PowerScale All-Flash F210, F710 & F910

World’s first Ethernet storage certified on NVIDIA DGX SuperPOD

May 2024

H19951.2

White Paper

Abstract

This document describes the PowerScale All-Flash F210, F710 and F910 nodes, the world first Ethernet-based storage solution certified on the NVIDIA DGX SuperPOD.

Dell Technologies

Copyright

Contents

Executive summary

Overview

PowerScale OneFS is the operating system that powers the Dell PowerScale scale-out NAS storage solutions. It is designed to handle large-scale, unstructured data with high performance, efficiency, and security. OneFS provides a unified, scalable file system that integrates data protection, eliminates volume management, and supports multiple protocols and workloads. OneFS also enables data mobility across edge, core, and cloud environments. OneFS 9.7.0.0 introduced support for the F210 and F710. OneFS 9.8.0.0 introduces support for the F910, the densest all-flash PowerScale node.

Revisions

Date	Description
February 2024	Initial release
May 2024	PowerScale F910 Updates

We value your feedback

Dell Technologies and the authors of this document welcome your feedback on this document. Contact the Dell Technologies team by email.

Author: Aqib Kazi

Introduction

Overview

The Dell PowerScale portfolio is a family of scale-out NAS storage solutions that can handle any type of unstructured data, such as files, images, videos, and analytics. PowerScale combines the performance, flexibility, and efficiency of the OneFS operating system with the latest hardware innovations from Dell Technologies, such as NVMe, all-flash, hybrid, and archive nodes. PowerScale enables customers to run diverse workloads across multiple protocols, clouds, and locations while ensuring data security, protection, and management at scale. PowerScale is designed to help customers unlock the potential of their data and accelerate their AI, ML, and big data projects.

PowerScale offers all-flash, hybrid, and archive hardware platforms to suit various performance, capacity, and cost requirements. The current all-flash hardware platforms are:

- **PowerScale F900**: An all-NVMe node that provides the highest performance and scalability for demanding workloads. It has 24 NVMe SSDs and can scale from 46 TB to 368 TB per node.
- **PowerScale F600**: An all-NVMe node that provides high performance and capacity in a compact form factor. It has 4 NVMe SSDs and can scale from 15.36 TB to 122.8 TB per node.
- **PowerScale F200**: An all-flash node that provides cost-effective performance for a wide range of workloads. It has 4 SAS SSDs and can scale from 3.84 TB to 30.72 TB per node.
- **Isilon F800 and F810**: All-flash nodes that provide high throughput and low latency for unstructured data. The F810 also includes deduplication for higher efficiency. They have 60 SAS SSDs and can scale from 96 TB to 924 TB per node.

The next-generation PowerScale platform is based on the PowerEdge platform, providing massive performance improvements across reads, writes, and throughput, as illustrated in the figure below.

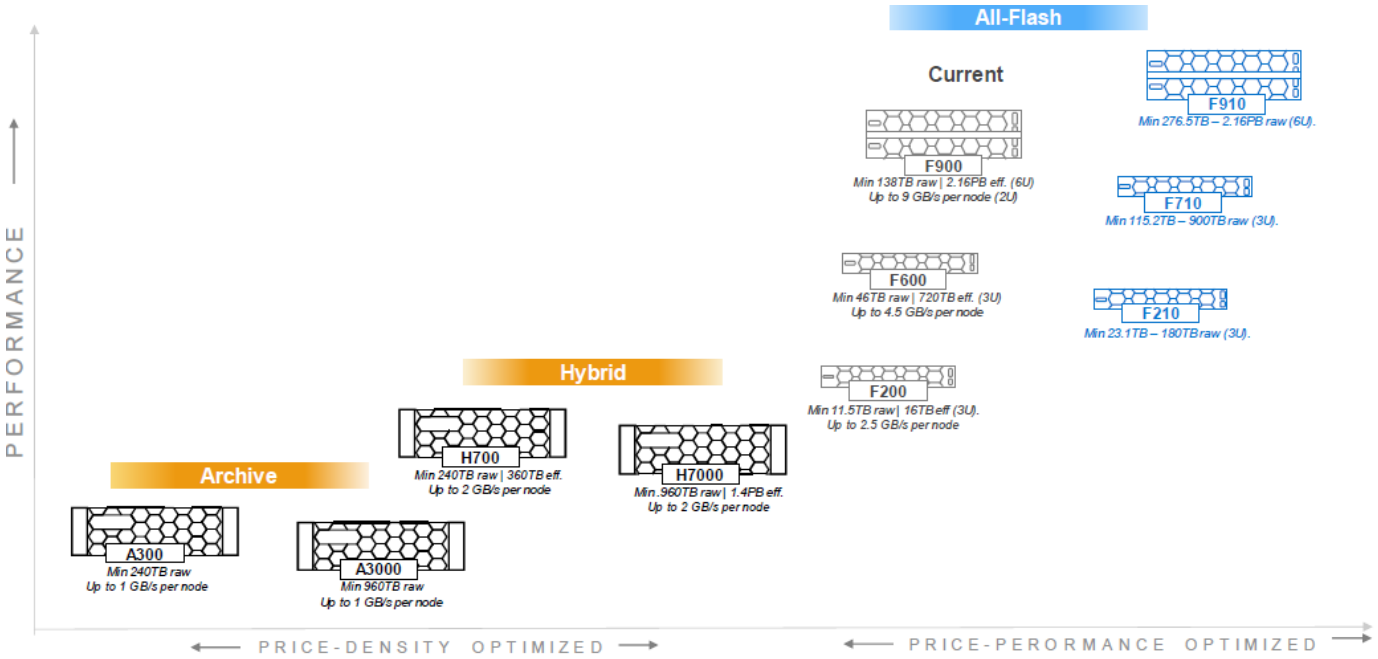

Figure 1. PowerScale family

Generative AI storage requirements

Overview

Generative AI systems thrive on vast amounts of unstructured data, which is essential for training algorithms to recognize patterns, make predictions, and generate new content. Unstructured data, such as text, images, and audio does not follow a predefined model, making it more complex and varied than structured data. For generative AI, the requirements for this data include diversity to avoid biases, high quality to ensure accurate outputs, and large volumes to improve the model's robustness. Additionally, the data must be preprocessed and cleaned to enhance the learning process. Ethical considerations are also paramount, ensuring that the data used does not infringe on privacy and is sourced responsibly. Ultimately, the unstructured data serves as the creative fuel that powers the generative capabilities of AI, enabling it to produce innovative and coherent outputs that mimic human-like creativity.

Preprocessing Unstructured data

Preprocessing unstructured data for generative AI is a crucial step that involves preparing the raw data for use in training AI models. This process includes several tasks such as cleaning, normalizing, and transforming the data into a format that the AI can understand and learn from. The goal is to enhance the quality and structure of the data to improve the performance of generative models. This can involve removing noise, handling missing values, labeling data for supervised learning, and augmenting the dataset to increase its size and diversity. By doing so, the AI models can generate more accurate and coherent outputs.

NVIDIA DGX SuperPOD certification

Overview

The NVIDIA DGX SuperPOD certification refers to the process where each DGX SuperPOD is fully racked, stacked, and configured by NVIDIA-Certified partners. These NVIDIA AI experts ensure that installations are easy, even when building out an AI infrastructure with dozens or hundreds of nodes connected by extensive cabling.

The DGX SuperPOD is an AI data center infrastructure that enables IT to deliver performance for every user and workload. NVIDIA extensively tests it, pushing it to the limits with enterprise AI workloads, so you don't have to worry about application performance. The DGX SuperPOD is powered by NVIDIA Base Command, a software that includes: AI workflow management, libraries that accelerate compute, storage, and network infrastructure, and an operating system optimized for AI workloads.

PowerScale Advantages for DGX SuperPOD

Dell PowerScale is the world's first Ethernet storage solution, which has been validated with NVIDIA's DGX SuperPOD, and includes powerful DGX H100 systems. This collaboration between Dell and NVIDIA is designed to help customers achieve faster and more efficient AI storage. The integration of these advanced technologies allows for handling vast amounts of data at unprecedented speeds, thereby accelerating the process of training AI models. Furthermore, the efficiency of AI storage is significantly improved, reducing required resources and optimizing performance. This innovative solution is a testament to Dell's commitment to pushing the boundaries of technology and delivering cutting-edge solutions to its customers. With the PowerScale storage solution, Dell is paving the way for the future of AI data storage.

PowerScale is the ultimate foundation for GenAI, including the following:

- **Refining GenAI Models with unparalleled adaptability and protection**: PowerScale's scalable architecture allows organizations to seamlessly increase their storage capacity as required, guaranteeing the necessary resources to enhance GenAI models and extract valuable insights. Furthermore, the robust security features of PowerScale, driven by the OneFS operating system, assure organizations that their data is safeguarded at all times.
- **Easy data access with high-speed Ethernet**: Many data centers still use Ethernet as the primary networking connectivity, allowing PowerScale to fit easily. NVIDIA Magnum IO, GPUDirect Storage, and NFS over RDMA are inherently integrated into NVIDIA ConnectX-6 NICs to expedite network access to storage. These technologies aid PowerScale in ensuring that data transfer times are reduced, resulting in quicker storage throughput for AI training, checkpointing, and inferencing. This flawless fusion of storage and computation enables organizations to optimize the performance of their AI workloads while reducing latency.
- **Boosting performance with intelligent scale-out features**: Beyond providing rapid data access, PowerScale introduces a novel Multipath Client Driver that boosts GPU usage and optimizes performance. This cutting-edge feature guarantees that organizations can reach the high-performance benchmarks set by the DGX SuperPOD, effortlessly facilitating the acceleration of AI model training and inference.

For more on the PowerScale NVIDIA DGX SuperPOD certification, see h19971-powerscale-ethernet-superpod-certification.pdf (delltechnologies.com).

Leveraging PowerEdge

Overview

Dell PowerEdge servers are the latest generation, offering superior performance, versatility, and efficiency for various workloads and environments. Compared to previous releases, Dell Servers have several advantages, such as:

- Supporting the new 4th-gen Intel® Xeon® Scalable Sapphire Rapids processors, which deliver higher core counts, faster memory speeds, and improved security features.
- Featuring PCIe 5.0 technology, which doubles the bandwidth and reduces the latency of the previous generation, enabling faster data transfers and more efficient use of accelerators.
- Introducing the first PowerEdge servers with CXL 1.1 (Compute Express Link) capabilities, a new interconnect standard that allows high-speed communication between the CPU and devices such as GPUs, DPUs, and memory.
- Offering more storage capacity and flexibility, with up to 60% more density in the 1U chassis of the R660, taking advantage of the new E3.S form factor SSDs that are smaller and cooler than the previous U.2/U.3 SSDs1.
- Incorporating the Smart Flow design, which is a new feature within the Dell Smart Cooling suite that allows for more airflow through the systems than previous generations, keeping them cool while performing at the highest levels for long periods of time.
- Designed with a focus on environmental sustainability, reducing power consumption and carbon footprint.

PowerScale all-flash F210, F710, and F910 platform

Overview

PowerScale's continuous innovation extends into the AI era by introducing the next generation of PowerEdge-based nodes, introducing the PowerScale F210, F710, and F910. The platforms take advantage of the Sapphire Rapids CPU on the performance front, providing 19% lower cycles-per-instruction. PCIe Gen 5 doubles throughput when compared to PCIe Gen 4. Further, the nodes take advantage of DDR5, offering greater speed and bandwidth. The new platforms also provide greater density and power efficiency.

OneFS 9.7

From a software perspective, PowerScale OneFS 9.7 introduces a significant leap in performance. OneFS 9.7 updates the protocol stack, locking, and direct-write.

Implementing a round-robin distribution strategy across thread groups has significantly reduced thread lock contention. Moreover, thread group numbers are optimized based on core count, which leverages higher core counts to enhance parallelism. This leads to an increase in OneFS's overall efficiency and performance.

Contention on turnstile locks has been reduced by increasing the value of Read-Write (RW) Lock retries. This adjustment permits locks to spin adaptively for an extended duration. It is particularly noteworthy that contention on turnstile locks for Non-Uniform

Memory Access (NUMA) nodes incurs a high cost, making these improvements even more valuable in optimizing system performance.

In the context of NVMe storage nodes, writing operations are strategically executed around the journal for newly allocated blocks. This method circumvents potential bottlenecks associated with journal flushers, thereby enhancing the storage system's efficiency and reliability. Such an approach is crucial for maintaining high performance and preventing data processing and access delays.

When combined, all of the software updates provide tremendous performance improvements across workloads. The software improvements are further expanded with the hardware upgrades.

OneFS 9.8

OneFS 9.8 further pushes the pure software performance envelope, optimizing the OneFS 9.7 updates to build on them even further. Additionally, OneFS 9.8 introduces enhancements to thread handling and lock management. Finally, general code updates have brought about a significant performance leap.

Licensing

The new all-flash nodes are part of a single license with a new tier identifier, “Tier 16”. The single license applies to the F210, F710, and F910

Data reduction

The PowerScale F210, F710, and F910 nodes further increase the density and capacity envelope with data reduction. All 3 platforms provide inline data reduction, which is enabled by default. Compression and dedupe are part of the inline data reduction. Combined, the overall density of a cluster is significantly increased. As the per Rack Unit (RU) density increases, the solution's Total Cost of Ownership (TCO) decreases and subsequently lowers the carbon footprint.

Journal enhancements

The OneFS journal in the new PowerScale all-flash nodes uses a 32 GB Dell Software Defined Persistent Memory (SDPM) technology configuration. Previous platforms used NVDIMM-n for persistent memory, which consumed a DIMM slot. The BIOS-managed SDPM is a vault solution with a single module comprised of the Battery Backup Unit (BBU) and the Vault Optimized Storage Solution (VOSS) NVMe M.2 drive. Data is copied from DDR5 DIMMs to the VOSS module during power loss. The F210 and F710 have a single BBU, while the F910 has dual BBUs.

Fault LEDs

The new PowerScale all-flash nodes use the “Fault” LEDs on the PowerEdge platform. The “Fault” LEDs are part of the front panel, with one at each end. An amber LED indicates a failed drive. For more information on the other LED indicators, refer to the *Dell PowerScale F210 and F710 Field Replaceable Unit Guide* or the *Dell PowerScale F910 Field Replaceable Unit Guide*.

Cooling

The Power Supply Units (PSU) on the F210, F710, and F910 nodes are strategically split on both ends of the node for maximum cooling optimization.

Performance

The PowerScale F210, F710, and F910 nodes capitalize on significant leaps in hardware and software from the previous generations. OneFS 9.7 and 9.8 introduce tremendous

performance-oriented updates, including the protocol stack, locking, and direct-write. The PowerEdge servers offer a substantial hardware leap from previous generations. The hardware and software advancements combine to offer enormous performance gains, particularly for streaming reads and writes.

Density

The new PowerScale all-flash platforms all capitalize on significant innovation in disk capacity. With the introduction of new drive capacities, each platform offers a significant jump in per Rack Unit (RU) density. The additional per RU density offers tremendous advantages for data center utilization. As data center utilization decreases, this has a direct impact on the environmental cooling requirements of the data center. Combined, the density and environmental factors lead to a reduced carbon footprint, allowing organizations to meet sustainability goals. The cluster density per RU is listed in the table below.

Table 1. Isilon and PowerScale cluster density per Rack Unit

Platform	Cluster Density per Rack Unit
PowerScale F200	30.72 TB
PowerScale F210	61 TB
PowerScale F600	245 TB
PowerScale F710	307 TB
Isilon F810	231 TB
PowerScale F910	360 TB

Given the additional density of each platform, organizations consume a smaller data center footprint. The density savings are further magnified for clusters that scale up to the 252-node maximum. In addition to the density improvements, there are also significant performance leaps, allowing organizations to consolidate further on overall data center utilization and reducing carbon footprints.

The F910 offers 20% more density than the F710, offering performance and density in a single package. The additional density is further magnified as the cluster grows.

Networking

The F210, F710, and F910 nodes provide options for the ConnectX-6 Dx dual port 100GbE PCIe Gen4.0 x16 and the ConnectX-6 Lx dual port 25GbE PCIe Gen3 and Gen4 x8.

Serial port

The PowerScale all-flash F210, F710, and F910 use an RIO/DB9 card for a serial port connection.

LAN on Motherboard

The PowerScale all-flash F210, F710, and F910 use a LAN on Motherboard (LOM) for the management port. The 1 Gb LOM card connects to the planar via PCIe.

Supported OneFS releases

The F210 and F710 require OneFS 9.7 at minimum, while the F910 requires OneFS 9.8 at minimum.

F210 and F710 platform

The new PowerScale F210 and F710 all-flash nodes leverage Dell PowerEdge R660 platform, unlocking the next generation of performance. On the software front, the F210 and F710 take advantage of significant performance improvements in PowerScale OneFS 9.7 and 9.8. Combining the hardware and software innovations, the F210 and F710 easily tackle the most demanding workloads. The F210 and F710 offer greater density in a 1U platform, with the F710 supporting 10 NVMe SSDs per node and the F210 offering a 15.36 TB drive option.

F910 platform

The new PowerScale F910 all-flash node leverages the Dell PowerEdge R760 platform, unleashing a combination of performance, density, and power efficiency. The F910 takes advantage of the software improvements in OneFS 9.7 and 9.8 to provide the next level of node performance. It combines software enhancements with hardware that is ready for the next generation of workloads. Utilizing a dual-socket CPU, the Intel Sapphire Rapids 6442Y (2.6G/24C), and 512 GB of DDR5 RAM, the F910 delivers blistering performance. Adding to the storage suite of the world's first Ethernet storage-certified NVIDIA DGX SuperPOD solution, the F910 accelerates the learning of GenAI models.

The F910 supports 24 NVME SSDs per node, with a maximum drive size of 30 TB, adding to the density equation while consuming power efficiently. In a single package, the F910 delivers density, performance, and power efficiency in one package.

PowerScale F210

Overview

The PowerScale F210 is a 1U chassis based on the PowerEdge R660. A minimum of three nodes is required to form a cluster and a maximum of 252 nodes. The F210 is node pool compatible with the F200. For more details, see the Brownfield Clusters section.

The front panel of the F210 has a bezel protecting the 4 NVMe SSD drives, as displayed in the image below. The bezel has an LCD panel providing system information, status, iDRAC IP address and error messages. Removing the bezel provides access to the 4 drives. The LED alarm indicator is located on the left of the front panel.

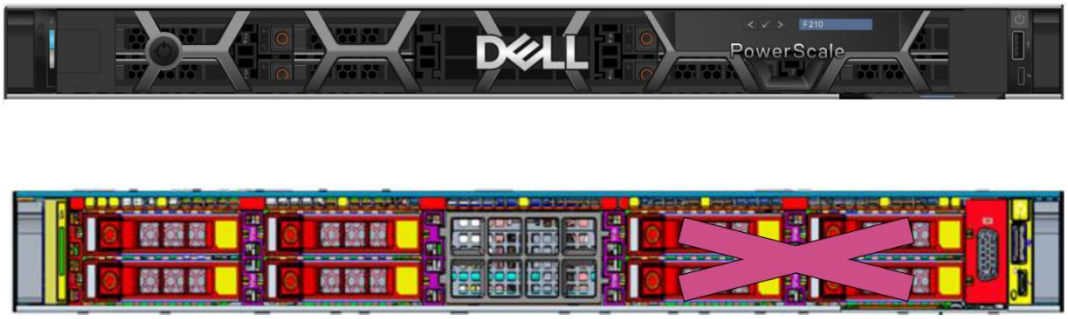

Figure 2. PowerScale F210

The following table lists the F210 specifications.

Table 2. PowerScale F210 node specifications

Attribute	PowerScale F210 Specification
Chassis	1U Dell PowerEdge R660
CPU	Single Socket – Intel Sapphire Rapids 4410Y (2G/12C)
Memory	Dual Rank DDR5 RDIMMs 128 GB (8 x 16 GB)
Journal	1 x 32 GB SDPM
Front-end networking	2 x 100 GbE or 25 GbE
Infrastructure networking	2 x 100 GbE or 25 GbE
NVMe SSD drives	4
Trusted Platform Module	Yes

The following figure illustrates the F210 node interfaces.

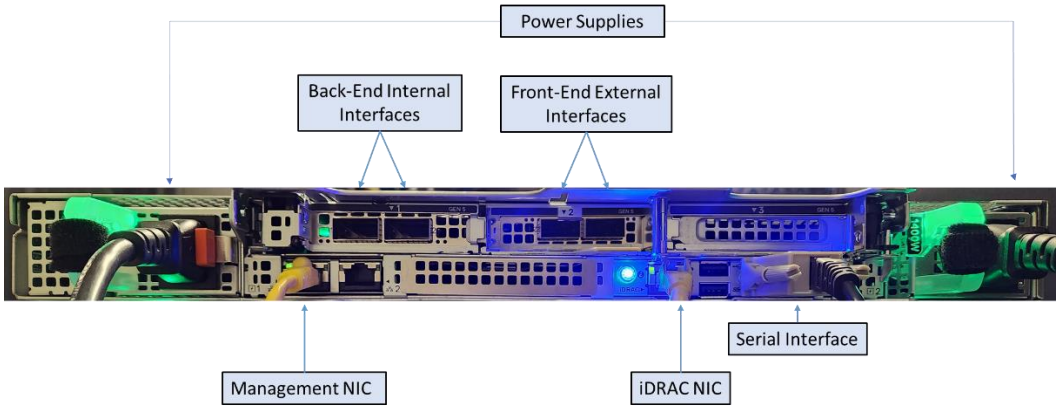

Figure 3. PowerScale F210 interfaces

Drive support & capacity

The F210 nodes use NVMe SSDs, where the previous generation F200 used SAS SSDs. Utilizing NVMe SSDs allows the F210 the advantage of the 15.36 TB QLC drive option. The F210 offers a 100% density improvement over the F200 with 2x the performance over the F200. In a 3 RU rack configuration of 3 nodes, the F210 raw capacity spans a minimum of 23 TB to a maximum of 184 TB. The available drive capacities for the F210 are listed in the following table.

Table 3. PowerScale F210 drive support & capacity

Non-SED Drive Capacities	SED-FIPS Drive Capacities	SED-Non-FIPS Drive Capacities
960 GB*	960 GB*	15.36 TB QLC
1.92 TB	1.92 TB	
3.84 TB	3.84 TB	
7.68 TB	7.68 TB	
15.36 TB QLC	15.36 TB QLC**	

*Short formatted 1.92TB SSDs for F200 node compatibility **Future availability

PowerScale F710

Overview

The PowerScale F710 is a 1U chassis based on the PowerEdge R660. A minimum of three nodes with a maximum of 252 nodes is required to form a cluster. The F710 is not node pool compatible with the F600.

The front panel of the F710 has a bezel protecting the 10 NVMe SSD drives, as displayed in the image below. The bezel has an LCD panel providing system information, status, iDRAC IP address and error messages. Removing the bezel provides access to the 10 drives. The LED alarm indicator is located on the left of the front panel.

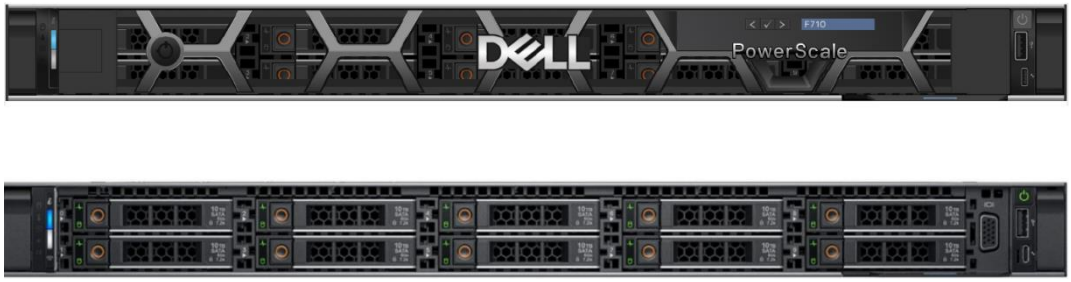

Figure 4. PowerScale F710

The following table lists the F710 specifications.

Table 4. PowerScale F710 node specifications

Attribute	PowerScale F710 Specification
Chassis	1U Dell PowerEdge R660
CPU	Dual Socket – Intel Sapphire Rapids 6442Y (2.6G/24C)
Memory	Dual Rank DDR5 RDIMMs 512 GB (16 x 32 GB)
Journal	1 x 32 GB SDPM
Front-end networking	2 x 100 GbE or 25 GbE
Infrastructure networking	2 x 100 GbE
NVMe SSD drives	10
Trusted Platform Module	Yes

The following figure illustrates the F710 node interfaces.

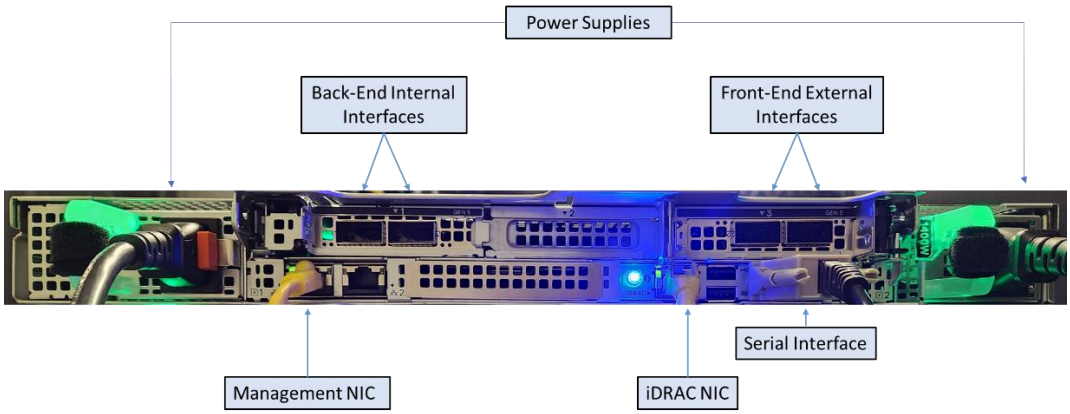

Figure 5. PowerScale F710 interfaces

Drive support & capacity

The F710 nodes use NVMe SSDs. The F710 offers a 25% density improvement over the F600 with 2x the performance over the F600. In a 3 RU rack configuration of 3 nodes, the F710 raw capacity spans a minimum of 115 TB to a maximum of 922 TB. The available drive capacities for the F710 are listed in the following table.

Table 5. PowerScale F710 drive support & capacity

Non-SED Drive Capacities	SED-FIPS Drive Capacities	SED-Non-FIPS Drive Capacities
3.84 TB	3.84 TB	15.36 TB QLC
7.68 TB	7.68 TB	30.72 TB QLC
15.36 TB QLC	15.36 TB QLC*	
30.72 TB QLC	30.72 TB QLC*	

*Future availability

PowerScale F910

Overview

The PowerScale F910 is a 2U chassis based on the PowerEdge R760. A minimum of three nodes and a maximum of 252 nodes is required to form a cluster. The F910 is node pool compatible with the F900.

The front panel of the F910 has a bezel protecting the 24 NVMe SSD drives, as displayed in the image below. The bezel has an LCD panel providing system information, status, iDRAC IP address and error messages. Removing the bezel provides access to the 24 drives. The LED alarm indicator is located on the left of the front panel.

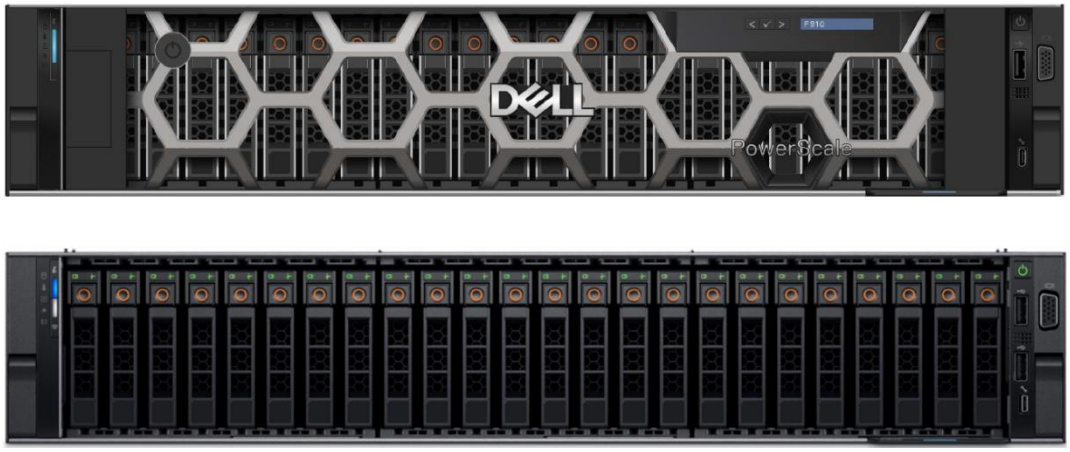

Figure 6. PowerScale F910

The following table lists the F910 specifications.

Table 6. PowerScale F910 node specifications

Attribute	PowerScale F910 Specification
Chassis	2U Dell PowerEdge R760
CPU	Dual Socket – Intel Sapphire Rapids 6442Y (2.6G/24C)
Memory	Dual Rank DDR5 RDIMMs 512 GB (16 x 32 GB)
Journal	1 x 32 GB SDPM
Front-end networking	2 x 100 GbE or 25 GbE
Infrastructure networking	2 x 100 GbE
NVMe SSD drives	24
Trusted Platform Module	Yes

The following figure illustrates the F910 node interfaces.

Figure 7. PowerScale F910 interfaces

Drive support & capacity

The F910 nodes use NVMe SSDs. In a 6 RU rack configuration of 3 nodes, the F710 raw capacity spans a minimum of 276 TB to a maximum of 2.16 PB. The available drive capacities for the F910 are listed in the following table.

Table 7. PowerScale F910 drive support & capacity

Non-SED Drive Capacities	SED-FIPS Drive Capacities	SED-Non-FIPS Drive Capacities
3.84 TB	3.84 TB	15.36 TB QLC
7.68 TB	7.68 TB	30.72 TB QLC
15.36 TB QLC	15.36 TB QLC*	
30.72 TB QLC	30.72 TB QLC*	

*Future availability

Greenfield clusters

Overview

Because a greenfield cluster is a new cluster deployment, a minimum of 3 of the F210, F710, or F910 nodes is required to form a cluster. All 3 of the nodes must be the same platform.

Brownfield clusters

Overview

Existing clusters in the field may add the F210, F710, and F910 nodes for additional capacity and performance. A minimum of 3 nodes is required to create a new node pool for the F210, F710, and F910.

Clusters that have F600s are not node pool compatible with the F710, requiring a separate node pool for the F710 with a minimum of 3 nodes. Clusters that have F900s are node pool compatible with the F900.

However, for clusters that have F200s, they have the option to increase the same node pool by adding a single F210. In this case, the only supported drive is the short formatted 1.92 TB option, resulting in a 960 GB capacity per drive. While the F210 is node pool compatible with the F200, a performance degradation is experienced where the F210 is not at full performance. When an F210 is added to the F200 node pool, the WebUI explains the soft restriction in the following image.

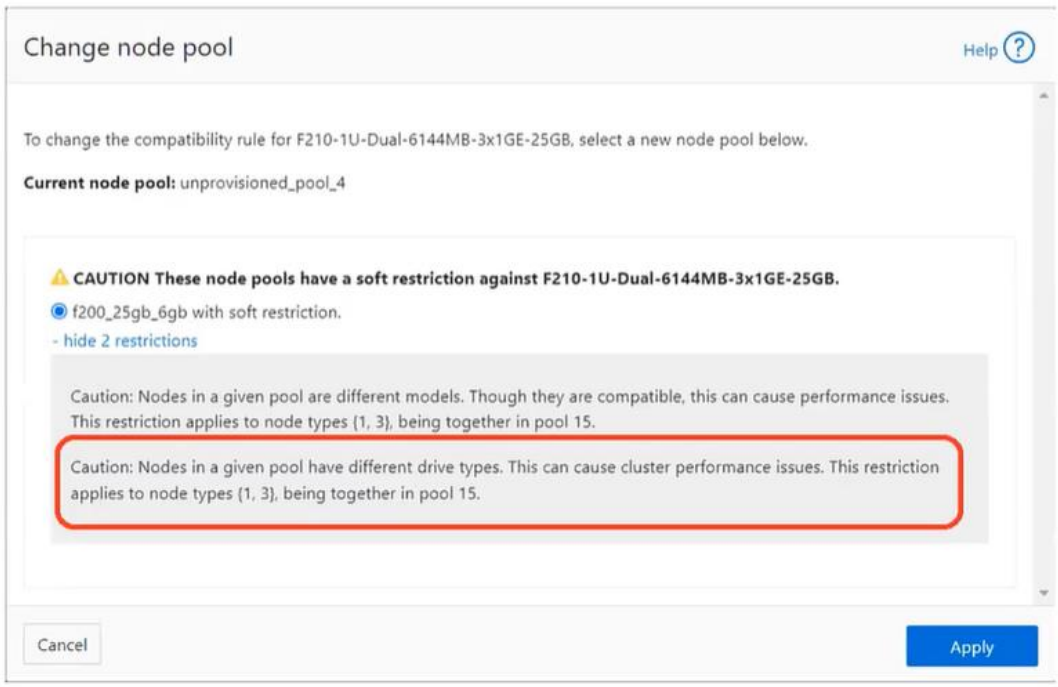

Figure 8. F210 WebUI node pool soft restriction

A similar soft restriction message is displayed in the CLI, as displayed in the following image.

```
Action can be performed successfully.
Caution: Nodes in a given pool are different models. Though they are compatible, this can cause performa
nce issues. This restriction applies to node types {1, 2}, being together in pool 13.
Caution: Nodes in a given pool have different drive types. This can cause cluster performance issues. Th
is restriction applies to node types {1, 2}, being together in pool 13.
Changing node pool membership from [1] to [1, 2].
Are you sure? (yes/[no]): yes
```

Figure 9. F210 CLI node pool soft restriction

References

Dell Technologies documentation

The following Dell Technologies documentation provides other information related to this document. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell Technologies representative.

- Dell Technologies Info Hub
- PowerScale OneFS Info Hubs | Dell US