The Dell EMC PowerMax Family Overview

Abstract
This white paper provides an overview of the Dell EMC PowerMax NVMe-based mission-critical data storage offering. It provides details on the theory of operation, packaging, and the unique features which make it a premier ultra-performing all flash storage product for the modern data center. This white paper is intended for Dell EMC customers and potential customers, Dell EMC Sales and Support Staff, Partners, and anyone interested in gaining a better understanding of the PowerMax storage array and its features.

Dell EMC Engineering
October 2018
Revisions

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 2018</td>
<td>Initial Release V1.0</td>
</tr>
<tr>
<td>October 2018</td>
<td>Updated for 5978 Service Release</td>
</tr>
</tbody>
</table>

The information in this publication is provided “as is.” Dell Inc. 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 software described in this publication requires an applicable software license.

Copyright © 2018 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Reference Number H17118.1

Dell believes the information in this document is accurate as of its publication date. The information is subject to change without notice.
# Table of Contents

Revisions ................................................................................................................................. 2

1 Executive Summary ........................................................................................................... 5
   1.1 Introduction ................................................................................................................ 5
   1.2 PowerMax Platform High Level Overview ................................................................. 5
   1.3 Primary PowerMax Benefits ...................................................................................... 6

2 PowerMax Overview ........................................................................................................... 7
   2.1 Background ................................................................................................................. 7
   2.2 PowerMax Terminology ............................................................................................ 8
   2.3 The PowerMax Family .............................................................................................. 10

3 PowerMax Key Value Propositions ................................................................................... 11
   3.1 Designed for Non-Volatile Memory Express (NVMe) .............................................. 11
   3.2 Expandable Modular Architecture - The Brick ......................................................... 11
   3.2.1 Engines ................................................................................................................ 13
   3.2.2 Drive Array Enclosures (DAE) and Drive Configurations ................................. 15
   3.2.3 Director Slot Layout and Connectivity Options .................................................. 25
   3.3 Streamlined Software Packaging ............................................................................. 28
   3.4 Flash Optimization ................................................................................................... 30
   3.4.1 Cache Architecture and Caching Algorithms ......................................................... 30
   3.4.2 Understanding Flash Cell Endurance .................................................................. 30
   3.4.3 PowerMax Write Amplification Reduction ......................................................... 31
   3.4.4 Boosting Flash Performance with PowerMaxOS FlashBoost ............................ 31
   3.5 Reliability, Availability, and Serviceability (RAS) .................................................. 32
   3.6 Data Services ............................................................................................................ 33
   3.6.1 Data Reduction using the Adaptive Compression Engine (ACE) ...................... 33
   3.6.2 Data Reduction using Inline Deduplication ......................................................... 34
   3.6.3 Remote Replication with SRDF .......................................................................... 35
   3.6.4 Local Replication with TimeFinder SnapVX ....................................................... 35
   3.6.5 PowerMaxOS Service Levels and Host I/O Limits ............................................. 36
   3.6.6 Consolidation of Block and File Storage Using eNAS ....................................... 37
   3.6.7 Embedded Management (eManagement) using Unisphere for PowerMax ........ 38
   3.6.8 Non-Disruptive Migration (NDM) ....................................................................... 39

4 PowerMax System Deployments ...................................................................................... 41
1 Executive Summary

1.1 Introduction

The PowerMax family is the first Dell EMC hardware platform with a storage back-end that uses Non-Volatile Memory Express (NVMe) for customer data. NVMe is a set of standards which define a PCI Express (PCIe) interface used to efficiently access data storage volumes based on Non-Volatile Memory (NVM) media, which includes today’s NAND-based flash along with future, higher-performing Storage Class Memory (SCM) media technologies such as 3D XPoint and Resistive RAM (ReRAM). The NVMe-based PowerMax was specifically created to fully unlock the bandwidth, IOPS, and latency performance benefits that NVM media offers to host based applications which are unattainable using the current generation all flash storage arrays.

1.2 PowerMax Platform High Level Overview

There are two primary hardware platforms included in PowerMax: Dell EMC PowerMax 2000 and Dell EMC PowerMax 8000. Both of these platforms use a modular “Brick” building concept, similar to that used in the VMAX All Flash family.

The PowerMax 2000 platform includes:

- 1 – 2 Bricks per system (details shown in Table 2)
- 2 x 12 core, 2.5 GHz Intel Broadwell CPUs yielding 48 cores per Brick
- 512 GB, 1 TB, or 2 TB of DDR4 cache per Brick
- Up to 64 FE ports per system (open systems only)
- Up to 1 PBe capacity per system of PCIe Gen3 NVMe storage in a half rack
- Open systems only
- 3rd party rack support
- 1.7M IOPS (8K RRH)

The PowerMax 8000 platform includes:

- 1 - 8 Bricks / zBricks per system
- 2 x 18 core, 2.8 GHz, Intel Broadwell CPUs yielding 72 cores per Brick
- 1 TB, 2 TB DDR4 cache per Brick
- Up to 256 FE ports per system (OS and MF)
- Up to 4 PBe capacity per system of PCIe Gen3 NVMe storage in two racks
- Open systems and/or mainframe support
- 10M IOPS (8K RRH)

Both PowerMax systems feature:

- 24 slot NVMe DAE using 2.5” form factor 1.92 TB, 3.84 TB, or 7.68 TB NVMe Drives
- Dual ported NVMe PCIe Gen3 (8 lane) backend I/O interface modules (4 per engine) delivering 8 GB/sec of bandwidth per module (32 GB/sec per engine) to the NVMe storage
- Powerful data reduction I/O modules which perform both inline hardware data compression and deduplication.
1.3 Primary PowerMax Benefits

The primary benefits that PowerMax platforms offer to Dell EMC customers are:

- An end-to-end NVMe storage design providing industry-leading IOPS density in a system with a small footprint
- Ready for next generation data storage media such as Storage Class Memory (SCM) and NVMe over Fabric (NVMe-oF) infrastructure
- Efficient workload consolidation for block, file, and mainframe workloads on a single platform
- Designed for the enterprise with ultra-high reliability, availability, and serviceability
- Best-in-class data services including advanced data reduction using inline deduplication and compression
2 PowerMax Overview

2.1 Background

For the past few years, the migration to flash storage from spinning disk has been a major focus in the enterprise data center as Exabytes (EB) of data have been migrated from spinning disk to flash storage. The eventuality of this migration should not be surprising as the potential performance gains offered by flash storage are orders of magnitude greater than that of spinning disk.

Unfortunately, the true performance potential of flash storage in the enterprise data center has been somewhat hamstrung by choke points around the input/output (I/O) path from the application to the storage. These choke points are centered on the bootstrapping of the historical storage interfaces of Serial Attached SCSI (SAS) and Serial AT Attachment (SATA) to flash storage media. SAS (based on the SCSI protocol) and SATA (based on the ATA protocol) are interfaces that were developed in the late 1980s. These interfaces and protocols were designed for spinning disks and do not have the capability to fully exploit the benefits that flash storage can provide.

Fortunately, choke points lead to innovations. The next disruptive innovation entering the data center is a new data storage interface specifically designed for the current NAND-based flash and the next generation data storage media. This new interface is called **Non-Volatile Memory Express (NVMe)**. NVMe is a command set (replacing SCSI) and associated storage interface standards (replacing SAS and ATA) that enable efficient access to storage devices and systems based on Non-Volatile Memory (NVM) media. NVMe is broadly applicable to NVM storage technology, including current NAND-based flash and higher-performance, Storage Class Memory (SCM) technologies such as 3D XPoint and Resistive RAM (ReRAM). The NVMe standards were created to fully exploit the bandwidth, IOPS, and latency performance benefits that NVM-based storage offers. The resulting levels of performance and parallelism for both drives and storage systems are unattainable using legacy storage interfaces.

The **PowerMax Family** is the first Dell EMC data storage product to fully use NVMe technology for customer application data. The innovative PowerMax is built using a 100% NVMe storage backend, allowing it to reach unprecedented I/O densities and performance by eliminating the flash media choke points found using traditional SAS and SATA interfaces. The PowerMax opens the door for customers to deploy innovative applications in the areas of real-time analytics, machine learning, and big data that demand lower latency and higher performance.
## 2.2 PowerMax Terminology

This white paper uses the following PowerMax terminology:

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Equivalent Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell EMC PowerMax family</td>
<td>PowerMax</td>
<td>The PowerMax family refers to Dell EMC's new NVMe-based mission-critical data storage offering.</td>
</tr>
<tr>
<td>Dell EMC PowerMax 2000</td>
<td>PowerMax 2000</td>
<td>PowerMax 2000 is the entry NVMe scale out array sold with the Essentials and Pro software packages.</td>
</tr>
<tr>
<td>Dell EMC PowerMax 8000</td>
<td>PowerMax 8000</td>
<td>PowerMax 8000 is the flagship NVMe scale out array sold with the Essentials and Pro software packages.</td>
</tr>
<tr>
<td>Drive Array Enclosure</td>
<td>DAE</td>
<td>DAE refers to the drive array enclosure used to store flash drives and SCM drives in PowerMax.</td>
</tr>
<tr>
<td>DAE24</td>
<td>DAE24</td>
<td>DAE24 refers to the drive array enclosure that is used to store up to 24 NVMe drives in PowerMax arrays.</td>
</tr>
<tr>
<td>Non-Volatile Memory Express (NVMe)</td>
<td>NVMe</td>
<td>NVMe is a command set and associated storage interface standards that specify efficient access to data storage devices and systems based on Non-Volatile Memory (NVM).</td>
</tr>
<tr>
<td>NVMe over Fabrics</td>
<td>NVMe-oF</td>
<td>NVMe over Fabrics (NVMe-oF) extends the NVMe command set and its benefits over data center fabrics, such as RoCEv2 (RDMA [Remote Direct Memory Access] over Converged Ethernet) and Fibre Channel.</td>
</tr>
<tr>
<td>NVMe flash drives / NAND / SSD</td>
<td>flash drives</td>
<td>NVMe flash drives are the latest devices used to store capacity in PowerMax arrays.</td>
</tr>
<tr>
<td>PowerMaxOS 5978</td>
<td>PowerMaxOS</td>
<td>The PowerMaxOS 5978 release supports PowerMax NVMe arrays, dedupe, and other software enhancements and is installable on legacy VMAX All Flash arrays.</td>
</tr>
<tr>
<td>Unisphere for PowerMax</td>
<td>Unisphere</td>
<td>Unisphere for PowerMax enables management and monitoring of PowerMax arrays along with legacy VMAX All Flash and VMAX3 arrays.</td>
</tr>
<tr>
<td>Smart RAID</td>
<td>Smart RAID</td>
<td>Smart RAID provides active/active shared RAID support for PowerMax arrays.</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Service levels</td>
<td>Service levels identify a specific performance tier in PowerMax arrays. This term replaces SLOs.</td>
<td></td>
</tr>
<tr>
<td>inline deduplication</td>
<td>Inline dedupe refers to the deduplication technology used with PowerMax arrays.</td>
<td></td>
</tr>
<tr>
<td>inline compression</td>
<td>Inline compression refers to the intelligent compression technology used with PowerMax arrays.</td>
<td></td>
</tr>
<tr>
<td>usable capacity (in Terabytes)</td>
<td>Refers to the amount of unique, non-compressed data that can be written into the array.</td>
<td></td>
</tr>
<tr>
<td>effective capacity (in Terabytes)</td>
<td>Includes the benefits of thin provisioning, inline compression, dedupe and space-efficient copies.</td>
<td></td>
</tr>
<tr>
<td>PowerMax Brick</td>
<td>A Brick is the building block for a PowerMax array. It includes an engine, two DAEs, and a fixed TBu of capacity.</td>
<td></td>
</tr>
<tr>
<td>Flash capacity pack</td>
<td>A Flash capacity pack includes NVMe flash drive capacity (TBu) that can be added to a PowerMax array.</td>
<td></td>
</tr>
<tr>
<td>PowerMax zBrick</td>
<td>zBrick is the PowerMax mainframe building block consisting of an engine, two DAEs, and fixed TBu of capacity.</td>
<td></td>
</tr>
<tr>
<td>zFlash capacity pack</td>
<td>A zFlash capacity pack includes NVMe flash drive capacity (TBu) that can be added to a PowerMax array for mainframe.</td>
<td></td>
</tr>
<tr>
<td>Essentials software package</td>
<td>The Essentials package is the software package sold as a PowerMax appliance.</td>
<td></td>
</tr>
<tr>
<td>Pro software package</td>
<td>The Pro package is the comprehensive software offering sold as a PowerMax appliance.</td>
<td></td>
</tr>
<tr>
<td>zEssentials software package</td>
<td>The zEssentials package is the software package sold as a PowerMax appliance for mainframe.</td>
<td></td>
</tr>
<tr>
<td>zPro software package</td>
<td>The zPro package is the comprehensive software offering sold as a PowerMax appliance for mainframe.</td>
<td></td>
</tr>
<tr>
<td>scale up</td>
<td>Scale up refers to adding zFlash Capacity Packs to a PowerMax array.</td>
<td></td>
</tr>
<tr>
<td>scale out</td>
<td>Scale out refers to adding zBrick s to grow performance and expansion on PowerMax systems.</td>
<td></td>
</tr>
</tbody>
</table>
### 2.3 The PowerMax Family

The Dell EMC PowerMax family consists of two models - the PowerMax 2000 and the flagship PowerMax 8000. The PowerMax 2000 is designed to provide Dell EMC customers with efficiency and maximum flexibility in a 20U footprint. The PowerMax 8000 is designed for massive scale, performance, and IOPS density all within a two-floor-tile footprint.

Both PowerMax arrays have at their foundation the trusted Dynamic Virtual Matrix architecture and a new version of HYPERMAX OS rewritten for the NVMe platform called PowerMaxOS 5978. PowerMaxOS can run natively on both PowerMax systems and on legacy VMAX All Flash systems as an upgrade. As with the previous generation VMAX All Flash, PowerMax systems are true all flash arrays – being products specifically targeted to meet the storage capacity and performance requirements of the all flash enterprise data center. The PowerMax products are feature-rich all flash offerings with specific capabilities designed to take advantage of the higher capacity NVMe flash drives used in the densest configuration possible. PowerMax offers enterprise customers trusted data services, along with improved simplicity, capacity, and performance that their highly virtualized environments demand, while still meeting the economic needs of the more traditional storage workloads. In addition, PowerMax now allows customers to deploy applications such as real-time analytics, machine learning, and big data that demand the lower storage latency and higher IOPS densities previously unattainable with legacy all flash offerings.

#### Figure 1  The PowerMax Family

![PowerMax 2000 and 8000]  

**PowerMax 2000**  
- 1.7M IOPS (8K RRH)  
- 64 FC/SCSI Ports  
- 1 -2 Bricks per system  
- Up to 1 PBs per system of NVMe storage  
- All in ½ Rack (20U)  
- Open Systems Workloads Only  

**PowerMax 8000**  
- 10M IOPS (8K RRH)  
- 256 FC/FCON/ISCSI Ports  
- 1 -8 Bricks per system  
- Up to 4 PBs per system of NVMe storage  
- All in 2 Racks (2 floor tiles)  
- Open Systems / Mainframe / Mixed Workloads
3 PowerMax Key Value Propositions

Although the PowerMax platform uses many of the technologies and data services found in legacy VMAX All Flash, PowerMax provides customers with a differentiating value as it is designed from the ground up to be the first platform in the industry to take full advantage of emerging data storage media. The PowerMax platform delivers the following value propositions to Dell EMC customers:

3.1 Designed for Non-Volatile Memory Express (NVMe)

PowerMax is the technology leader providing a full NVMe flash storage backend for storing customer data. The PowerMax NVMe architecture provides:

- **I/O Density with Predictable Performance** – PowerMax has been designed to deliver industry leading I/O density – capable of delivering over 10 Million IOPS in a two rack system (two floor tiles) predictably, regardless of workload and storage capacity utilization.
- **NVMe Storage Density** – Using commercially available, high capacity, dual-ported enterprise NVMe flash drives, PowerMax delivers industry leading NVMe TB / floor tile. PowerMax support for high capacity commercially available NVMe flash drives provides a differentiated capability as compared to many all flash alternatives which use a proprietary flash drive design. This allows PowerMax to leverage the increases in flash drive densities, economies of scale, and fast time to market provided by the industry flash drive suppliers.
- **Future Proof Design** - The PowerMax NVMe design is future proof as it is ready for Storage Class Memory (SCM) flash and future NVMe-oF SAN connectivity options.

3.2 Expandable Modular Architecture - The PowerMax Brick

PowerMax configurations consist of modular building blocks called PowerMax Bricks (Bricks). The modular Brick architecture reduces complexity and allows for easier system configuration and deployment. This architecture also allows the system to scale while continuing to deliver predictable high performance.

There are two types of Bricks available for PowerMax:

- The open systems Brick supports configurations with Fibre Channel and / or iSCSI connectivity with FBA device formatting. The Brick also can be configured for file storage using embedded NAS.
- The mainframe zBrick supports configurations with FICON connectivity and CKD device formatting.

**Note:** In this document, the term “Brick” will be used when discussing features and functions applicable to both the open systems and the mainframe. When discussing features specific to mainframe, the term zBrick will be specifically called out.

The initial system Brick includes a single engine comprised of two directors, two system power supplies (SPS), and two 24-slot 2.5” NVMe Drive Array Enclosures (DAE24) pre-configured with an initial total usable capacity. The PowerMax 2000 comes with an initial capacity of 11 or 13 TBu depending upon RAID configuration. The PowerMax 8000 comes with an initial capacity of 53 TBu for open systems, 13 TBu for mainframe, and 66 TBu for mixed systems. Each Brick comes preloaded with PowerMaxOS.
The Brick concept allows PowerMax to scale up and scale out. Customers can scale up by adding Flash Capacity Packs. Each Flash Capacity Pack for the PowerMax 8000 has 13 TBu of usable storage, and 11 TBu or 13 TBu for the PowerMax 2000 model, depending upon the RAID protection type selected. PowerMax scales out by aggregating up to two Bricks for the PowerMax 2000, and up to eight for the PowerMax 8000 in a single system with fully shared connectivity, processing, and capacity resources. Scaling out a PowerMax system by adding additional Bricks produces a predictable, linear performance improvement regardless of the workload.

The following table provides details about the Bricks used in the PowerMax 2000 and PowerMax 8000:

<table>
<thead>
<tr>
<th>Component</th>
<th>Specification</th>
<th>PowerMax 2000</th>
<th>PowerMax 8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Layout</td>
<td>Floor Tile Space Required</td>
<td>1</td>
<td>1 - 2</td>
</tr>
<tr>
<td>Compute</td>
<td># of Bricks per System</td>
<td>1 - 2</td>
<td>1 - 8</td>
</tr>
<tr>
<td></td>
<td>Support for mainframe zBrick</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Brick CPUs</td>
<td>12 core, 2.5 GHz Intel Broadwell (2 per director)</td>
<td>18 core, 2.8 GHz Intel Broadwell (2 per director)</td>
</tr>
<tr>
<td></td>
<td># of cores per Brick</td>
<td>48</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Maximum # of cores per system</td>
<td>96</td>
<td>576</td>
</tr>
<tr>
<td>Cache</td>
<td>Brick Cache Options</td>
<td>512 GB, 1 TB, and 2 TB (DDR4)</td>
<td>1 TB and 2 TB (DDR4)</td>
</tr>
<tr>
<td></td>
<td>Maximum cache per system</td>
<td>4 TB</td>
<td>16 TB</td>
</tr>
<tr>
<td></td>
<td>Mixed cache support</td>
<td>Yes</td>
<td>Yes, for 4+ engine systems</td>
</tr>
<tr>
<td>Ports and Modules</td>
<td>Maximum FE modules per Brick</td>
<td>8 (32 total FE ports per Brick)</td>
<td>6 (24 total FE ports per Brick – open systems / Mixed)</td>
</tr>
<tr>
<td></td>
<td>Maximum FE modules per zBrick</td>
<td>NA</td>
<td>6 – 8 (24 or 32 total FICON ports per zBrick)</td>
</tr>
<tr>
<td></td>
<td>Maximum FE ports per system</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>Drives and Capacity</td>
<td>Maximum # of drives per system</td>
<td>96</td>
<td>288</td>
</tr>
<tr>
<td></td>
<td>Maximum open systems effective capacity per system</td>
<td>1 PBe</td>
<td>4 PBe</td>
</tr>
<tr>
<td></td>
<td>Maximum mainframe usable capacity per system</td>
<td>NA</td>
<td>1.7 PBu</td>
</tr>
<tr>
<td></td>
<td>Starter Brick usable capacity</td>
<td>11 or 13 TBu</td>
<td>53 TBu (OS)</td>
</tr>
<tr>
<td></td>
<td>Flash Capacity Pack increment size</td>
<td>11 or 13 TBu</td>
<td>13 TBu</td>
</tr>
<tr>
<td></td>
<td>RAID Options</td>
<td>RAID 5 (3+1), RAID 5 (7+1), RAID 6 (6+2)</td>
<td>RAID 5 (7+1), RAID 6 (6+2)</td>
</tr>
<tr>
<td></td>
<td>Supported Brick NVMe Flash Drive Sizes</td>
<td>1.92 TB, 3.84 TB, 7.68 TB</td>
<td>1.92 TB, 3.84 TB, 7.68 TB</td>
</tr>
<tr>
<td>Supported zBrick NVMe Flash Drives Sizes</td>
<td>NA</td>
<td>1.92 TB, 3.84 TB, 7.68 TB</td>
<td></td>
</tr>
</tbody>
</table>

(1) Default zBrick comes with 2 FICON modules. Extra FICON modules can be ordered individually
(2) A zBrick can support up to 32 FE ports if SRDF compression is not used in the configuration
(3) The PowerMax 2000 starter Brick and capacity increments can be either 11 TBu (RAID 5 3+1) or 13 TBu (RAID 5 7+1, RAID 6 6+2)
(a) Dell EMC uses PBu (and TBu) to define usable storage capacity in the absence of compression, referring to the amount of usable physical storage in the box. Dell EMC uses PBe (and TBe) to define effective storage capacity in the presence of compression. For example, if a system has 50 TBu of physical storage, and it is compressible on a 2:1 basis, then the system has 100 TBe (effective storage).

3.2.1 Engines

The core of the Brick is the engine. The engine is the central I/O processing unit, redundantly built for high availability. Each Brick consists of:

- Redundant directors that contain multi-core CPUs and memory modules
- Interfaces to universal I/O modules, such as front-end, back-end, InfiniBand, and flash I/O modules

The communication backbone of the Brick is the trusted Dynamic Virtual Matrix Architecture. Fundamentally, the virtual matrix enables inter-director communications over redundant internal InfiniBand fabrics. The InfiniBand fabric provides a foundation for a highly scalable, extremely low latency, and high bandwidth backbone which is essential for an all flash array. This capability is also essential for allowing the PowerMax to scale upwards and scale outwards in the manner that it does.

Figure 2   Brick Engine Director

Brick CPU Core Configurations
Each Brick engine has two directors, with each director having dual CPU sockets which can support multi-core, multi-threaded Intel processors. The following table details the engine CPU core layout for each PowerMax model:

<table>
<thead>
<tr>
<th>PowerMax Model</th>
<th>Engine CPU Type</th>
<th>CPU Cores</th>
<th>Cores per Brick Engine</th>
<th>Max Cores per System</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerMax 2000</td>
<td>Dual Intel Broadwell, 12 core, 2.5 GHz</td>
<td>24</td>
<td>48</td>
<td>96 (2 Bricks max.)</td>
</tr>
<tr>
<td>PowerMax 8000</td>
<td>Dual Intel Broadwell, 18 core, 2.8 GHz</td>
<td>36</td>
<td>72</td>
<td>576 (8 Bricks max.)</td>
</tr>
</tbody>
</table>

The Brick engine uses a core pooling mechanism which can dynamically load-balance the cores by distributing them to the front end, back end, and data services (such as SRDF, eNAS, and embedded management) running on the engine. The core pools can be tuned to shift the bias of the pools at any time to front-end heavy or back-end heavy workloads to further optimize the solution for a specific use case.

**Note:** Due to the advanced cooling dynamics of the PowerMax engine, the Intel CPUs can run in “Turbo” mode, providing additional performance capabilities

**Brick Cache Configurations**

Every director has 16 memory slots which can be populated with 32 GB and 64 GB DDR4 DIMMS to achieve up to 1 TB cache per director (2 TB cache maximum per Brick engine).

<table>
<thead>
<tr>
<th>PowerMax Model</th>
<th>Cache per Brick</th>
<th>Max Cache per System</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerMax 2000</td>
<td>512 GB, 1 TB, 2 TB</td>
<td>4 TB (2 Bricks max.)</td>
</tr>
<tr>
<td>PowerMax 8000</td>
<td>1 TB or 2 TB</td>
<td>16 TB (8 Bricks max.)</td>
</tr>
</tbody>
</table>

On single engine PowerMax 2000 systems, cache is mirrored within the engine across the directors. This is also true for multi-engine PowerMax 2000 systems and single engine PowerMax 8000 systems. On multi-engine PowerMax 8000 systems, cache is mirrored across directors in different engines for added redundancy.

Both the PowerMax 2000 and PowerMax 8000 can support engine configurations with differing cache sizes (mixed cache). For dual engine PowerMax 2000 models, the system can use engines with different cache sizes between the engines which are one cache size smaller or larger than the other engine in the system. For example, cache on engine 1 can be 1 TB while the cache on engine 2 is 512 GB. This would yield a total cache size of 1.5 TB for the system. Valid mixed cache configurations for the PowerMax 2000 are shown in the following table:

**Table 5  Supported PowerMax 2000 Mixed Cache Configurations**

<table>
<thead>
<tr>
<th>Mixed Cache Configuration</th>
<th>Smallest Engine Cache Size</th>
<th>Largest Engine Cache Size</th>
<th>Total System Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration 1</td>
<td>512 GB</td>
<td>1 TB</td>
<td>1.5 TB</td>
</tr>
<tr>
<td>Configuration 2</td>
<td>1 TB</td>
<td>2 TB</td>
<td>3 TB</td>
</tr>
</tbody>
</table>
Mixed cache configurations are available on the PowerMax 8000; but require a minimum of four Bricks or zBricks in the system. The following table details the supported mixed cache configurations available for the PowerMax 8000:

Table 6  Supported PowerMax 8000 Mixed Cache Configurations

<table>
<thead>
<tr>
<th>Number of Bricks in System</th>
<th>Mixed Cache Configuration</th>
<th>Smallest Engine Cache Size</th>
<th>Largest Engine Cache Size</th>
<th>Total System Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Not Supported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Not Supported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Configuration 4.1</td>
<td>2 engines at 1 TB</td>
<td>2 engines at 2 TB</td>
<td>6 TB</td>
</tr>
<tr>
<td>5</td>
<td>Configuration 5.1</td>
<td>2 engines at 1 TB</td>
<td>3 engines at 2 TB</td>
<td>8 TB</td>
</tr>
<tr>
<td>5</td>
<td>Configuration 5.2</td>
<td>3 engines at 1 TB</td>
<td>2 engines at 2 TB</td>
<td>7 TB</td>
</tr>
<tr>
<td>6</td>
<td>Configuration 6.1</td>
<td>2 engines at 1 TB</td>
<td>4 engines at 2 TB</td>
<td>10 TB</td>
</tr>
<tr>
<td>6</td>
<td>Configuration 6.2</td>
<td>4 engines at 1 TB</td>
<td>2 engines at 2 TB</td>
<td>8 TB</td>
</tr>
<tr>
<td>7</td>
<td>Configuration 7.1</td>
<td>2 engines at 1 TB</td>
<td>5 engines at 2 TB</td>
<td>12 TB</td>
</tr>
<tr>
<td>7</td>
<td>Configuration 7.2</td>
<td>5 engines at 1 TB</td>
<td>2 engines at 2 TB</td>
<td>9 TB</td>
</tr>
<tr>
<td>7</td>
<td>Configuration 7.3</td>
<td>3 engines at 1 TB</td>
<td>4 engines at 2 TB</td>
<td>11 TB</td>
</tr>
<tr>
<td>7</td>
<td>Configuration 7.4</td>
<td>4 engines at 1 TB</td>
<td>3 engines at 2 TB</td>
<td>10 TB</td>
</tr>
<tr>
<td>8</td>
<td>Configuration 8.1</td>
<td>2 engines at 1 TB</td>
<td>6 engines at 2 TB</td>
<td>14 TB</td>
</tr>
<tr>
<td>8</td>
<td>Configuration 8.2</td>
<td>6 engines at 1 TB</td>
<td>2 engines at 2 TB</td>
<td>10 TB</td>
</tr>
<tr>
<td>8</td>
<td>Configuration 8.3</td>
<td>4 engines at 1 TB</td>
<td>4 engines at 2 TB</td>
<td>12 TB</td>
</tr>
</tbody>
</table>

Note: Cache within an engine can be upgraded (capacity added), but cache cannot be downgraded (capacity removed).

**PowerMaxOS**

Each PowerMax engine comes with PowerMaxOS 5978 installed. PowerMaxOS is derived from the trusted and proven HYPERMAX OS used by the legacy VMAX3 and VMAX All Flash arrays; however, PowerMaxOS has been re-written to take advantage of NVMe architectures. PowerMaxOS continues to provide industry-leading high availability, I/O management, quality of service, data integrity validation, data movement, and data security within an open application platform. PowerMaxOS uses a real-time, non-disruptive storage hypervisor that manages and protects embedded services by extending high availability to services that traditionally would have run external to the array. The primary function of PowerMaxOS is to manage the core operations performed on the array, which include:

- Processing I/O from hosts
- Implementing RAID protection
- Optimizing performance by allowing direct access to hardware resources
- Managing and monitoring the system

### 3.2.2 Drive Array Enclosures (DAE) and Drive Configurations

Each Brick comes with two 24-slot, dual-ported, 2.5” PCIe NVMe DAEs (DAE24). These DAEs use redundant, hot-swappable Link Control Cards (LCCs) which provide PCIe I/O connectivity to the NVMe flash drives. Aside from redundant LCCs, the DAE24 features redundant power supplies with separate power
feeds, providing N+1 power and cooling, resulting in an energy efficient consumption of 25 watts per drive slot. The DAE24 is 2U high and 19" deep.

Figure 3  Brick NVMe DAE24

PowerMax Drive Options and Universal Sparing

Both PowerMax 2000 and PowerMax 8000 support 1.92 TB, 3.84 TB, and 7.68 TB NVMe flash drive capacities. All the drive sizes are 2.5” and feature a dual ported U.2 form factor PCIe interface. These drive capacities can be intermixed on the system. PowerMax uses universal sparing: systems with mixed drive configurations can use a single spare drive with the largest capacity size in the system. For example, if a system uses both 3.84 TB and 7.68 TB drives in the configuration, only one 7.68 TB drive needs to be configured as a spare, because it can replace either the 3.84 TB or 7.68 TB drives.

Figure 4  Universal Sparing Example

The Universal Spare can be also used in mixed systems with multiple storage resource pools (SRP). For example, if the mixed system uses 7.68 TB drives for the FBA SRP and 3.84 TB drives for the CKD SRP, the system will still only need a single 7.68 TB drive spare.

There are some rules for mixing drive capacities and universal sparing on PowerMax:

- There can be only two drive size types on the system and the drive sizes need to be one sequential drive size apart. For example, the system can have a mix of 1.92 TB and 3.84 TB drives, but not 1.92 TB and 7.68 TB sizes.
- On mixed systems, drive mixing rules apply only within the individual SRPs. For example, the CKD SRP can have 1.92 TB and 3.84 TB drives, while the FBA SRP can have 3.84 TB and 7.68 TB drives. In this case, both SRPs would require a unique spare. The CKD SRP could use a single 3.84 TB spare, while the FBA SRP could use a single 7.68 TB spare. A single 7.68 TB drive could not be used as a universal spare for both the CKD and FBA SRPs in this case because it is beyond one drive size apart from the 1.92 TB drives in the CKD SRP.
PowerMax Smart RAID

PowerMax uses a new active/active RAID group accessing scheme called Smart RAID. This allows RAID groups to be shared across directors, giving each director active access to all drives on the Brick or zBrick.

The use of Smart RAID on PowerMax provides customers with performance benefits as both directors on an engine can drive I/O to all the flash drives. This creates balanced configurations in the system regardless of the number of RAID groups. Smart RAID also allows for increased flexibility and efficiency as customers can order PowerMax systems with a single RAID group allowing for a minimum of 9 drives per engine with RAID 5 (7+1) or RAID 6 (6+2 and 1 spare) and as little as 5 drives per system for a PowerMax 2000 with RAID 5 (3+1 and 1 spare). This leaves more drive slots available for capacity upgrades in the future. When the system is scaled up, customers have more flexibility because flash capacity pack increments can be a single RAID group. **PowerMax 2000 DAE connectivity and drive allocation schemes**

Smart RAID and Universal Spare allow flexible connectivity and drive allocation schemes to occur with the PowerMax DAE. With the PowerMax 2000, each engine director has two NVMe I/O Modules. Each I/O module has two redundant paths. One path connects to either Link Control Canister (LCC)A or LCC B in DAE 1 while the other path connects to either LCC A or LCC B in DAE 2. Each path from the NVMe I/O module to the LCC is a four lane PCIe Gen3 connection (4 GB/sec).

The following diagrams detail the DAE connectivity layout and drive allocation schemes for the PowerMax 2000:
The PowerMax 2000 can use the RAID 5 (3+1), RAID 5 (7+1), or RAID 6 (6+2) protection schemes. Only one RAID protection scheme can be applied on the system. When populating the PowerMax 2000 DAEs, each engine requires a minimum of 1 RAID group including spare drives. There are two spare drive slots in a PowerMax 2000 system (slot 24 in each DAE). However, there can be only one spare drive for each Brick or zBrick. When populating the drives into the system, the drives are alternately placed in DAE1 and DAE2.
Figure 8  PowerMax 2000 DAE Drive Slot Allocations for a single Brick

DAE 2

Slots 1 – 24 accessed by Brick 1

Figure 9  PowerMax 2000 DAE Drive Slot Allocations for a dual Brick

DAE 4

Slots 1 – 24 accessed by Brick 2

DAE 3

Slots 1 – 24 accessed by Brick 2

DAE 2

Slots 1 – 24 accessed by Brick 1

DAE 1

Slots 1 – 24 accessed by Brick 1

Spare Drive Slot
Using Smart RAID, the PowerMax 2000 can be configured for a minimum capacity 11.28 TBu using RAID 5 (3+1) or RAID 6 (6+2), and 13.16 TBu using RAID 5 (7+1) while using only a single RAID group. Depending on drive size and RAID protection scheme used, these minimum configurations can be achieved using only 4 drives. The maximum number of usable drives which can be used with a single PowerMax 2000 Brick is 40 plus 1 spare drive for RAID 5 (7+1) or RAID 6 (6+2) configurations; and 44 usable drives plus 1 spare using a RAID 5 (3+1) configuration. The table below details the supported PowerMax 2000 drive configurations needed to achieve the minimum required system capacity:

<table>
<thead>
<tr>
<th>Supported Drive Capacity</th>
<th>RAID 5 (3+1)</th>
<th>RAID 5 (7+1)</th>
<th>RAID 6 (6+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>Raw Formatted</td>
<td>Raw Formatted</td>
<td>Raw Formatted</td>
</tr>
<tr>
<td>1.92 TB</td>
<td>1.88 TB</td>
<td>11.28 TBu (8 Drives)</td>
<td>11.28 TBu (8 Drives)</td>
</tr>
<tr>
<td>3.84 TB</td>
<td>3.76 TB</td>
<td>11.28 TBu (4 Drives)</td>
<td>11.28 TBu (4 Drives)</td>
</tr>
<tr>
<td>7.68 TB</td>
<td>7.52 TB</td>
<td>22.56 TBu (4 Drives)</td>
<td>22.56 TBu (4 Drives)</td>
</tr>
</tbody>
</table>

(1) Minimum PowerMax 2000 RAID 5 (3+1) configurations, which use a single RAID group of 4 drives, are not intended for high performance. To achieve higher performance, use two or more RAID groups of 4 drives each in a minimum PowerMax 2000 RAID 5 (3+1) configuration.

**PowerMax 2000 DAE and Drive Allocation Summary:**

- PowerMax 2000 supports Universal Sparing and Smart RAID.
- Mixed drive sizes can be used in the system. Drive sizes need to be one size increment apart (for example, 1.92 TB and 3.84TB, or 3.84 TB and 7.68 TB).
- Only one spare drive per Brick is required. The spare needs to be the same size as the largest drive size used in the system.
- Every PowerMax 2000 system requires at least one RAID group.
- DAEs are not shared by the engines in a dual Brick PowerMax 2000 configuration.
- RAID groups are associated with a single Brick engine.
- Only one RAID protection scheme per PowerMax 2000 system is allowed.
- RAID 5 (3+1) requires a minimum of 4 drives while RAID 5 (7+1) and RAID 6 (6+2) require a minimum of 8 drives.
PowerMax 8000 DAE connectivity and drive allocation schemes

The PowerMax 8000 uses Smart RAID and Universal Sparing to achieve the densest possible engine and flash drive capacity configurations in the industry. In order to achieve these high densities, the PowerMax 8000 uses a different DAE connectivity and drive allocation schemes than those used in the PowerMax 2000. In systems using a single Brick, the DAE connectivity is like the PowerMax 2000; however, drive slots 15-24 in the DAE 2 are reserved for future scale out of a second Brick.

Figure 10  PowerMax 8000 Single Engine DAE Connectivity

When a second Brick is added into the system, a third DAE is also added, and drive slots 15-24 of the DAE 2 on the first Brick can be populated and accessed by the second Brick. This is made possible as the 3rd and 4th Mini-SAS HD PCIe I/O ports on the LCCs in DAE 2 are used by the second Brick as shown in the following diagram:

Figure 11  PowerMax 8000 Dual Engine DAE Connectivity
The PowerMax 8000 can use the RAID 5 (7+1) or RAID 6 (6+2) protection schemes. Like the PowerMax 2000, only one RAID protection scheme can be applied on the system, even on systems that have both mainframe and open systems SRPs. When populating the PowerMax 8000 DAEs, each Brick engine must have at least 1 RAID group including spare drives. For single Brick configurations, drives can be added in slots 1-24 of DAE 1, and in slots 1-12 on DAE 2. Slots 13 and 14 in DAE 2 are reserved for spare drives. This results in a maximum of 32 usable drive slots plus spares in a single Brick system. As with the PowerMax 2000, only one spare drive is required per Brick.

Figure 12  PowerMax 8000 drive slot allocations for a single Brick

A third DAE (DAE 3) is added to the system when adding a second Brick into the system. The second Brick uses slots 1-24 of DAE 3 and shares DAE 2 with the first Brick, using slots 17-24 in DAE 2. Slots 15 and 16 in DAE 2 are reserved for the second Brick spare drives. The diagram below shows how drive slots are allocated in a dual Brick PowerMax 8000 system:

Figure 13  PowerMax 8000 drive slot allocations for dual Bricks
A PowerMax 8000 can be configured for open systems, mainframe, or mixed open systems and mainframe workloads. Like the PowerMax 2000, it uses Universal Sparing plus Smart RAID, while providing support for 1.92 TB, 3.84 TB, and 7.68 TB capacity NVMe drives. Drives can be mixed in the system but must be one drive size apart (for example, 1.92 TB and 3.84 TB). The resulting minimum drive configuration and minimum capacities will differ depending on the drive size, RAID protection, and the system configuration chosen. For example, an open systems PowerMax 8000 requires a minimum initial capacity of 56.24 TBu for RAID 5 (7+1) protection, while mainframe configurations support 13.16 TBu of RAID 5 (7+1) initial capacity. The following tables detail the various PowerMax 8000 minimum capacity and capacity pack increments for the different system configurations.

Table 8  Open Systems PowerMax 8000 Minimum Capacity and Capacity Pack Increments

<table>
<thead>
<tr>
<th>Supported Drive Capacity</th>
<th>RAID 5 (7+1)</th>
<th>RAID 6 (6+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw</td>
<td>Formatted</td>
<td>Minimum Capacity</td>
</tr>
<tr>
<td>1.92 TB</td>
<td>1.88 TB</td>
<td>52.64 TBu (32 Drives)</td>
</tr>
<tr>
<td>3.84 TB</td>
<td>3.76 TB</td>
<td>52.64 TBu (16 Drives)</td>
</tr>
<tr>
<td>7.68 TB</td>
<td>7.52 TB</td>
<td>52.64 TBu (8 Drives)</td>
</tr>
</tbody>
</table>

(1) Although supported, Dell EMC does not recommend this configuration, because all available drive slots are taken, leaving no room for future capacity expansion within the original Brick.
(2) This initial minimum capacity is achieved by using a mixed configuration of 16 x 3.84 TB and 8 x 1.92 TB drives. The system universal spare must be a 3.84 TB drive.
Table 9  Mainframe PowerMax 8000 Minimum Capacity and Capacity Pack Increments

<table>
<thead>
<tr>
<th>Drive Capacity</th>
<th>RAID 5 (7+1)</th>
<th>RAID 6 (6+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum Capacity</td>
<td>Capacity Pack</td>
</tr>
<tr>
<td>Raw</td>
<td>Formatted</td>
<td></td>
</tr>
<tr>
<td>1.92 TB</td>
<td>1.88 TB</td>
<td>13.16 TBu (8 Drives)</td>
</tr>
<tr>
<td>3.84 TB</td>
<td>3.76 TB</td>
<td>26.32 TBu (8 Drives)</td>
</tr>
<tr>
<td>7.68 TB</td>
<td>7.52 TB</td>
<td>52.64 TBu (8 Drives)</td>
</tr>
</tbody>
</table>

Table 10  Mixed open systems and mainframe PowerMax 8000 Minimum Capacity and Capacity Pack Increments

<table>
<thead>
<tr>
<th>Drive Capacity</th>
<th>RAID 5 (7+1)</th>
<th>RAID 6 (6+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MF Capacity Pack</td>
<td>OS Capacity Pack</td>
</tr>
<tr>
<td>Raw</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.92 TB</td>
<td>52.64 TBu (32 Drives)</td>
<td>13.16 TBu (8 Drives)</td>
</tr>
<tr>
<td>3.84 TB</td>
<td>52.64 TBu (16 Drives)</td>
<td>26.32 TBu (8 Drives)</td>
</tr>
<tr>
<td>7.68 TB</td>
<td>52.64 TBu (8 Drives)</td>
<td>52.64 TBu (8 Drives)</td>
</tr>
</tbody>
</table>

(1) Although supported, Dell EMC does not recommend this configuration as all available 32 drive slots are taken leaving no room for future capacity expansion within the original Brick.

(2) This initial minimum capacity is achieved by using a mixed configuration of 16 x 3.84 TB and 8 x 1.92 TB drives. The system universal spare must be a 3.84 TB drive.

PowerMax 8000 DAE and Drive Allocation Summary:

- PowerMax 8000 supports Universal Sparing and Smart RAID.
- Mixed drive sizes can be used in the system. Drive sizes need to be one size increment apart (for example, 1.92 TB and 3.84 TB, or 3.84 TB and 7.68 TB).
- Only one spare drive per Brick is required. The spare needs to be the same size as the largest drive used in the system.
- Every PowerMax 8000 system requires a minimum of one RAID group.
- RAID groups are associated to a single Brick engine.
- Only one RAID protection scheme is allowed per PowerMax 8000 system.
- RAID 5 (7+1) and RAID 6 (6+2) protection schemes require a minimum of 8 drives plus one spare.
- Every even numbered Brick will share a DAE with the previous odd numbered Brick.
- Odd numbered Bricks will have 24 plus 12 drives. Even numbered Bricks will have 24 plus 10 drives.
### 3.2.3 Director Slot Layout and Connectivity Options

The Brick engine architecture uses a series of hot swappable modules that plug into slots in the engine directors. These modules include:

- Engine cooling fans and power supplies in slots accessible from the front of the engine director.
- I/O modules, management modules, and control stations in slots accessible from the rear of the engine director.

The following table describes the module components used in a Brick engine director:

<table>
<thead>
<tr>
<th>Director Component</th>
<th>QTY per Director</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Supply</td>
<td>2</td>
<td>Provides redundant power to director</td>
</tr>
<tr>
<td>Fan</td>
<td>5</td>
<td>Provides director cooling</td>
</tr>
<tr>
<td>Management Module</td>
<td>1</td>
<td>Manage environmental functionality</td>
</tr>
<tr>
<td>NVMe Flash I/O Module</td>
<td>Up to 4</td>
<td>The flash I/O modules use NVMe technology to safely store data in cache during the vaulting sequence (800 GB)</td>
</tr>
<tr>
<td>Front-end I/O Module</td>
<td>Up to 4</td>
<td>Provide front-end connectivity to the array. There are different types of front-end I/O modules that allow connectivity to various interfaces including Fibre Channel, iSCSI, FICON, SRDF, and embedded NAS (eNAS)</td>
</tr>
<tr>
<td>NVMe PCIe Back-end I/O Module</td>
<td>2</td>
<td>Dual-ported PCIe 4x Gen3 interface to the NVMe storage (8 GB/sec)</td>
</tr>
<tr>
<td>Data Reduction Module</td>
<td>1</td>
<td>Performs inline data compression and deduplication, as well as SRDF compression</td>
</tr>
<tr>
<td>Fabric I/O Module</td>
<td>1</td>
<td>Provides connectivity between directors. In multi-engine PowerMax 8000 systems the fabric I/O modules are connected to an internal InfiniBand switch</td>
</tr>
</tbody>
</table>
The following diagram shows the director module layouts for the PowerMax 2000:

Figure 14  PowerMax 2000 director module layout by slot number

Both single-engine and multi-engine PowerMax 2000 systems use the same director module layout. Both configurations use two NVMe flash modules residing in slots 0 and 6 on each director. Slots 7 houses the data reduction module. Slots 2, 3, 8, and 9 are used for front-end connectivity modules. Slots 4 and 5 contain the NVMe PCIe back-end connectivity modules. Slot 10 houses the fabric modules. Slot 1 is reserved for future use.

The following diagrams detail the director module layouts for single engine and multi-engine PowerMax 8000 systems:

Figure 15  PowerMax 8000 director module layout by slot number – Single Engine System
Unlike the PowerMax 2000, there are differences in the director module layouts between single-engine and multi-engine PowerMax 8000 systems. Single-engine PowerMax 8000 systems use four NVMe Flash modules. These modules occupy director slots 0, 1, 6, and 7. The data reduction module resides in slot 9. Slots 2, 3, and 8 are used for front-end connectivity modules.

Multi-engine PowerMax 8000 systems use three NVMe flash modules, occupying slots 0, 1, and 6. The data reduction module occupies slot 7. This leaves an additional slot for a front-end connectivity module allowing multi-engine PowerMax 8000 systems to have four front-end connectivity modules, occupying director slots 2, 3, 8, and 9.

Notes:
For PowerMax 8000 systems that only had a single engine originally, the single-engine configuration of three slots available for front-end modules is applied to each additional engine added to the system when the system is scaled out. When additional engines are added to PowerMax 8000 systems that were originally multi-engine systems, these engines can have up to four slots available for front-end modules. On multi-engine systems, the compression module must use the same director slots on each engine. Data compression and deduplication are not available on the mainframe PowerMax 8000, but SRDF compression is available. On Mainframe PowerMax 8000 systems (zBricks) which use SRDF compression only, place a compression module on the director with ports configured for SRDF. On single-engine configuration systems, place the SRDF compression module in slot 9; while on multi-engine configuration systems, place the SRDF compression module in slot 7.

Both the PowerMax 2000 and the PowerMax 8000 provide multiple front-end connections that implement several protocols and speeds. The table below highlights the various front-end connectivity modules available for a PowerMax system:
Table 12  Supported Brick Front-End Connectivity Modules

<table>
<thead>
<tr>
<th>Connectivity Type</th>
<th>Module Type</th>
<th>Number of Ports</th>
<th>Mix With Protocols</th>
<th>Supported Speeds (Gbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibre Channel</td>
<td>16 Gbps FC</td>
<td>4</td>
<td>SRDF</td>
<td>2 / 8 / 16</td>
</tr>
<tr>
<td>SRDF</td>
<td>10 GbicE</td>
<td>4</td>
<td>iSCSI</td>
<td>10</td>
</tr>
<tr>
<td>iSCSI</td>
<td>10 GbicE</td>
<td>4</td>
<td>SRDF</td>
<td>10</td>
</tr>
<tr>
<td>FICON (1)</td>
<td>16 Gbps FICON</td>
<td>4</td>
<td>Single / Multi Mode</td>
<td>4 / 8 /16</td>
</tr>
<tr>
<td>eNAS</td>
<td>10 GbicE</td>
<td>2</td>
<td>None</td>
<td>10</td>
</tr>
<tr>
<td>eNAS</td>
<td>10 GbicE (Copper)</td>
<td>2</td>
<td>None</td>
<td>10</td>
</tr>
<tr>
<td>eNAS Tape Backup</td>
<td>8 Gbps FC</td>
<td>4</td>
<td>None</td>
<td>2 / 4 / 8</td>
</tr>
</tbody>
</table>

(1) Supported on PowerMax 8000 only

Notes:

- Each Brick engine has at least one front-end module pair (one front-end module per director)
- Since the number of front-end modules used in the Brick engine depends on the customer’s requirements, some director slots may not be used.
- Front-end modules for Fibre Channel support both multi-mode (MM) and single mode (SM) optics. Front-end modules for 10 GbE support only MM optics.

3.3 Streamlined Software Packaging

Software for the PowerMax is available in different packages, each with additional, optional features. There are two packages for open systems named the Essentials package and the Pro package. These are available on all PowerMax arrays. There are two additional packages available for PowerMax 8000: the zEssentials and zPro packages. These packages are available for mainframe environments only.

The open system packages are:

Table 13  PowerMax open systems Software Packaging Options

<table>
<thead>
<tr>
<th>Feature</th>
<th>Essentials Package Included</th>
<th>Essentials Package Options</th>
<th>Pro Package Included</th>
<th>Pro Package Options</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerMaxOS</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes Migration Tools, VVOLS, QoS (3)</td>
</tr>
<tr>
<td>Embedded Management</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>Includes Unisphere for PowerMax, Database Storage Analyzer, Solutions Enabler, REST APIs, SMI-S</td>
</tr>
<tr>
<td>Advanced Data Reduction</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes inline compression and deduplication</td>
</tr>
<tr>
<td>Local Replication</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes Timefinder SnapVX</td>
</tr>
<tr>
<td>AppSync Starter Pack</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The mainframe software packages and options are shown in the table below:

Table 14  PowerMax Mainframe Software Packaging Options (PowerMax 8000 only)

<table>
<thead>
<tr>
<th>Feature</th>
<th>zEssentials Package Included</th>
<th>zEssentials Package options</th>
<th>zPro Package Included</th>
<th>zPro Package options</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerMaxOS</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes Migration Tools, QoS</td>
</tr>
<tr>
<td>Embedded Management</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes Unisphere for PowerMax, Database Storage Analyzer, Solutions Enabler, REST APIs, SMI-S</td>
</tr>
<tr>
<td>Local Replication</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes Timefinder SnapVX</td>
</tr>
<tr>
<td>Mainframe Essentials</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes Compatible High Performance FICON (zHPF) and Compatible PAV (Dynamic, Hyper, and SuperPAV) support</td>
</tr>
<tr>
<td>Remote Replication Suite</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>Includes SRDF/S/A/STAR</td>
</tr>
<tr>
<td>Unisphere 360</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4 Flash Optimization

All flash-based storage systems demand the highest levels of performance and resilience from the enterprise data storage platforms that support them. The foundation of a true all flash array is an architecture that can fully leverage the aggregated performance of modern high-density flash drives while maximizing their useful life. Many features are built into the architecture of PowerMax to maximize flash drive performance and longevity. This section discusses these features in detail.

3.4.1 Cache Architecture and Caching Algorithms

PowerMax is built upon a very large, high-speed DRAM cache-based architecture, driven by highly complex and optimized algorithms. These algorithms accelerate data access by avoiding physical access to the back end whenever possible. Dell EMC has spent many years developing and optimizing caching algorithms. The algorithms used by PowerMax optimize reads and writes to maximize I/Os serviced from cache and minimize access to back-end flash drives. The system also monitors I/O patterns and proactively populates cache based on access to increase the chances of cache hits.

Some of the techniques used by the cache algorithms to minimize disk access are:

- 100% of host writes are cached
- More than 50% of reads are cached
- Recent data is held in cache for long periods, as that is the data most likely to be requested again
- Intelligent algorithms de-stage in a sequential manner

3.4.2 Understanding Flash Cell Endurance

Write cache management is essential to improving performance, but it is also the key part of how PowerMax helps extend the endurance of flash drives. Flash drive longevity and endurance are most impacted by writes, particularly small block random writes. Writing to a flash cell requires that any old data is first erased from the cell, and then the cell is programmed with the new data. This process is called the Program and Erase Cycle (P/E Cycle). Each flash cell has a finite number of P/E Cycles that it can endure before it wears out (can no longer hold data). Most modern flash cells can endure several thousand P/E Cycles.

One of the peculiarities of flash is that writes are spread out across a flash page (typically KBs in size); however, prior to the write operation, the existing data in the entire flash block (typically MBs in size) that the page is located in must be erased. Prior to erasing the page, the flash controller chip finds an empty (erased) location on the drive and copies (writes) any existing data from the page to that location.
flash writes data, a simple 4 KB write from a host could result in many times that amount of data being written internally on the drive causing P/E cycling on many cells. This write multiplying effect is called Write Amplification and is detrimental to flash cell endurance. This effect is even more dramatic with small block random write workloads. In this situation, many small block random writes tend to “buckshot” across the drive, impacting an even greater number of cells and invoking P/E cycling on a much larger cell area. Write amplification is not nearly as significant with larger sequential writes as this data is written sequentially to a single flash block, thereby aligning better with flash page sizes and containing the P/E cycling to a smaller area.

3.4.3 PowerMax Write Amplification Reduction

Write amplification must be properly controlled to ensure the longevity of flash devices. Controlling flash cell write amplification is one of PowerMax’s greatest strengths and what truly sets it apart from other flash arrays. Aside from intelligent caching algorithms, which keep data in cache as long as possible, the PowerMax uses additional methods to minimize the number of writes to flash. These methods are:

- **Write Folding** – Write Folding avoids unnecessary drive I/Os when hosts re-write to a particular address range. This re-written data is simply replaced in cache and never written to the flash drive. Write folding can reduce writes to the flash drives by up to 50%.

- **Write Coalescing** – Write Coalescing merges subsequent small random writes from different times into one large sequential write. These larger writes to the flash drives align much better with the page sizes within the flash drive itself. Using write coalescing, PowerMax can take a highly random write host I/O workload and make it appear as a sequential write workload to the flash drives.

- **Advanced Wear Analytics** – PowerMax also includes advanced drive wear analytics optimized for high capacity flash drives to make sure writes are distributed across the entire flash pool to balance the load and avoid excessive writes and wear to particular drives. Not only does this help manage the flash drives in the storage pools, it makes it easy to add and rebalance additional storage into the system.

All of the write amplification reduction techniques used by PowerMax result in a significant reduction in writes to the back end, which in turn significantly increases the longevity of the flash drives used in the array.

3.4.4 Boosting Flash Performance with PowerMaxOS FlashBoost

Dell EMC is always striving to improve performance in its products. With every new hardware platform and release of software, the company makes strong efforts to remove potential bottlenecks which can impede performance in any way. One feature that Dell EMC introduced and has made standard as a part of PowerMaxOS is FlashBoost.

FlashBoost maximizes PowerMaxOS efficiency by servicing read requests directly from the back-end flash drives. This approach eliminates steps required for processing I/O through global cache and reduces the latency for reads, particularly for flash drives. Customers with heavy read miss workloads residing on flash can see up to 100% greater IOPS performance.
3.5 Reliability, Availability, and Serviceability (RAS)

PowerMax arrays are based on a revolutionary design and include key enhancements that improve the reliability, availability, and serviceability of the new systems – ideal choices for critical applications and 24x7 environments demanding uninterrupted access to information.

PowerMax 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 systems to remain online and operational during component repair. All critical components are fully redundant, including director boards, global memory, internal data paths, power supplies, battery backup, and all NVMe back-end components. Periodically, the system tests all components. PowerMaxOS reports errors and environmental conditions to the host system as well as to the Customer Support Center.

PowerMaxOS validates the integrity of data at every possible point during the lifetime of the data. From the point at which data enters an 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 subsystem, allowing the array to provide true end-to-end integrity checking and protection against hardware or software faults.

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

PowerMax’s industry leading reliability, availability, and serviceability (RAS) make it the ideal platform for environments requiring always-on availability. These arrays are designed to provide six-nines of availability in the most demanding, mission-critical environments. Some of the key PowerMax RAS features are summarized below:

- No single points of failure—all components are fully redundant to withstand any component failure.
- Completely redundant and hot-pluggable field-replaceable units (FRUs) ensure repair without taking the system offline.
- Choice of RAID 5 or RAID 6 deployment options to provide the highest level of protection as desired.
- Mirrored cache, where the copies of cache entries are distributed to maximize availability.
- PowerMaxOS Flash Drive Endurance Monitoring – The nature of flash drives is that their NAND flash cells can be written to a finite number of times. This is referred to as flash drive endurance and is reported by drive firmware as a “percentage of life used”. PowerMaxOS periodically collects and monitors this information and uses it to trigger alerts back to Dell EMC Customer Support when a particular drive is nearing its end of useful life.
- Vault-to-flash with battery backup allows for cache-destage-to-flash and an orderly shutdown for data protection in the event of a power failure.
- Active/active remote replication via SRDF/Metro with read/write access to both Site A and Site B ensures instant data access during a site failure.
- Fully non-disruptive upgrades, including loading PowerMaxOS software from small updates to major releases.
• Continuous system monitoring, call-home notification, and advanced remote diagnostics.
• Data at Rest Encryption (D@RE) with integrated RSA® key manager, FIPS 140-2 compliant to meet stringent regulatory requirements.
• T10 DIF data coding, with extensions for protection against lost writes.
• Detailed failure mode effects analysis (FMEA) during design of each component to ensure failure conditions can be handled gracefully.
• Extensive fault detection and isolation, allowing early wear-out detection and preventing the passing of bad data as good.
• Service defined and scripted to ensure success, including color-coded cabling, cable positioning, scripted steps, and checks of key parameters in those scripts.
• All flash cache data vault capable of surviving two key failures, ensuring that the system comes back even when something has failed before the vault and something else fails when returning from the power cycle.
• Support for thermal excursions with graceful shutdown if, for example, a data center loses air conditioning.
• Integrated data protection via Dell EMC ProtectPoint backup and rapid restore, combining the gold standards in backup with industry leading SRDF replication technology.

Note: For more information on PowerMax RAS capabilities, please see the Dell EMC PowerMax Reliability, Availability, and Serviceability Technical Note found in the reference section of this document.

3.6 Data Services
PowerMax Data Services help protect, manage, and move customer data on the array. These services run natively, or embedded inside the PowerMax itself using the PowerMaxOS hypervisor to provide a resource abstraction layer. This allows the data services to share array resources – CPU cores, cache, and bandwidth. Doing this optimizes performance across the entire system and also reduces complexity in the environment as resources do not need to be dedicated. Some of the most sought-after data services that are offered with the PowerMax product line are:

• Advanced data reduction using inline compression and inline deduplication
• Remote replication with SRDF
• Local replication with TimeFinder SnapVX
• Embedded NAS (eNAS)
• Embedded Unisphere for PowerMax (eManagement)

3.6.1 Data Reduction using the Adaptive Compression Engine (ACE)
PowerMax employs inline hardware compression using the Adaptive Compression Engine (ACE). ACE provides PowerMax and VMAX All Flash customers with a data reduction method that provides negligible performance impact while delivering the highest space saving capability. The following design factors make the Dell EMC Adaptive Compression Engine unique:
• **Intelligent compression algorithms** – Intelligent compression algorithms determine the best compression ratios to be used and provide the ability to dynamically modify storage backend layout for the highest data compression efficiencies.

• **Inline hardware data compression** – Inline hardware data compression greatly inhibits the compression function from consuming critical PowerMax system core resources.

• **Activity Based Compression** – Activity Based Compression (ABC) focuses the compression function on the least busy data in the system, while allowing the most active data in the system to bypass the compression workflow. This ensures that all data in the system receives the appropriate compression focus while maintaining optimal response time.

• **Fine Grain Data Packing** – Fine Grain Data Packing which includes a zero reclaim function that prevents the allocation of buffers with all zeros or no actual data.

• **Enhanced Compression** – There is an additional compression algorithm found in PowerMaxOS called Enhanced Compression (EC). The EC algorithm scans already compressed data which the system determines to have been not accessed for a long period of time. The EC algorithm then tries to further reduce this data, to a larger compression ratio, to realize additional capacity savings.

The Adaptive Compression Engine is available to all open systems PowerMax and VMAX All Flash customers at no additional charge. ACE is not currently available for mainframe environments. Glacial Compression is not supported on VMAX All Flash systems.

### 3.6.2 Data Reduction using Inline Deduplication

PowerMax employs inline hardware deduplication to identify repeated data patterns on the array and store those repeated patterns using a single instance in the array’s usable capacity. Depending on customer workloads, inline deduplication along with inline compression give PowerMax the ability to achieve an industry leading data reduction ratio of 4:1 with negligible performance impact.

The following are the important design factors for deduplication on PowerMax:

• **Inline hardware data deduplication** – Inline hardware-based data deduplication prevents the consumption of critical PowerMax system core resources, limiting performance impact. The deduplication and compression functions are performed on the same hardware module in the PowerMax system.

• **Deduplication Algorithm** – PowerMax uses the SHA-2 algorithm to perform the deduplication function. The SHA-2 algorithm produces a unique data identifier for each item of data which is processed through the compression engine. These unique identifiers are stored in a Hash ID table on the PowerMax system.

• **Hash ID Table** – The Hash ID table stores all of the Hash IDs for the data processed through the compression engine. When a new write enters the compression engine, the Hash ID created for the write is compared to the Hash IDs already in the table. If the Hash ID is found to already exist in the Hash ID table, the write is not written to the storage.

• **Dedupe Management Object (DMO)** – The DMO is a 64-byte object which serves as the connection (pointer) between the devices and the single instance of actual data. DMOs are stored in the PowerMax global cache.
Inline deduplication is available to all open systems PowerMax customers at no additional charge. Deduplication is not currently supported for mainframe environments or on VMAX All Flash systems.

3.6.3 Remote Replication with SRDF

SRDF is perhaps the most popular data service in the enterprise data center because it is considered a gold standard for remote replication. Up to 70% of Fortune 500 companies use this tool to replicate their critical data to geographically dispersed data centers throughout the world. SRDF offers customers the ability to replicate tens of thousands of volumes, with each volume being replicated to a maximum of four different locations globally.

PowerMax runs an enhanced version of SRDF specific for all flash use cases. This version uses multi-core, multi-threading techniques to boost performance; and powerful write folding algorithms to greatly reduce replication bandwidth requirements along with source and target array back-end writes to flash.

There are three types of SRDF:

- **SRDF Synchronous (SRDF/S)** – SRDF/S delivers zero data loss remote mirroring between data centers separated by up to 60 miles (100 km).
- **SRDF Asynchronous (SRDF/A)** – SRDF/A delivers asynchronous remote data replication between data centers up to 8000 miles (12875 km) apart. SRDF/S and SRDF/A can be used together to support three or four site topologies as required by the world’s most mission-critical applications.
- **SRDF/Metro** – SRDF/Metro delivers active-active high availability for non-stop data access and workload mobility within a data center, or between data centers separated by up to 60 miles (100 km). SRDF/Metro allows for storage array clustering, enabling even more resiliency, agility, and data mobility. SRDF/Metro allows hosts or host clusters access to LUNs replicated between two different sites. The hosts can see both views of the Metro Replicated LUN (R1 and R2), but it appears to the host OS as if it were the same LUN. The host can then write to both the R1 and R2 devices simultaneously. This use case accounts for automated recovery and the seamless failover of applications thus avoiding recovery scenarios altogether. Other key features of SRDF Metro are:
  - It provides concurrent access of LUNS /storage groups for non-stop data access and higher availability across metro distances
  - It delivers simpler and seamless data mobility
  - It supports stretch clustering which is ideal for Microsoft and VMware environments

SRDF software is included in the PowerMax Pro and zPro software packages, with no capacity-based licensing. It can be ordered as an addition to the Essentials and zEssentials software packages. Any hardware needed to support SRDF must be purchased separately.

3.6.4 Local Replication with TimeFinder SnapVX

Every PowerMax array comes with the local replication data service Timefinder SnapVX, which is included as part of the Essentials and zEssentials packages. SnapVX creates very low-impact snapshots. SnapVX supports up to 256 snapshots per source volume and up to 16 million snapshots per array. Users can assign names to identify their snapshots, and they can set 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-copy clones.

Local replication with SnapVX starts out as efficiently as possible by creating a snapshot: a pointer-based structure that preserves a point-in-time view of a source volume. Snapshots do not require target volumes. They share back-end allocations with the source volume and other snapshots of the source volume, and only consume additional space when the source volume is changed. A single source volume can have up to 256 snapshots.

Each snapshot has a user-defined name and can optionally have an expiration date, both of which can be modified later. New management interfaces provide the user with the ability to take a snapshot of an entire storage group with a single command.

A point-in-time snapshot can be accessed from a host by linking it to a host-accessible volume referred to as a target. Target volumes are standard thin volumes. Up to 1024 target volumes can be linked to the snapshot(s) of a single source volume. This limit can be achieved either by linking all 1024 target volumes to the same snapshot from the source volume, or by linking multiple target volumes to multiple snapshots from the same source volume. However, a target volume may only be linked to a single snapshot at a time.

By default, targets are linked in a no-copy mode. This no-copy linked target functionality greatly reduces the amount of writes to the back-end flash drives as it eliminates the requirement of performing a full volume copy of the source volume during the unlink operation in order to continue to use the target volume for host I/O. This saves the back-end flash devices from enduring a large amount of write activity during the unlink operation, further reducing potential write amplification on the PowerMax array.

3.6.5 PowerMaxOS Service Levels and Host I/O Limits

Service Levels for PowerMaxOS provide open systems customers with the ability to separate applications based on performance requirements and business importance. PowerMaxOS provides the ability to set specified Service Levels to ensure the highest priority application response times are not impacted by lower priority applications.

Service Levels address the requirements of customers to insure that applications have a predictable, and consistent, level of performance while running on the array. The available Service Levels are defined in PowerMaxOS and can be applied to an application’s storage group at any time. This allows for the Storage Administrator to initially set, as well as change, the performance level of an application as needed. A Service Level can be applied to a storage group using the PowerMax management tools (Unisphere for PowerMax, REST API, Solutions Enabler, and SMI-S).

Service Levels can be used along with Host I/O Limits to make application performance more predictable while enforcing a specified service level. Setting Host I/O Limits allows a user to define front end port performance limits on a storage group. These front end limits can be set by IOPS, host MB per host, or a combination of both. Host I/O Limits can be set on storage group that has a specified Service Level to throttle IOPS on applications that are exceeding expected Service Level desired performance.

There are six PowerMaxOS service levels to choose from:
Table 15  PowerMaxOS Service Levels

<table>
<thead>
<tr>
<th>Service Level</th>
<th>Expected Average Response Time(1)</th>
<th>Penalizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond (Highest Priority)</td>
<td>.6 ms</td>
<td>Platinum, Gold, Silver, Bronze</td>
</tr>
<tr>
<td>Platinum</td>
<td>.8 ms</td>
<td>Gold, Silver, Bronze</td>
</tr>
<tr>
<td>Gold</td>
<td>1 ms</td>
<td>Silver, Bronze</td>
</tr>
<tr>
<td>Silver</td>
<td>3.6 ms</td>
<td>Bronze</td>
</tr>
<tr>
<td>Bronze (lowest Priority)</td>
<td>7.2 ms</td>
<td></td>
</tr>
<tr>
<td>Optimized (2)</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

(1) Diamond, Platinum, and Gold service levels have an upper limit but no lower limit, ensuring IO will be serviced as fast as possible. Silver and Bronze service levels have both an upper and lower limit designed to allow higher priority IOPS to be unaffected.

(2) Storage groups that are set to Optimized will be throttled for higher priority IOPS on all service levels with the exception of Bronze.

Note: For more information on PowerMaxOS Service Levels, see the Dell EMC Service Levels for PowerMaxOS White Paper listed in the reference section of this document

PowerMaxOS Service Levels and Host IO Limits are available at no additional cost for both PowerMax systems and VMAX All Flash systems which are running PowerMaxOS 5978.

3.6.6  Consolidation of Block and File Storage Using eNAS

Embedded NAS (eNAS) data service extends the value of PowerMax to file storage by enabling customers to leverage vital enterprise features including flash level performance for both block and file storage, as well as simplify management, and reduce deployment costs. PowerMax, with the eNAS data service, becomes a unified block and file platform, using a multi-controller, transactional NAS solution. It is designed for customers requiring hyper consolidation for block storage combined with moderate capacity, high performance file storage in mission-critical environments. Common eNAS use cases include running Oracle on NFS, VMware on NFS, Microsoft SQL on SMB 3.0, home directories, and Windows server consolidation.

eNAS uses the hypervisor provided in PowerMaxOS to create and run a set of virtual machines within the PowerMax array. These virtual machines host two major elements of eNAS: software data movers and control stations. The embedded data movers and control stations have access to shared system resource pools so that they can evenly consume PowerMax resources for both performance and capacity.

Aside from performance and consolidation, some of the benefits that PowerMax with eNAS can provide to a customer are:

- Scalability – easily serve over 6000 active SMB connections
- Meta-data logging file system ideally suited for an all flash environment
- Built-in asynchronous file level remote replication with File Replicator
- Integration with SRDF/S
• Small attack surface – not vulnerable to viruses targeted at general purpose operating systems

The eNAS data service is included in the Pro software package. It can be ordered as an additional item with the Essentials software package. All hardware required to support eNAS on PowerMax must be purchased separately.

3.6.7 Embedded Management (eManagement) using Unisphere for PowerMax

PowerMax customers can take advantage of simplified array management using embedded Unisphere for PowerMax. Unisphere for PowerMax is an HTML5 based management interface that allows IT managers to maximize productivity by dramatically reducing the time required to provision, manage, and monitor PowerMax data storage assets.

Embedded Unisphere enables customers to simplify management, reduce cost, and increase availability by running PowerMax management software directly on the array. Embedded management is configured in the factory to ensure minimal setup time on site. The feature runs as a container on a director, eliminating the need for a customer to allocate their own equipment to manage their arrays. Aside from Unisphere, other key elements of the eManagement data service include Solutions Enabler, Database Storage Analyzer, and SMI-S management software.

Unisphere for PowerMax delivers the simplification, flexibility, and automation that are key requirements to accelerate the transformation to the all flash datacenter. For customers who frequently build up and tear down storage configurations, Unisphere for PowerMax makes reconfiguring the array even easier by reducing the number of steps required to delete and repurpose volumes. With PowerMax, storage provisioning to a host or virtual machine is performed with a simple four step process using the default Diamond class storage service level. This ensures all applications will receive sub-ms response times. Using Unisphere for PowerMax, a customer can set up a multi-site SRDF configurations in a matter of minutes. In addition, Unisphere for PowerMax provides a full REST API, enabling customers to fully automate the delivery, monitoring, and protection of storage services from their enterprise storage. REST API also enables organizations to integrate their PowerMax storage with their own DevOps environment or with third party tools.

Embedded Unisphere for PowerMax is a great way to manage a single PowerMax array; however, for customers who need to view and manage their entire datacenter, Dell EMC provides Unisphere 360. Unisphere 360 aggregates and monitors up to 200 PowerMax, VMAX All Flash, and legacy VMAX arrays across a single datacenter. This solution is a great option for customers running multiple PowerMax and VMAX All Flash arrays with embedded management (eManagement) who are looking for ways to facilitate better insights across their entire datacenter. Unisphere 360 provides storage administrators the ability to view site-level health reports for every PowerMax and legacy system VMAX or coordinate compliance to code levels and other infrastructure maintenance requirements. Customers can leverage the simplification of PowerMax management at datacenter scale.

Embedded Unisphere and Database Storage Analyzer are available on every PowerMax array as they are included in the Essentials and zEssentials software packages. Unisphere 360 is included in the Pro and zPro software packages or can be ordered with the Essentials and zEssentials software packages. Unisphere 360 does not run in an embedded environment and requires additional customer-supplied server hardware.
3.6.8 Advanced Data Analytics with CloudIQ

CloudIQ is a cloud-based monitoring and storage analytics application that can be used to proactively monitors PowerMax arrays. The value of CloudIQ is centers on its ability to give users new and valuable insights into the health of the storage system. It proactively monitors and measures overall health using intelligent, comprehensive, and predictive analytics—and that makes it easier for IT to identify storage issues quickly and accurately. These analytics (which admins can access from anywhere through a web interface or mobile app) can drive business decisions that could lower the organization’s total cost of ownership associated with the array. CloudIQ delivers several key values to customers:

- **Reduce Total Cost of Ownership:** CloudIQ provides an easy single pane of glass from which you can monitor your Unity and SC systems, all from the web so you can access anytime, anywhere.
- **Expedite Time to Value:** Because it is deployed from the EMC Cloud, customers can simply log into their CloudIQ account and immediately access this valuable information. There is nothing to set up, no licenses, no burdens.
- **Drive Business Value:** CloudIQ’s Proactive Health Score provides an easy way to identify and understand potential vulnerabilities in the storage environment. With these proactive and targeted guidelines, the result is a more robust and reliable storage environment, resulting in higher uptime and optimized performance and capacity.

CloudIQ is free and included with all PowerMax and VMAX All Flash arrays.

3.6.9 Non-Disruptive Migration (NDM)

Data migrations have always been challenging in an enterprise environment. The complexity and size of very large data storage environments makes planning for, scheduling, and executing migrations extremely difficult. Migrations also often involve applications that cannot be taken offline, even briefly, for cutover to a new data storage array. Dell EMC’s Non-Disruptive Migration, or NDM, allows customers to perform online data migrations that are simple and completely non-disruptive to the host and application.

NDM is designed to help automate the process of migrating hosts and applications to a new PowerMax array with no downtime. Non-Disruptive Migration leverages SRDF replication technologies to move the application data to the new array. It also uses auto-provisioning, in combination with PowerPath or a supported host multi-pathing solution, to manage host access to the data during the migration process.

NDM offers PowerMax customers the following benefits:

- Allows migration from VMAX, VMAX3 to PowerMax or VMAX All Flash arrays with hosts and applications completely online
- Designed for ease of use with control operations that automate the setup and configuration of the migration environment
- Managed by familiar and simple user interfaces via Solutions Enabler and Unisphere
- Migrations can be easily cancelled and failed back to the source array for any reason prior to commit
- Completely built in and requires no additional software or licensing costs

**Notes:**
Migrations should take place during low I/O activity to minimize performance impact.
NDM currently does not support mainframe CKD devices
3.7 The Dell EMC Future-Proof Loyalty Program

The Dell EMC Future-Proof Loyalty Program gives customers additional peace of mind with guaranteed satisfaction and investment protection for future technology changes. This program includes the entire Dell EMC Storage Portfolio including the flagship PowerMax, VMAX All Flash, XtremIO X2, SC Series, Dell EMC Unity, Data Domain, Integrated Data Protection Appliance (IDPA), Isilon, and Elastic Cloud Storage (ECS) appliance. This program provides Dell EMC customers with the following benefits:

- Full Portfolio Program:
  - 3 Year Satisfaction Guarantee – Dell EMC guarantees 3 years of storage and data protection appliance satisfaction
  - Hardware Investment Protection – Trade in existing or competitive systems for credit towards next generation Dell EMC data storage systems, data protection appliances, or Hyper Converged Infrastructure product offerings
  - Predictable Support Pricing – consistent and predictable maintenance pricing and services for your storage appliances
  - 4:1 All Flash Storage Efficiency Guarantee – Get effective logical capacity at least 4X the purchased physical capacity
  - Never-Worry Data Migrations – Use built-in data migration tools with seamless upgrades to move to next generation data storage systems
  - All-inclusive software packages

For more information about the Dell EMC Future Proof Loyalty Program, contact Dell EMC sales.
4 PowerMax System Deployments

The Dell EMC PowerMax family offers customers an all NVMe storage platform that provides industry leading IOPS density per system in a single and dual floor tile footprint. This section describes the deployable system layouts for the PowerMax 2000 and PowerMax 8000 systems. See Expandable Modular Architecture - The PowerMax Brick for available drive configurations and system usable capacities.

4.1 PowerMax 2000 System Configurations

The PowerMax 2000 brings unmatched efficiency and flexibility to the data center, providing customers with over 1.7 million IOPS (8K RRH) and up to 1 PB of effective capacity in just 20U of total space.

4.1.1 PowerMax 2000 Configurations

The PowerMax 2000 can be configured using either one or two Bricks in a single standard Dell EMC Titan rack. Each Brick consumes 10U of rack space (20U max for dual Brick PowerMax 2000 systems). The initial Brick occupies the bottom 10U of the rack when shipped from Dell EMC manufacturing. The second Brick occupies the 10U directly above the initial Brick. This is applicable for systems ordered as dual Bricks or scale out systems. An additional PowerMax 2000 system can be added into the remaining 20U in the rack.

Figure 17  Single and Dual Brick PowerMax 2000 configurations

The PowerMax 2000 does not feature a system tray, KVM, or internal Ethernet or InfiniBand switches. It uses direct InfiniBand connections between engines on dual Brick systems.
**Note:** The PowerMax 2000 can be installed in third party racking. The third-party rack must be a standard NEMA 19-inch rack and meet Dell EMC standards for power, cable access, and cooling. For more information about PowerMax 2000 third party racking options please see the Dell EMC PowerMax Family Site Planning Guide.

4.2 Converged Solutions using PowerMax 2000

Due to PowerMax 2000’s small maximum footprint of 20U per system, the rack which houses the PowerMax 2000 can be used to house additional hardware such as Dell network switches and Dell servers. This allows customers to use the PowerMax 2000 as a building block to create true Dell “converged” infrastructures.

Converged infrastructure is an approach to datacenter management that packages compute, networking, storage, and hardware virtualization into an integrated solution. These converged solutions are typically included in scalable reference architectures for larger enterprise applications such as Oracle, SAP, Microsoft SQL Server, Exchange, and SharePoint. One of the primary benefits of converged solutions is that they simplify management as the storage, compute, networking, and virtualization can be automated and managed using far fewer interfaces. Also, all of the system components typically come from a single vendor; therefore the components have much higher degrees of integration than disparate off-the-shelf components from multiple vendors. The increased integration and simpler management of the hardware ultimately leads to lower operating costs, which is the primary reason customers deploy converged infrastructure solutions.

4.2.1 Mission-Critical Requirements

Customers using converged infrastructure often choose a standalone data storage array over server-based SAN offerings when their applications are mission critical. When a Dell EMC customer selects the PowerMax 2000 as the storage backbone for a Dell converged solution, they can be assured that they are selecting an array that is built on an architecture which sets the industry standard for reliability, uptime, and performance.

Mission-critical applications often need to be replicated both locally and remotely. PowerMax 2000’s native data services such as SRDF and TimeFinder SnapVX provide highly scalable remote and local replication capabilities which are difficult to obtain using standard application-based or hypervisor-based tools.

In mission-critical converged infrastructures, the goal is to shrink the distance and latencies between the storage and compute functions. A PowerMax 2000 converged solution places NVMe flash storage capabilities and the compute nodes within the same rack (single floor tile). Customers selecting PowerMax 2000 for their Dell converged solutions can also take comfort that the platform is future ready for the next generation high performance Storage Class Memory (SCM) media and next generation connectivity of NVMe-oF.

4.2.2 PowerMax 2000 Converged Solution

A converged solution featuring PowerMax 2000 frequently includes the latest Dell EMC PowerEdge 14G servers such as the R640 and R740; redundant Dell EMC Networking top of rack switches such as the 48 port 10 GbE S-4048-ON for the IP SAN (iSCSI), and the Dell Networking S3048 switch for the management network and connectivity to the corporate LAN. The selection of SAN connectivity can depend upon workload characteristics and customer preference. Since the PowerMax 2000 only consumes a maximum of 20U, there is also room in the rack for optional Fibre Channel switches such as the Dell EMC Connectrix DS-6510B with up to forty-eight 16 Gb ports. This gives a customer additional options when determining SAN connectivity.
The PowerMax 2000 provides a customer with maximum flexibility in this area as it supports up to 32 front-end connections per engine which can be exclusively, or a mix of, 16 Gb Fibre Channel and 10 GbE iSCSI.

The following diagram shows a possible converged solution using the above component in the same rack with a single Brick PowerMax 2000 system:

Figure 18  Single Brick PowerMax 2000 in a converged solution featuring Dell 14G Servers and Dell Network Switches

The combination of Dell networking, servers, and storage in a single converged solution for mission-critical applications produces many benefits. The ability to run applications such as Oracle, SAP, and Microsoft SQL Server in a single virtualized environment which uses PowerMax 2000 as its storage element safeguards availability, lowers costs, and increases productivity for Dell customers.

### 4.3 PowerMax 8000 System Configurations

PowerMax 8000 is the flagship of the PowerMax family and provides Dell EMC customers with unmatched scalability, performance, and IOPS density. It can consolidate disparate workloads on a mass scale as 8 Bricks can support over 10 million IOPS (8K RRH) and provide up to 4 PB of effective capacity in just two floor tiles of space.
The PowerMax 8000 is a highly configurable data storage array as it can support configurations from 1 to 8 Bricks within two standard Dell EMC Titan racks. Each rack can support up to four Bricks. Bricks 1 – 4 always occupy a single rack. PowerMax 8000 only requires a second rack when the Brick count is greater than four.

4.3.1 Single Rack PowerMax 8000 configurations

The diagram below shows a PowerMax 8000 single and dual Brick configurations:

Figure 19  PowerMax 8000 single and dual Brick configurations

The PowerMax 8000 uses redundant 16 port Dell X1018 Ethernet Switches for the internal management network. This network connects to every engine and to the two internal InfiniBand fabric switches. The InfiniBand switches are required when two or more Bricks are configured in the system. The redundant 18 port InfiniBand fabric switches connect to every director in the system.

DAE 3 is added with the second Brick. As mentioned earlier in the document, DAE 2 is shared by Brick 1 and Brick 2. In DAE 2, drive Slots 1 – 14 are used by Brick 1 while slots 15 – 24 are used by Brick 2. A PowerMax 8000 configuration rule of thumb is that every even numbered Brick shares a DAE with the previous odd-numbered Brick.

Note: The PowerMax 8000 supports the use of third-party racking. The third-party rack must be a standard NEMA 19-inch rack and meet Dell EMC standards for power, cable access, and cooling. For more
The following diagram shows a three and four Brick configuration for the PowerMax 8000:

Figure 20  PowerMax 8000 three and four Brick configurations

4.3.2  **Dual Rack PowerMax 8000 Configurations**

PowerMax 8000 systems that have more than four engines require a second rack (system bay). Bricks are added to the second rack in the same manner and order as Bricks 1 – 4 are added in the first rack. The engines for the Bricks in the second rack are connected to the Dell X1018 Ethernet management switches. The Brick engine directors in the second rack are cabled to the InfiniBand switches in the first rack also. No additional InfiniBand or Dell X1018 switches are required for the second rack.

The PowerMax 8000 supports rack 2 dispersion configurations of up to 82 feet (25 meters) from the fabric switches in rack 1. Dispersed configurations require optical connections between the InfiniBand switches in rack 1 and the Brick engine directors in rack 2. Adjacent rack configurations can use Copper for connections between the Brick engines in rack 2 and the InfiniBand switches in rack 1.
The following diagrams show the various PowerMax 8000 two-rack configurations:

**Figure 21** PowerMax 8000 five and six Brick configurations

**Figure 22** PowerMax 8000 seven and eight Brick configurations
Summary

The Dell EMC PowerMax provides Dell EMC customers with a platform that is ready for current and next-generation data center data storage requirements. The PowerMax Family is the first Dell EMC data storage system to fully use NVMe technology for customer application data. The innovative PowerMax is built using a 100% NVMe storage back-end, allowing it to reach unprecedented IOPS densities by eliminating the flash media choke points found using traditional SAS and SATA interfaces.

There are two PowerMax models:

- **PowerMax 2000** - designed to bring unmatched efficiency and flexibility to the datacenter, providing Dell EMC customers with 1.7 million IOPS (8K RRH) and up to 1 PB of effective capacity in just 20U total space.

- **PowerMax 8000** - designed to provide Dell EMC customers unmatched scalability, performance, and IOPS density. It can consolidate disparate workloads on a mass scale as 8 Bricks can support 10 million IOPS (8K RRH) and provide up to 4 PB of effective capacity in just two floor tiles of space.

The design of PowerMax is future-proof as it is ready for next generation SCM media such as 3D XPoint and next generation connectivity NVMe-oF infrastructure. The design is also built to provide not only the highest levels of performance, but also to deliver the highest levels of enterprise class reliability, availability, and serviceability. PowerMax customers can consolidate their workloads as block, file, and mainframe can all run on the platform. PowerMax comes complete with optimized data services which include advanced data reduction using inline deduplication and compression, the industry standard remote and local replication technologies of SRDF and TimeFinder SnapVX, and embedded array management with Unisphere for PowerMax.
### References

<table>
<thead>
<tr>
<th>Document Title</th>
<th>Collateral Type</th>
<th>Part Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell EMC PowerMax Family Product Guide</td>
<td>Product Guide</td>
<td></td>
</tr>
<tr>
<td>Dell EMC Service Levels for PowerMaxOS</td>
<td>White Paper</td>
<td>H17108</td>
</tr>
<tr>
<td>Dell EMC Embedded Management on PowerMax, VMAX All Flash, and VMAX3</td>
<td>White Paper</td>
<td>H16856</td>
</tr>
<tr>
<td>Data Reduction with Dell EMC PowerMax</td>
<td>White Paper</td>
<td>H17072</td>
</tr>
<tr>
<td>Dell EMC® PowerMax™ Reliability, Availability, and Serviceability Technical</td>
<td>White Paper</td>
<td>H17064</td>
</tr>
<tr>
<td>Dell EMC PowerMax and VMAX All Flash: GDPS and Advanced Copy Services Compatibility</td>
<td>White Paper</td>
<td>H16124</td>
</tr>
<tr>
<td>Dell EMC PowerMax Family Site Planning Guide</td>
<td>Technical Guide</td>
<td></td>
</tr>
<tr>
<td>Dell EMC Timefinder SnapVX Local Replication Technical Note</td>
<td>Technical Guide</td>
<td>H13697</td>
</tr>
<tr>
<td>Dell EMC SRDF/Metro Overview and Best Practices</td>
<td>Technical Guide</td>
<td>H14556</td>
</tr>
<tr>
<td>Transform SAP Landscape Challenges with Dell EMC PowerMax</td>
<td>Solution Overview</td>
<td>H17093</td>
</tr>
<tr>
<td>Consolidate Microsoft SQL Server with Dell EMC PowerMax</td>
<td>Solution Overview</td>
<td>H17092</td>
</tr>
<tr>
<td>Accelerate and Simplify Oracle Databases with Dell EMC PowerMax</td>
<td>Solution Overview</td>
<td>H16732</td>
</tr>
<tr>
<td>PowerMax Data Sheet</td>
<td>Data Sheet</td>
<td>H16891</td>
</tr>
<tr>
<td>Dell EMC PowerMax Software</td>
<td>Data Sheet</td>
<td>H16748</td>
</tr>
<tr>
<td>PowerMax 2000 / 8000 Specification Sheet</td>
<td>Specification Sheet</td>
<td>H16739</td>
</tr>
<tr>
<td>Future of Enterprise Storage (PowerMax)</td>
<td>Infographic</td>
<td>H16864</td>
</tr>
<tr>
<td>Top Ten Reasons Why Customers Deploy PowerMax</td>
<td>Handout/FAQ</td>
<td>H16738</td>
</tr>
<tr>
<td>Top Ten Reasons Why Customers Deploy Dell EMC PowerMax for Microsoft SQL Server</td>
<td>Top Reasons Handout</td>
<td>H17091</td>
</tr>
<tr>
<td>Top Ten Reasons Why Customers Deploy Dell EMC PowerMax for SAP Landscapes</td>
<td>Top Reasons Handout</td>
<td>H17090</td>
</tr>
<tr>
<td>Top Ten Reasons Why Customers Deploy Dell EMC PowerMax for VMware</td>
<td>Top Reasons Handout</td>
<td>H17074</td>
</tr>
<tr>
<td>Top 10 Reasons Why Dell EMC PowerMax for Oracle</td>
<td>Top Reasons Handout</td>
<td>H16725</td>
</tr>
</tbody>
</table>
Appendix A: Understanding Non-Volatile Memory Express (NVMe)

Summary:

**NVMe is a new interface for Non-Volatile Memory (NVM) storage technologies:** NVMe is a new interface to PCIe SSDs (PCI Express Solid State Drives). The two primary components of NVMe are a new command set and a PCIe interface to SSDs that is based on shared memory between the host processor and the SSD. There is also a management interface, NVMe-MI, to manage SSDs. NVMe SSDs are used in both servers and storage systems, including servers that provide software-defined storage systems such as VSAN and ScaleIO.

**NVMe achieves high levels of performance and parallelism:** Multiple shared memory queues ensure that sufficient commands can be outstanding to use all of an SSD’s available bandwidth and enable each processor core to handle I/O on its own without additional processor instructions and delays for coordination with other cores. NVMe’s small command set separates administrative commands from I/O commands, with the result that high-performance I/O paths can be optimized to support a small number of commands – at a minimum, only three I/O commands are required - Read, Write, and Flush. These performance characteristics and the use of parallelism enable flash-based NVMe SSDs to outperform their SAS and SATA counterparts, and are crucial for SSDs based on new low-latency Storage Class Memory technologies such as 3D XPoint.

**NVMe over Fabrics (NVMe-oF) extends NVMe benefits to data center fabrics:** NVMe over fabrics extends NVMe over data center fabrics such as Ethernet, Fibre Channel, and InfiniBand. Fabric applicability varies by class of infrastructure. For example, InfiniBand is primarily applicable to embedded use in storage systems. Fibre Channel is primarily applicable to external use between servers and storage. Ethernet fabrics are amenable to both external and embedded use. Fabrics protocol support for NVMe exists or is in development for a wide range of infrastructure classes.

**NVMe adoption is well underway:** Single port NVMe SSDs are in widespread usage, for example, in servers. Dual-port NVMe SSDs have recently become available that include the enterprise functionality necessary for use in high availability storage systems with no single point of failure. NVMe over fabrics technology is at earlier lifecycle stages – the initial focus has been on Remote Direct Memory Access (RDMA) fabrics such as InfiniBand and RDMA over Converted Ethernet version 2 (RoCEv2), which is now being joined by Fibre Channel and non-RDMA use of TCP over Ethernet. The drivers for adoption of NVMe vary by usage, and NVMe technologies are expected to co-exist with existing storage technologies for a number of years; displacement of existing technologies will be gradual.

**Dell EMC has been a pioneer in NVMe architecture/design:** Dell EMC server engineers were key architects and designers of the original NVMe interface and are playing a similar role in the development of NVMe over Fabrics. Dell EMC engineers continue to drive NVMe design and standards forward in cooperation with the rest of the industry in areas that include data integrity, security, fabric multi-pathing, and enclosure management.

**Dell EMC already uses NVMe in products, with more to come:** NVMe SSDs are supported in Dell EMC servers, and have been for several years. PowerMax uses a full NVMe storage back end for storing customer
data. Legacy VMAX storage arrays use NVMe SSDs as vault drives to quickly store cache contents on power failure. Extensive plans are in place for Dell EMC storage systems to expand the use of enterprise-grade dual port SSDs and fabrics (both embedded in storage systems and for external connection to servers). Dell EMC will continue to innovate and lead the industry in this important area of technology.

1. What is Non-Volatile Memory Express (NVMe) and what are its key benefits?

NVMe is a command set and associated storage interface standards that specify efficient access to storage devices and systems based on Non-Volatile Memory (NVM) media. NVMe is broadly applicable to NVM storage technology, including current NAND-based flash and higher-performance, Storage Class Memory (SCM) technologies such as 3D XPoint and Resistive RAM (ReRAM). The NVMe standards were created to fully exploit the bandwidth, IOPS, and latency performance benefits that NVM-based storage offers. The resulting levels of performance and parallelism for both drives and storage systems are unattainable using current storage interfaces.

NVMe over Fabrics (NVMe-oF) extends the NVMe command set and its benefits over data center fabrics, such as RoCEv2 (RDMA [Remote Direct Memory Access] over Converged Ethernet) and Fibre Channel. NVMe-oF can also be used with fabrics such as InfiniBand that are embedded in storage systems, taking advantage of the increased connectivity and resilience to failures of such fabrics by comparison to SSD interfaces.

2. What are the core components of NVMe?

NVMe consists of a storage command set, interface specifications for two environments, and a management interface specification:

- **NVMe storage command set**: Commands to administer NVMe interfaces and perform I/O.
- **NVMe over PCIe**: A register and shared memory interface for PCIe SSDs. This interface applies to both host and storage system use of NVMe SSDs.
- **NVMe over Fabrics**: A message and data interface for data center fabrics that applies to both host-to-storage access and to fabric interconnects within storage systems.
- **NVMe Management Interface (NVMe-MI)**: A management interface for NVMe PCIe SSDs that is being extended to enclosure (e.g., rack or cabinet) management.

In more detail, the four core components of NVMe are:

**A storage command set**: A modern streamlined storage command set used to communicate with NVMe drives and storage systems, replacing SCSI and ATA. The NVMe command set contains only three mandatory I/O commands: Read, Write, and Flush, with additional optional commands, for example, for persistent reservations. The standard specifies ten mandatory administrative commands which are used for configuration and management, plus additional optional commands, for example to manage firmware and PCIe I/O virtualization (SR-IOV). Host or server implementations of the NVMe command set use significantly
less CPU instructions to process I/O as compared to the SCSI and ATA command sets, for example, as used with SAS and SATA drives. This results in much higher IOPS and lower I/O latency.

**A storage interface for PCIe SSDs**: The NVMe standard defines an interface to PCIe SSDs (PCI Express Solid State Drives) based on registers and queues in shared memory. Command and completion communication is based on

- Paired submission and completion queues.
- Circular buffers that decouple host and drive interactions. This increases concurrency.

Multiple pairs of queues and associated interrupts ensure that I/O completions are processed on the same processor core that issued the original command. This takes full advantage of the parallel processing abilities of modern multi-core processors. I/O data is transferred over PCIe via DMA (Direct Memory Access) between the SSD and main memory. There are deep queues that can hold thousands of commands each, ensuring that sufficient I/O commands can be queued to fully utilize SSD bandwidth.

**A storage interface for data center fabrics (NVMe over Fabrics or NVMe-oF)**: The NVMe over Fabrics standard specifies how NVMe command set usage is extended over data center fabrics such as RoCEv2 (RDMA over Converged Ethernet) and Fibre Channel. This is completed by fabric transport specifications for use of specific fabrics with NVMe-oF. In place of shared memory queues, NVMe over Fabrics exchanges messages (called capsules) between the host and NVM subsystem, such as a storage array. The key fabric types are:

- RDMA-based: Any fabric that supports RDMA, including RoCEv2 (over Ethernet) and InfiniBand. The RDMA fabric transport is part of the NVMe-oF standard itself.
- Fibre Channel: The NVMe fabric transport for Fibre Channel is a separate standard developed by T11, the Fibre Channel standards body.
- TCP/IP (typically over Ethernet): This non-RDMA fabric transport specification is under development as part of a forthcoming version of the NVMe-oF standard.

**A management interface and protocol (NVMe-MI)**: The NVMe Management Interface (NVMe-MI) is an interface that enables management functionality such as:

- Discovery of NVMe drives that are present and capabilities of each drive
- Store data about the host environment (e.g., operating system and device driver versions) for subsequent retrieval
- Monitoring of drive health, including operational temperature.

NVMe-MI access to an SSD can be realized over multiple physical interfaces, for example SMbus, I2C and PCIe. NVMe-MI is being extended to enclosure (cabinet) management via extensive reuse of existing SCSI functionality (SCSI Enclosure Services, aka SES).

3. How is NVMe impacting our industry?
The NVMe and NVMe-oF specifications, along with complementary ecosystem support such as drivers for Linux and Windows, plus inter-operability test suites, has enabled a rapid rate of product implementation, both of NVMe SSDs and related systems. Single port SSDs in multiple form factors are in widespread use in servers, laptops and smaller mobile devices. Additional NVMe devices are emerging including dual-port SSDs suitable for high availability storage arrays and SSDs that use very low latency Storage Class Memory technologies (e.g., 3D XPoint) where NVMe is the only interface available.

NVMe benefits a wide range of applications. Its higher performance enables faster analysis on larger data sets producing increasingly beneficial business outcomes. NVMe also satisfies client and mobile device demands for balancing I/O performance, storage capacity, and energy consumption. Along with other supporting technologies, NVMe and NVMe-oF will rapidly improve existing storage architectures to enhance scalability, connectivity, and performance.

4. Where is NVMe used in Dell EMC products currently and in the future?

Dell and EMC have been industry leaders in NVMe since long before the formation of Dell EMC. Between the server and storage divisions, Dell EMC engineers have been among the original architects and key designers of all aspects of NVMe – NVMe over PCIe, NVMe-MI and NVMe-oF, as far back as 2006. Not only has Dell EMC invested in developing the specifications, our product portfolio has taken advantage of NVMe technologies for years. PowerEdge servers have been available with PCIe SSDs since 2013 and Dell EMC now offers servers that are fully populated with PCIe SSDs. The Dell EMC storage flagship PowerMax is the first Dell EMC array to use a full NVMe back end for storing customer data. The legacy VMAX All Flash uses NVMe SSDs as vault drives to store array cache in the event of power failure. Along with hardware readiness for usage of NVMe technologies, we are also investing in the software defined datacenter of the future with ScaleIO and ScaleIO Ready Nodes that support NVMe SSDs. Dell EMC is also planning for future connectivity via NVMe over Fabrics (NVMe-oF) that will enable customers to take full advantage of Storage Class Memory in networked environments based on fabrics such as Ethernet and Fibre Channel. Dell EMC is committed to the future of NVMe technology through the continued development of the specification and support in our products.

5. What are some areas of NVMe related innovation at Dell EMC?

To evolve the NVMe and NVMe-oF standards, Dell EMC is working to ensure that NVMe technology will carry forward some of the crucial robustness and usability functionality developed in the SCSI standards such as data integrity, discovery and multi-pathing to support and simplify transition to NVMe based systems. Dell EMC is also working to improve security and manageability aspects of these standards and NVMe-MI, e.g., enclosure management.

Cloud based systems and the wide use of virtualization has led to increasing use of containers for easy and flexible application deployment and management. NVMe’s high performance and high parallelism is a good match to the intensive IO requirements and scaling needs of container-based infrastructure in cloud and other environments. Dell EMC is working to create infrastructure that supports these applications at scale, including the ability to deliver different classes of storage quality of service on common infrastructure and manage the environment in an automated fashion.

The emergence of Storage Cass Memory with lower latency compared to NAND and higher density compared to DRAM strongly favors use of the NVMe and NVMe-oF. Dell EMC is exploring the nature and
capabilities of these new media types and investigating how to combine this new media with existing technologies (SSD, HDD) in tiered and efficient ways.

6. When will IT infrastructure products that use “NVMe” be widely available?

We are in the midst of those transitions now. Servers and clients (e.g., laptops, tablets) that use NVMe are already widely available. NVMe SSDs that use Storage Class Memory (faster than flash media such as 3D XPoint) and dual-port NVMe SSDs that enable highly available storage systems are now entering the market. NVMe over Fabrics is a nascent technology for which a few early products are available, with widespread availability expected in the next year or two, both embedded in storage systems and used externally for connectivity between hosts/servers and storage.