

RAID 50 Interleave White Paper

Reference	Title
Date	22/9/06
Version	1.00

Document History.

Author	DATE	ACTION	Rev
BB	22/9/06	Document Creation	1.00

Contents

1	Overview.....	4
1.1	RAID 50 Interleave Buffered Tiered Storage	4
2	How RAID 50 Interleave Works	4
2.1	Traditional RAID 5.....	4
2.2	RAID 50 – RAID 5 with spanned or striped drive strata	5
2.3	RAID 50 Interleave.....	7
2.4	RAID 50 Interleave Fault Tolerant Examples	7
2.4.1	Single Drive Failure.....	7
2.4.2	Dual Drive Failure – Single Drive Carrier.....	9
2.4.3	Dual Drive Failure – Separate Drive Carriers.....	9
2.4.4	Dual Drive Failure with Spares available	10
3	Summary.....	11

1 Overview

1.1 RAID 50 Interleave Buffered Tiered Storage

Traditional Redundant Array of Independent Disk (RAID) Systems have tended to use RAID 5 as the most cost effective means of providing fault tolerant protection for large volumes of data. The emergence of large capacity configurations from applications including Disk to Disk Back up, Virtual Tape disk farms, Continuous Data Protection CDP, Content Addressable Storage, and Regulatory driven extended retention times, and Archive are driving the push to tiered storage. Configurations increasingly incorporate a mixture of Enterprise, Nearline, and Desktop drives. The consolidation of storage on to Storage Networks and the addition of mixed use and archive information management to online mission critical systems are stretching the RAID 5 approach. Several factors are combining creating the need to rethink the overall approach including:

- Increasing Use of Drives with Lower MTBF & Duty Cycle – Lower Quality of Service QOS
- Larger Capacity Drives both Enterprise Class as well as Nearline and Desktop
- Larger Number of Drives per Array/ Storage Domain

The combined impact of these trends have increased the likelihood of drive failures, extended the drive rebuild times and are leading to ever increasing probability of a second drive failure spelling a disastrous loss of the RAID set. Dead RAID sets potentially cause data loss, server downtime, and at minimum lengthy and expensive recovery and restore procedures. Many customers are asking suppliers to deliver RAID 6 to mitigate the increased risk and exposure. However, most attempts at RAID 6 to date carry costly processor overhead and performance penalties.

RAID 50 Interleave is a new advanced multi-drive failure tolerant implementation of RAID 5 from MPSTOR. RAID 50 Interleave goes beyond simple RAID 5 or even RAID 50 to provide an interleaved set of RAID 5 stripe groups or strata. The interleaved strata can be distributed across multiple enclosures comprising a single RAID set on a Storage Domain (A Storage Domain as used here means the entire group of drives or block devices under the control of one or more Storage Server Blades). Multiple RAID sets can then be aggregated into Virtualized Logical Volumes that are then presented as host Logical Units or LUNs. The benefits of RAID 50 Interleave include:

- Ability to tolerate multiple drive failures within a RAID set- only one drive failure within a stratum
- Ability to take advantage of High Density Dual or Multi-Drive Carrier Enclosure Designs
- Unique Labels on each registered drive within a Storage Set to allow automatic tracking, detection and Predictive Failure management.
- Cached Down Stripe Writes & Partial Rebuilds

2 How RAID 50 Interleave Works

For years Storage System developers have been using the concept of multiple drive strata within an overall RAID 5 storage group to more practically handle large numbers of drives within a single RAID 5 set. For example a set of 48 drives may be subdivided into more efficient sets of 8 or 16 drives, known as stripe groups or strata, to optimize performance and simplify management of the overall pool. This subdivision is handled transparently by a hardware RAID controller and has been called RAID 50 by some vendors and analysts.

RAID 50 Interleave is a next generation architectural approach that builds on the stripe group concept adding unique labels, down stripe caching, and interleaving to maximize the performance and expand the fault tolerance to allow multiple drive failures.

2.1 Traditional RAID 5

RAID 5 has become widely used to protect company information in business and government institutions of all sizes. Originally conceived as a means to more cost effectively protect mission critical data than traditional mirroring, RAID 5 uses the concept of distributing data across a group of drives with the addition

of parity data. The combined data and parity allows RAID 5 to be single fault tolerant. That is, any single drive in the group can fail and the remaining drives can regenerate the missing data using the parity information.

Mirroring had a one for one or 50 percent overhead which proved too expensive for many applications and budgets. RAID 5 only carried a 1 out of n overhead, where n is the number of drives in the RAID 5 group. With drive capacities exploding driving the cost from dollars to pennies per MB, and the introduction of SAN and NAS network storage allowing networks of all types and sizes to take advantage of the protection and cost efficiencies, RAID 5 based storage spread from large corporate data centers out through small business and ultimately to home office use.



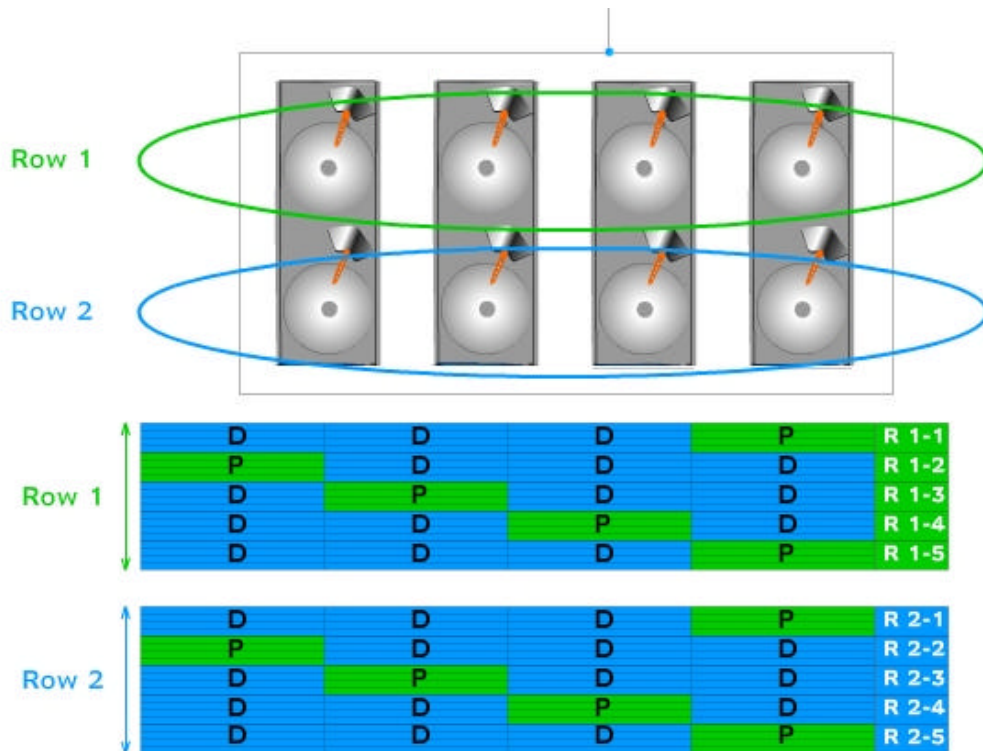
4 Drive RAID 5 Group and Logical Map

The spread of network storage led to ever increasing storage array sizes making managing RAID 5 groups more difficult and lead to the hybrid now known as RAID 50.

2.2 RAID 50 – RAID 5 with spanned or striped drive strata

The advent of increasingly larger storage arrays began to strain the practicality of single stratum RAID 5 implementations. With drive capacities well below 100GB and spindle speeds at 7200 rpm or less, 10-20 or more drives increased performance and lowered cost by taking advantage of the 1 out n overhead of RAID 5. As spindle speeds increased and individual drive capacities grew, optimum performance required progressively fewer drives per group. The overhead and penalty of very large numbers of high capacity drives severely impacted RAID 5 rebuild times and performance during non-fault tolerant operation, also known as degraded operation. There needed to be an alternative to allow smaller numbers of drives per group while maintaining overall RAID set size and performance.

Advances in controller design and caching led developers to incorporate multiple RAID 5 strata into what was presented externally as a single RAID 5 set. Initially the RAID sets would include multiple RAID 5 strata that were spanned together to form a single RAID 5 set. Later striping across the strata further increased IO performance while maintaining throughput bandwidth.



2 Rows of 4 Drive RAID 5 Groups Presented as a Single RAID 50 Set

RAID 50 with a striping implementation offers the benefits of:

- Scalable Capacity
- Improved IO performance
- Reduced number of drives per stripe group or strata = shorter rebuild times

Many RAID 50 Implementations are still the spanned version. Even with the stripe implementation RAID 50 has several limitations especially in larger configurations including:

- RAID 50 is still single fault tolerant – only one drive failure per set (despite multiple strata)
- Increased use of Near Line and Desktop Drives – lower MTBF and duty cycle
- Larger capacity drives increases rebuild times & reduces performance

The net impact is increased likelihood of a multi-drive failure with potentially disastrous consequences including:

- Dead RAID set
- Loss of data
- Potential downtime
- Lengthy & costly reconstruction and restoration

Many customers and vendors are looking to multi-drive failure tolerant implementations to mitigate the risk. The alternatives most frequently pursued recently are either RAID 51, mirrored RAID 5 sets, which is very costly in number of drives required and reduces write performance, or RAID 6/ 60 which uses duplex parity to allow 2 drive failure tolerant operation.

RAID 6/ 60 implementations suffer:

20-50% performance decrease versus RAID 5/ 50 during normal operation

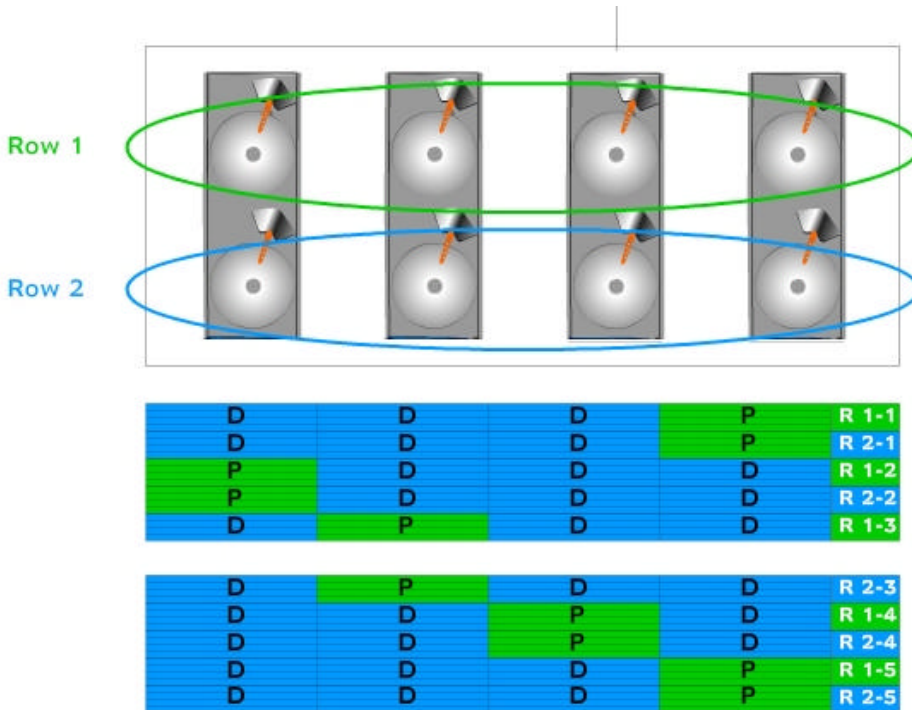
Performance penalty increases substantially more in degraded mode, i.e. with a single drive failure

RAID 6/ 60 is still only 2 drive failure tolerant across the entire set

2.3 RAID 50 Interleave

MPSTOR engineers have developed an exciting alternative, RAID 50 Interleave, as depicted physically and logically in the diagram below. RAID 50 interleave allows:

- Performance optimization
- IO throttling
- Multiple drive failure (only 1 drive failure within a stratum)
- QOS based Storage Tiers
- Cached drive removal and partial rebuild
- Use with High Density Tiered Storage Enclosure Designs
- Limited additional Drive Overhead – Same as RAID 50
- No RAID 6/ 60 Performance Penalty



2 Rows of 4 Drive RAID 5 Groups Presented as a Single RAID 50 Interleaved Set

All drives within a Storage Domain, defined as under the control of one or more Storage Server Blades, carry unique labels that allow the Storage Server Blade to identify:

- Drive ID & classification – used to classify drive Quality of Service (QOS)
- RAID Set and Strata membership
- RAID Set, Strata, and Drive Status
- Performance and Error Tracking and Predictive Failure Analysis
- Storage Domain Wide Drive Roaming

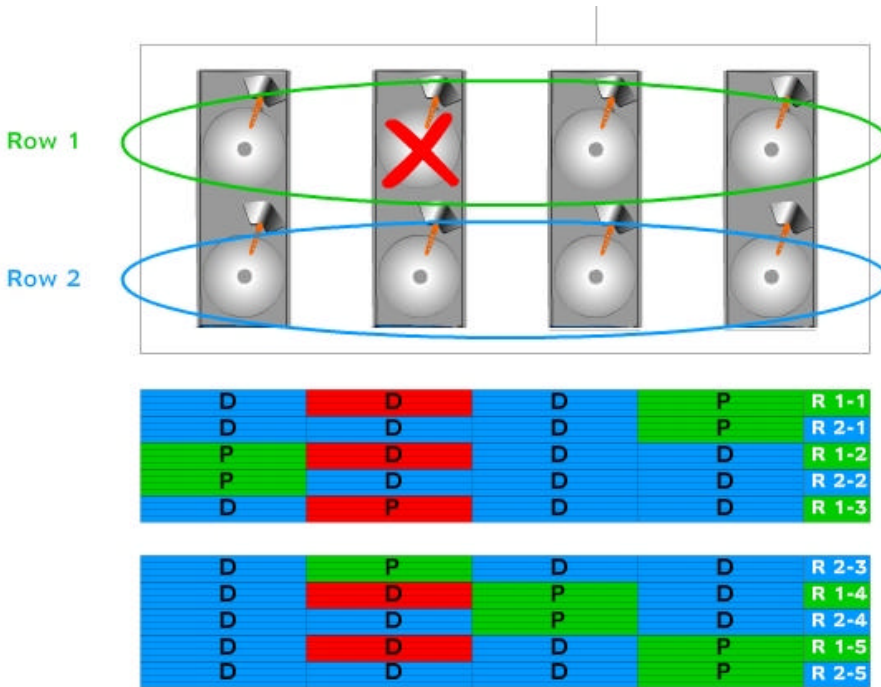
2.4 RAID 50 Interleave Fault Tolerant Examples

2.4.1 Single Drive Failure

Failure of a single drive in a RAID 50 interleave configuration is better handled than with traditional RAID 5. The drive can be in a high density enclosure, The Storage Server Blades track and identify the failed drive and may have proactively flagged it as failed due to excessive error rates or other degraded behaviors. The system allows the removal of the drive pair and caches writes that would have gone to the failed drive as well as to the absent good drive. When the good drive and replacement drive are

reintroduced, the system identifies the good drive and the new drive. Cached writes are transferred to the respective drives and the remaining portion of the new drive is reconstructed as in a traditional RAID 5. The benefits are

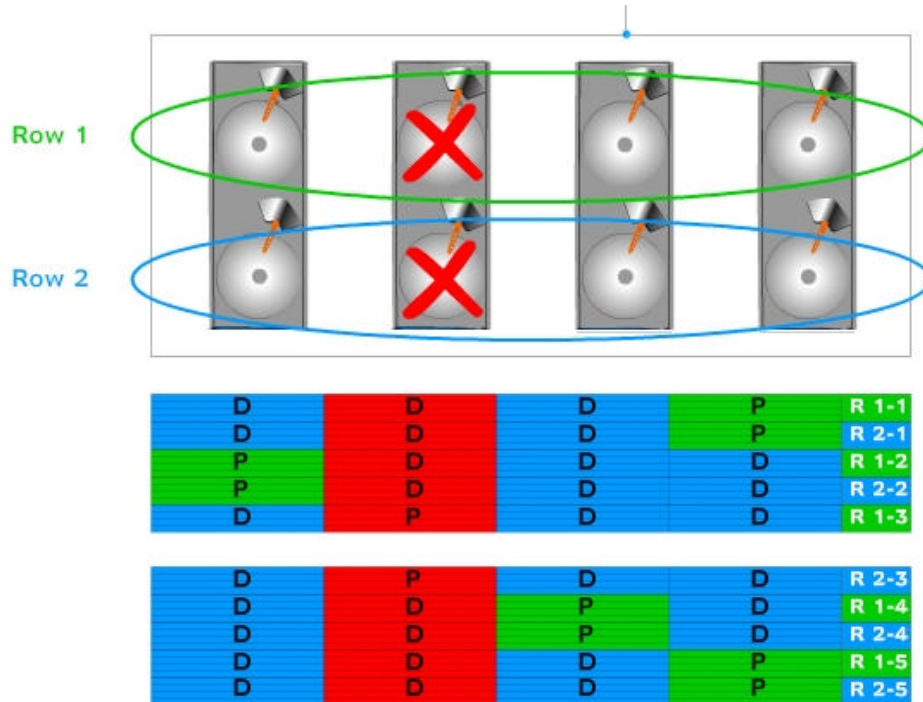
- Improved Performance in Degraded Mode
- Faster Rebuilds = Shorter Rebuild Times
- Lower likelihood of second drive failure
- No RAID 6 Duplex Parity Calculation Required



RAID 50 Interleaved Set in Dual Drive HD Enclosure with single Drive Failure Flagged

2.4.2 Dual Drive Failure – Single Drive Carrier

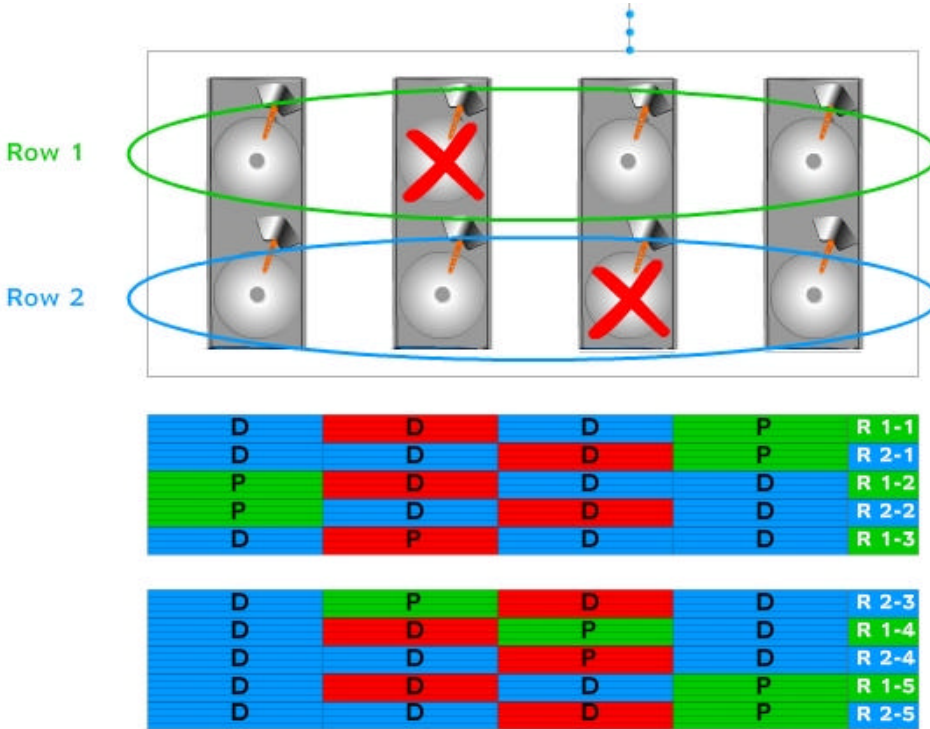
RAID 50 Interleave allows multiple drive failure fault tolerance so long as the failures occur in separate strata within the RAID 50 set. There can be a number of strata across multiple enclosures. There can also be multiple strata arrayed across groups of drives in high density Tiered Storage enclosure configurations.



RAID 50 Interleaved Set in Dual Drive HD Enclosure with Dual Drive Failure Flagged (Same Drive Carrier)

2.4.3 Dual Drive Failure – Separate Drive Carriers

RAID 50 Interleave allows multiple drive failure in separate Carriers within the same storage domain. The unique labels allow the good drives to be tracked even if they were reintroduced in different physical locations. The advance caching and partial rebuild architecture minimize performance impact and dramatically reduce rebuild times.



RAID 50 Interleaved Set in Dual Drive HD Enclosure with Dual Drive Failure Flagged (Separate Carriers)

2.4.4 Dual Drive Failure with Spares available

Adding online hot spares extends the benefits of RAID 50 Interleave further allowing cached writes to be written to the spare set while drives are removed and replaced allowing:

- Extended Drive Removal and Replacement Window
- Minimal Performance Impact in Degraded mode
- Faster Recovery Interval



RAID 50 Interleaved Set in Dual Drive HD Enclosure with Spares Available



3 Summary

RAID 50 Interleave is one of many new architectural concepts MPSTOR is developing to bring about truly Automated Tiered Storage Solutions. MPSTOR is committed to delivering a complete Tiered Storage Architecture that:

- Delivers on the promise of Information Lifecycle Management ILM
- Automates and simplifies Storage Management and Administration
- Dramatically Reduces Storage TCO
- Modular, Portable, and Partner Friendly Software

RAID 50 Interleave delivers benefits including:

- Multiple Drive Failure Tolerant
- Improved Performance over Traditional RAID 5 or RAID 50
- Improved Performance in Degraded Mode
- Faster Rebuilds/ Shorter Rebuild Times
- Lower likelihood of critical second drive failure
- No RAID 6 Duplex Parity Overhead or Performance Penalty

MPSTOR's vision is synchronized with the vision and demands of the evolving Enterprise Storage Market. The industry knowledge and experience attained by the founders established the foundation for the vision. That foundation has been extensively researched and validated with leading industry experts as well as prospective customers. MPSTOR has also put in place a strong team with the core competencies to execute the vision.

**People In Partnership Have the Edge.
Be With Us at the Edge!**

Disclaimer

This document contains internal analysis and opinions as well as forward looking statements that may not occur as anticipated. The Company expressly disclaims any and all liability which may be based on such information, errors therein or omissions there from or liability which may be based on any other written or oral communications transmitted to any party in the course of its evaluation of or entering into any Transaction with the Company. The recipient shall be entitled to rely solely on the representations and warranties made to it in a definitive agreement, when, as and if executed and subject to limitations and restrictions as may be specified in such agreement.

In furnishing this Document, the Company does not undertake any obligation to provide the recipient with access to any additional information. This Document shall not be deemed an indication of the state of affairs of the Company after the distribution hereof and the Company does not intend to update or otherwise revise this Memorandum following its distribution. The Company expressly reserves the right, without giving any reasons therefore, at any time and in any respect, to terminate discussions with the recipient and any other prospective parties to any Transaction or to negotiate with any party with respect to a Transaction involving the Company.