StoreVault™
A NetApp Division

StoreVault and RAID-DP™

RAID data protection levels explained

Limitations of conventional RAID

How RAID-DP adds an extra layer of protection for your data

introduction

The increased dependence on disk storage and the need for higher levels of reliability, availability and performance has led to a dramatic increase in the use of Redundant Array of Independent Disks or RAID technology. The redundancy allows for a disk array to be fault tolerant. The failure of a drive does not lead to loss of data and the array can be rebuilt from the surviving drives.

This paper provides an overview of some RAID techniques and explains how RAID-DP overcomes the shortcomings of conventional RAID schemes to provide a significantly higher level of data protection without adding significant cost.

Further reading may be found at www.netappp.com.

what is RAID?

Overview

In the late 1980's, a team at the University of California, Berkeley, proposed a number of schemes that used an array of inexpensive disk drives to increase performance and reliability when compared to a single large expensive disk drive. The Redundant Array of Inexpensive Disks or RAID technology has since been widely adopted across the storage industry. As the price of disk drives plummeted, and the concept of "expensive disk drive" diminished, the RAID acronym later became interpreted as Redundant Array of Independent Disks.

The original paper specified five levels of RAID, of these; levels 1, 3, 4 and 5 are in common use today. Additional RAID levels have been defined, including RAID 0 (which had no redundancy) and compound RAID where multiple RAID levels are combined to improve performance or reliability and availability. The latest addition is the NetApp implementation, RAID-DP, which allows recovery from a dual disk drive failure.

Before we look at this, let us review how the existing levels of RAID work and how they compare in cost, protection and performance.

RAID 0

Otherwise known as disk striping, RAID 0 is not strictly RAID as there is no redundancy. Data is split up and written to multiple disks at the same time. Performance for both read and write are dramatically improved, but the failure of a single disk can result in the loss of the entire data set.

RAID 0 is typically only used where performance is more important than the data itself, for example, in a test environment or perhaps video editing on a copy of the data. There is no capacity impact for RAID 0, the total capacity of the drives in the array is available for data.

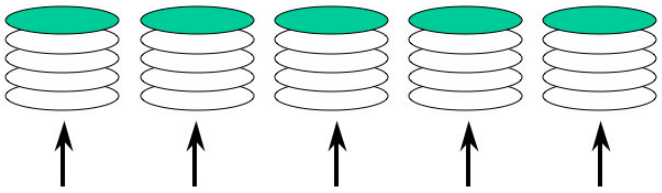

Figure 1: RAID 0 Striping

RAID 0: data is split up and striped across all drives in the array. This significantly improves performance but offers no redundancy. The loss of a single drive can result in the loss of all data.

RAID 1

Disk mirroring or RAID 1 provides very secure redundancy but at double the cost in terms of capacity. Data sent to a disk is also sent to a mirror disk, in the event of any disk failure, the data can be read from the surviving disk. The failed disk can be replaced and the mirror can be rebuilt from the good drive in the pair.

Write performance is the same as for a single disk but read performance can be higher if the host is able to read from both drives in the pair.

RAID 1 is commonly used where the value of the data is more important than the cost considerations of paying double for the cost of the disk drives.

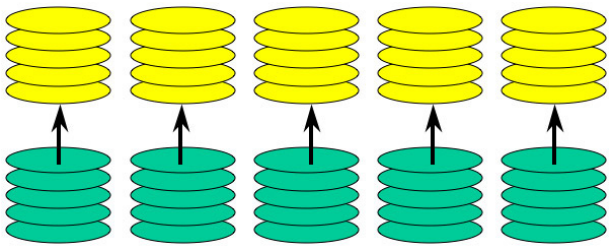

Figure 2: RAID 1 Mirroring

RAID 1: data on each drive is mirrored to another drive and the host addresses each pair individually. If one of the drives in any pair fails, the data may be read from the surviving drive.

RAID 3

RAID 3 has a dedicated parity drive in the array that allows data to be recovered if one fails. Data is striped across the drives at the byte level, parity is calculated for the stripe and written to the parity drive.

There is a significant write performance penalty because data is written to all the disks simultaneously and there is no ability to overlap writes but read performance is very good. Although RAID 3 is rarely found, it is common in streaming large files such as video and scientific applications.

The capacity penalty is one disk drive out of a group so for a RAID group of ***n*** drives, ***n***-1 are available for data.

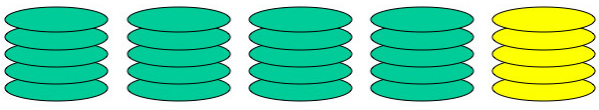

Figure 3: RAID 3 Dedicated Parity

RAID 3: Byte-level data is divided among the data drives, parity is calculated and written to a dedicated parity drive. Typically only used when streaming large files because of poor write performance.

RAID 4

As with RAID 3, RAID 4 uses a dedicated parity disk to provide redundancy for the RAID group. The difference is that RAID 4 distributes and calculates parity based on block-level data. For those systems that have no RAID awareness, there is a penalty in write performance, which is why RAID 4 is seldom found in other manufacturers' systems.

However the NetApp implementation is based on the tight implementation of the Write Anywhere File System (WAFL), the Non-Volatile memory (NVRAM) and efficient use of RAID 4. The write penalty normally associated with continuous writing to the parity disk is eliminated because writes are aggregated in the NVRAM and parity is calculated and written only once for each stripe.

The compromise in capacity to provide redundancy is one drive per RAID group, the same as for RAID levels 3 and 5.

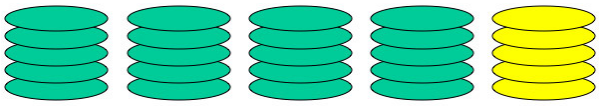

Figure 4: RAID 4 Dedicated Parity Drive

RAID 4 is similar to RAID 3 but data is distributed at the block level. The implementation in a NetApp environment ensures high performance on both read and writes. The array can be expanded on the fly with no impact on system performance.

RAID 5

RAID 5 is similar to RAID 4, in that it is a single parity scheme, but instead of writing the parity to a dedicated drive in the group, the parity is rotated amongst all the drives. This was done to enable files systems that are not RAID aware to be able to overcome the bottleneck of a single parity drive. The cost of redundancy is the same as for RAID levels 3 and 4; the capacity equivalent of *n*-1 drives in the array is available for data.

Figure 5: RAID 5 Distributed Parity

RAID 5 is similar to RAID 4 but parity is distributed across all the drives in the RAID group. For file systems that have no RAID awareness RAID 5 provides a good compromise between redundancy and cost.

Compound RAID

RAID levels can be combined to improve reliability or performance. For example, RAID levels 10 and 0+1 both combine striping for performance with mirroring for redundancy, RAID 51 combines RAID 5 with mirroring to provide the capability to survive and recover from multiple drive failures. The capacity compromise reflects the combined redundancy and there is generally also an adverse performance impact.

Recovery

When a disk drive fails in a RAID group, the defective drive is replaced with a good one and the drive is rebuilt using the information contained in the surviving drives. In the case of a mirrored pair, the entire contents are read from the good drive and copied to the new drive. In the case of parity (either dedicated or distributed) there is a calculation required for each block of data. If all the data blocks are intact and a parity block is missing it is simply recalculated from the data blocks. If a data block is missing then it is calculated by doing a reverse calculation using the good data blocks and the parity information. The time taken for recovery will depend on a number of factors, including the workload on the system, the speed and capacity of the drive, and the amount of data on the array.

the problems with conventional RAID

Data Protection during Recovery

Conventional single parity RAID (e.g. RAID 5) allows for continued system operation and for recovery of the data and rebuilding the array using the surviving drives. This has provided a good basis for system reliability in the past, but there are several factors that are making that reliability less robust.

- At the disk drive level we have seen continued growth in magnetic areal density, which has enabled disk drive capacities to grow into the hundreds of gigabytes.
- The bit error rate inside the drives has not improved at the same rate as capacity and so unrecoverable errors per drive are becoming more likely.
- Disk I/O performance has also failed to keep pace with increases in capacity so the rebuilding of a drive or an array takes much longer.
- While a recovery is underway, system performance is slowed.
- A recovery and rebuild can take over 8 hours during which time your data is unprotected and susceptible to a second disk drive failure.

In all instances of single parity RAID, a second disk drive failure results in loss of data. Any time a disk drive failure occurs in an array, the data is exposed to a second disk drive failure while the recovery operation takes place.

the solution: RAID-DP

How it Works

RAID-DP is a new level of data protection for an array of drives. It enables recovery from simultaneous failure of 2 drives in a single RAID group. RAID-DP offers four orders of magnitude more reliability than RAID 5. Where RAID 4 uses a dedicated drive for parity to reconstruct missing blocks, RAID-DP uses two dedicated drives and uses alternate views of the data to calculate a different but complementary parity.

The first parity disk is calculated in exactly the same way as for RAID 4 across the horizontal stripe. The second parity is calculated by looking at a "diagonal stripe". By using these two algorithms together, data can be read and reconstructed (much as missing squares are completed in a Sudoku puzzle!) even in the event of a dual drive failure. It does not matter if the failure occurs in data or parity drives; RAID-DP is resilient against failure of any 2 drives in the array.

A more detailed, technical explanation may be found at www.netapp.com.

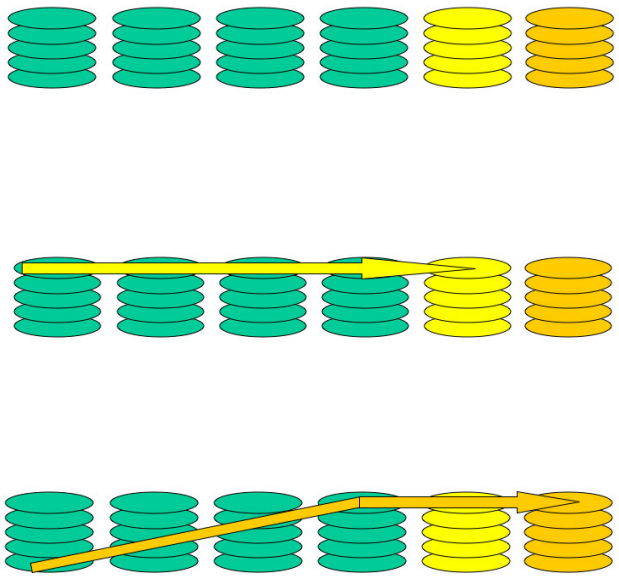

Figure 6: RAID 5 Distributed Parity

RAID-DP is like RAID 4 but with an additional parity disk.

Parity is calculated for the first parity disk exactly like RAID 4 by looking at a horizontal stripe of data.

For the second parity disk, a diagonal view of the data is used to calculate parity.

summary

We have seen how RAID was developed to provide varying levels of protection, performance and availability to data storage. With the addition of RAID-DP, NetApp is delivering a significantly higher level of protection to your environment without the significant cost burden that additional redundancy usually brings. RAID-DP is currently offered in all NetApp products, including the new StoreVault S500, which was developed to satisfy the needs of smaller to medium-sized businesses.

The new StoreVault S500 from Network Appliance is an all-in-one scalable network storage appliance that allows first-time NAS and SAN users to simplify not only the installation process, but also the day-to-day protection and management of their data. StoreVault provides access to the powerful data storage management tools and capabilities that have previously been only available to large IT shops. With Snapshots built in, backup and recovery are virtually instantaneous, delivering unprecedented simplicity and reliability.

To find out more about how the StoreVault S500 and RAID-DP can provide more protection for your business data, please visit www.netapp.com

examples of RAID efficiencies

RAID Level	Capacity Impact	Data Protection	Write Performance	Read Performance
RAID 0	No impact	None	Very Good	Very Good
RAID 1	50%	Good	OK	OK
RAID 3	n-1	Good	Poor	Very Good
RAID 4	n-1	Good	Very Good*	Very Good
RAID 5	n-1	Good	Good	Very Good
RAID-DP	n-2	Very Good	Very Good	Very Good

*NetApp implementation with Data ONTAP and WAFL

NetApp®

Network Appliance, Inc.
495 East Java Drive
Sunnyvale, CA 94089

About Network Appliance

Network Appliance is a world leader in network storage solutions for today's data-intensive world. Since its inception in 1992, Network Appliance has delivered technology, product, and partner firsts that simplify data management. Information about Network Appliance solutions and services is available at **www.netapp.com.**

For more information on StoreVault, a NetApp division, go to **www.storevault.com.**

20-195-102 06/16/2006