Dell PowerScale

Overview

Dell 's PowerScale systems are a series of scale out NAS systems to address various usages in Information Technology, Media & Entertainment, and high-performance computing markets. PowerScale is a successor to Isilon with continuation of the underlying OneFS software. Using clustering of multiple nodes, PowerScale is a scale out NAS architecture to address customer needs in performance, capacity, and price. Multiple models provide different performance and capacity scaling while using the same underlying architecture and software.

Scaling of multiple nodes in an N-Way cluster architecture provides for redundancy and scaling of capacity and performance. A minimum of three nodes are required in the clustering configuration with a maximum of 252 when Ethernet is used for the interconnect (144 if InfiniBand is used). The OneFS operating system enables the clustering and all features to provide the performance and capabilities to meet customer needs across multiple markets.

Key security features include use of Self Encrypting Drives and internal key management to protect from obtaining data in the case of drive thefts and cryptographic erase for secure deletion of data. Audit trails of access control changes are also supported with the integration of third-party software. Role-based access control, access zones, and OpenSSL add to the security features. Compliance features include data retention and WORM mode.

Transparent tiering to public or private clouds is a native feature of OneFS.

Highlights

- Scale Out NAS
- High Performance Distributed File System – OneFS
- F600 and F900 NVMe systems
- Up to 252 nodes when Ethernet interconnect used
- Single Namespace
- Smartlock for WORM
- Multiple models for different perf / price
- NDMP support
- Object support with S3
- HDFS protocol
- Single file system with isolated multi-tenancy
- CloudIQ SaaS based monitoring and reporting
- DataIQ for discovery, indexing, and searching

PowerScale OneFS is offered as a software only version meant to be installed at remote locations, connected to a central PowerScale system. The OneFS Edge runs as a VM in VMware ESXi.

Three families of systems are defined:

- H - Hybrid
- A – Archive
- F - Flash

Usage and Deployment

The Dell PowerScale and earlier Isilon systems have been targeted at high capacity, high performance storage requirements from markets such as media & entertainment, bioinformatics, and other high performance computing requirements. In addition, the single name space capability allows for usage in consolidation projects, very large numbers of files applications, and for what Dell terms Big Data and AI with large capacity amounts.

The PowerScale F600 and F900 all flash systems have embraced full use of NVMe devices for additional high performance with the support for up to 8 or 24 NVMe SSDs per node, respectively. The use of NVMe devices adds additional performance for more demanding workloads including high performance computing and analytics. Added support for QLC SSDs in the F600 and F900 systems provides the option for all flash storage with the high-capacity points of QLC devices.

- Characteristics
  - Performance – The system with the 252-node scaling capability is a very high-performance system with standard benchmark numbers published by Dell EMC. Load balancing is performed across nodes. Ethernet interconnect is required for 252 nodes. InfiniBand interconnect supports up to 144 nodes.
  - Availability – N-way scaling provides a high availability configuration. The minimum number of nodes is three ensuring HA. Interconnects between nodes are InfiniBand or Ethernet (the H5600 model is 40GbE only) and must be one or the other. Data protection is accomplished by distributing data across nodes. Forward Error Correction is implemented and called FlexProtect with the use of Reed Solomon codes to protect against up to four failures of nodes or devices.
  - Replication for BC/DR – Currently remote replication is asynchronous.
  - Object access through S3 API over RESTful interface.
  - There is an optional backup accelerator node with FC ports to connect directly to tape systems offered to improve the backup operations for the large amount of capacity across multiple nodes.
  - Single file system with single namespace – isolation supported by user defined access zones.
- Applications

- The Dell PowerScale systems are targeted at any file-based applications (unstructured data). The performance and capacity with the scale out implementation would be needed with demanding applications.
- The PowerScale system is capable of effectively supporting block large files and small files at the same time, allowing it to be used for multiple applications. Object is supported with S3.
- VMware VAAI, VASA, and SRM SRA are supported
- Audit trails are supported for compliance with 3rd party software

- System environments
  - The NAS support includes NFS and native CIFS/SMB implementations with an additional system software to support file sharing and locking for simultaneous access from the two remote file system protocols.
  - At the client/application layer, PowerScale supports a wide range of operating system environments including Windows, Linux/Unix, and Apple OS.
  - PowerScale also supports HDFS (Hadoop File System), HTTP, S3, and FTP as additional protocols.
- Deployment and Administration
  - PowerScale systems are installable by a customer as a NAS system with moderate networking and storage skills. Larger, more complex environments may be more efficiently deployed using professional services.
  - Currently the management software is delivered as part of the system. Both a GUI and a CLI is available.
  - Performance and capacity monitoring is available through InsightIQ, which also retains historical data on performance and capacity.
  - CloudIQ is the SaaS-based monitoring and reporting system for Dell EMC storage and PowerScale is included in the systems supported.
  - DataIQ is additional software running on an external virtual machine that can index, tag, and search both file and object data stored in a PowerScale system.

Key Capabilities

Architecture and Deployment

The original company Isilon announced the NAS system in October of 2003 and since then, Isilon, , and now Dell have continued to add capabilities, models, and advanced features. Scaling of multiple nodes in an N-Way cluster architecture provides for redundancy and scaling of capacity and performance. A minimum of three nodes are required in the clustering configuration with a maximum of 252 when Ethernet is used and 144 when InfiniBand is used. The OneFS operating system of the Dell PowerScale enables the clustering and all features to provide the performance and capabilities to meet customer needs across multiple markets. To further meet target market needs, PowerScale has developed three separate product lines:

- H – Hybrid with a mixture of SSDs and HDDs.
- A – Archive with high capacity HDDs
- F – All flash system with SSDs

Earlier models of Isilon can be intermixed with the models. Prior models were:

- S-Series – is targeted at high I/O's per second (IOPs) requirements and random access data
- X-Series – focuses on sequential access with high throughput for large sequential transfers
- NL-Series – is for lower cost storage using large capacity but less expensive disk drives
- HD-Server – high capacity system for big data, archives and data lakes.

Hardware Architecture

The Dell PowerScale series consists of both hardware, based on industry standard Intel servers, with direct attached devices and custom software to provide the scalable NAS function along with additional software enabled features.

The clustering of nodes is the heart of PowerScale. Each node contains storage and processing functions. From 4 to 252 nodes for Ethernet interconnect (144 maximum for InfiniBand) may be clustered to provide scalability for redundancy, performance and capacity. To maintain a single filesystem across the nodes, two InfiniBand links running at 25, 40, or 100 Gigabit Ethernet depending on the model are used for intra-cluster communication and synchronization. The high performance, dedicated back channel provides a consistent method for communicating control information.

Depending on the model, PowerScale supports SAS disk drives, SSDs, and SATA disks. The F600 and F900 models support NVMe SSDs, including high capacity QLC devices. The protection is really node-level protection with the N-Way scaling. A failed drive can be replaced and reconstructed with the parallel operations of the other nodes reconstructing the files contained on the failed node (failed drive). The protection feature is called **FlexProtect** and provides several options. A file may be protected with basic parity which is termed +1 in the FlexProtect nomenclature. Additional protection levels are +2, +3, and +4 which means in the +4 case that 4 simultaneous failures of nodes or disks can be tolerated. The protection level is selectable by the administrator on a per-file, directory, or file system level. Forward error correction using Reed Solomon codes are used for FlexProtect.

Evaluator Group Comment: While really a forward error correction using information dispersal implementation, the protection is fundamentally different than traditional RAID arrays. This protection method provides more protection and enables more parallelism of access. The desire to compare RAID levels has been created by vendors as a means of

explaining their offerings and abetted by analysts and others. The FlexProtect implementation is better but there will still be some desire for comparatives on RAID levels.

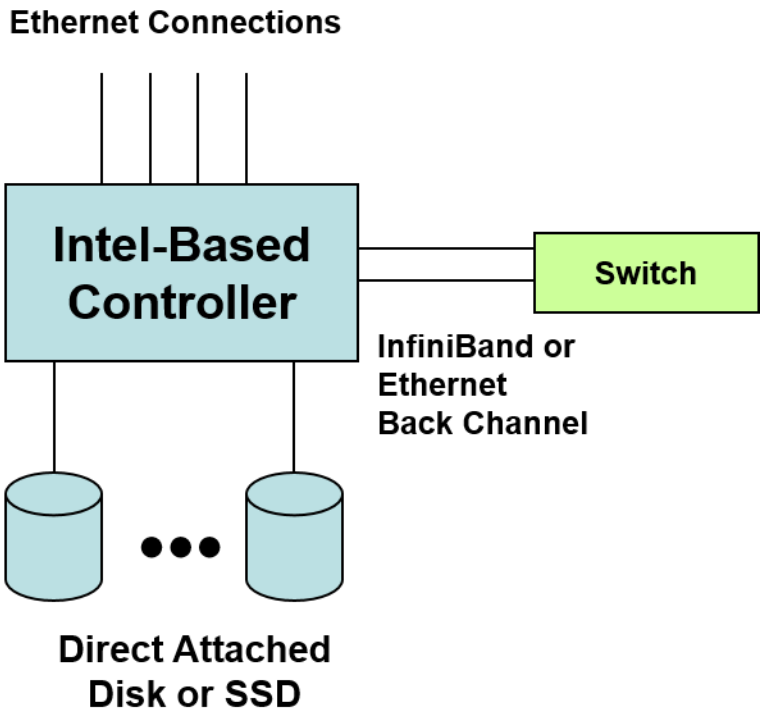

Figure 1: Basic Hardware Configuration

When additional nodes are connected over Ethernet, the OneFS software will recognize the addition of another node, automatically configure it, and begin distributing data to balance capacity.

OneFS is a single file system on PowerScale. Administrators isolate access to data for multi-tenancy support by the Secure Authentication Zone feature. The administrator creates isolated pools within the single file system and associates SMB shares, NFS exports and HDFS in order to segregate data. Secure Authentication Zone integrates with SmartConnect to enable the designation of IP pools for additional security. File audit is granular with the authentication zones as well.

Encryption

Self-Encrypting Drives (SEDs) are used to provide encryption of data at rest in PowerScale systems. An internal key manager is used for self-contained keys. The encryption is on a per device basis and is either all on or all off within PowerScale. Once turned on, it cannot be turned off. F600 and F900 systems do not support SEDs.

SmartFlash

Up to 1 PB of SSD capacity may be configured to use as cache for file data. Metadata may also be cached in SSDs. Individual files may be pinned as well.

CloudPools

Object pools can be allocated from cloud service providers or from Dell ECS object storage system. Amazon S3, Microsoft Azure, Google Cloud Service, Alibaba and VirtuStream are supported. A cloud pool becomes another tier of storage in the PowerScale system in which files can be tiered based on policy controls. When data is archived to a cloud pool, the file is replaced with a stub file called a SmartLink. This allows the file to remain visible to the system after it has been moved to a remote system. When the SmartLink file is accessed, the system retrieves the remotely stored data and any changes made file be made to the remotely store file.

Software Architecture

The key to the differential value provided by PowerScale is the OneFS software. OneFS is both the operating environment, based on BSD Unix, and the file system which includes the volume manager and RAID protection into a single layer that is optimized for NAS functionality. The optimized OneFS is different than most NAS systems that use a traditional program stack where each function is separate and requires the protocols and independent operations that may introduce additional overhead. OneFS controls placement of data on disks by managing redundancy on a file-by-file basis. Files greater than 128K in size will be striped across all nodes (up to a maximum of 20) in a cluster providing parallelism for performance. Files less than 128K will be mirrored in multiple locations (from 2 to 8 depending on the administrative settings).

Other software functions are layered on top of OneFS to provide the competitive advanced features in a NAS system. The network stack software executes in User Space and can be restarted without having to reboot the system.

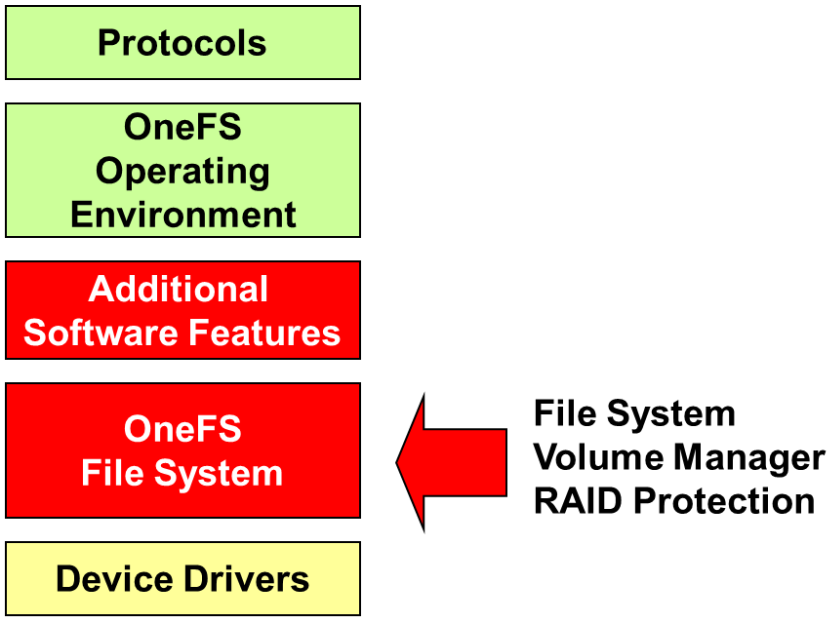

Figure 2: Basic Software Configuration

Object Storage access

An S3 API is used to store and retrieve object on PowerScale. Files may be accessed as object through the same interface. All file system and file operations that are available with NFS and CIFS/SMB are also supported through the object interface. PowerScale is the preferred solution when the primary access is for files. Dell recommends ECS as the preferred solution when the primary access is for objects.

Audit Trail Support

The OneFS file system supports auditing of configurations and NFS, SMB, and HDFS activity in order to achieve various governance and compliance use cases. Audit data is organized by audit topic and stored within the cluster file system where it can then be exported to third party applications via the Dell Common Event Enabler. Third party applications include Varonis DatAdvantage, Symantec Data Insight, and STEALTHbits stealthAUDIT.

Data Reduction

Data reduction is either on or off on a per cluster basis. The default is on and can only be turned off at the install time for the cluster. There are three components for data reduction. The first is zero block detect where data is not stored if a full 8KB block of zeroes are detected.

Deduplication is done inline. Prior to OneFS release 8.2.1, deduplication was done post storage and called SmartDedupe. Now, the deduplication is done as data is written.

Compression is done inline using an offload adapter for hardware assist with an FPGA. The adapter is the backend Ethernet adapter (back channel card) plugged into the PCIe card slot. Compression is done by chunking data into 128K chunks organized as 16 8K blocks. Compression must be positive (data does compress) for at least 8K or the chuck is passed over as uncompressible.

The following diagrams illustrate the write and read process flow for data compression.

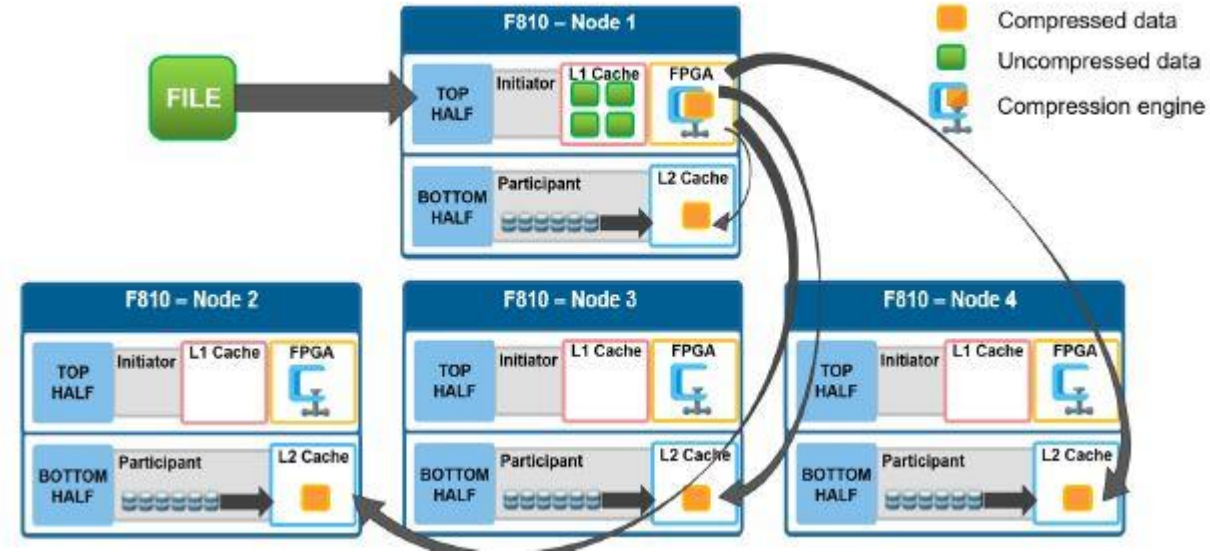

Figure 3: Compression Write Example Source: Dell

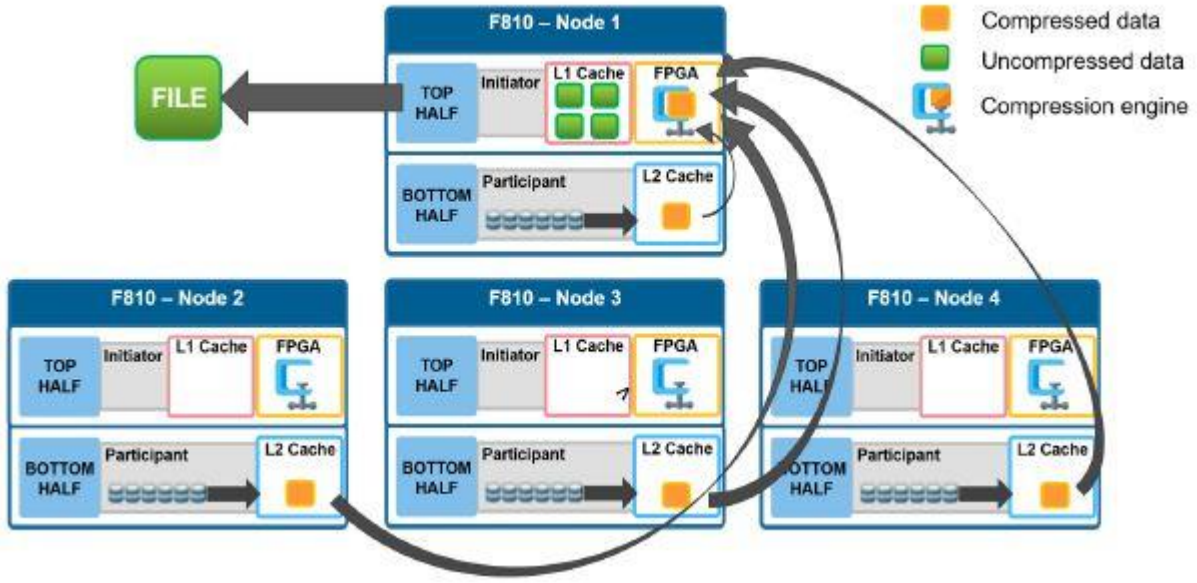

Figure 4: Compression Read Example Source: Dell

Cryptographic Erase

Files by be securely deleted with the cryptographic erase feature included with PowerScale.

OneFS for Cloud (previously called vOneFS)

OneFS for Cloud is a software only version that is meant to be installed on public clouds to provide the PowerScale functions. Replication features for remote connections and other advanced capabilities are available with OneFS.

Protocol and File System Support	
Network protocol	CIFS native – SMB1, 2, 3.0 including multi-channel NFS v4.0 (TCP and UDP) NDMP FTP HTTP HDFS 2.3
Management protocol	HTTP, LDAP, ADS, NIS, SMTP alerts
File system	OneFS
Protocols	Simultaneous CIFS and NFS S3

Maximum file system size	58PB
Maximum number of files per filesystem	
Client Support	Microsoft Windows Linux Unix systems Apple Macintosh

Table 1: Dell PowerScale Protocol and File System Support

Advanced Software Features and Functions

Some additional software features are available with the PowerScale system. These are additional licensed capabilities.

SyncIQ

Asynchronous replication is supported with the additional feature called SyncIQ. SyncIQ allows multiple nodes to concurrently push data elements of files for file-based remote replication. With enough bandwidth allocated, this parallelism can greatly speed replication. Throttling of the replication is included in the feature to control the amount of bandwidth used with time-based criteria. SyncIQ also supports replication of differential file changes only with full integration of SnapShotIQ

SmartConnect

Load balancing and failover software for NFS is an optional feature called SmartConnect. Load balancing is accomplished by spreading a file operation across multiple nodes which is possible because of the PowerScale clustering. The N-way failover and failback minimizes the impact of a single node failing.

SnapShot IQ

As the name implies, SnapShot IQ provides snapshots for clusters, directories, and sub-directories and operates across nodes in a cluster. There can be any number of snapshots within a cluster and up to 1,024 within a directory. The snapshots taken are read-only and copy-on-write methods are used for changed data. PowerScale includes a snapshot scheduling wizard in the management software. Individual file can also be copied with the File Clone capability of SnapShot IQ.

SmartQuota

An administrator can assign quotas based on capacity to users or groups with the optional SmartQuota feature. By assigning based on capacity, the administrator can present more capacity than actually exists in a system. Quota violations are notified by alerts and emails and through the management GUI.

Insight IQ

Visibility into the PowerScale system is accomplished through InsightIQ. The system analytics software provides capacity planning and utilization history reporting. The user interface has filters, timeframe, and results settings for analyzing operational information.

Smartlock

Smartlock is a means of providing WORM (write once read many) immutability at a file level. A directory can be set up with default retention periods and all files stored in that directory will have those retention rules. The retention period can be customized at the file level for more granularity. Although the data is immutable from the user standpoint (cannot be altered or deleted until the retention period expires), a privileged user can delete the file. There is no auto delete of files when the retention period expires. Smartlock version has SEC 17a-4 compliance capabilities, which include the audit trail logging required for file access and administrative actions.

Included with Smartlock is the ability to set legal holds on files so the retention period expiration (allowing deleting or alteration) is held off until the legal hold is removed. Smartlock is integrated with replication and backup.

Multi-Tenancy

Dell PowerScale supports multi-tenancy through a secure access zone for all protocols. Management and access are isolated based on the protocol.

CloudIQ

Dell uses telemetry data from storage systems including PowerScale for monitoring and storage analytics. Functions include proactive support with health checks and reporting such as capacity consumption and performance. Telemetry data is collected and analyzed with a SaaS product called CloudIQ.

Significant Announcements

- Dec 2023 – start with announcement coverage
- Dec 2023 – InsightIQ updated with version 5.0

Futurum Group EvaluScale – NAS

The Futurum Group product review methodology "EvaluScale" assesses each product within a specific technology area. The evaluation of each product is based on its capabilities, with capabilities for each technology segment grouped into distinct categories. The products are evaluated based on the following 4 criteria categories:

- Performance / Capacity

- Basic Functionality
- Advanced Capabilities
- Ability to Execute

The full NAS Storage EvaluScale can be found **here.**

The Futurum Group Opinion and Outlook for Dell PowerScale

The Dell PowerScale and Isilon systems have proven their scaling capability in performance and capacity in different vertical market uses. As more unstructured data is stored in traditional IT, the scaling and performance will make it one of the considerations. The advanced features currently available are important in the enterprise. Adding synchronous replication will be useful for some.

Embedded tiering (tiering within a node with different types of devices) is another capability that would be useful. As with the Unity systems, the ability to leave a symbolic link or stub file behind when files are tiered (such as to an ECS system) is an important and competitive capability.

There are obvious overlaps in the NAS solutions from Dell but PowerScale stands out with scaling and performance. The selection of different models may seem unnecessary because of the system flexibility but it is useful in helping customers arrive at the best configuration. The addition of NVMe devices in the higher end F600 and F900 models can provide a powerful performance boost for unstructured data workloads. PowerScale is really a great complement to ECS and can be positioned effectively across many vertical markets and enterprises.