

# EMC® VMAX3™ RELIABILITY, AVAILABILITY, AND SERVICEABILITY

## Technical Note

P/N H13807  
REV A04  
October, 2015

This technical notes document contains information on these topics:

|                                                                           |           |
|---------------------------------------------------------------------------|-----------|
| <b>Executive Summary.....</b>                                             | <b>4</b>  |
| Audience.....                                                             | 5         |
| <b>Introduction.....</b>                                                  | <b>6</b>  |
| <b>Remote support.....</b>                                                | <b>6</b>  |
| Supportability through the management module control station .....        | 7         |
| Secure Service Credential (SSC), Secured by RSA.....                      | 8         |
| Error detection.....                                                      | 8         |
| Block CRC error checks.....                                               | 8         |
| Data integrity checks.....                                                | 9         |
| Drive monitoring and correction .....                                     | 9         |
| Physical memory error verification and error correction.....              | 9         |
| Reliable components .....                                                 | 10        |
| <b>VMAX3 component-level redundancy .....</b>                             | <b>10</b> |
| Redundant engine components .....                                         | 10        |
| Redundant director boards.....                                            | 12        |
| Management module control station (MMCS) and management module (MM) ..... | 13        |
| Management module control station (MMCS) .....                            | 13        |
| Management module (MM) .....                                              | 14        |
| Flash I/O module.....                                                     | 15        |
| Front-end I/O module .....                                                | 15        |

|                                                    |           |
|----------------------------------------------------|-----------|
| Back-end I/O module .....                          | 15        |
| InfiniBand (IB) module .....                       | 15        |
| Shared physical memory .....                       | 16        |
| Global memory technology overview .....            | 16        |
| Physical port numbering.....                       | 16        |
| Logical port numbering .....                       | 17        |
| Channel front-end redundancy .....                 | 17        |
| SAS back-end redundancy .....                      | 18        |
| Redundant back-end director connectivity .....     | 18        |
| Redundant cable paths .....                        | 19        |
| Redundant drive paths .....                        | 19        |
| Point-to-point back-end .....                      | 19        |
| Drive Array Enclosure (DAE) .....                  | 19        |
| Dual-initiator feature .....                       | 22        |
| Dynamic virtual matrix.....                        | 22        |
| Internal environmental Ethernet connectivity ..... | 23        |
| Redundant power subsystem .....                    | 24        |
| Battery backup unit modules .....                  | 24        |
| Vaulting.....                                      | 25        |
| Vault triggers.....                                | 25        |
| Power-down operation .....                         | 26        |
| Power-up operation .....                           | 26        |
| <b>Data protection methods .....</b>               | <b>26</b> |
| RAID 1 (mirroring).....                            | 27        |
| RAID 5 .....                                       | 27        |
| RAID 6 .....                                       | 27        |
| Local RAID .....                                   | 28        |
| Thin provisioning.....                             | 28        |
| Drive sparing .....                                | 29        |
| Spare Drive Replenishment Process .....            | 30        |
| Local replication using TimeFinder .....           | 31        |
| Remote replication using SRDF.....                 | 31        |

|                                                   |           |
|---------------------------------------------------|-----------|
| Data at Rest Encryption (D@RE) .....              | 34        |
| <b>VMAX3 component-level serviceability .....</b> | <b>35</b> |
| Nondisruptive HYPERMAX OS upgrades .....          | 36        |
| Flashing blue rack lights.....                    | 37        |
| <b>Conclusion .....</b>                           | <b>37</b> |
| References .....                                  | 37        |
| <b>Appendix A: Resiliency Testing.....</b>        | <b>38</b> |
| EMC Internal QE Testing .....                     | 38        |
| On-site Resiliency Testing .....                  | 39        |

# Executive Summary

The IT industry has been growing and changing at a rapid pace, with a large area of focus on cloud computing and cloud-like consumption models. IT organizations are finding that they must be able to cater to both internal and external customers' storage needs, while maintaining alignment with the budget that is provided to them.

VMAX3 is a platform that can accommodate all of these needs. It provides a simple, service level-based provisioning model that changes the way users consume storage, taking the focus away from the back-end configuration steps and allowing them to concentrate on other key roles.

While simplifying storage consumption is critical, there are other features that also make VMAX3 a powerful platform that should be considered. Through redundant hardware components and intelligent software architecture, VMAX3 is able to deliver extreme performance while also providing high availability levels approaching 6 nines of availability. This combination makes the system exceptionally reliable overall, while also leveraging components in new and innovative ways which will decrease the total cost of ownership of each system.

Important functionality such as local and remote replication of data used to deliver business continuity must cope with more data than ever before, without impacting production activities. Furthermore, at the end of the day, all of these challenges must be met while continually improving economics.

Highlights of VMAX3 include:

- **Service level-based provisioning:** The requirements of a storage consumer revolve around capacity and performance. It is important for them to have the right amount of capacity to fulfill their needs as well as the proper performance levels. Service level-based provisioning gives users the opportunity to select a predefined service level, provision storage to it, and get the performance they need, without worrying about all of the back-end configurations.
- **Performance:** The processing power, memory, and bandwidth available in VMAX3 deliver high levels of predictable performance. This means providing the performance to meet the IOPS and MB/s needs of the world's most demanding applications. VMAX3 also provides the ability to absorb I/O bursts due to unpredictable workloads, such as end-of-quarter processing and performance intensive *ad hoc* queries, while continuing to guarantee consistent application-response times.
- **Information-centric security, built-in:** VMAX3 and integrated RSA technology provide industry-leading, information-centric security to secure people, infrastructure, and data. VMAX3 security features reduce risk and protect information. Organizations can authenticate, authorize, and audit activities on VMAX3 systems and devices.
- **Always-on availability (HA/RAS):** The VMAX3 family's reliability, availability, and serviceability (RAS) feature makes it the ideal platform for enterprise environments requiring always-on availability. The VMAX3 family is architected to

provide six-nines (99.9999%) of availability in the most demanding, mission-critical environments. The many redundant features of the VMAX3 architecture discussed in this document are taken into account in the calculation of overall system availability. This includes redundancy in the back-end, cache memory, front-end and fabric, as well as the types of RAID protections given to volumes on the back-end, and local and remote replication software features such as TimeFinder and Symmetrix Remote Data Facility (SRDF). Calculations may also include time to replace failed or failing FRUs (field replaceable units) which also take into account customer service levels, replacement rates of the various FRUs, and hot sparing capability in the case of drives.

### Always on with VMAX3

| Eliminate Costly Downtime                                                                                                                   | Exceed Stringent Replication SLAs (RTO, RPO)                                                                             | Eliminate Planned Downtime                                                                                                     | Ensure 100% Data Integrity                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
|                                                            |                                         |                                             |              |
| <b>Proven 6 Nines of Availability</b><br>Advanced Fault Isolation, map-out faulty memory DIMMs, mirrored memory no single points of failure | <b>Gold Standard in Multi-Site Replication</b><br>Proven Disaster Recovery and rapid restart; 2-site, 3-site replication | <b>Non-Disruptive HW and SW Upgrades</b><br>Continuous IO through parallel microcode NDUs, upgrade HYPERMAX O/S within seconds | <b>T10 DIF Data Coding</b><br>Single Bit Error Correction, validation checksum through T10 DIFF |

VMAX3 has raised customer expectations for high-end storage in terms of availability, performance, replication, scalability, and management. High-end availability is more than just redundancy; it means nondisruptive operations and upgrades, and being “always online.” Delivering high-end performance means handling all workloads, predictable or not, under all conditions. High-end replication means any amount of data anytime and sending it any distance. High-end scalability means more than capacity; it means having the flexibility to handle any service level or application, cost-effectively, no matter how your business changes. And high-end storage management goes beyond monitoring storage arrays; it means managing the service levels the business expects, from provisioning to business continuity. Just like its predecessors, VMAX3 has redefined the standard in high-end storage.

## Audience

This technical notes document is intended for anyone who needs to understand how the components and technology in VMAX3 systems provide a highly reliable and available platform. It is also intended for individuals who would like to know more about the serviceability aspects of VMAX3 systems. This document specifically targets EMC customers, sales, and field technical staff who are working with or plan to be working with a VMAX3 system.

# Introduction

Binding service-level agreements (SLAs) commit IT organizations to deliver agreed-to and measurable support metrics for application performance, end-user response time, and system availability. In the event of a component failure, such as a faulty disk drive or power supply, these organizations are required to provide the highest levels of performance, availability, and serviceability – without compromise.

IT executives around the globe have recognized that downtime is not only measured in minutes or hours, but is also calculated as potential revenue loss, missed opportunities, or dissatisfied customers. Any operational impacts resulting from a component failure must be completely transparent to the individual applications and users relying on information availability to drive the business.

Today's mission-critical environments require a high-end storage solution that guarantees uncompromised levels of service, backed by a vendor that will provide quality support and a smooth migration path to new technologies. This solution must include a robust architectural design to withstand any and all potential failures without impacting data availability.

The VMAX3 series is based on a revolutionary design and includes key enhancements that improve the reliability, availability, and serviceability of the new systems – making VMAX3 the ideal choice for critical applications and 24x7 environments demanding uninterrupted access to information.

The objective of this technical note is to provide an overview of the reliability, availability, and serviceability (RAS) features within the VMAX3 architecture.

## Remote support

Remote support is an important and integral part of EMC Customer Support. Every VMAX3 unit has two integrated Management Module Control Stations (MMCS) that continuously monitor the VMAX3 environment. The MMCS's can communicate with the EMC Customer Support Center through a network connection to the EMC Secure Remote Support (ESRS) Gateway.

Through the MMCS, the VMAX3 system actively monitors all I/O operations for errors and faults. By tracking these errors during normal operation, the VMAX3 system can recognize patterns of error activity and predict a potential hard failure before it occurs. This proactive error tracking capability can often prevent component failures by fencing off, or removing from service, a suspect component before a failure occurs. The call home capabilities allow the VMAX3 system to automatically notify EMC Customer Support of potential issues before a failure actually occurs. An EMC Technical Support Engineer handles these calls and can dispatch a local Customer Service Engineer to replace a component without disrupting access to data.

To provide remote support capabilities, the VMAX3 system is configured to call home and alert EMC Customer Support of a potential failure. An authorized EMC Technical Support Engineer is able to run system diagnostics remotely for further

troubleshooting and resolution. Configuring EMC products to allow inbound connectivity also enables EMC Customer Support to proactively connect to the systems to gather needed diagnostic data or to attend to identified issues. The current connect-in support program for the VMAX3 system uses the latest digital key exchange technology for strong authentication, layered application security, and a centralized support infrastructure that places calls through an encrypted tunnel between EMC Customer Support and the MMCS located inside the VMAX3 system.

Before anyone from EMC Customer Support can initiate a connection to a system at the customer site, that person must be individually authenticated and determined to be an appropriate member of the EMC Customer Support team. Field-based personnel who might be known to the customer must still be properly associated with the specific customer's account.

An essential part of the design of the connectivity support program is that the connection to the customer's VMAX3 MMCS must originate from one of several specifically designed Remote Support Networks at EMC. Within each of those EMC Support Centers, the necessary networking and security infrastructure has been built to enable both the call-EMC and call-device functions.

## **Supportability through the management module control station**

The previous generation of VMAX had a single service processor in each system that was used for fault monitoring, as well as remote connectivity and maintenance provided by EMC Customer Support. In VMAX3, this has been extended to two management module control stations (MMCS) in the first engine of each system (one per director). Each MMCS is powered by the redundant power supplies located in its respective director board.

The MMCS located in director 1 is known as the primary MMCS, and the MMCS located in director 2 is known as the secondary MMCS. The primary MMCS will provide all control station functionality when it is operating normally, while the secondary MMCS will provide a subset of this functionality. If the primary MMCS fails, then the secondary MMCS will be put in an elevated secondary state, which will allow more functionality for the duration of this state. Both MMCS will be connected to the customer network, giving the system the redundant ability to report any errors to EMC Customer Support, as well as allow EMC Customer Support to connect to the system remotely.

The MMCS is involved with the following support and maintenance tasks:

- Troubleshooting by EMC Customer Support
- Component replacement scripts
- Code loads and configuration & installation scripts
- Internal scheduler tasks that monitor the health of the system
- Error collection and logging
- Error reporting and remote connectivity

For more information on MMCS and component-level redundancy, see the [Management module control station \(MMCS\)](#) section of this document.

## Secure Service Credential (SSC), Secured by RSA

The Secure Service Credential technology applies exclusively to service processor activities and not host-initiated actions on VMAX3 devices. These service credentials describe who is logging in, what capabilities they have, a time frame that the credential is good for, and auditing of what actions the service personnel performed which can be found in the symaudit logs. If these credentials are not validated by VMAX3, the user will not be able to log in to the MMCS or other internal functions. SSC covers both on-site and remote login to VMAX3.

Some of the security features are transparent to the customer, such as service access authentication and authorization by EMC Customer Support and SC (user ID information) restricted access (MMCS and EMC Customer Support internal functions). Access is definable at a user level, not just at a host level. All VMAX3 user ID information will be encrypted for secure storage within the VMAX3.

MMCS-based functions will honor Solutions Enabler Access Control settings per authenticated user in order to limit view/control of non-owned devices in shared environments such as VMAX3 or SRDF-connected systems.

## Error detection

VMAX3 provides a suite of error and integrity checks to ensure that data integrity is maintained in the event of a system failure or power outage. The VMAX3 system is designed with these data integrity features:

- Block CRC error checks
- Data integrity checks
- Drive monitoring and correction
- Physical memory error

## Block CRC error checks

VMAX3 arrays support and provide:

- Industry standard T10 Data Integrity Field (DIF) block cyclic redundancy code (CRC) for track formats.  
  
For open systems, this enables host generated DIF CRC's to be stored with user data by the VMAX3 arrays and used for end-to-end data integrity validation.
- Additional protections for address/control fault modes for increased levels of protection against faults. These protections are defined in user definable blocks supported by the T10 standard.
- Address and write status information in the extra bytes in the application tag

and reference tag portion of the block CRC.

## **Data integrity checks**

VMAX3 arrays validate the integrity of data they hold at every possible point during the lifetime of the data. From the point at which data enter a VMAX3 array, the data is continuously protected by error detection metadata. This protection metadata is checked by hardware and software mechanisms any time data is moved within the VMAX3 subsystem, allowing the array to provide true end-to-end integrity checking and protection against hardware or software vaults.

The protection metadata is appended to the data stream, and contains information describing the expected data location as well as CRC representation of the actual data contents. The expected values to be found in protection metadata are stored persistently in an area separate from the data stream. The protection metadata is used to validate the logical correctness of data being moved within the VMAX3 array any time the data transitions between protocol chips, internal buffers, internal data fabric endpoints, system cache, and system disks.

## **Drive monitoring and correction**

VMAX3 arrays monitor medium defects by both examining the result of each disk data transfer and proactively scanning the entire disk during idle time. If a block on the disk is determined to be bad, the director:

1. Rebuilds the data in the physical memory if necessary.
2. Remaps the defected block to another area on the disk set aside for this purpose.
3. Rewrites the data from physical memory back to the remapped block on the disk.

The director maps around any bad block(s) detected, thereby avoiding defects in the media. The director also keeps track of each bad block detected on a disk. If the number of bad blocks exceeds a predefined threshold, the primary MMCS invokes a sparing operation to replace the defective drive and then automatically alerts EMC Customer Support to arrange for corrective action.

## **Physical memory error verification and error correction**

The HYPERMAX OS which runs on VMAX3 arrays can correct single-bit errors and report an error code once the single-bit errors reach a predefined threshold. When a multi-bit error occurs, the VMAX3 array fences the physical memory segment (removes it from service) and retrieves the data from the mirrored memory (if it was unwritten) or from the physical drive. In the unlikely event that physical memory replacement is required, the VMAX3 array notifies EMC support, and a replacement is ordered. The failed FRU is then sent back to EMC for failure analysis.

## Reliable components

VMAX3 systems use components that have a mean time between failure (MTBF) of several hundred thousand to millions of hours for a minimal component failure rate. A redundant design allows VMAX3 systems to remain online and operational during component repair.

A periodic system check tests all components as well as HYPERMAX OS 5977 integrity. VMAX3 systems report errors and environmental conditions to the host system as well as to the EMC Customer Support Center.

## VMAX3 component-level redundancy

All critical components of VMAX3 are fully redundant, including director boards, global memory, internal data paths, power supplies, battery backup, and all SAS back-end components. The following is an overview of the redundancy of each of these components.

### Redundant engine components

The engine is a critical building block of a VMAX3 system. It primarily consists of two redundant director boards, which house global memory, front-end connectivity, back-end connectivity, and internal network communications components. VMAX3 single engine configurations are fully redundant.

The different models of VMAX3 systems support different engine counts. The table below displays the supported engine count for each VMAX3 model.

**Table 1. Supported engine count per VMAX3 model**

| VMAX3 Model | Supported Engine Count |
|-------------|------------------------|
| 100K        | 1-4                    |
| 200K        | 1-4                    |
| 400K        | 1-8                    |

Figure 1. Front view of engine



Figure 2. Rear view of engine



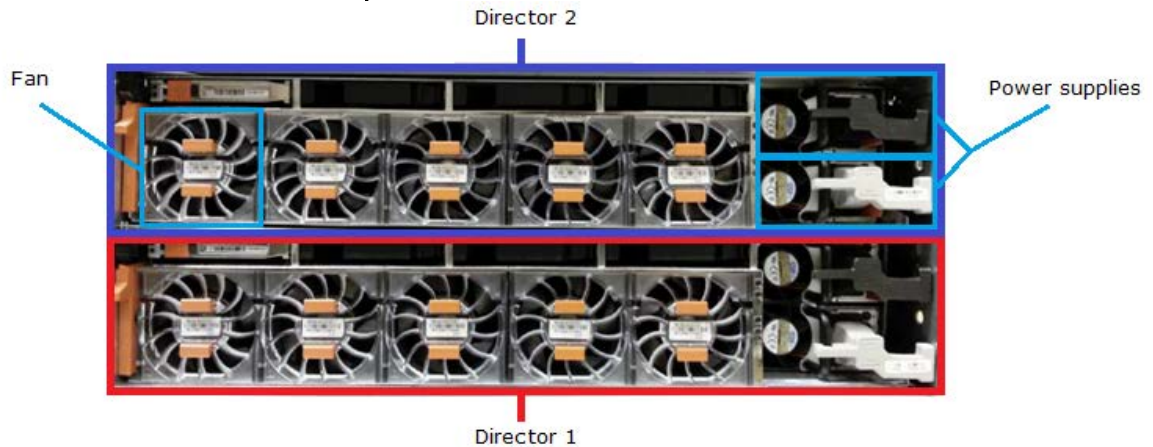
## Redundant director boards

As mentioned above, each engine consists of two director boards. Each director board has a dedicated power and cooling system. The following table lists the components within a director, their counts, and what their purposes are.

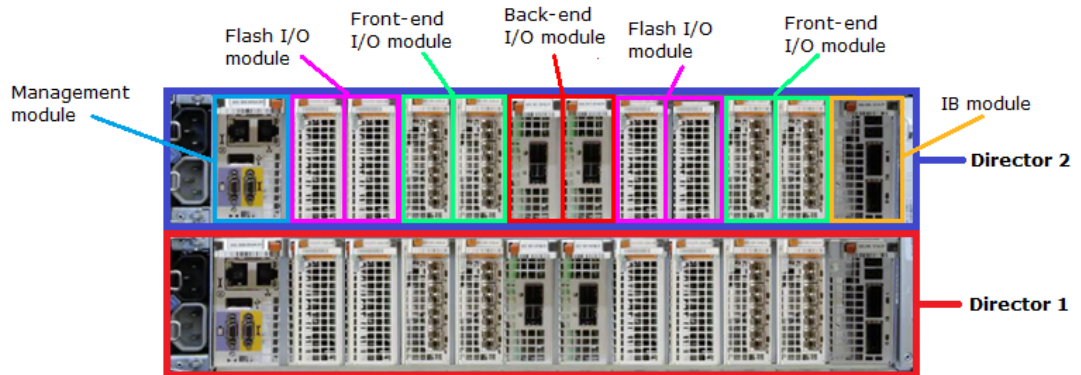
Table 2. Director components

| Director Component     | Count (per director) | Purpose                                     |
|------------------------|----------------------|---------------------------------------------|
| Power Supply           | 2                    | Provide redundant power to director         |
| Fan                    | 5                    | Provide cooling for director                |
| Management Module      | 1                    | Manage environmental functionality          |
| Flash I/O Module       | 4                    | Enables vault to flash capability           |
| Front-end I/O Module   | Up to 4              | Provide front-end connectivity to the VMAX3 |
| Back-end I/O Module    | 2                    | Provide back-end connectivity to drives     |
| InfiniBand (IB) module | 1                    | Connects to IB interconnect switches        |
| Memory Module          | 16                   | Global memory component                     |

Figure 3. Front view of director components



**Figure 4. Rear view of director components**



## **Management module control station (MMCS) and management module (MM)**

There are two different types of management modules in VMAX3 systems, which are management module control stations (MMCS) and standard management modules (MM). The first engine of each VMAX3 system will be deployed with an MMCS in each director. Each subsequent engine (engines 2-8) will be deployed with a management module in each director, in place of an MMCS.

### **Management module control station (MMCS)**

The MMCS combines the management module and control station (service processor) hardware into a single module. It provides environmental monitoring capabilities for power, cooling, and connectivity. Each MMCS monitors one of the system standby power supplies (SPS) via RS232 connection. Each MMCS is also connected to both internal Ethernet switches on the VMAX3 system as part of the internal communications and environmental control system.

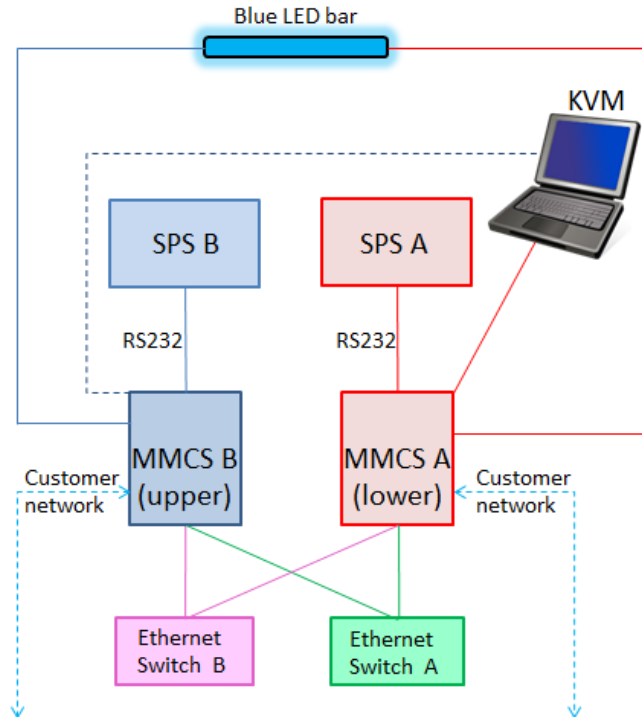
For more information on internal communications and environmental controls, see the [Internal Ethernet connectivity](#) section of this document.

The MMCS also provide support functionality. Each MMCS will connect to the customer's local area network (LAN) to allow monitoring of the system, as well as remote connectivity for the EMC Customer Support team. The primary MMCS will be connected to the keyboard/video/mouse (KVM) in the VMAX3 system and the secondary MMCS will have these connection capabilities as a backup option. These can also be connected to an external laptop or KVM source.

For more information on supportability through the MMCS, see the [Supportability through the management module control station \(MMCS\)](#) section of this document.

The MMCS also controls the blue LED bars on the front and back of each bay. These can be used for system identification purposes by EMC Customer Support. For more information, see the [System identification blinking bay LED feature](#) section of this document.

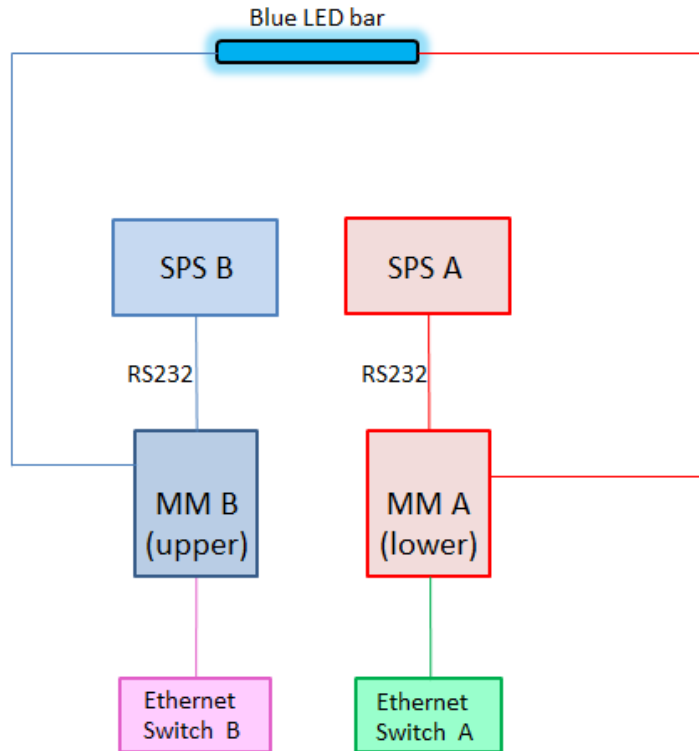
Figure 5. Management module control station connectivity



## Management module (MM)

The management module has a subset of features of the MMCS, which does not include the control station functionality. Each management module will have an RS232 connection to an SPS. Management module A will connect only to Ethernet switch A, and management module B will connect only to Ethernet switch B. Like the MMCS, each management module is responsible for monitoring and reporting any environmental issues, such as power, cooling, or connectivity problems.

Figure 6. Management module connectivity



## Flash I/O module

The VMAX3 flash I/O modules are utilized during the vaulting sequence. For more information, see the [Vaulting](#) section of this document.

## Front-end I/O module

The VMAX3 front-end I/O modules are used for channel connectivity. There are different types of front-end I/O modules, which allow connectivity to various different interfaces. These include SAN, SRDF, and embedded NAS (eNAS). For more information, see the [Channel front-end redundancy](#) section of this document.

## Back-end I/O module

The VMAX3 back-end I/O modules are used to connect the director boards to the back-end of the system, allowing I/O down to the system's drives. For more information, see the [SAS back-end redundancy](#) section of this document.

## InfiniBand (IB) module

The VMAX3 InfiniBand (IB) modules provide connectivity to the matrix interface board enclosure (MIBE), as part of the Dynamic Virtual Matrix. For more information, see the [Dynamic Virtual Matrix](#) section of this document.

## Shared physical memory

In VMAX3, global memory is accessible by any director within the array.

- If an array has a single engine, physical memory pairs are internal to the engine
- If an array has multiple engines, physical memory is paired across engines

Dual-write technology is maintained by the array. In the event of a director or memory failure, the data continues to be available from the redundant copy.

## Global memory technology overview

Global memory is a very crucial component in the VMAX3 array. All read and write operations are transferred to or from global memory. The VMAX3 allows any transfers between the host processor and channel directors, allowing them to be processed at much greater electronic speeds than transfers involved with physical disks. VMAX3 uses complex statistical prefetch algorithms which can adjust to proximate conditions on the VMAX3 array. EMC intelligent algorithms automatically adjust to the workload by constantly monitoring, evaluating and optimizing cache decisions.

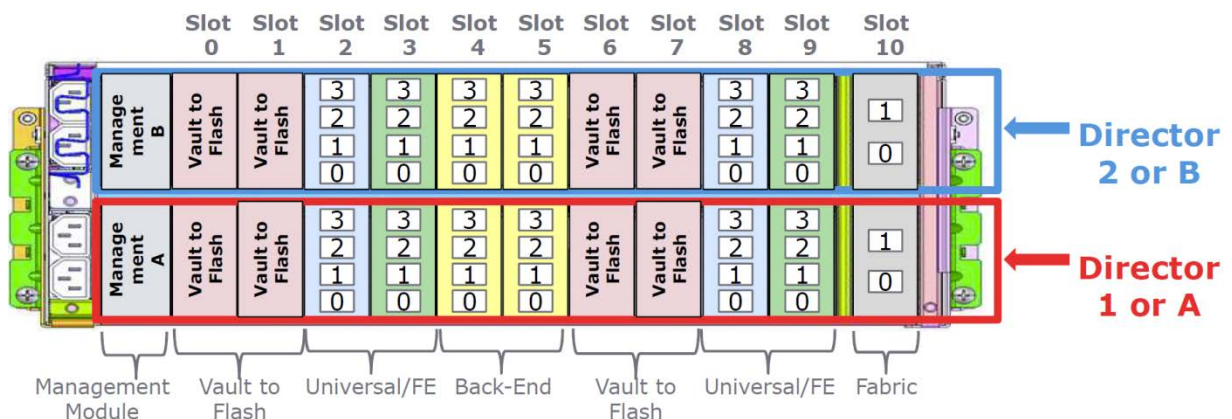
The VMAX3 array uses DDR3 DRAM technology (16 slots per director), which allows the VMAX3 array to be configured up to 2TB of mirrored memory per engine, and up to 16TB mirrored per array.

## Physical port numbering

Physical port numbering within the VMAX3 engine is used to determine how cables will be connected to the front-end, back-end, and fabric I/O modules. The physical port numbering on each VMAX3 director board follows the same rules. The following is a list of these rules:

- I/O module slots are numbered 0-10 from left to right
- I/O modules that have 4 ports are numbered 0-3 from bottom to top
- I/O modules that have 2 ports are numbered 0-1 from bottom to top

Figure 7. Rear view of engine with physical port numbering

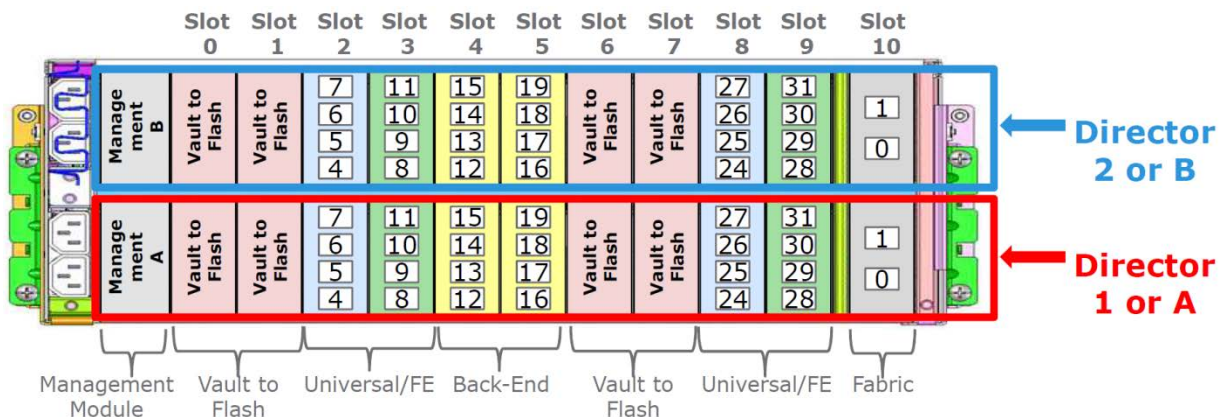


## Logical port numbering

Logical port numbering within the VMAX3 engine is used to determine how the HYPERMAX OS will interpret what is connected to the front-end, back-end, and fabric ports. The logical port numbering on each VMAX3 director board follows the same rules. The following is a list of these rules:

- Supports 32 logical ports (ports 0-31)
- Logical ports are numbered from left to right, bottom to top
- Ports 0-3 and 20-23 are internally dedicated to the vault I/O modules

Figure 8. Rear view of engine with logical port numbering



## Channel front-end redundancy

Channel redundancy is provided by configuring multiple connections from the host servers (direct connect) or Fibre Channel switch (SAN connect) to the VMAX3 system. With SAN connectivity, through Fibre Channel switches, each VMAX3 port can support multiple host attachments, enabling storage consolidation across a large number of host platforms. The multiple connections should be distributed across separate directors to ensure uninterrupted access in the event of a channel failure. A minimum of two connections per server or SAN is required to provide full redundancy.

Host connectivity to the front-end director ports should be spread across physical components for the most efficient form of redundancy.

It is recommended that when connecting a host or cluster to VMAX3:

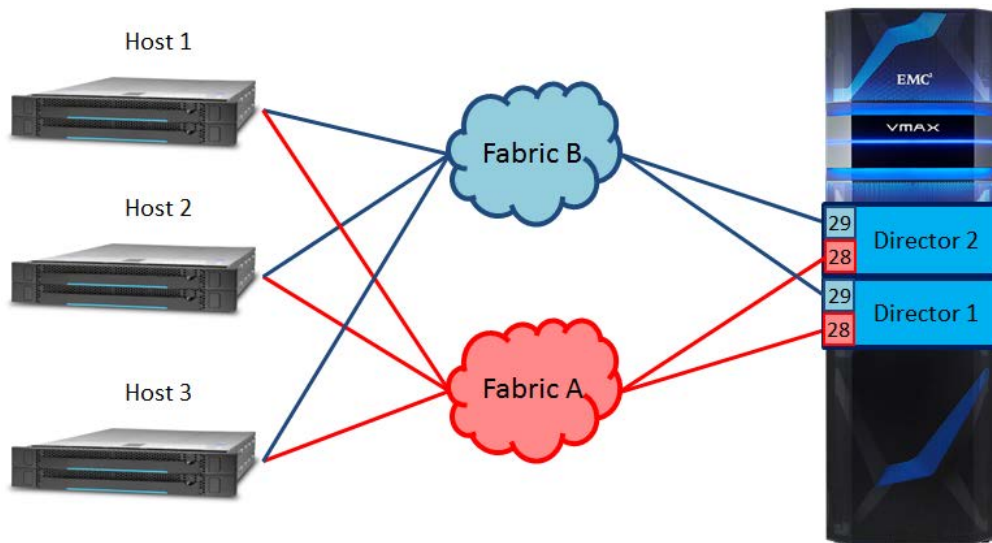
- 2-4 front-end paths are configured in the port group for masking and zones to the host (single initiator zoning is recommended).
- For cabling options, one approach is to connect all even numbered ports to fabric A and all odd numbered ports to fabric B.
- In single engine systems with this approach, select 2 I/O ports spanning both SAN fabrics on each director.

**Example:** Port 28 & 29 on both directors 1 and 2.

- In a multi-engine system, if an odd and even port on at least 2 different directors spanning different engines is selected, this will spread the load for performance and ensure fabric redundancy.

**Example:** Port 28 & 29 in directors 1 and 3.

**Figure 9. SAN connectivity in a single engine VMAX3 environment**



## SAS back-end redundancy

The VMAX3 architecture incorporates a 6 Gb/s SAS (Serial attached SCSI) back-end design to ensure high performance and full redundancy. SAS is a reliable high end protocol that uses a connectionless tree structure with unique paths to individual devices. The paths are stored in routing tables which are built during a discovery phase and are used to route I/O to the desired end point.

The SAS back-end subsystem provides independent redundant paths to the data stored on physical drives to provide seamless access to information, even in the event of a component failure and/or replacement.

The following is an overview of the redundant components of the VMAX3 back-end subsystem.

## Redundant back-end director connectivity

A pair of directors within the same engine is used to access each drive. One director is connected to the primary physical path to the drive, and the other director is connected to the secondary physical path to the drive. Directors are connected through a pair of independent back-end I/O modules and cabling that allow data to be moved back and forth between global memory and the drives. Each director is connected to global memory through primary and secondary paths, to eliminate possible single points of failure. For more information on single points of failure, see

the [Dual-initiator feature](#) section of this document.

## **Redundant cable paths**

Each back-end I/O module is connected to its associated link control card (LCC) and drive array enclosure (DAE) chain through a completely independent cable assembly. Each connection carries four paths, providing an increase in reliability and throughput.

## **Redundant drive paths**

Each SAS drive has two ports which are connected via a fully independent path. In the VMAX3 architecture, each of these ports is connected through separate directors, cables, and LCCs.

## **Point-to-point back-end**

VMAX3 systems employ a SAS connectionless, point-to-point network that has an independent relationship with each drive. This relationship between the back-end controller and each drive allows the VMAX3 to analyze drive health and improve serviceability.

The SAS connectionless, point-to-point network consists of SAS controllers in the back-end I/O modules and SAS expanders in the LCCs (VMAX DAE60 ICM/LCC and VMAX DAE120 LCC). These controlling points of the SAS network contain routing tables necessary to move data from point-A to point-B within the network. The routing tables are built in the discovery phase as the SAS network initializes or in an event of a topology change. For example, a topology change can consist of pull/plug of a drive, or connect/disconnect of a cable, or pull/plug of an LCC.

## **Drive Array Enclosure (DAE)**

The VMAX3 can be equipped with two different types of drive array enclosures (DAE). The VMAX DAE60 supports 3.5" drives and 2.5" drives in a 3.5" carrier. The VMAX DAE120 supports 2.5" drives. Both types can be mixed together in a single system. The DAE houses the physical drives as well as the LCCs.

### **DAE components**

The DAE60 can be populated with a maximum of 60 drives and relies on both Interconnect modules (ICM) and link control cards (LCCs) for providing communication to and from the physical drives to the back-end I/O modules. The ICM contains expanders that interface to the rest of the system, including cables to the back-end I/O modules, as well as expansion DAEs. The LCCs contain two expanders that each interface to the ICM and the drives. LCC A will connect to a set of 60 drives and LCC B will connect to the second port of the same drives. The two drive expanders in each LCC A and LCC B connecting to 30 drives each will give a total of 60 drives maximum connected to the dual ports of each drive for redundancy.

**Figure 10. VMAX DAE60**



The DAE120 can be populated with a maximum of 120 drives. There are four drive expanders on each LCC. Each of the drive expanders connects to 30 drives. A total of four drive expanders connecting to 30 drives each will give a total of 120 drives. LCC A and LCC B will connect to the dual ports on each drive for redundancy.

**Figure 11. VMAX DAE120**



### **Fault zones for DAEs**

VMAX3 DAEs are designed to have complete fault zone protection in the event of an unforeseen power related failure. Both types of DAEs have multiple fault zones which effectively provide four virtual drive enclosures within a single DAE.

Figure 12. Fault zones for DAE60

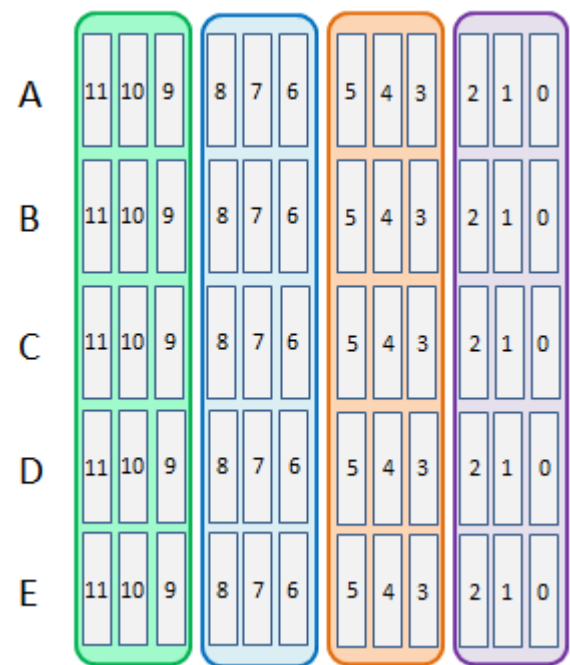


Figure 13. Fault zones for DAE120



## Dual-initiator feature

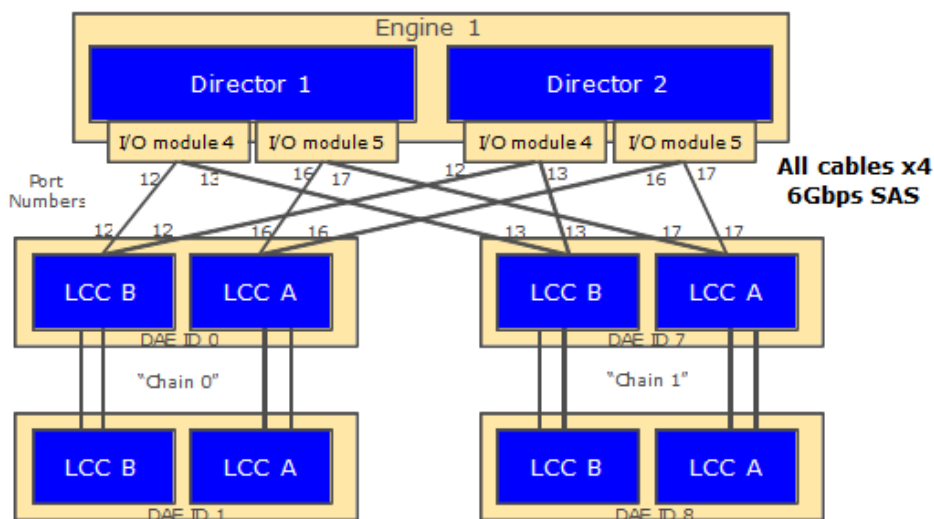
VMAX3 systems have a dual-initiator feature that ensures continuous availability of data in the unlikely event of a drive management hardware failure. The dual-initiator does not provide data availability in the event of a drive failure. Note that drives are configured in RAID groups which provide protection for them. For more information on RAID, see the [Data Protection Methods](#) section of this document.

This feature works by having both directors in a single engine connect to redundant paths in the DAE. A single director can connect to both LCC A and LCC B of the same DAE. The other director in the engine, the dual-initiator, can connect to the other side of both LCC A and LCC B. This cabling scheme will provide complete redundancy for availability purposes.

If the sophisticated fencing mechanisms of the VMAX3 system detect a failure on the back-end director, the system can process reads and writes to the drives through a completely independent path from the dual-initiator without interruption.

**Figure 14. SAS cabling example for VMAX3**

Single engine with 4 DAEs



## Dynamic virtual matrix

The dynamic virtual matrix uses Infiniband (56 Gbps) technology to carry control, metadata, and user data through the VMAX3 system. This technology connects all of the engines in the system to provide a powerful form of redundancy and performance. This allows the engines to share resources and act as a single entity while communicating. In any VMAX3 systems that have two or more engines, there are two redundant matrix interface board enclosures (MIBE) that connect to the fabric I/O modules of each director board. The purpose of the dynamic virtual matrix is to create a communication interconnection between all of the engines, and therefore a single engine system does not require a dynamic virtual matrix or MIBEs.

The VMAX 100K and 200K platforms contain two 12 port MIBEs with redundant, non-

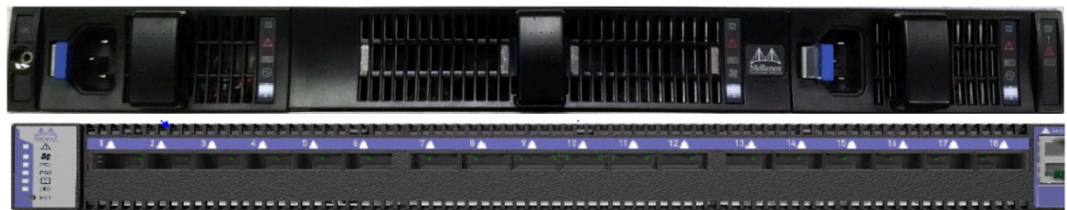
FRU power supplies.

**Figure 15. Front view of VMAX 100K and 200K MIBE**



The VMAX 400K platform contains two 18 port MIBEs with redundant, hot pluggable power supplies, as well as fans that are replaceable.

**Figure 16. Front (top) and rear (bottom) view of VMAX 400K MIBE**



## Internal environmental Ethernet connectivity

The VMAX3 system carries environmental information through two redundant Ethernet switches. Each MMCS connects to both Ethernet switches, and each management module connects to one switch (odd directors to one, even directors to the other). These provide low-level system-wide communications and environmental control for running application software, monitoring, and diagnosing the system from the MMCS.

---

Note: These are separate from the Infiniband MIBEs that are part of the dynamic virtual matrix.

---

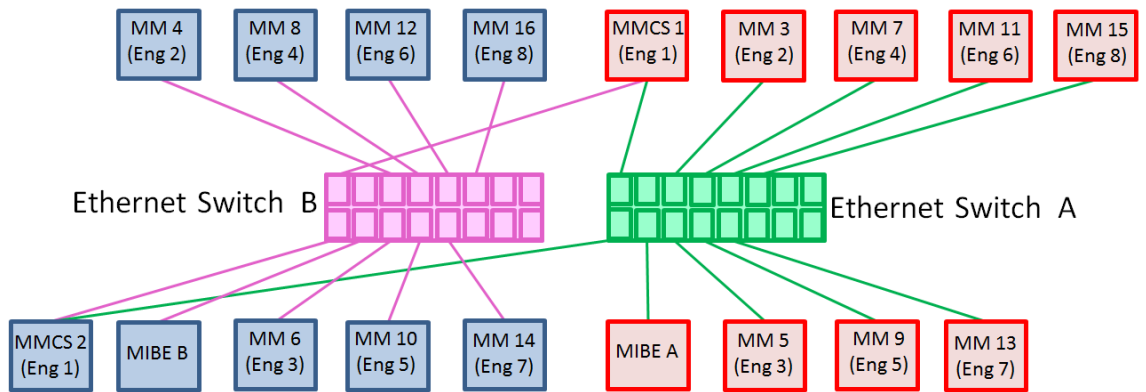
The internal Ethernet connectivity network monitors and logs environmental events across all critical components and reports any operational problems. Critical components include director boards, global memory, power supplies, power line input modules, fans, and various on/off switches. This network's environmental control capability is able to monitor each component's local voltages, ensuring optimum power delivery. Temperature of director boards and memory are also continuously monitored. The AC power main is checked for the following:

- AC failures
- Transfer to auxiliary
- DC failures
- Current sharing between DC supplies

- DC output voltage
- Specific notification of overvoltage condition
- Current from each DC supply
- Voltage drops across major connectors

Failing components can be detected and replaced before a failure occurs.

**Figure 17. Internal Ethernet connectivity**



## Redundant power subsystem

VMAX3 has a modular power subsystem featuring a redundant architecture that facilitates field replacement of any of its components without any interruption in processing.

The VMAX3 power subsystem has two power zones for redundancy. Each power zone connects to a separate dedicated or isolated AC power line. If AC power fails on one zone, the power subsystem continues to operate through the other power zone. If any VMAX3 single power supply module fails, the remaining power supplies continue to share the load. The VMAX3 system senses the fault and reports it as an environmental error.

## Battery backup unit modules

VMAX3 now leverages lithium-ion standby power supply (Li-Ion-SPS) modules for battery backup purposes. These are lighter and have a longer shelf life than lead acid SPS modules. VMAX3 Family 100K, 200K, and 400K system bays can include up to six Li-Ion-SPS modules. The number and location of SPS modules depends on the type of system bay and the number of engines in the array. The following rules apply to SPS configurations:

- SPS 3A and 3B provide back-up power to the first (odd) engine in that bay. They will also provide back-up power to both matrix interface board enclosures (MIBE) if they are configured in the system.
- SPS 2A and 2B power the second (even) engine in a dual engine system bay,

and are required on all dual engine system bays.

- SPS 1A and 1B provide back-up power to both MIBEs and the first (odd) engine. They are only required in configurations that have MIBEs, which include dual engine configurations and single engine configurations with two or more bays.

## Vaulting

As cache size has grown, the time required to move all array data to a persistent state has also increased. Vaulting is designed to limit the time needed to power off the system if it needs to switch to a battery supply. Unlike previous platforms, the VMAX3 platform does not vault to disks. It now vaults to dedicated I/O modules, known as flash I/O modules, saving disk space. Vaulting to flash also expedites the vaulting process, and centralizes it to the engine components, removing the need for battery backup to disks, creating an overall denser configuration than previous systems.

## Vault triggers

State changes in the VMAX3 system that require the system to vault are referred to as vault triggers. There are two types of vault triggers, which are internal availability triggers and external availability triggers.

### Internal availability triggers

Internal availability triggers are initiated when global memory data becomes compromised due to component unavailability. Once these components become unavailable, the system will trigger the Need to Vault (NTV) state, and vaulting will occur. There are three internal triggers:

1. Vault flash availability – The flash I/O modules are used for storage of meta data under normal conditions, as well as storing any data that is being saved during the vaulting process. When the overall available flash space in the flash I/O modules becomes the same size as *N* copy of global memory, the NTV process will be triggered. This is to ensure that all of the data will be saved before a potential further loss of vault flash space occurs.
2. Global memory (GM) availability – When any of the mirrored director pairs are both unhealthy either logically or environmentally, NTV will be triggered because of GM unavailability.
3. Fabric availability – When both the fabric switches are environmentally unhealthy, NTV will be triggered because of fabric unavailability.

### External availability triggers

External availability triggers are initiated under circumstances when global memory data is not compromised, but it is determined that the system preservation will be improved by vaulting. Vaulting in this context is used as a mechanism to stop host activity, facilitate easy recovery or act as an attempt to proactively take action to prevent potential data loss. There are two external triggers:

1. Engine trigger – When an entire engine fails, the VMAX3 system will vault.
2. DAE trigger – If the system has lost access to the whole DAE or DAEs, including dual-initiator failure, and loss of access will cause configured RAID members to become non-accessible, the system will vault.

## Power-down operation

When a VMAX3 system is powered down or transitioned to offline, or when environmental conditions trigger a vault situation, a vaulting procedure is initiated. During power-down or power loss, the part of global memory that is saved first reaches a consistent image (no more writes). The directors then write the appropriate sections of global memory to the flash I/O modules, saving three copies of the logical data. The battery backup unit (BBU) modules maintain power to the system during the power-down process for up to 5 minutes.

## Power-up operation

During power-up, the data is written back to global memory to restore the system. When the VMAX3 system is powered on the startup program does the following:

- Initializes the hardware and the environmental system
- Restores the global memory from the saved data while checking the integrity of the data. This is accomplished by taking sections from each copy of global memory that was saved during the power-down operation and combining them into a single complete copy of global memory. If there are any data integrity issues in a section of the first copy that was saved, then that section will be extracted from the second copy during this process.
- Performs a cleanup, data structure integrity, and reinitialization of needed global memory data structures

At the end of the startup program, the system resumes normal operation when the BBUs are recharged enough to support another vault. If any condition is not safe, the system will not resume operation and will call Customer Support for diagnosis and repair. In this state, Customer Support will be able to communicate with the VMAX3 system and find out the reason for not resuming normal operation.

## Data protection methods

Although the VMAX3 system has standard features that provide a higher level of data availability than conventional DASDs, the following data protection options ensure an even greater level of data recoverability and availability:

- RAID 1 (mirroring)
- RAID 5
- RAID 6

- Local RAID
- Thin provisioning
- Drive sparing
- Local replication with TimeFinder
- Remote replication with Symmetrix Remote Data Facility (SRDF)

These data protection options (excluding disk sparing) are configurable at the physical volume level, so that different levels of protection can be applied to different datasets within the same VMAX3 system.

## RAID 1 (mirroring)

RAID 1 configurations have higher performance in most applications because each drive has a copy of the data. Accordingly, it is possible for the system to be satisfying two I/O requests simultaneously by sending or receiving one from either copy. Mirrored volumes will also have lower response times due to the VMAX3 Dynamic Mirror Service Policy (DMSP), which automatically determined the optimal disk to read to achieve maximum performance. Mirrored volumes will also have lower response times since the VMAX3 system can access the data on either of the mirrored drives.

Mirrored configurations also provide higher performance if there is a disk failure since another complete copy of the data is immediately available. Furthermore, since only two disks are used for the mirror, the chance of failure on multiple drives containing the same data is reduced.

## RAID 5

The VMAX3 family supports RAID 5 data protection. This is an implementation of the industry-standard RAID 5 data protection technique with rotating parity across all members of the RAID 5 set. In the event of a physical drive failure, the missing data is rebuilt by reading the remaining drives in the RAID group and performing XOR calculations.

RAID 5 provides cost-effective data protection against drive failures. While the most demanding environments continue to opt for mirrored storage for maximum performance, RAID 5 configurations offer an extremely attractive alternative for information storage where price is more important than performance.

For VMAX3, RAID 5 is available in two configurations:

- RAID 5 (3+1) – Data and parity are striped across 4 drives (3 data, 1 parity)
- RAID 5 (7+1) – Data and parity are striped across 8 drives (7 data, 1 parity)

## RAID 6

Protection schemes such as RAID 1 and RAID 5 can protect a system from a single physical drive failure within a mirrored pair or RAID group. RAID 6 supports the ability

to rebuild data in the event that two drives fail within a RAID group.

EMC's implementation of RAID 6 calculates two types of parity. This is important during events when two drives within the same RAID group fail, as it still allows the data in this scenario to be reconstructed. Horizontal parity is identical to RAID 5 parity, which is calculated from the data across all of the disks in the RAID group. Diagonal parity is calculated on a diagonal subset of data members. For applications without demanding performance needs, RAID 6 provides the highest data availability.

## Local RAID

Local RAID is a new concept that has been introduced in VMAX3 and is implemented in all VMAX3 environments. Local RAID puts all members of a RAID group behind a single engine. This improves back-end performance, while still supporting the dual-initiator failover/failback model. The following member distribution rules apply to Local RAID configurations:

- No two members of the same RAID group can be placed on the same disk
- Disks will reside in disk groups
- Each disk group supports only one RAID type
- One disk group per disk type is created by default
- Each disk group will be provisioned with its own spare disks

Aside from improved performance, there are other benefits associated with Local RAID as well, including the following:

- Elimination of cross-bay cabling for direct/daisy chain DAE cabling
- Dispersion at the engine/bay level
- Configuration of new systems or upgrades with any combination of contiguous or dispersed engines/bays
- Can position engine/bay on either side of system bay #1
- Most flexible floor planning capability in the industry

## Thin provisioning

Virtual Provisioning enables the ability to increase capacity utilization by enabling more storage to be presented to a host than is physically consumed, and by allocating storage only as needed from a shared virtual pool. Virtual Provisioning also simplifies storage management by making data layout easier through automated wide striping, and by reducing the steps required to accommodate growth.

Virtual Provisioning uses a type of host-accessible device called a virtually provisioned device, also known as a thin device, which does not need to have physical storage completely allocated at the time the devices are created and presented to a host. The physical storage that is used to supply drive space for a virtually provisioned device comes from a shared storage pool, also known as a

storage resource pool (SRP). The SRP is comprised of one or more data pools containing internal devices called data devices. These data devices are dedicated to the purpose of providing the actual physical storage used by virtually provisioned devices.

When a write is performed to a portion of the virtually provisioned device, the VMAX3 array allocates a minimum allotment of physical storage from the pool and maps that storage to a region of the virtually provisioned device, including the area targeted by the write. The storage allocation operations are performed in small units of storage called virtually provisioned device extents. Extents may also be called chunks.

The virtually provisioned device extent size is 1 track (128 KB).

When a read is performed on a virtually provisioned device, the data being read is retrieved from the appropriate data device in the SRP to which the virtually provisioned device is bound. Reads directed to an area of a virtually provisioned device that has not been mapped do not trigger allocation operations. The result of reading an unmapped block will be return null data. When more storage is required to service existing or future virtually provisioned devices, data devices can be added to existing virtually provisioned data pools within the SRP.

For more information on Virtual Provisioning on VMAX3, refer to *EMC® VMAX3 Service Level Provisioning with Fully Automated Storage Tiering™ (FAST)* Technical Notes.

## Drive sparing

VMAX3 systems running HYPERMAX OS support Direct Sparing to automatically replace a failing drive with a spare drive. Direct Sparing on VMAX3 is supported with all protection types, including RAID 6 (14+2).

Two traditional sparing factors, vault drives and specific back-end director within the engine, do not apply to VMAX3 systems. VMAX3 systems vault to flash I/O modules on the engine rather than vault drives, and HYPERMAX OS can dynamically relocate spare drives across back-end directors within the same engine.

The major factor with Direct Sparing in VMAX3 is the power zone within the disk enclosure where the spare drive, failing drive, and other RAID members are located. RAID 1 and RAID 5 are only allowed to have one member per power zone, and RAID 6 is allowed to have up to two members per power zone.

HYPERMAX OS uses the same categories for spare drives; Preferred, Regular, and Non-Preferred. The disk sparing script selects the best available drive by using the following rules as described in the table below.

**Table 3. VMAX3 Drive Sparing Categories**

| Spare category                          | Description                                                                                                                                                                                                                                                                                                                                            | Spare placement behavior         |                                             | Script handling behavior                              |                                   |
|-----------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|---------------------------------------------|-------------------------------------------------------|-----------------------------------|
|                                         |                                                                                                                                                                                                                                                                                                                                                        | Same DAE power zone distribution | Keeps high availability configuration rules | Failing drive gets restored to its original positions | Spare Drive Replenishment Process |
| <b>Preferred Spare</b><br>(Level 1)     | <p>The spare is configured in either the same DAE power zone as the failing drive, or in a different DAE power zone where another RAID member does not already exist.</p> <p>Use of this spare will result in a valid RAID configuration.</p> <p>This spare does not create a configuration with multiple RAID members in the same DAE power zone.</p> | Yes                              | YES                                         | NO                                                    | YES                               |
| <b>Regular Spare</b><br>(Level 2)       | <p>The spare is not configured in the same DAE power zone as the failing drive.</p> <p>Use of this spare will result in a valid RAID configuration.</p> <p>This creates a legal configuration with multiple RAID members in the same DAE power zone.</p>                                                                                               | NO                               | YES                                         | YES                                                   | YES                               |
| <b>Non-Preferred Spare</b><br>(Level 3) | <p>The spare is not configured in the same DAE power zone as the failing drive.</p> <p>Use of this spare will result in a configuration that breaks the rules for RAID member/DAE power zone distribution.</p>                                                                                                                                         | NO                               | NO                                          | YES                                                   | NO                                |

## Spare Drive Replenishment Process

The Spare Drive Replenishment Process calls out spare drives for replacement when any drive in the system no longer has a spare available following a direct sparing operation. This allows customers to schedule drive replacement maintenance with EMC Customer Support less frequently, as multiple drives will be replaced simultaneously.

For more information on drive sparing, refer to *Drive Sparing in EMC® Symmetrix® VMAX® Family Systems* White Paper.

## Local replication using TimeFinder

EMC TimeFinder® software delivers point-in-time copies of volumes that can be used for backups, decision support, data warehouse refreshes, or any other process that requires parallel access to production data.

Previous VMAX families offer several different TimeFinder offerings, each with their own characteristics and ideal use cases. These offerings also have several similarities, the main one being that they each require a target volume to retain snapshot or clone data.

TimeFinder in HYPERMAX OS 5977 for VMAX3 introduces TimeFinder SnapVX which combines the best aspects of the previous TimeFinder offerings, adds some new ease-of-use features, and increases scalability.

SnapVX provides very low impact snapshots and clones for VMAX3 data volumes. SnapVX supports up to 256 snapshots per source volume, which are tracked as versions with less overhead and simple relationship tracking. Users can assign names to identify their snapshots, and they have the option of setting automatic expiration dates on each snapshot.

SnapVX provides the ability to manage consistent point-in-time copies for storage groups with a single operation. Up to 1024 target volumes can be linked per source volume, providing read/write access as pointers or full copies.

TimeFinder in HYPERMAX OS also provides compatibility modes for users who rely on their TimeFinder Mirror, Snap, Clone, or VP Snap command scripts. This will allow users to use their existing scripts while learning how to take advantage of the new features of SnapVX.

For more information on TimeFinder SnapVX, refer to *EMC® VMAX3 Local Replication Suite* Technical Notes.

## Remote replication using SRDF

EMC Symmetrix Remote Data Facility (SRDF) solutions provide disaster recovery and data mobility solutions for VMAX and VMAX3 arrays. SRDF services are provided by the following operating environments:

- HYPERMAX OS for VMAX 100K, 200K, and 400K arrays
- Enginuity for VMAX 10K, 20K, and 40K arrays

SRDF replicates data between 2, 3 or 4 arrays located in the same room, on the same campus, or thousands of kilometers apart.

- SRDF synchronous (SRDF/S) maintains a real-time copy at arrays located within 200 kilometers. Writes from the production host are acknowledged from the local array when they are written to cache at the remote array.
- SRDF asynchronous (SRDF/A) maintains a dependent-write consistent copy at

arrays located at unlimited distances. Writes from the production host are acknowledged immediately by the local array, thus replication has no impact on host performance. Data at the remote array is typically only seconds behind the primary site.

SRDF disaster recovery solutions use “active, remote” mirroring and dependent-write logic to create consistent copies of data. Dependent-write consistency ensures transactional consistency when the applications are restarted at the remote location. You can tailor your SRDF solution to meet various Recovery Point Objectives/Recovery Time Objectives.

Using only SRDF, you can create complete solutions to:

- Create real-time (SRDF/S) or dependent-write-consistent (SRDF/A) copies at 1, 2, or 3 remote arrays.
- Move data quickly over extended distances.
- Provide 3-site disaster recovery with zero data loss recovery, business continuity protection and disaster-restart.

You can integrate SRDF with other EMC products to create complete solutions to:

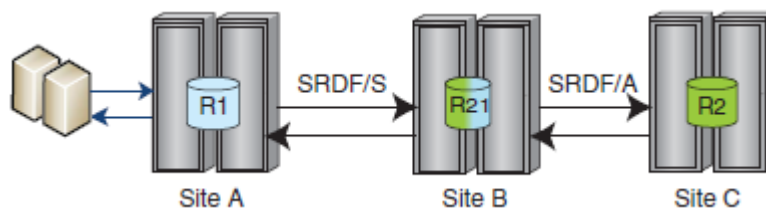
- Restart operations after a disaster with zero data loss and business continuity protection.
- Restart operations in cluster environments. For example Microsoft Cluster Server with Microsoft Failover Clusters.
- Monitor and automate restart operations on an alternate local or remote server.
- Automate restart operations in VMware environments.

### SRDF/Cascade and SRDF/Star support

The HYPERMAX OS 5977 Q1 2015 Service Release introduces support for SRDF/Cascaded and SRDF/Star configurations.

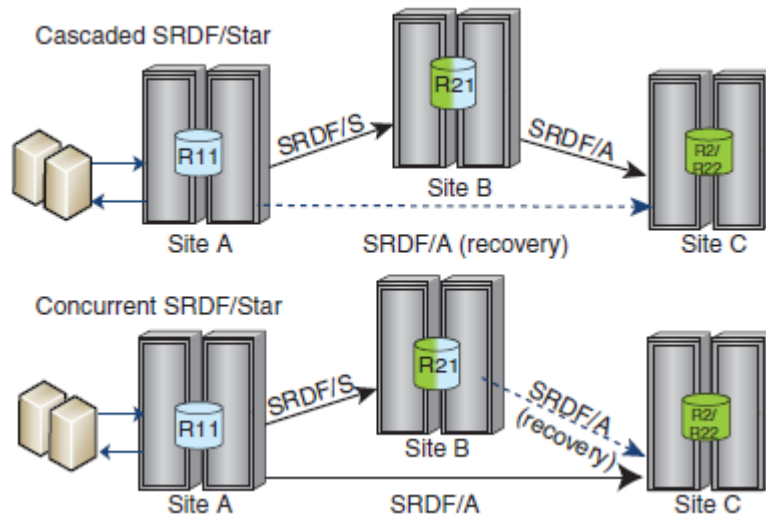
SRDF/Cascade configurations use 3-site remote replication with SRDF/A mirroring between sites B and C, delivering additional disaster restart flexibility. Figure 18 shows an example of an SRDF/Cascade solution.

**Figure 18. SRDF/Cascade**



SRDF/Star is commonly used to deliver the highest resiliency in disaster recovery. SRDF/Star is configured with 3-sites enabling resumption of SRDF/A with no data loss between the two remaining sites, providing continuous remote data mirroring and preserving disaster-restart capabilities. Figure 19 shows examples of Cascaded and Concurrent SRDF/Star solutions.

**Figure 19. SRDF/Star**



### SRDF/Metro support

The HYPERMAX OS 5977 Q3 2015 Service Release introduces support for SRDF/Metro configurations.

SRDF/Metro significantly changes the traditional behavior of SRDF® Synchronous mode with respect to the remote (R2) device availability to better support host applications in high-availability environments. With SRDF/Metro, the SRDF R2 device will now be Read/Write accessible to the host and will take on the federated personality of the primary R1 device (geometry, device WWN, etc.). By providing this federated personality on the R2 device, both R1 and R2 devices may then appear as a single virtual device across the two SRDF paired arrays for host presentation. With both the R1 and R2 devices being accessible, the host or hosts (in the case of a cluster) are allowed to read and write to both R1 and R2 devices with SRDF/Metro insuring that each copy remains current, consistent, and addressing any write conflicts which may occur between the paired SRDF devices.

Figure 20. SRDF/Metro



On the left, we have an SRDF/Metro configuration with a standalone host which has visibility to both VMAX3 arrays (R1 and R2 devices) using multi-pathing software such as PowerPath, to enable parallel reads and writes to each array. This is enabled by federating the personality of the R1 device to insure that the paired R2 device appears, via additional paths to host, as a single virtualized device. On the right, we have a clustered host environment where each cluster node has dedicated access to an individual VMAX3 array. In either case, writes to the R1 or R2 devices are synchronously copied to its SRDF paired device. Should a conflict occur between writes to paired SRDF/Metro devices, the conflicts will be internally resolved to insure a consistent image between paired SRDF devices are maintained to the individual host or host cluster.

SRDF/Metro may be selected and managed through Solutions Enabler SYMCLI or Unisphere for VMAX 8.1 or greater client software and will require a separate SRDF/Metro license to be installed on both VMAX3 arrays to be managed.

For more information on SRDF, refer to the *EMC® VMAX3 Family with HYPERMAX OS Product Guide*, and *Introduction to SRDF/Metro EMC Engineering White Paper*.

## Data at Rest Encryption (D@RE)

The HYPERMAX OS 5977 Q1 2015 Service Release introduces support for Data at Rest Encryption (D@RE) which protects data confidentiality by adding back-end encryption to the entire array. D@RE provides hardware-based, on-array, back-end encryption for VMAX3 systems. Back-end encryption protects information from unauthorized access when drives are removed from the system.

D@RE provides encryption on the back-end using Fibre Channel I/O modules that incorporate XTS-AES 256-bit data-at-rest encryption. These I/O modules encrypt and decrypt data as it is being written to or read from a drive. All configured drives are encrypted, including data drives, spares, and drives with no provisioned volumes.

D@RE incorporates RSA® Embedded Key Manager for key management. With D@RE, keys are self-managed, and there is no need to replicate keys across volume

snapshots or remote sites. RSA Embedded Key Manager provides a separate, unique DEK for each drive in the array, including spare drives.

By securing data on enterprise storage, D@RE ensures that the potential exposure of sensitive data on discarded, misplaced, or stolen media is reduced or eliminated. As long as the key used to encrypt the data is secured, encrypted data cannot be read. In addition to protecting against threats related to physical removal of media, media can readily be repurposed by destroying the encryption key used for securing the data previously stored on that media.

D@RE is compatible with all VMAX3 features, allows for encryption of any supported local drive types or volume emulations, and delivers powerful encryption without performance degradation or disruption to existing applications or infrastructure.

For more information on D@RE, refer to the *EMC® VMAX3 Data at Rest Encryption White Paper*.

## VMAX3 component-level serviceability

VMAX3 systems implement a modular design with a low parts count that improves serviceability by allowing nondisruptive component replacements, should a failure occur. This low parts count minimizes the number of failure points.

VMAX3 systems feature nondisruptive replacement of their major components, which can be replaced while the system is powered on, including:

- Engine components:
  - Director boards
    - I/O Modules
      - Fibre Channel (front-end)
      - Embedded NAS (eNAS)
      - SAS (back-end)
      - Flash (Vault)
      - Compression (SRDF Compression)
    - Management modules/management module control stations
    - InfiniBand (IB) module
    - Power supplies
    - Fans
- Disk Array Enclosure (DAE) components:

- o Link Control Cards (LCC)
- o Power supplies
- o SAS & flash drives
- o Fans
- o System Status Card (SSC)
- Matrix interface board enclosures (MIBE)
  - o Fans (400K only)
  - o Power supplies (400K only)
- Ethernet switches
- Standby Power Supplies (SPS) - Lithium Ion batteries
- Power Distribution Units (PDU)

VMAX3 systems provide full component-level redundancy to protect against a component failure and ensure continuous and uninterrupted access to information. This nondisruptive replacement capability allows the EMC Customer Support Engineer to install a new component, initialize it if necessary, and bring it online *without*:

- Disrupting access to unaffected volumes
- Powering down the VMAX3 unit
- Stopping the operating system
- Taking unaffected channel paths offline
- Taking devices offline (other than the affected device)

## Nondisruptive HYPERMAX OS upgrades

Interim updates of the HYPERMAX Operating System can be performed remotely by the EMC Remote Change Management (RCM) group. These updates provide enhancements to performance algorithms, error recovery and reporting techniques, diagnostics, and HYPERMAX OS fixes. They also provide new features and functionality for HYPERMAX OS. During an online HYPERMAX OS code load, a member of the EMC RCM team downloads the new HYPERMAX OS code to the MMCS. The new HYPERMAX OS code loads into the EEPROM areas within the directors, and remains idle until requested for a hot load in the control store as described in the next paragraph.

The VMAX3 system does not require customer action during the performance of this function. All directors remain in an online state to the host processor, thus maintaining application access. The VMAX3 system will load executable HYPERMAX OS code within each director hardware resource until all directors have been loaded. Once the executable HYPERMAX OS code is loaded, internal processing is

synchronized and the new code becomes operational.

## Flashing blue rack lights

Each VMAX3 bay has a blue LED bar on the front and rear that perform a valuable function during service activities. Procedures such as installation, component replacement, and cable connectivity verification will send a signal to the appropriate LED bar(s) initiating a blinking of the LED, helping to guide EMC service personnel to the correct bay. Additional indicator LEDs inside the bay, along with detailed instructions provided by the service procedure on the MMCS, will further guide service personnel to the correct component(s). This is a very valuable feature in today's data centers, especially on dispersed bay systems.

The state of the LED bar on a bay can also be changed by way of the Solutions Enabler command `symcfg set -led` and through Unisphere for VMAX under the System Hardware menu.

Each light bar is connected to both power zones for redundant power. The light bars themselves as well as their cabling can be replaced nondisruptively.

## Conclusion

VMAX3 is a platform that improves upon every aspect of the already extraordinary VMAX product line. This is a system with highly redundant hardware components, creating a remarkably reliable environment that has also been condensed into a configuration that minimizes carbon footprint in the data center and increases total cost of ownership. The introduction of HYPERMAX OS enhances the customer's experience through new technologies such as service level-based provisioning, making storage management easier while also increasing upon availability of data through improvements to concepts such as vaulting, disk sparing, and RAID. VMAX3's local and remote replication suites also bring the system to an elevated level of availability, through TimeFinder SnapVX and SRDF, respectively. The serviceability aspect of VMAX3 is just as distinguished, making the part replacement process quick and easy.

VMAX3 includes key enhancements that improve the reliability, availability, and serviceability of the systems, making VMAX3 the ideal choice for critical applications and 24x7 environments demanding uninterrupted access to information.

## References

Reference information and product documentation can be found at [EMC.com](http://EMC.com) and [support.EMC.com](http://support.EMC.com), including:

- EMC® VMAX3 Local Replication Suite
- EMC® VMAX3 Family with HYPERMAX OS Product Guide
- EMC® VMAX3 Service Level Provisioning with Fully Automated Storage Tiering™ (FAST)

## Appendix A: Resiliency Testing

### EMC Internal QE Testing

EMC's Quality Engineering (QE) Teams perform thorough testing of all VMAX3 FRUs. Each FRU is tested multiple times for each code level with very specific pass/fail criteria.

Standard tests perform verification of the GUI-based scripted replacement procedures that are used by EMC field personnel. These tests are designed to verify the replaceability of each FRU without any adverse effects on the rest of the system, and to verify the functionality and ease-of-use of the scripted procedures. These tests are straightforward replacement procedures performed on operational components.

Non-standard tests are also performed on components that have failed either by error injection or hot removal of the component or its power source. These tests also incorporate negative testing by intentionally causing different failure scenarios during the replacement procedure. These negative tests are designed to make sure that the replacement procedure properly detects the error and that the rest of the system is not affected.

Some examples of these negative tests include:

- Replacing the wrong component
- Replacing component with an incompatible component
- Replacing component with a faulty component
- Replace component with a new component that has lower code that needs to be upgraded
- Replace component with a new component that has higher code that needs to be downgraded
- Replacing component with the same component and make sure script detects and alerts the user that the same component is being used
- Improperly replace a component (miscable, unseated, etc)
- Initiate a system vault save (system power loss) operation during a replacement procedure

Both the standard and non-standard tests are performed on all system models and various configurations with customer-like workload running on the array. Tests are also performed repeatedly to verify there are no residual issues left unresolved that could affect subsequent replacements of same or different component(s).

Components that are known to fail more frequently in the field, drives for example, as well as complex component replacements are typically tested more frequently.

## On-site Resiliency Testing

---

Important! Before running any replacement procedure, the EMC Customer Support Center must be consulted to verify that the specific scripts to be performed are fully supported at the specific code level that will be running on the array at the time of testing.

---

The supported method to demonstrate system resiliency on a customer system is through the replacement scripts. While VMAX3 arrays are designed to withstand hard failures and live component removal, that type of testing belongs in the QE environment where manual error will not cause any adverse short or long term effects in a live environment.

The main concern with performing resiliency testing outside of the replacement scripts is human error. Specifically, not detecting an unrelated pre-existing condition that may cause the test to fail, and not properly recovering the system after a test.

The replacement scripts have thorough pre-checks at the beginning of a procedure, and thorough post-checks at the end of a procedure. The pre-checks will ensure that there are no other issues in the system that may cause the test to fail or cause major issues across the array, and the post-checks will ensure that the system has been properly recovered. The combination of the pre-checks and post-checks allow for confidence in performing consecutive replacements during the resiliency demonstration.

Scripted replacement procedures are far more real-world scenarios than hot component removal or error injections alone. The scripted replacement procedures are the procedures that will be used on the customer's system while the system is in production. Demonstrating the end-to-end replacement procedure is more valuable than demonstrating that the system can simply withstand the initial component failure. The replacement procedures will, by default, place the selected component in a failed state as part of the script, while greatly reducing the chance of human error.

The scripted replacement procedures provide a very deep level of detail to the field engineer performing the replacement to greatly reduce the chance of manual error during the procedure. The details include written instructions, pictures, and animations to help the field engineer identify the correct component and replace it properly. The details should be shown to customers during any resiliency demonstration to make them aware of the lengths taken to reduce manual error and make them comfortable that VMAX3 FRU replacements will be successful.

Copyright © 2015 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, the EMC logo, FAST, SRDF, TimeFinder, Unisphere, VMAX, and VMAX3 are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners.

For the most up-to-date regulatory document for your product line, go to EMC Online Support.