IBM TotalStorage Network Attached Storage
October 2001

RAID technology and IBM TotalStorage™ NAS products

By Janet Anglin and Chris Durham
Storage Networking Architecture, SSG

Contents

2 RAID 1, RAID 3, RAID 4
3 RAID 5, RAID 5E
4 RAID Summary, Hardware vs. Software RAID
6 IBM Total Storage NAS products
7 Summary

Highlights

Data striping interleaves blocks of data across the disks.

RAID 0 enables you to create a large logical disk drive and provides performance acceleration.

Introduction

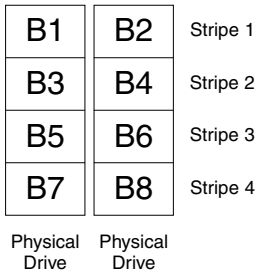
RAID technology plays a key role in IBM TotalStorage™ NAS products. This paper explains the primary RAID levels in the industry, along with their strengths and weaknesses. It also provides a comparison between the RAID technology used in IBM TotalStorage NAS products and the software RAID options.

RAID overview

Redundant Array of Independent Disks (RAID) technology was developed to increase the performance and reliability of data storage by distributing data across multiple, inexpensive drives. This technology has evolved over time and has proven itself to be a cost-effective solution for business-critical storage applications.

RAID 0

RAID 0 allows multiple physical drives to be logically concatenated into a single logical disk drive. A technique called data striping is applied to the physical disk drives. This technique interleaves blocks of data across the disks. The layout is such that a sequential read of data on the logical drive results in parallel reads to each of the physical drives. RAID 0 requires a minimum of two drives.



RAID 0 enables you to create a large logical disk drive and provides performance acceleration due to paralleling I/O accesses among multiple disks.

RAID 0 provides no redundancy protection such as parity protection or data mirroring. If a single disk fails, all data is lost and all disks must be reformatted.

Highlights

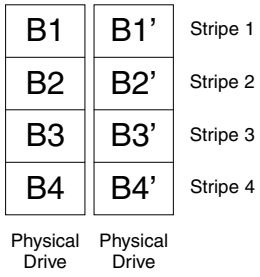
RAID 1 uses data mirroring, which duplicates the data from a single logical drive across two physical drives.

RAID 1 provides the lowest level of storage efficiency because it maintains a complete duplicate of the user data.

The use of a parity makes more efficient use of the disk drives and eliminates the need for mirroring to provide redundancy.

RAID 4 uses block-level striping, which improves the write performance over RAID 3.

RAID 1



RAID 1 uses the concept of data mirroring, which duplicates the data from a single logical drive across two physical drives. Data written to the logical drive is written to both physical disk drives. This creates a pair of drives that contain the same data. If one of these physical drives fails, the data is still available from the remaining disk drive. RAID 1 also supports the use of a hot-spare disk. The hot-spare disk takes the place of the failed drive in the array, so that data redundancy and performance can be recovered while the failed disk is being replaced. RAID 1 requires a minimum of two drives.

RAID 1 provides the lowest level of storage efficiency because it maintains a complete duplicate of the user data. The effective storage capacity available is half of the actual physical storage available in a RAID 1 configuration.

RAID 3

RAID 3 stripes data across all the data drives, writing a single block across all drives. This type of striping is referred to as byte-level striping. Parity data is then stored on a dedicated drive. Parity data can be used to reconstruct the data if a single disk drive fails. RAID 3 requires a minimum of three drives (two data disks and one parity disk).

The use of a parity makes more efficient use of the disk drives and eliminates the need for mirroring to provide redundancy. However, there is a performance penalty because the single parity drive creates a bottleneck during write operations. RAID 3 is most appropriate for workloads that are read-oriented or where write performance is not essential, such as Web-site file serving, or video streaming applications.

RAID 4

RAID 4 is very similar to RAID 3, except that it uses block-level striping instead of byte-level striping. With block-level striping, a complete block is written to a single disk. The use of larger stripes improves the write performance over RAID 3. It still maintains the use of a dedicated parity drive and requires a minimum of three drives, as does RAID 3.

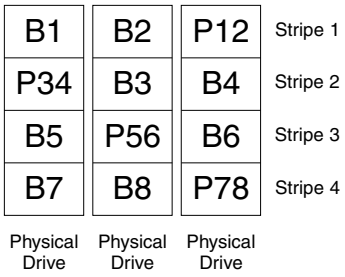
Highlights

The dedicated parity drive becomes a bottleneck in a RAID 4 system due to the excessive traffic and seek times of that drive. RAID 4 also incurs a performance penalty for the read-modify-write cycle when writing data smaller than a full stripe because the system needs to process the whole stripe to recalculate parity.

RAID 5

Block-level striping and distributed parity eliminates the bottleneck of writing to the dedicated parity drive.

RAID 5 uses block-level striping and distributed parity. This eliminates the bottleneck of writing to the dedicated parity drive and does not require the duplicate disk drives of RAID 1. Both the data and parity information are spread across the disks one block at a time. RAID 5 requires a minimum of three drives.



As with RAID 4, the one performance penalty is in the read-modify-write cycle for writes smaller than a full stripe.

A RAID array operating with a failed drive is said to be in degraded mode. RAID 5 arrays synthesize the requested data for the failed drive by reading the parity information for the corresponding data stripes from the remaining drives in the array. A failed drive in a RAID 1 or RAID 5 array can be replaced by physically swapping in a new drive or by a designated hot spare.

RAID 5E

With RAID 5E, you receive the advantages of RAID 5, but with additional performance provided by putting the hot spare to work.

RAID 5E (Enhanced) puts hot spares to work to improve reliability and performance. A hot spare is normally inactive during array operation and is not used until a drive fails. By utilizing unallocated space on the drives in the array, a "virtual" hot spare is created. By putting the hot spare to work, performance improves because more "heads" are writing the data. In the event of a drive failure, the RAID controller will start rearranging the data from the failed disk into the spare space on the other drives in the array. Thus, with RAID 5E, you receive the advantages of RAID 5, but with additional performance provided by putting the hot spare to work.

Highlights

RAID summary

RAID level	Mirroring	Striping	Parity	Min. drives	Key features
0	No	Block	No	2	Fastest, but lacks data protection.
1	Yes	No	No	2	Requires double capacity, but fastest protected solution.
3	No	Byte	Dedicated	3	Distributes each block across disks.
4	No	Block	Dedicated	3	Larger blocks improve performance. Dedicated parity disk is potential bottleneck.
5	No	Block	Distributed	3	Eliminates parity bottleneck.

Hardware vs. Software RAID

Software RAID solutions are closely tied to the operating system and are typically hardware-independent.

Software RAID solutions rely on standard host bus adapter (HBA) cards that receive I/O commands directly from the host computer CPU. Software RAID solutions are closely tied to the operating system and are typically hardware-independent.

The major advantage of software RAID is that it is considered "free" with the purchase of an operating system. However, with software RAID, you are often limited in the types of RAID levels that can be supported, and there may be performance tradeoffs due to the additional workload on the host CPU. This workload impact for the operating system is the additional operating system context switching to the software RAID code running on the host CPU, and additional cycles required for reading, checking, computing and writing the parity data. Software RAID also requires additional host resources such as memory bus bandwidth and extra I/O bandwidth.

There are two forms of hardware RAID controllers: internal and external.

Hardware RAID solutions offload the host CPU to improve overall system performance. There are two forms of hardware RAID controllers: internal and external.

Highlights

Internal RAID controllers are cards installed in the host bus (typically PCI) of a server. They resemble SCSI or Fibre Channel adapters and contain additional logic to perform the RAID functionality for calculating parity and striping data.

External RAID controllers move the controller outside the server into a separate chassis. Depending on the implementation, this RAID controller chassis may also house the actual disk drives. This RAID controller handles all RAID functions and presents all the logical drives to the system. External RAID controllers typically contain more memory and provide greater performance than internal controllers. External controllers also typically support a much larger number of disk drives than internal controllers.

Another benefit of the hardware RAID solution is the incorporation of NVRAM.

Another benefit of the hardware RAID solution is the incorporation of nonvolatile memory (NVRAM). This contains RAID configuration parameters and allows write data to be cached in nonvolatile storage for immediate acknowledgements (ACKs) for network file system (NFS) writes.

Products that use software RAID solutions can address the performance concerns by using hardware assistance. For example, Network Appliance uses NVRAM to cache write requests and write those requests out to disk periodically. This allows for immediate acknowledgments for NFS writes and helps eliminates excessive disk writes.

There is a significant difference in performance between software and hardware RAID when a RAID 1 or RAID 5 configuration is running in degraded mode.

There is a significant difference in performance between software and hardware RAID when a RAID 1 or RAID 5 configuration is running in degraded mode. The RAID controller rebuilds the data for the failed drive on the new drive or hot spare. This rebuild operation occurs online while normal host reads and writes are being processed by the array. On a hardware RAID solution, this rebuilding occurs with no additional workload required by the host CPU. Conversely, these operations would require significant host CPU cycles for a software RAID implementation. Significant reduction in system performance can be experienced for software RAID solutions when in degraded mode.

Highlights

The IBM TotalStorage NAS 200 and the IBM TotalStorage NAS 300 can provide you with a low-cost entry into network-attached storage.

Controllers supporting nine different levels of RAID allow you the flexibility to tailor the RAID level based on your typical workload.

The 300 series is designed to meet a host of requirements in demanding environments.

The NAS 300 has a battery-backed cache that will protect any unwritten data … for up to 72 hours.

IBM TotalStorage NAS products

IBM TotalStorage network-attached storage (NAS) products are designed to leverage hardware RAID solutions to increase overall system performance. Both internal and external hardware RAID solutions are used. The entry-level IBM TotalStorage NAS 200 uses an internal RAID controller, while the IBM TotalStorage NAS 300 uses an external RAID controller for increased performance and redundancy of the RAID subsystem. These products can provide you with a low-cost entry into network-attached storage if you are reducing your use of general-purpose servers.

The IBM NAS 200 consists of two models—a tower model for work groups and a rack-mount departmental model. The tower model uses the IBM ServeRAID™-4L Ultra160 SCSI internal RAID controller. The rack model uses the ServeRAID-4H UltraSCSI internal RAID controller for more performance and storage capacity support. These controllers support nine different levels of RAID, including RAID 0, 1, 5 and 5E. This allows you the flexibility to tailor the RAID level based on your typical workload. Support for Logical Drive Migration allows various complex RAID setup and maintenance operations to be administered in a simple, straightforward manner.

The 300 series with its rack configuration is targeted toward departmental and small enterprise applications. It is designed to meet a host of requirements in demanding environments. This NAS appliance is well suited for applications such as file serving in a traditional office environment and file serving or backup storage in a Web-hosting environment. The NAS 300 incorporates the IBM TotalStorage 5191 external RAID controller. The NAS 300 has a battery-backed cache that will protect any unwritten data (that was still in cache when the failure occurred) for up to 72 hours. It supports RAID levels 0, 1, 3 and 5. The 5191 is designed for high-availability applications requiring a high degree of component redundancy. It features two hot-plug RAID controllers and two hot-plug power supplies and redundant fans.

Highlights

When you use a hardware RAID solution, the range of RAID levels enables you to trade off reliability, performance and cost.

Summary

RAID is an excellent and proven technology for protecting data against the possibility of hard disk failure; in addition, data striping offers improved system performance. When you use a hardware RAID solution, the range of RAID levels enables you to trade off reliability, performance and cost. This capability is lost with software RAID, where you are typically limited to only a single level of RAID.

IBM Corporation
PO Box 12195
Research Triangle Park, NC 27709
U.S.A.

Produced in the United States of America
10-01

G522-2535-00