EMC²
where information lives®

White Paper

EMC CLARiiON RAID 6 Technology

A Detailed Review

Abstract

This white paper discusses the EMC® CLARiiON® RAID 6 implementation available in FLARE® 26 and later, including an overview of RAID 6 and the CLARiiON-specific implementation, when to use RAID 6, and reliability and performance comparisons to existing RAID technologies.

July 2007

Part Number H2891

Table of Contents

Executive summary

RAID 6 is generally described as data striped across a number of drives within a RAID group, with two independent parity fields maintained for redundancy. Thus RAID 6 is also referred to as *N+2* or *double parity*. Key points regarding the general concept of RAID 6 are as follows:

- Data is still available even if two drives within the RAID group cannot be read.
  - This means RAID 6 can tolerate two drives that are physically inaccessible for reading.
  - RAID 6 can tolerate a hard media error during a single drive rebuild.
  - RAID 6 can tolerate a second drive failure during a single drive rebuild.
  - RAID 6 can tolerate hard media errors from two drives on the same logical block address.
- Two independent parity fields are maintained.
  - This is required to provide the level of protection noted previously.
  - This N+2 redundancy yields a price point that falls between RAID 5 (N+1) and RAID 1/0 (2N), when considering the number of disk drives that are required to achieve an equivalent usable capacity.
  - RAID 6 may have some performance impact compared to either RAID 5 or RAID 1/0, though not necessarily significant.

The remainder of this paper examines the points just noted. The key points of the paper are as follows:

- RAID 6 provides a more redundant RAID solution than RAID 5 or RAID 1/0 technologies.
- There is a hardware cost (extra drive's worth of parity) and a performance cost associated with implementing RAID 6.
- EMC® CLARiiON® has unique data integrity technologies built into the product that give it an availability advantage over other midrange storage systems regardless of the RAID technology used.

Introduction

The purpose of this white paper is to introduce the reader to CLARiiON RAID 6 technology for CX series systems. This paper:

- Describes the EMC implementation of RAID 6 technology on CLARiiON and how it protects the user from data loss.
- Describes the features of CLARiiON RAID 6 technology and how it interacts with existing CLARiiON software.
- Compares the reliability and performance of RAID 6 with RAID 5 and RAID 1/0 to help determine when RAID 6 is most effective.

Audience

This white paper is intended for EMC customers, partners, and employees who are considering the use of RAID 6. It is assumed that the reader is familiar with CLARiiON RAID technology in general. A good introductory paper is *EMC CLARiiON Fibre Channel Storage Fundamentals*. The examples also assume that the reader is familiar with bitwise XOR operations.

Terminology

Even and odd parity – With even parity, parity bits are set so that there is an even number of 1 bits across the data and parity bits. Odd parity sets the parity bits so that there is an odd number of 1 bits. Both provide the same level of protection.

Remaining consistent with other CLARiiON white papers[1], this paper will use the following terms:

Stripe element – The amount of data written to disk by FLARE before moving to the next disk in the RAID group. The default and recommended value for the stripe element size is 128 blocks or 64K.

Stripe – A stripe is the collection of one stripe element per disk across each disk in the raid group. For a RAID 6 group with four data disks, the stripe size is 4 * 64K, or 256K.

Specific to RAID 6, this paper introduces a new term:

Block-stripe – 512 bytes of data per disk, striped across all disks in the RAID group. It can be thought of as a stripe with a stripe element size of 512 bytes. There are 128 block-stripes in a stripe.

There are several types of errors that may be encountered when attempting to read data from disk:

Soft media error – An error at the drive level that is correctable by the drive. This may indicate that the drive or sector is marginal, but does not indicate any data loss.

Hard media error (unreadable sector) – An error at the drive level that is not correctable by the drive, but can be corrected if a redundant RAID type is used, in which case no data is lost.

Uncorrectable error – An error that is not correctable at the RAID level and results in the loss of information stored on the affected sector(s).

If any of these three errors are encountered, the CLARiiON remaps the affected sector(s) to a new location on disk.

Advantages of CLARiiON RAID technology

RAID 6 at its most basic level protects against up to two disk failures. The CLARiiON implementation of RAID 6 has many unique features that give it an advantage over other RAID 6 implementations. RAID 6 is the only CLARiiON RAID type to offer parity of checksums, while the other features listed here benefit all RAID types within the CLARiiON.

- Parity beyond RAID – CLARiiON sectors are formatted to 520 bytes, 8 bytes more than a normal 512-byte sector. The extra 8 bytes contain checksums of the sector and proprietary RAID stamps. This information allows the CLARiiON to protect against data corruption. These 8 bytes always travel with the data while it is in the CLARiiON. This protects against memory and media errors as well as a power failure during a multistep write operation.
- Parity of checksums – RAID 6 takes parity beyond RAID one step further by recording a parity of the checksums that are kept on each sector of data. This allows CLARiiON to maintain parity on a bad block, allowing for the preservation of data on good disk sectors within the same stripe as a bad sector.
- Proactive hot sparing – Starting with FLARE® release 24, CX series storage systems include proactive hot sparing, which invokes a proactive copy of a disk that is detected to be failing. The intention is to copy all or most of the data from the failing drive to a hot spare before a rebuild is necessary.
- SNiiFFER – Hard media errors are often latent; if the sector is never read, the error is never reported. They are always detected during a RAID 5 rebuild because every sector on every disk needs to be read. At this point, it is too late to reconstruct the data. SNiiFFER is a CLARiiON background process that verifies all the data on all the RAID groups in a system and proactively corrects errors. It scans all RAID groups in parallel at a constant rate. For 300 GB drives, the scan completes about once weekly and the rate is still low enough that there is no noticeable effect on host I/O.
- Distributed parity – Unlike other RAID 6 implementations in the industry where parity is stored on dedicated drives, CLARiiON RAID 6 rotates the parity evenly among all drives in the RAID group so

[1] *EMC CLARiiON Fibre Channel Storage Fundamentals* provides further explanation of stripes and stripe elements.

that parity drives are not a bottleneck for write operations. Parity rotation mainly benefits small, random writes, but CLARiiON has also designed the rotation scheme to benefit large, sequential transfers.

- Write coalescing and full stripe writes – If a write is as large as or larger than the RAID stripe size and properly aligned, it is written to the disks as a single operation, eliminating the need to read existing data and recalculate parity. For example, a RAID 6 group with four data disks has a RAID stripe size of 256K, so writes of that size or larger are a single operation. CLARiiON also has the ability to coalesce adjacent writes within write cache before writing to disk. This allows the CLARiiON to perform full stripe writes and minimize disk I/O even when host writes are smaller than the stripe size or misaligned.

CLARiiON RAID 6 implementation

CLARiiON RAID 6 technology is based on the EVENODD[2] algorithm. The algorithm uses exactly two disk drives' worth of parity to tolerate the loss of data on up to two disk sectors at the same logical block address. The algorithm is extremely efficient with parity calculations, requiring only simple XOR logic. It does not require any additional hardware or recursive computations (for example when diagonal parity depends on row parity) as with other implementations. CLARiiON RAID 6 distributes parity among all drives in the RAID group so that performance is uniform. Other implementations of RAID 6 use two dedicated parity drives, and those two drives become a bottleneck because they have to service multiple I/Os for each write to the RAID group. Distributed parity and many other optimizations that have been built into CLARiiON RAID 5 technology for years have been translated to RAID 6 to optimize both reliability and performance.

RAID 6 overview

RAID 6 requires two disk drives' worth of parity. One is row parity (RP), which is horizontal parity of the data for a block location within one block-stripe; this is similar to RAID 5. The other is diagonal parity (DP), which is unique to RAID 6. Diagonal parity is the parity of the data diagonally across bits in a block-stripe. RP and DP are completely independent of one another and contained within a stripe. Figure 1 shows a simplified example of how RAID 6 parity is stored with six disks in a RAID group.

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	
RP	DP	D	D	D	D	} 8 stripes
D	D	RP	DP	D	D	
D	D	D	D	RP	DP	

Figure 1. RAID 6 parity placement for a six-disk RAID group

Note that after every 8 stripes of data, parity is rotated to a different pair of disks. This is in contrast to having dedicated parity disks, which can often become a bottleneck when the RAID group is servicing many write operations. Figure 2 zooms in a level to show how parity information is stored within a block-stripe. The rest of the examples in this section represent data within a single block-stripe.

[2] M. Blaum, J. Brady, J. Bruck, and J. Menon, "EVENODD: An Efficient Scheme for Tolerating Double Disk Failures in RAID Architectures," *IEEE Transactions on Computers*, Vol. 44, No. 2, February 1995

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP	
1	0	1	0	**0**	**0**	
0	0	1	0	**1**	**0**	1 block-stripe
1	1	0	0	**0**	**1**	
0	1	0	1	**0**	**1**	

Figure 2. RAID 6 parity within a stripe

Figure 2 roughly represents the bits in a block-stripe of data in the RAID group (an actual block-stripe has many more bits and would not be possible to illustrate here). Each bit in these examples actually represents many bits on disk. Each block-stripe maintains its own RP and DP independent of the parity of other block-stripes. The diagonal parity is calculated across the 512-byte blocks on each disk. By calculating parity based on a diagonal contained within the block-stripe, no additional disk blocks must be read from the drive than are necessary for calculating traditional row parity. Additionally, CLARiiON RAID 6 maintains the ability to perform full stripe writes[3] for optimized performance.

Row parity is always even for each row of data as illustrated in Figure 3. For example, the RP calculation for the first row (orange) is $1 \oplus 0 \oplus 1 \oplus 0 = 0$. Row parity alone is essentially RAID 5 and can be used to tolerate the loss of data from any single drive in the RAID group.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
1	0	1	0	**0**	
0	0	1	0	**1**	
1	1	0	0	**0**	
0	1	0	1	**0**	

Figure 3. Row parity calulations

Diagonal parity is a little more difficult to visualize. The diagonal parity may be even or odd depending on the parity of a particular diagonal that is purple in Figure 4. This parity is denoted *S* and is an integral part of the EVENODD algorithm. In this example, $S = 1 \oplus 0 \oplus 0 = 1$, so all diagonal parity is odd for this stripe. The value of *S* is not directly stored on disk, but is factored into each diagonal parity calculation. For example, the parity for the yellow diagonal is calculated as $1 \oplus 0 \oplus 0 \oplus S = 0$. That 0 is then stored in the corresponding yellow DP field.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
1	0	1	0	**0**	**0**
0	0	1	0	**1**	**0**
1	1	0	0	**0**	**1**
0	1	0	1	**0**	**1**

Figure 4. Diagonal parity calulations

[3] The *EMC CLARiiON Best Practices for Fibre Channel Storage* white paper on EMC.com provides an explanation of full stripe writes and their benefits.

In these figures, the orange diagonal is the only one with four data elements. In reality, the other diagonals are padded to the same number of data elements with null data. This is part of the EVENODD algorithm, but does not require any additional storage or computations. The EVENODD paper provides more detailed information on the parity calculations.

RAID 6 data recovery

RAID 6 uses two independent parity sets so that data can be recovered when any two drives in the RAID group are unreadable. This maintains data availability in the following situations:

- Two disk drives are physically inaccessible for reading.
- A hard media error occurs during a single drive rebuild.
- A second drive failure occurs during a single drive rebuild.
- Hard media errors occur from two drives on the same logical block address.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
	0		0	**0**	**0**
	0		0	**1**	**0**
	1		0	**0**	**1**
	1		1	**0**	**1**

Figure 5. Two disk drives in a RAID group are unreadable

For example, suppose two data drives in the same RAID group are unreadable as in Figure 5. All the data on these drives can be reconstructed using the independent row and diagonal parity. The key to data recovery with the EVENODD algorithm is that the parity of all RP and DP bits is always equal to *S*. This is proven mathematically in the EVENODD paper. In this example, the XOR of all the RP and DP bits in the block-stripe is odd, therefore *S* is odd (*S* = 1). Since *S* is also the parity of the purple diagonal in Figure 6, the one unknown bit can be calculated. $S = 1 = 1 \oplus unknown \oplus 0$, therefore the unknown bit must be 0.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
	0		0	**0**	**0**
	0		0	**1**	**0**
	1		0	**0**	**1**
	1		1	**0**	**1**

Figure 6. Using the *S* value to recover the first diagonal

Once that piece of data is recovered, it is straightforward to use the row parity and recover the missing data for the blue row in Figure 7. Since row parity is always even, the missing data must be a 1.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
	0		0	**0**	**0**
	0		0	**1**	**0**
	1	0	0	**0**	**1**
	1		1	**0**	**1**

Figure 7. Recovering the first row

Once the first row is recovered, there is now exactly one diagonal (green) with exactly one missing data element. Parity for the green diagonal is calculated as $1 \oplus 0 \oplus unknown \oplus S = 1$. Since S=1, the unknown data must be a 1.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
	0		0	**0**	**0**
	0		0	**1**	**0**
1	1	0	0	**0**	**1**
	1		1	**0**	**1**

Figure 8. Recovering the next diagonal

Once this diagonal is recovered, there is exactly one row with one missing data element. This pattern continues until all data is recovered. The order of recovery is indicated by the lettering in Figure 9.

Disk 0	Disk 1	Disk 2	Disk 3	RP	DP
D	0	C	0	**0**	**0**
H	0	G	0	**1**	**0**
B	1	A	0	**0**	**1**
F	1	E	1	**0**	**1**

Figure 9. Order of recovery

There are other situations, such as when parity is lost, where the recovery method is slightly different. The EVENODD paper discusses these situations.

Features and functionality

CLARiiON RAID 6 technology is compatible with the storage-system features that enhance RAID on CLARiiON. RAID 6 can co-exist with other RAID types (such as RAID 5, or RAID 1/0) within the same storage system or enclosure. CLARiiON virtual LUN technology can be used to migrate data between these varied RAID types while maintaining host access to the LUN.

Basic features

CLARiiON RAID 6 technology for CX and CX3-series storage systems is available in FLARE release 26 and later. It supports RAID groups with 2, 4, 6, 8, 10, 12, and 14 data disks. The equivalent of two drives' worth of space is added for parity. Shorthand for RAID group sizes is denoted as 2+2, 4+2, 6+2, etc., where

the "+2" indicates that there are two drives' worth of parity data. RAID 6 is treated similarly to the other RAID technologies by Navisphere® Manager and Secure CLI. Binding a RAID 6 LUN is an option for any RAID group created with an even number of disks.

Hot sparing

CLARiiON RAID 6 technology supports global hot sparing and follows the same hot spare rules as the other RAID types. With RAID 6, up to two drives that have been failed by FLARE may be rebuilding at the same time, while the RAID group's LUNs remain online for host access.

For proactive hot sparing, one drive in a RAID 6 group may be performing a proactive copy. A second drive may be rebuilding during that proactive copy.

MetaLUNs

CLARiiON RAID 6 technology supports metaLUN expansion via concatenated or striped expansion. All component LUNs of the metaLUN must be RAID 6 LUNs. CLARiiON does allow mixed RAID types for concatenated LUNs in general. However, because RAID 6 is much more reliable than other RAID types, EMC does not support metaLUNs that mix RAID 6 with other RAID types.

Limitations

As of the writing of this paper, the following limitations apply to the CLARiiON RAID 6 implementation. These have been put in place to ensure the highest code reliability possible at initial release. RAID group expansion and defragmentation are targeted for a future release.

- RAID 6 is only supported for CX series platforms that are capable of running FLARE 26 or later.
- RAID group expansion (adding new disks to an existing RAID group) is not supported. You may easily add more spindles or capacity to existing LUNs by using metaLUN technology.
- RAID group defragmentation is not supported. If there is not enough free contiguous space in the RAID group to bind a LUN of desired capacity, a concatenated metaLUN may be built to utilize the fragmented space. RAID group fragmentation only affects the free capacity available to bind new LUNs and has no negative impact on performance.
- Command line management is only supported by Navisphere Secure CLI. Classic CLI and Java CLI are being phased out in favor of the Secure CLI and are not able to manage RAID 6 technology.

Reliability

Reliability is the key criteria for selecting RAID 6 versus other RAID technologies. RAID 5 and RAID 1/0 have good reliability, while RAID 6 has great reliability. Reliability can be modeled through probabilistic calculations, which are useful for profiling a new technology. However, it is also important to look at the bigger picture of how RAID fits in as one of many data protection mechanisms.

Theoretical reliability considerations for RAID

This section examines the theoretical reliability of RAID 6 as compared to RAID 5 and RAID 1/0. This paper investigates the mean time to data loss (MTTDL) due to two different failure modes:

- Losing data due to multiple disks failing during a rebuild – This is based on the predicted mean time to failure (MTTF) for disk drives. Drive vendors typically quote MTTF values of 500,000 to 1,000,000 hours. For RAID 6, a third drive must fail while two other drives are rebuilding for data loss to occur. The MTTDL due to multiple disks failing during a rebuild depends on the MTTF of the individual disk drives, the size of the RAID group, and the mean time for a rebuild to complete.
- Losing data upon encountering an uncorrectable error during a rebuild – A hard media error at the drive level is correctable by RAID unless two drives are rebuilding (RAID 6), or one rebuilding (RAID

5). Drive vendors publish uncorrectable error rates (UER) for their drives. UER is defined as the average number of bits read without encountering a hard media error. For all SATA and Fibre Channel drives used in CLARiiON CX storage systems, the UER is at least 10^{15} bits. The MTTDL that is due to an uncorrectable error during a rebuild depends on the MTTF of the individual disk drives, the size of the RAID group, the mean time for a rebuild to complete (for RAID 6 only), the UER, and the capacity of the drives.

When comparing the MTTDL calculations[4] for different inputs, the theoretical models produce the following results:

- The MTTDL due to an uncorrectable error during a rebuild is much less than the MTTDL due to multiple disk failures during a rebuild. This means an uncorrectable error during a rebuild causes data loss more often than multiple disk failures.
- As disk capacity increases, the MTTDL due to an uncorrectable error during a rebuild decreases. This means that RAID groups of larger drives are more likely to encounter an uncorrectable error than RAID groups of smaller drives (purely because there are more bits to be read).
- As the number of disks in a RAID group increases, the MTTDL due to an uncorrectable error during a rebuild of that RAID group decreases. Again, this is because there are more bits to read during the rebuild. The exception is RAID 1/0, which only needs the other disk in the mirrored pair to rebuild.
- As rebuild time increases, the MTTDL decreases. The longer the rebuild, the higher the probability of a second drive failure during the rebuild.

While these theoretical models are useful for comparing different RAID technologies, they are just models. They rely on two predicted values, MTTF and UER. If the drive behavior in the field is substantially different, then the models lose their accuracy. The models also do not account for factors such as human error, nor do they account for the extra data protection mechanisms discussed in the "Advantages of CLARiiON RAID technology" section.

CLARiiON proactive hot sparing is specifically designed to avoid data loss due to encountering a hard media error during a rebuild. By copying data to a hot spare before the drive fails completely, proactive hot sparing can avoid a rebuild completely. If a hard media error is encountered during the copy operation, it is simply rebuilt from redundant RAID information.

The SNiiFFER process is designed to detect and rebuild hard media errors on otherwise healthy drives before a rebuild occurs. These CLARiiON-specific features are not modeled in reliability calculations, but work to further reduce the probability of data loss.

A holistic view of reliability

Data reliability is more than just RAID. RAID still leaves a single copy of data on a single set of disks, which makes it susceptible to human error, software failures, site failures, and other factors. Hardware failures are only a subset of all data loss incidents and only a subset of those hardware failures are related to disk drives and RAID. To protect the most critical data, it is important to look beyond RAID to other factors that impact overall data reliability and availability. Multiple, complementary technologies are often used to ensure the highest levels of reliability and availability. EMC offers many ways, above and beyond RAID, to protect data on CLARiiON storage systems, such as:

- Backup to disk with CLARiiON or EMC Disk Library – Backups are an important form of data protection. They protect against human error, and software and hardware failures that RAID cannot protect against, by restoring data back to a previous point in time. However, restores, even from disk, can often be quite lengthy. Therefore, avoiding the need for restores by protecting against drive failures at the source, via technology such as RAID, is still important.
- Business Continuity offerings such as SnapView™, SAN Copy™, MirrorView™, RecoverPoint, and RepliStor® – These products offer methods to maintain a standby copy of production data so that in

[4]Supporting mathematics are provided in: P. Chen, et al., "RAID: High-Performance, Reliable Secondary Storage," *ACM Computing Surveys*, Vol. 26, Issue 2, June 1994, pp. 145-185.

the event of a failure, operations can be switched to the standby copy. The remote replication products that maintain a copy on a second storage system (usually at another site) have the ability to minimize application downtime even during an entire site outage. These failover-type solutions avoid the cost of restoring backups because the data is already where it needs to be.

In addition to having multiple data protection mechanisms in place, it is also important to follow EMC best practices for managing and servicing a CLARiiON storage system. For example, the Navisphere Service Taskbar (NST) guides users through common operations such as replacing a disk drive or upgrading FLARE software. The goal of the NST is to standardize common service procedures and minimize human error as much as possible. The *EMC CLARiiON CX3 Best Practices for Achieving "Five Nines" Availability* white paper provides recommendations on achieving the highest levels of availability and reliability from a CLARiiON storage system.

Performance

Customer solutions with RAID systems are normally sized and designed based on:

- Reliability – This paper has established that RAID 6 is more redundant than RAID 5 or RAID 1/0 technologies. Looking at only the reliability perspective, RAID 6 is an obvious choice.
- Capacity – To obtain the same amount of user capacity in a RAID group, RAID 6 requires one more drive than RAID 5 and N-2 less drives than RAID 1/0. Looking at only reliability and capacity, RAID 6 does cost more than RAID 5 for the same capacity, but has an advantage in both over RAID 1/0.
- Performance – The most difficult factor in a design is performance since storage-system performance always depends on application workload. This section compares RAID 6 with other CLARiiON RAID technologies.

I/O performance

Performance in an intelligent RAID system such as a CLARiiON depends primarily on the disks, but also depends on other effects such as controllers, caching, data layout, and application / file system effects. This section explores performance in two ways:

- Single RAID group performance mitigates the effects of controllers and caching to focus on the performance of the disks and the RAID technology.
- Overall system performance looks at a fully configured system of each RAID type to include factors such as caching and controllers.

All results presented in this section are based on internal EMC testing. All testing was performed on a CX3-80 with Fibre Channel disk drives. The results are intended as rules of thumb. For in-depth performance assistance, ask your EMC Sales Representative to engage a local CLARiiON SPEED resource.

Single RAID group performance

The single RAID group tests for RAID 5 and RAID 6 were designed to focus on the performance of the RAID technology. Comparisons are based on the same number of data disks (for example a 4+1 RAID 5 was compared directly with a 4+2 RAID 6). Since caching can mask disk performance characteristics, all caching was disabled to isolate the effects of the individual RAID technologies. The general results of these tests can be summarized as follows:

- RAID 6 read performance is slightly better than RAID 5 for the same number of data disks. This can be attributed to the extra disk required for RAID 6, and the fact that parity is rotated so all drives hold data. For example, comparing a 4+1 RAID 5 to a 4+2 RAID 6, the 4+2 has an extra disk to service reads.
- When write cache is disabled for both RAID types, RAID 6 performance for small, random writes is lower (in terms of IOPS) than RAID 5. With caching enabled (which is normal operation for the

CLARiiON), the CLARiiON coalesces adjacent writes and acts as a buffer so performance is much less dependent on the RAID type.

- RAID 6 performance for large writes (as large as the stripe size and properly aligned) is similar to RAID 5. Full stripe writes are performed in both cases, so there is a negligible write penalty for the extra parity calculation of RAID 6.

Overall system performance

The overall system performance tests compared a 480-drive CX3-80 configured with 4+2 RAID 6 groups to one configured with 4+1 RAID 5 groups. This test utilizes read and write caching and pushes the limits of the system. The workloads were such that they often pushed the limits of the storage processors rather than the limits of the disks. The findings of these tests were:

- RAID 6 read performance is similar to RAID 5. The system limits for read bandwidth and IOPS are more dependent on the SPs and cache than on the disks.
- RAID 6 write performance is less than that of RAID 5. The extent of the difference varies depending on write sizes and RAID group sizes, but the extra parity information that has to travel on the back-end buses to disk results in lower system bandwidth for a system with all RAID 6.

Rebuild performance

EMC conducted internal testing of drive rebuilds within RAID 6 groups. Testing was performed on a CX3-40 with 73 GB, 15k rpm Fibre Channel disk drives and 500 GB, 7200 rpm SATA disk drives. All tests were performed with no application load. Rebuild times depend on a number of factors, including the rebuild rate, hot spare type and location, disk types, disk sizes, bus speed, application load, and RAID type. Compared to a single drive failure on a RAID 5 4+1 (the relative comparisons hold for either all SATA or all fibre configurations):

- A RAID 6 4+2 rebuild can take up to 10 percent longer.
- A RAID 6 8+2 rebuild can take up to 60 percent longer (up to 10 percent longer than a RAID 5 8+1).
- A RAID 6 14+2 rebuild can take up to 200 percent (three times) longer (up to 10 percent longer than a RAID 5 14+1).

To summarize, a RAID 6 single drive rebuild will take up to 10 percent longer than a RAID 5 single disk rebuild if all other parameters remain constant.

If two disk drives in the same RAID 6 group fail at the same time, the two rebuilds occur simultaneously. Test results for the scenario where two drives fault at the same time show that a full rebuild of both drives takes up to 100 percent (two times) longer than a single drive rebuild for same size RAID 6 group.

The rebuild process includes only the rebuild of data to the hot spare. Once the failed drive(s) that were originally members of the RAID group are replaced, there is an equalize operation that copies data from the hot spare back to the original drive. Since this is a disk to disk copy, the RAID type does not have a significant impact on the time to complete the operation. Similar to rebuilds, equalizes can happen in parallel for two drive failures.

Migrating to RAID 6

With FLARE release 26 and later, users can configure new RAID groups as RAID 6. Existing LUNs can be easily converted to RAID 6 protection using CLARiiON virtual LUN technology to move existing LUNs to new RAID 6 LUNs[5]. The move is performed by the CLARiiON without disruption to host applications. There is no need for downtime or reconfiguration. The new RAID 6 LUN assumes the exact identity of the existing LUN during the migration. With virtual LUN technology, users can easily move existing data to

[5] LUN migration may require some free disk space. The exact amount depends on the size of the LUNs and the raw capacity requirements of the new RAID 6 configuration compared to the raw capacity usage of the existing configuration.

RAID 6 and maintain the flexibility of several different RAID protection levels. *EMC Virtual LUN Technology* provides more information.

Conclusion

CLARiiON RAID 6 technology offers more reliable RAID protection than other RAID technologies. It can tolerate two failed disks in a RAID group and more importantly hard media errors encountered during a single disk rebuild. CLARiiON offers many built-in protection mechanisms beyond basic RAID technology, so users have an added level of protection regardless of the RAID type used.

The added reliability may come at the cost of performance when compared with other RAID types. RAID 6 has a disadvantage to RAID 5 and RAID 1/0 for small, random writes and system write bandwidth performance, but other I/O profiles are not affected as significantly.

RAID is one of many ways to protect data. Even with good RAID protection, it is still important to evaluate threats to data loss and implement proper backup and replication strategies to ensure that your most important data is protected in all situations.

References

The following white papers can be found on EMC.com and Powerlink®, EMC's password-protected customer- and partner-only extranet:

- *EMC CLARiiON Fibre Channel Storage Fundamentals*
- *EMC CLARiiON CX3 Best Practices for Achieving "Five Nines" Availability*
- *EMC CLARiiON Global Hot Spares and Proactive Hot Sparing*
- *EMC Virtual LUN Technology*

Other references include the following:

- M. Blaum, J. Brady, J. Bruck, and J. Menon, "EVENODD: An Efficient Scheme for Tolerating Double Disk Failures in RAID Architectures," *IEEE Transactions on Computers*, Vol. 44, No. 2, February 1995
- P. Chen, et al., "RAID: High-Performance, Reliable Secondary Storage," *ACM Computing Surveys*, Vol. 26, Issue 2, June 1994, pp. 145-185