the complexity of dual distributed parity, degradation can be substantial after a failure and during rebuilding. Dual redundancy may allow rebuilding to be delayed to avoid a performance hit.

Random Read Performance: Very good to excellent; generally better for larger stripe sizes.

Random Write Performance: Poor, due to dual parity overhead and complexity.

Sequential Read Performance: Good to very good; generally better for smaller stripe sizes.

Sequential Write Performance: Fair.

Cost: Higher.

Special Considerations: Requires special implementation; not yet widely available.

Recommended Uses: RAID 6 is well suited to the same applications as RAID 5, but in situations where additional fault tolerance is required.

Summary

RAID 6 is conceptually very much like RAID 5 with one additional key feature: the ability to tolerate more than one drive failure without data access interruption or even worse—data loss. Overland Storage offers RAID 6 in the ULTAMUS RAID family of appliances. ULTAMUS RAID offers industry-leading features such as:

Technology	Benefits	Overland Leadership
RAID 6 Hardware Accelerated	<ul style="list-style-type: none"> Higher Availability 	<ul style="list-style-type: none"> Better for mission-critical application availability. Overland's hardware accelerated RAID 6 allows storage to remain online in the event of two drives failing simultaneously.
SAS Drives	<ul style="list-style-type: none"> Higher Performance Lower Costs Investment Protection Higher Availability 	<ul style="list-style-type: none"> Ideal for email, web serving, database and other performance-intensive applications provides the performance and reliability of Fibre Channel drives at 30% lower cost SAS and SATA II drives can be combined in the same RAID chassis to support applications with vastly different capacity and performance requirements.
Enterprise SATA II Drives	<ul style="list-style-type: none"> Higher Availability Higher Performance Better Price/TB Investment Protection 	<ul style="list-style-type: none"> Better performance for D2D backup and media Built for 7 x 24 operating environments 1.4 million hour MTBF at 100% duty cycle—more than double the reliability of desktop SATA II drives Directed Offline Scan monitors every write and performance diagnostic tests when drives are idle, enhancing data availability Perpendicular Recording reduces the number of moving parts, for higher availability
SAS expansion & scalability	<ul style="list-style-type: none"> Grows with your application requirements 	<ul style="list-style-type: none"> Effortlessly scales up to 60 drives in 12-drive/2U increments No performance penalty
4Gb Fibre Channel	<ul style="list-style-type: none"> Higher Performance Investment Protection 	<ul style="list-style-type: none"> Better performance for all applications. 4Gb FC improves IOP intensive applications with reduced latency and improves streaming applications with higher data rates
Snapshots	<ul style="list-style-type: none"> Disaster Recovery Data Protection 	<ul style="list-style-type: none"> Snapshot technology offers the best value in data protection by creating low-impact copies of data that support point-in-time recovery
Pricing	<ul style="list-style-type: none"> Low Acquisition Cost 	<ul style="list-style-type: none"> Excellent price/TB for RAID storage Integrated RAID ASIC technology reduces controller costs and an optimized supply chain reduces platform costs.

Save

Worldwide Headquarters

4820 Overland Avenue
San Diego, CA 92123 USA
TEL 1-800-729-8725
TEL 1-858-571-5555
FAX 1-858-571-3664
EMAIL sales@overlandstorage.com

United Kingdom (EMEA Office)

Overland House
Ashville Way
Wokingham, Berkshire
RG41 2PL England
TEL +44 (0) 118-9898000
FAX +44 (0) 118-9891897
EMAIL europe@overlandstorage.com

France Office

Overland Storage
126 rue Gallieni
92643 Boulogne Cedex France
TEL +33 (0) 1 55 19 23 93
FAX +33 (0) 1 55 19 25 02
EMAIL europe@overlandstorage.com

Germany Office

Humboldtstr. 12
85609 Dornach Germany
TEL +49-89-94490-214
FAX +49-89-94490-414
EMAIL europe@overlandstorage.com

Asia Pacific Office

Level 44, Suntec Tower 3
8 Temasek Boulevard
Singapore, 038988
TEL +65.6866.3847
FAX +65.6866.3838
EMAIL asia@overlandstorage.com

www.overlandstorage.com

Overland
STORAGE®
Simply Protected