Moving Beyond RAID –
DXi and Dynamic Disk Pools

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The shortcomings and failures of traditional RAID are well known, and become more acute as disk capacities continue to increase. Numerous alternate methods and technologies exist for primary storage, but these have not been used with secondary storage such as disk-based backup appliances. Quantum is leading the charge 'beyond RAID' with the DXi®-Series, in which Dynamic Disk Pools (DDP) technology is leveraged for its ability to provide more consistent performance, improved data protection, and extreme versatility.

DDP is the latest example of Quantum's continued investment into next-generation data storage technology, adding to a deep portfolio including deduplication, the StorNext®[1] high-performance file system and data manager, Lattus™[2] Object Storage solutions, and more.

This Technology Brief contrasts DDP with traditional RAID, and outlines the superior business benefits of DDP when applied to a disk-based backup appliance such as the Quantum DXi6900.

INTRODUCTION

Almost since the term "RAID" was publicized by Patterson, Gibson, and Katz in their 1988 UC Berkeley paper "A Case for Redundant Arrays of Inexpensive Disks (RAID)"[3], RAID techniques have been the de-facto standard for enterprise data storage. Today, more than twenty years later, the realities of current hard disk technology have magnified RAID's limitations, and other technologies have arisen that improve on RAID in many ways. This Technology Brief describes one of these 'beyond RAID' technologies—Dynamic Disk Pools, or DDP—implemented in the new Quantum DXi6900 Disk-Based Deduplication Appliance.

RAID IS NO LONGER 'GOOD ENOUGH'

Anyone who has worked in data storage long enough probably remembers their first RAID 5 array, and how liberating it was not to have to restore from tape every time a disk failed. They probably also remember—vividly—the first time RAID 5 betrayed them, through a double-disk failure, unrecoverable read error, or some other problem that sent them down to the vault for the backup tapes yet again.

The fact that RAID technology can fail is not news, and the ways it can fail have been catalogued extensively. These facts led Henry Newman to conclude "RAID's Days May Be Numbered" in 2009[4], and posit that techniques such as de-clustered RAID[5] would be increasingly necessary.

Many primary storage vendors understand the risks of traditional RAID, as seen in the proliferation of nested or hybrid RAID implementations such as RAID levels 10, 50, 60, and vendor-proprietary variants. So-called scale-out NAS vendors have embraced RAIN, essentially RAID at the node level. Vertical markets making heavy use of 'big data' are already migrating to Wide-Area-Storage-type architectures such as Quantum Lattus[2] that are far more resilient and available than RAID. Secondary storage—including disk used as a backup target—is subject to the same risks as primary storage, but secondary storage vendors have not acknowledged or addressed the limitations of RAID—until now.

Bigger Disks, Longer Rebuilds, and More Errors

3TB disks are common in enterprise applications today, and 4TB consumer disks are shipping now. This increasing density is making RAID's problems worse in several important ways. As disk capacity has increased, so have RAID rebuild times. Longer rebuilds result in a greater window of exposure to potential data loss, and performance can be significantly affected when a rebuild is in progress.

Another side effect of higher density is the increased likelihood of encountering an unrecoverable read error (URE) during the rebuild process. Disk reliability improvements have not kept pace with density improvements[2], making it more likely than ever that a URE will be hit during a rebuild. Best case: this will delay the rebuild; worst case: all data on the RAID set will be lost. Additionally, disk manufacturer's hard error specifications do not account for errors that may occur in the controller hardware, software, or data path, so the reliability of the system is lower than indicated by disk reliability figures alone. Simply put, the density vs. reliability plot is trending in the wrong direction.

RAID 6 is a Band-Aid®

The use of RAID 6 and its double parity has become common precisely because of the issues outlined above. If there is a problem with reading the first copy of parity due to corruption or a disk failure, the second copy saves the day. Having a RAID array that can withstand two simultaneous disk failures without data loss is a good idea, but simply pushes the problem into the future. As disk density continues to increase, eventually a third copy of the parity information is needed, or a different strategy altogether.

Hot Spares Waste Resources

In traditional RAID architectures, 'hot spare' disks are used to minimize the vulnerability window after a disk failure. When a failure occurs, the array controller immediately initiates a rebuild onto a hot spare disk. When the failed unit is replaced, either it becomes a hot spare, or the reconstituted data is copied back from the hot spare, depending on how the array is configured. Typically, the larger the array, the more hot spare disks are assigned to counter the increased risk of disk failure. Five, ten, or more hot spare disks may be present in a large array, simply waiting for an on-line disk to fail.
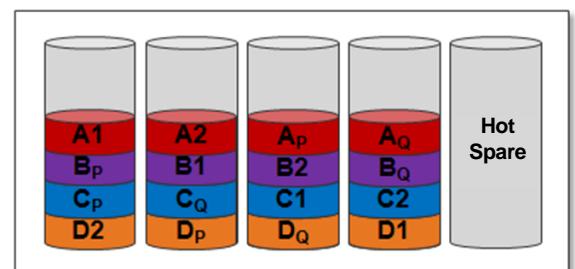
Hot spare disks represent a waste of resources. They sit idly by, consuming space, consuming power, and producing heat but not contributing their performance to the production workload. What if there was a way to leverage those extra spindles for performance, but still maintain hot spare capacity to quickly deal with the inevitable disk failures?

RAID 6 Reference

In order to understand how DDP provides benefits, it is necessary to understand a bit about how it operates, and how it is different from traditional RAID. This paper will focus on RAID 6, the industry standard for disk-based backup appliances such as the Quantum DXi.

A RAID 6 group (see Figure 1) uses block-level striping with double distributed parity. Data is broken into blocks and written across n-2 disks, where n equals the number of disks in the array (exclusive of hot spares). On the other two disks, two copies

Figure 1. 4-Disk RAID 6 group with Hot Spare

of parity information are recorded. For example, in Figure 1, A1 and A2 represent data, and AP and AQ represent the parity information for data A1 and A2. Similarly the next stripe of data is represented by data blocks B1, B2, with parity BP, BQ, and so on through C and D. Note that the data and parity is distributed across all disks equally. Due to the dual parity, this RAID group can sustain two simultaneous disk failures without data loss.

This RAID group also includes a hot spare, which is online and consuming power, but does not contain any data or parity unless called into action after a failure of one of the primary disks. At that time, the array controller will use the data and parity on the remaining disks to calculate the missing information and populate the hot spare.

While the array is running in degraded mode, there is always a performance penalty, often severe. After a single disk failure, the array controller and remaining disks are busy performing reads, writes, and parity calculations to rebuild onto the hot spare (if present). If two disks have failed, read requests are served by reconstituting missing data on the fly, which further impacts performance. It is important to note that, regardless of the number of disks in the RAID group, all disks are under additional load during degraded mode operation. There is a tradeoff between the speed of the rebuild and the performance penalty while the rebuild is in progress, usually selectable by modifying rebuild priority parameters on the array controller. Even when set at the highest priority, the rebuild speed is limited by how fast a single disk—the hot spare—can write.

BEYOND RAID – DYNAMIC DISK POOLS

One technique that can help overcome the limitations of RAID is Dynamic Disk Pools, or DDP. DDP is a form of de-clustered RAID[3] that uses concepts familiar from traditional RAID (such as striping data and parity across an array of disks and hot sparing), but leveraged in a different way. This provides business benefits while avoiding many problems inherent in traditional RAID. In general, DDP can provide improved data protection, more consistent performance, and greater configuration versatility vs. simple RAID 6, the method used in almost all secondary storage devices today. The rest of this paper will compare DDP technology to traditional RAID 6 and explain the business benefits of DDP in a secondary storage device such as the Quantum DXi-Series.

At a high level, RAID 6 and DDP are similar: both are techniques for striping data and parity information across a set of disks to provide fault-tolerance. It is the specifics of how this goal is attained that are different, and the details make all the difference.

Volume Layout

Consider how information is striped across disks in a Dynamic Disk Pool. Data, parity, and spare capacity are striped across a set of disks. With RAID 6, hot spare disks are idle loafers. With DDP, there is no such thing as a hot spare disk. Hot spare capacity equivalent to the capacity of some number of physical disks is striped across all disks in the pool. Because all disks contain data and parity (see Figure 2), all share in the load and contribute their performance to the array. There is no wasted performance.
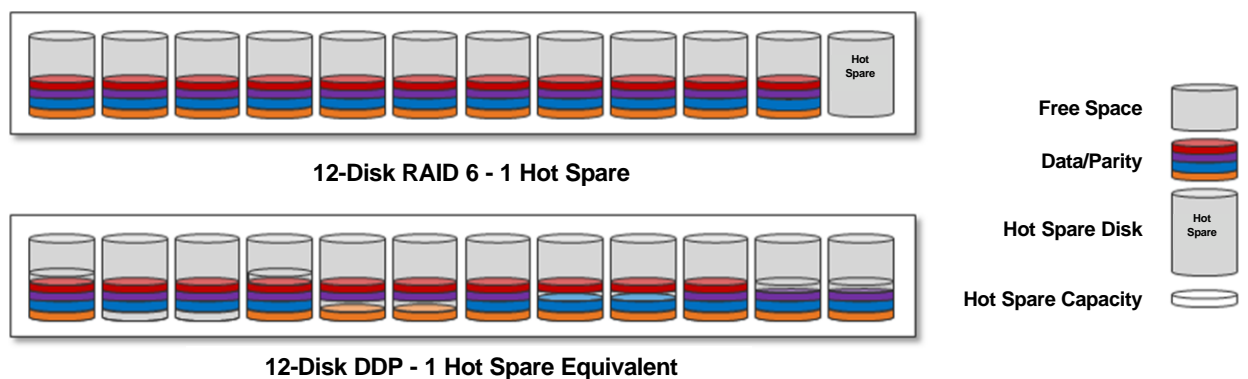

12-Disk RAID 6 - 1 Hot Spare

12-Disk DDP - 1 Hot Spare Equivalent

Free Space
Data/Parity
Hot Spare Disk
Hot Spare Capacity

Figure 2.   RAID 6 vs. DDP Volume

Dynamic Stripes

Each volume in a Dynamic Disk Pool consists of a set of dynamic stripes known as 'D-Stripes'. D-Stripes are allocated at volume creation by the DDP algorithm, and each D-Stripe is always distributed across exactly ten disks, regardless of the number of disks in the pool. Each D-Stripe is distributed across a different set of ten disks to balance allocation across the entire pool. The portion of each D-Stripe that resides on a particular disk is called a 'D-Piece'. In contrast, as shown in Figure 2, RAID 6 stripes data across all active disks. As a result, while a RAID 6 group with one hot spare may contain a minimum of five disks, the minimum size for a DDP (including one hot spare equivalent) is eleven disks.
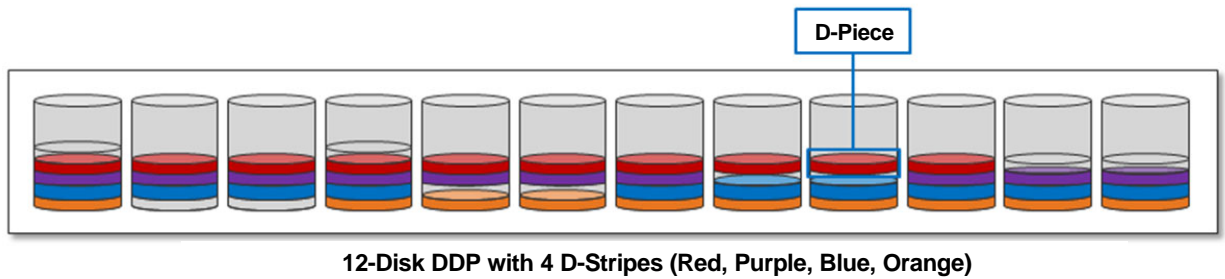

12-Disk DDP with 4 D-Stripes (Red, Purple, Blue, Orange)

Figure 3.   Dynamic Stripes

Inside Dynamic Stripes

Data within each D-Stripe is written using a large number of traditional RAID 6 stripes, evenly spread across the ten disks occupied by that D-Stripe. At its core, DDP is using traditional RAID techniques of striping data and parity, but in a different way and with much greater granularity.
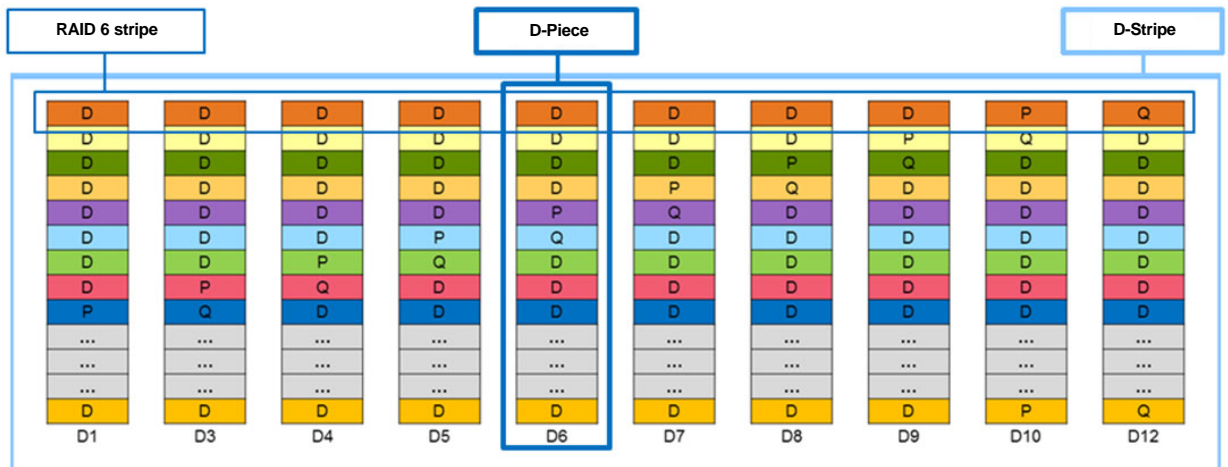

Figure 4.   Inside a Dynamic Stripe

Disk Failure and Recovery

Both RAID 6 and DDP must perform the same basic steps for recovery from a failed disk, summarized in Table 1 below. This is not to say the details are the same, just the overall process.

Table 1.

Recovery Step	RAID 6 Detail	DDP Detail
------------------ Disk Failure Detected -------------------		
Identify Missing Data & Parity	Every block of data (or its parity) always affected as data is striped across all active disks	Only D-Stripes with a D-Piece on the failed disk are affected – identify affected D-Stripes
Read Request Redirect	Reads are served from remaining data, with reconstruction from parity on the fly as required	Reads are served from remaining data, with reconstruction from parity on the fly as required
Write Request Redirect	Writes go to remaining disks, written with single parity (RAID 5)	Writes go to remaining whole D-Stripes
Reconstruct Missing Data & Parity	Read from all disks, perform parity calculations. Write to a single hot spare disk	Read from remaining 9 disks in the D-Stripe, perform parity calculations. Write each reconstructed D-Piece to a different disk in parallel
----------- Full Data Protection Restored -----------		
Failed Disk Replaced, Return to Original Configuration	Initiate copy back from hot spare to new disk (if copy back configured)	Initiate copy back of reconstructed D-Pieces in parallel from multiple disks to new disk
--------- Original Configuration Restored ---------		

The fact that each D-Stripe is resident on only ten disks is the key to a potential rebuild performance advantage of DDP over RAID 6. With a RAID 6 group, all remaining active disks must execute reads to support a rebuild, and all write activity is focused on a single hot spare disk. This is true no matter how many disks are in the RAID group. The read activity may be in contention with inbound I/O requests, and the write to a single disk bottlenecks the process, extending time to recovery.

Due to DDP's D-Stripe construct, a rebuild involves reading from a number of disks—perhaps all of them, depending on the location of the D-Stripes affected and the size of the pool. But writes are not focused on a single disk; rather, they are spread across many disks and occur in parallel. The result is potentially much faster recovery from a failed disk vs. traditional RAID 6. These benefits are magnified with larger disk pools, with a smaller percentage of the total number of disks affected by the rebuild process.

Finally, reconstruction of data in a DDP is intelligently handled via patented prioritized reconstruction technology. With this method, D-Stripes experiencing multiple disk failures are reconstructed first. For example, in a two-disk failure scenario, any D-Stripes with a D-Piece resident on both failed disks are reconstructed first. D-Stripes with a D-Piece residing on only one of the failed disks are reconstructed next. This prioritization means the most vulnerable data always has the highest priority for reconstruction, further reducing the chance of data loss should another disk fail before the full reconstruction is complete.

BENEFITS OF DDP

There are many benefits of deploying DDP in the Quantum DXi solutions. With the DXi-Series, Quantum is beginning a shift to this technology where it makes sense. The magnitude and nature of DDP benefits will depend on the exact configuration, but in general they may be grouped into three categories:

- Consistent Performance
- Improved Data Protection
- Extreme Versatility

Consistent Performance

Traditional RAID architectures are highly tunable for the type of performance required—high-bandwidth streaming, high transaction rates, or a balance of the two. But when disks fail, throughput takes a nosedive. As the number of disks in a system increases, it becomes more likely that a disk failure and rebuild will be happening at any given moment. This makes real-world performance inconsistent, which can negatively impact business operations. With DDP, the performance impact can be much lower, and of a shorter duration. This means less disruption to operations, and in the case of a backup appliance, can make the difference between making and missing the backup window.

Figure 5.

Much like next-gen object storage systems, DDP was designed with the reality of failure in mind. Because parity, data, and spare capacity are spread across the entire pool, every disk participates in the reconstruction process when a disk fails. As a result, a Dynamic Disk Pool can heal itself up to eight times faster than a traditional RAID group and maintain higher performance while reconstruction is in progress.

Improved Data Protection

As noted earlier, with increased disk capacities come longer RAID rebuild times, resulting in an increased probability of data loss due to multiple disk failures. The shorter rebuild times possible with DDP reduce the exposure window to a minimum. In addition, the patented prioritized reconstruction technology used by DDP ensures that the most vulnerable data is reconstructed first, further reducing risk. And, of course, DDP features work in concert with T-10 PI data integrity, background scrubbing, and all of the other capabilities of the array, to provide comprehensive data protection.

Extreme Versatility

DDP is one tool in the extensive technology portfolio that Quantum draws from to create the DXi appliances. Because DXi systems are sold as appliances, the versatility benefit provided by DDP may not be readily apparent, but it is present and relevant.

Tuning a stack of hardware and software to perform as a high-performance target for backup and recovery at a palatable cost point is not a trivial task. In the DXi line, Quantum leverages storage technologies such as fast SAS, SSD, RAID, the Quantum StorNext file system, and now DDP, in order to provide an appliance with the maximum performance and protection for the best cost in any market segment.
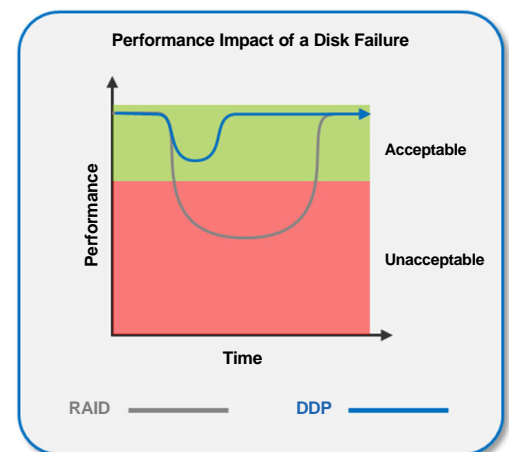
DDP helps by removing some of the limitations of traditional RAID. Disk Pools may be small to extremely large, spanning shelves as required, and they may be easily expanded on the fly. File system stripe size may be optimized without concern for the number of disks in a shelf. Different disk types may be used to create storage tiers. Hot spare capacity is integrated and configurable. And Dynamic Disk Pools may co-exist with traditional RAID groups for those cases where RAID is the best fit for part of the workload.

The additional versatility provided by DDP is one factor enabling the leading performance and startling price-performance of the DXi-Series, and opens additional avenues for innovation in future DXi models.

CONCLUSION

With the introduction of DDP in the DXi-Series, Quantum is the first deduplication appliance vendor to move beyond traditional RAID, reaping benefits in performance, efficiency, flexibility, and improved data protection. Due to the flexible and extensible nature of DDP technology, these benefits will continue to accrue as Quantum expands the use of DDP.

For more information on the DXi-Series, visit **www.quantum.com/dxi**

1 http://www.quantum.com/stornext
2 http://www.quantum.com/products/BigDataManagement/Lattus/index.aspx
3 http://www.cs.cmu.edu/~garth/RAIDpaper/Patterson88.pdf
4 http://www.enterprisestorageforum.com/technology/features/article.php/3839636/RAIDs-Days-May-Be-Numbered.htm
5 http://www.pdl.cmu.edu/PDL-FTP/Declustering/ASPLOS.pdf

Quantum®

BE CERTAIN

www.quantum.com • 800-677-6268

BE CERTAIN