

## **Introducing RAID 5 on Symmetrix DMX**

Date 01/19/2004

---

Copyright © 2004 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC<sup>2</sup>, EMC, Symmetrix, AViiON, CLARiiON, CLARAlert, DG, DG/UX, Navisphere, PowerPath, ResourcePak, VisualSAN, and The EMC Effect are registered trademarks and EMC Automated Networked Storage, EMC ControlCenter, EMC Developers Program, EMC Enterprise Storage, EMC Enterprise Storage Network, EMC Enterprise Storage Specialist, EMCLink, EMC OnCourse, EMC Orbit, EMC Proven, The EMC Effect Alliance, the EMC Information Orb, Access Logix, Application Transparent Failover, AutoIS, Automated Resource Manager, AVALONidm, C-Clip, CacheStorm, Celerra, Celerra Replicator, Centera, CentraStar, CLARevent, Connectrix, CopyCross, CopyPoint, CrosStor, Direct Matrix, Direct Matrix Architecture, EDM, E-Infostructure, E-Lab, Enginuity, FarPoint, FLARE, GeoSpan, HighRoad, InfoMover, MirrorView, OnAlert, PowerVolume, RepliCare, SafeLine, SAN Manager, SDMS, SnapSure, SnapView, SnapView/IP, SRDF, StorageScope, SymmAPI, SymmEnabler, TimeFinder, Universal Data Tone, WideSky, and where information lives are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

Part Number: NL-01-2004

H1114

## Table of Contents

---

|                                                           |           |
|-----------------------------------------------------------|-----------|
| <b>Introduction .....</b>                                 | <b>3</b>  |
| <b>Symmetrix DMX Data Protection Options .....</b>        | <b>4</b>  |
| <b>Symmetrix DMX RAID 5 Implementation Overview .....</b> | <b>5</b>  |
| RAID 5 vs. Parity RAID.....                               | 6         |
| <b>Symmetrix DMX RAID 5 Implementation Details .....</b>  | <b>6</b>  |
| RAID 5 Components.....                                    | 6         |
| Hypervolume .....                                         | 6         |
| Layout of a RAID 5 Device .....                           | 6         |
| RAID 5 3+1 .....                                          | 7         |
| RAID 5 7+1 .....                                          | 7         |
| RAID 5 Modes of Operation .....                           | 7         |
| Normal Mode .....                                         | 7         |
| Regeneration .....                                        | 7         |
| Fast Copy Back.....                                       | 8         |
| Writing Data in a RAID 5 Device.....                      | 8         |
| Reading Data in RAID 5.....                               | 9         |
| RAID 5 Performance Optimization .....                     | 9         |
| Data Recovery with RAID 5.....                            | 9         |
| Write Requests during Regeneration.....                   | 10        |
| Read Requests during Regeneration .....                   | 10        |
| Fast Copy Back after Replacing a Disk .....               | 10        |
| <b>Migrating to RAID 5.....</b>                           | <b>10</b> |
| <b>Conclusion.....</b>                                    | <b>11</b> |

## Introduction

Symmetrix® DMX defines high-end storage by delivering new levels of performance, availability, functionality, and economics. It provides the foundation to implementing Information Lifecycle Management by providing flexibility and choice in meeting different service level requirements with a single, integrated storage infrastructure.

Symmetrix DMX™ offers a wide range of configuration options and features providing users a tiered storage platform to meet different service levels. With the initial introduction Symmetrix DMX, EMC delivered high-performance Parity RAID support as an alternative to RAID 1 storage protection. Parity RAID was architected around hyper-volume-based multi-spindle parity protection, and required fewer disk drives to be configured to provide protected usable capacity than mirrored protection, reducing acquisition costs. Parity RAID offered high performance for many workloads and could be mixed with RAID 1 protection within a single systems to provide flexible data protection option to meeting different service level requirements.

To provide even more flexibility and choice, the Symmetrix DMX now offers RAID 5 as a third storage protection alternative. RAID 5 offers the same economic advantages as Parity RAID, and is architected around block-based multi-spindle parity protection. It implements data striping and rotating parity across all hypervolumes of a RAID 5 device, providing customers equal or, depending on the specific workloads, potentially better performance than Parity RAID. In addition, RAID 5 can require fewer global hot spares within a single system to compensate for drive failures. Symmetrix Optimizer is also supported for RAID 5 (as well as RAID 1) protected devices to further enhance the performance capabilities of RAID 5.

This paper describes the implementation of native RAID 5 on Symmetrix DMX systems with Enginuity™ 5670+.

## Symmetrix DMX Data Protection Options

Symmetrix DMX provides multiple data protection options to ensure uninterrupted access to data in the event of a disk failure or replacement. Within the same Symmetrix DMX system, data volumes can be protected with RAID 1, Parity RAID, and RAID 5 protection options. Each of these options utilizes different implementations, which can impact the performance, operation, and cost of the storage capacity. Global dynamic hot spares can be employed with any of these data protection options for further protection in the event of a drive failure.

A summary of the Symmetrix data protection options is provided in the table below.

| Data Protection       | Description                                                                                     | Features                                                                                                                                                                                                                                                                                                                                                |
|-----------------------|-------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| RAID 1<br>(Mirroring) | Data on one volume is protected by another identical mirrored volume to create a RAID 1 device. | <ul style="list-style-type: none"><li>• Highest performance – I/Os automatically distributed across both volumes</li><li>• Highest availability – remaining volume available without rebuild</li><li>• Efficient failed-drive recovery – single hot spare; drive copied upon replacement</li><li>• Symmetrix Optimizer support</li></ul>                |
| Parity RAID           | Data on (3 or 7) volumes in Parity RAID are protected by a separate parity volume.              | <ul style="list-style-type: none"><li>• High performance dependent upon volume layout and external striping</li><li>• High availability – data from lost volume is regenerated from remaining members</li><li>• Failed drive recovery requires N (3 or 7) hot spares; drive rebuilt upon replacement</li><li>• No Symmetrix Optimizer support</li></ul> |
| RAID 5                | Data and parity are striped across all (4 or 8) hypervolumes to create a RAID 5 device.         | <ul style="list-style-type: none"><li>• High performance with automatic striping across hypervolumes</li><li>• High availability – lost hypervolume regenerated from remaining members</li><li>• Efficient failed-drive recovery – single hot spare; drive copied upon replacement</li><li>• Symmetrix Optimizer support</li></ul>                      |

The following table details the features and capabilities supported by each of the data protection alternatives offered by Symmetrix DMX.

|                                | RAID 1 | RAID 10 aka<br>mirrored<br>striped CKD | RAID 5<br>3+1 or 7+1 | Parity RAID<br>3+1 or 7+1 | Unprotected |
|--------------------------------|--------|----------------------------------------|----------------------|---------------------------|-------------|
| STD                            | √      | √                                      | √                    | √                         | RPQ         |
| Metavolume - concatenated      | √      | N/A                                    | √                    | √                         | RPQ         |
| Metavolume - striped           | √      | N/A                                    | √                    | √                         | RPQ         |
| TimeFinder/Mirror Source       | √      | √                                      | √                    | √                         | RPQ         |
| TimeFinder/Mirror Target       | √      | √                                      | N/A                  | N/A                       | √           |
| TimeFinder/Snap Source         | √      | √                                      | √                    | √                         | RPQ         |
| TimeFinder/Snap Target         | √      | √                                      | √                    | √                         | RPQ         |
| TimeFinder/Mirror Clone Source | √      | √                                      | √                    | √                         | RPQ         |
| TimeFinder/Mirror Clone Target | √      | √                                      | √                    | √                         | RPQ         |
| SRDF/S Source                  | √      | √                                      | √                    | √                         | √           |
| SRDF/S Target                  | √      | √                                      | √                    | √                         | √           |
| SRDF/A Source                  | √      | √                                      | √                    | √                         | √           |
| SRDF/A Target                  | √      | √                                      | √                    | √                         | √           |
| Dynamic SRDF                   | √      | √                                      | √                    | X                         | √           |
| PPRC                           | √      | √                                      | √                    | X                         | √           |
| ECRC                           | √      | √                                      | √                    | √                         | √           |
| Optimizer                      | √      | X                                      | √                    | X                         | N/A         |
| DMSP                           | √      | √                                      | N/A                  | N/A                       | N/A         |
| Mirror Positions Required      | 2      | 2                                      | 2                    | 2                         | 1           |
| Number of Hot Spares           | 1      | 1                                      | 1                    | 3 or 7                    | 1           |

## Symmetrix DMX RAID 5 Implementation Overview

RAID 5 on Symmetrix DMX is an implementation of the industry-standard RAID 5 data protection algorithms optimized for the Direct Matrix Architecture. Symmetrix DMX RAID 5 provides block-based protection with parity rotating across all hypervolumes of a RAID 5 device. A RAID 5 block on Symmetrix DMX constitutes of four Symmetrix logical tracks. This block size is specifically chosen to optimize RAID 5 performance on the Symmetrix DMX architecture. In Enginuity 5670+, the track size is 32KB for open systems and 57KB for mainframes.

With Enginuity 5670+, RAID 5 is implemented as a feature of the operating environment, and is available for all models of the Symmetrix DMX series. RAID 5 is available in one of two configuration choices per array, RAID 5 3+1 or RAID 5 7+1 in which data and parity is striped across four or eight physical disks (respectively). These drives can be configured on a single drive channel/DA pair, or they can be striped across multiple drive channels/DA pairs. Either approach affords approximately the same level of performance as a result of the Symmetrix DMX Direct Matrix Architecture™ and its global memory design. The use of global memory improves data availability in Symmetrix DMX systems with a rotating parity scheme by allowing all read and write operations to be overlapped.

All parity-based RAID implementations, including Symmetrix DMX Parity RAID and RAID 5, use a Boolean XOR calculation to allow reconstruction of data if a drive fails. This method of data protection was first described in a 1988 paper delivered at ACM SIGMOD<sup>1</sup>. The following sections describe how RAID 5 functions in Symmetrix DMX systems.

<sup>1</sup> David A. Patterson, Garth Gibson, and Randy H. Katz. A case for redundant arrays of inexpensive disks (RAID). ACM SIGMOD International Conference on Management of Data (Chicago, IL), pages 109--116, 1--3 June 1988.

## ***RAID 5 vs. Parity RAID***

Parity RAID is implemented by deploying (three or seven) hypervolumes containing data that is protected by using a dedicated hypervolume containing parity information corresponding to the data (three or seven) hypervolumes.

Like Parity RAID, RAID 5 provides cost-effective data protection against drive failure. While the most demanding environments continue to opt for mirrored storage for maximum performance, Parity RAID and RAID 5 offer cost-effective data protection alternatives to meet a wider range of service levels requirements. As compared to the Parity RAID implementation, RAID 5 offers Symmetrix DMX customers the added benefits of:

- High performance: RAID 5 automatically stripes data across all members in the stripe set at the block level. For maximum performance, Parity RAID requires manual striping across hypervolumes. Parity RAID performance has been significantly enhanced since its initial introduction. RAID 5, depending on the specific workload, can deliver performance equal to or better than that of a Parity RAID configuration.
- Single hot spare: RAID 5 uses a single global hot spare to temporarily replace a failing drive. Parity RAID requires either three or seven hot spares as temporary mirrors for the surviving drives. Both ensure that customers' data is continuously protected. RAID 5 requires fewer global hot spare drives, and can reduce the cost for spare disks.
- EMC ControlCenter™ Symmetrix Optimizer support: The allocation of RAID 5 (and RAID 1) devices across physical drives can be automatically optimized by implementing Symmetrix Optimizer. This reallocation maximizes performance and throughput of all volumes based on actual usage patterns.

In addition, RAID 5 devices can be used to support dynamic SRDF® and PPRC, which are not supported with Parity RAID. And while neither Parity RAID nor RAID 5 can be used to protect a TimeFinder™/Mirror BCV, both can be used to protect TimeFinder/Mirror clones, TimeFinder/Snap replicas and all SRDF family R1 and R2 devices.

## **Symmetrix DMX RAID 5 Implementation Details**

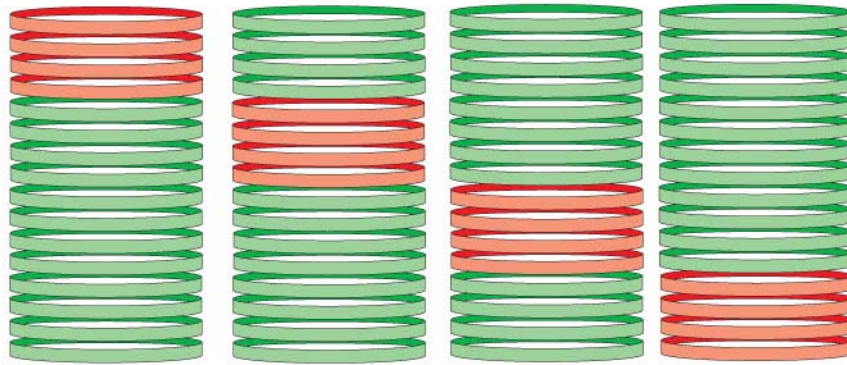
### ***RAID 5 Components***

#### **Hypervolume**

A RAID 5 device is made of four or eight hypervolumes each containing both data and rotating parity. A hypervolume is a unit of storage implemented on a single Symmetrix disk drive. Up to 255 hypervolumes of varying sizes can exist on a physical volume. With Symmetrix Parity RAID, these volumes would either be “Data”, or “Parity” volumes. Symmetrix DMX RAID 5 does not have separate data and parity volumes; RAID 5 stripes both data and parity across all hypervolumes of the RAID 5 device.

#### **Layout of a RAID 5 Device**

A RAID 5 device is composed of four or eight hypervolumes. In a RAID 5 device, all hypervolumes containing data and rotating parity are related to each other for protection. The entire set of RAID 5 hypervolumes is presented to the host as a single LUN or RAID 5 device. The following figure shows a RAID 5 device consisting of four Symmetrix hypervolumes on four separate physical drives with one RAID 5 3+1 device defined.



### **RAID 5 3+1**

Each ring indicates a Symmetrix track. RAID 5 stripe size is set at four tracks. Red rings indicate parity and green rings indicate data. Data and parity are striped across four physical disks.

### **RAID 5 3+1**

A RAID 5 3+1 configuration consists of four Symmetrix devices with data and parity striped across each device. With this approach, effectively 75 percent (or  $3/4^{\text{th}}$ ) of the total storage capacity of a RAID 5 device is available for storing data.

### **RAID 5 7+1**

A RAID 5 7+1 configuration consists of eight Symmetrix devices with data and parity striped across each device. With this approach, effectively 87.5 percent (or  $7/8^{\text{th}}$ ) of the total storage capacity of each RAID 5 device is available for storing data.

Symmetrix RAID 5 3+1 and RAID 5 7+1 cannot be configured in the same Symmetrix DMX system although each can be configured together with RAID 1 volumes and/or similarly configured Parity RAID volumes. A Symmetrix DMX system allows intermixing of different capacity physical disk devices within a single Symmetrix unit. However, all hypervolumes participating in a Symmetrix RAID 5 device must have identical raw storage capacity.

## ***RAID 5 Modes of Operation***

This section describes the following RAID 5 modes of operation:

- Normal Mode
- Regeneration
- Fast Copy Back

### **Normal Mode**

When a RAID 5 device operates with all hypervolumes functioning, it is operating in normal mode. In normal mode, the Symmetrix DMX system accomplishes data redundancy by using the standard EXCLUSIVE OR (XOR) logic to generate and store parity that can then be used to reconstruct the information stored on a failed/failing member.

### **Regeneration**

In the event that a RAID 5 hypervolume fails, the RAID 5 device operates with the surviving members. It is then running in reduced mode. The data on the failing/failed member is reconstructed by XOR'ing the information from the surviving members in the rank. This process is called regeneration.

## Fast Copy Back

If a hot spare is configured in the Symmetrix, it will be integrated into a RAID 5 device when a RAID 5 member disk fails. Data from regeneration is stored on the hot spare. When the failed device is replaced as part of a service action, the information from hot spare is copied back to the newly inserted disk and the original hot spare is returned to the hot spare pool. This process is called *fast copy back*.

## Writing Data in a RAID 5 Device

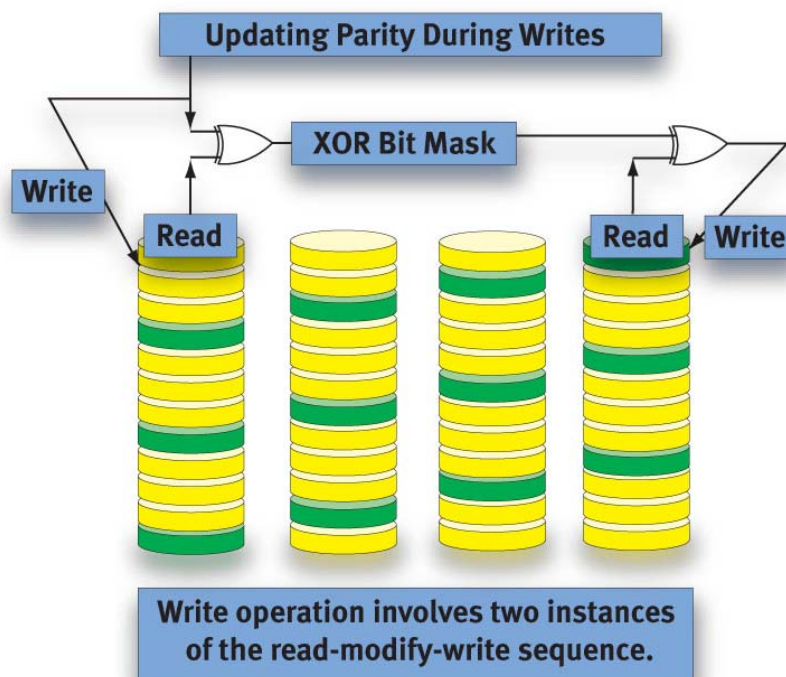
The RAID 5 implementation requires that the controller execute two Read-Modify-Write sequences executed concurrently in time.

| Drive containing data | Drive containing parity |
|-----------------------|-------------------------|
| Read old information  | Read old information    |
| Compute XOR bit mask  | Receive XOR bit mask    |
| Write new information | Write new information   |

The Symmetrix DMX implementation of RAID 5 distributes the work of computing parity between the disk director and the disk drives using the XOR chip located on the disk drive and the disk-level buffer.

Symmetrix DMX native RAID 5 optimizes performance for large sequential writes workloads as there is no need to read the parity from disks. Since many sequential tracks are written, they are all in Symmetrix global memory and the parity is calculated in global memory and information is written to the disks in one stroke without requiring the use of an “expensive” disk-level read-XOR-write operation.

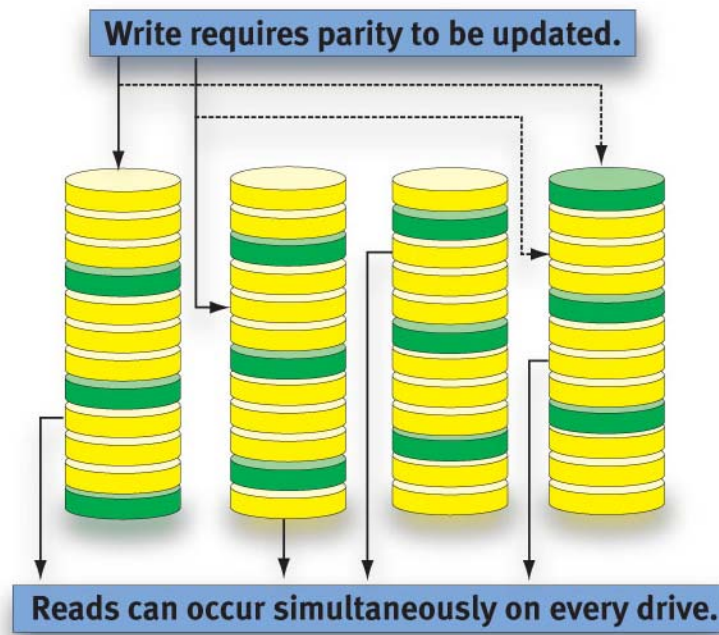
### Example: Parity 3+1 XOR Operation



The green rings represent a four-track stripe of parity and the yellow rings each represent a stripe of four tracks of data

## Reading Data in RAID 5

“Read hit” operations in a Symmetrix RAID 5 are processed through Symmetrix global memory as in normal Symmetrix system processing. Read misses are directed to the disk drive and processed as normal Symmetrix system “read misses”. There is no need of XOR’ing and only one disk drive is involved in servicing the request. With Symmetrix RAID 5 implementation, since data and parity are striped across all members of the RAID 5 rank, up to four (or eight) read operations can be executed concurrently.



## RAID 5 Performance Optimization

The Enginuity operating environment deploys advanced I/O optimization techniques to enhance RAID 5 performance on Symmetrix DMX. On sequential writes to a RAID 5 device, there is no need to read the parity from disk. Since many sequential tracks are written, they are all in Symmetrix DMX global memory and hence parity is calculated in memory and is written to disk in one shot without the “expensive” read-modify-write operation.

Symmetrix Optimizer deployed with RAID 5 on Symmetrix DMX collects statistics on I/O access on individual RAID 5 hypervolumes as well as on RAID 5 devices over extended periods of time (typically a week). Symmetrix Optimizer provides a comprehensive analysis of the statistical data collected to identify over- and under-utilized system resources. For example, Symmetrix Optimizer can determine if hot spots (places of high I/O activity) or cold spots (places of low I/O activity) exist. In line with the Pareto principle, it is typically observed that 20 percent of physical disks are doing 80 percent of the I/O activity. Symmetrix Optimizer observes and analyzes this unbalanced workload and makes recommendations for swapping the hot spindles with the cold spindles. If policy-based automation is chosen by the customer, Symmetrix Optimizer implements the recommendations automatically and transparently. This results in a uniformly distributed workload across all physical spindles which in turn results in enhanced performance and improved resource utilization across the entire Symmetrix DMX system.

## Data Recovery with RAID 5

Symmetrix RAID 5 provides continuous availability for all data if any single physical disk in the RAID 5 device fails or becomes unavailable. If a media error occurs, the affected tracks are regenerated. When a device fails or

becomes unavailable, the Symmetrix RAID 5 device is put in regeneration mode. All data is still available to the host, but is unprotected against additional failures unless protected by a global hot spare.

If a physical device reports too many errors (even recoverable errors), or fails completely, that device is taken offline by the Symmetrix. This condition causes the Symmetrix DMX system to place a remote service call to the Customer Support Center. The Product Support Engineer (PSE) at the support center can determine if a disk drive should be identified for replacement and dispatches a Customer Engineer (CE) who starts the Symmetrix non-disruptive interactive drive replacement process. The logical devices on the disk being replaced are placed in a not-ready state and the associated ranks will begin operating in regeneration mode. After the new drive is in place, the regeneration process begins.

If the Symmetrix DMX system is configured with a global hot spare, it copies the data from the failing device to the global hot spare before the device fails, and reconstructs the data if necessary from the other surviving members of the RAID 5 device. After a hot spare is activated, it behaves as if it is a full fledged member of the RAID 5 device.

## Write Requests during Regeneration

As write requests are made during regeneration, the new data is written to the RAID 5 device consisting of the surviving members and the global hot spare.

## Read Requests during Regeneration

Read requests for data in global memory (read hits) are serviced immediately from global memory. During the regeneration process, when read requests not found in global memory (read misses) occur, the required data is automatically regenerated by reading surviving members and using an XOR calculation, and is delivered to applications that request it.

## Fast Copy Back after Replacing a Disk

After the failed disk is replaced, data is rapidly copied back from the hot spare to the newly installed drive. This is a fast process of simply copying track by track. After copy back is completed, the original hot spare is returned to the global hot spare pool.

## Migrating to RAID 5

Customers who have already deployed Symmetrix DMX systems may want to take advantage of some of the benefits offered by the RAID 5 implementation. EMC offers an extensive variety of tools and services to aid in the migration. Many of these conversion methods are non-disruptive or require only minimal application downtime, ensuring that customers can easily migrate to the new technology.

The following list outlines many of the EMC offerings that support migration to RAID 5:

- TimeFinder clones
- Fast Mirror Migrator for Mainframe (non-disruptive)
- Host-Based LVM Mirroring for UNIX and Windows
- Veritas Volume Manager and VVR
- SDMS™ (mainframe)
- FDRPAS (mainframe)
- TDMF (mainframe and open systems)
- DRU (Windows)
- SAN Copy™ (CLARiiON® source)

## Conclusion

EMC's vision is to help our customers get the maximum value from their information at the lowest total cost of ownership at every point in the information lifecycle. Information Lifecycle Management is a way to map the right service level to the right application at the right cost... at the right time.

The Symmetrix DMX provides flexible data protection options to allow users to meet different performance, availability, functionality, and economic requirements. The ability to support a wide range of service levels with a single storage infrastructure provides a key building block to implementing ILM by deploying a tiered storage strategy.

No matter what requirement or service level you need to meet, EMC delivers a range of offerings and choice to provide the foundation to ILM.