



insideHPC

insideHPC Special Report

A Practical Guide to Networking Dell EMC Ready Nodes

Written by Peter ffoulkes



BROUGHT TO YOU BY

DELLEMC



Contents

Introduction.....	2	Network switch configuration with Dell EMC Ready Nodes	8
How do I network vSAN Ready Nodes?	3	Dedicated storage networks.....	8
Dell EMC networking for vSAN.....	4	Disaggregation network architecture.....	8
Network security with vSAN.....	5	Hybrid dedicated/leaf-spine network architecture.....	9
How do I network Dell EMC Microsoft Storage Spaces Direct Ready Nodes?	6	Where Ready Nodes are the optimal choice	9

Introduction

The nature of Information Technology (IT) has been transformed by the advent of software-defined infrastructure based on virtualized services for servers, networking, and storage to deliver capabilities that can expand and contract as needed to meet the demands of any workload. Together these virtualized services can simplify and speed deployment, automate the management of agile and highly scalable environments, and can deliver [cost savings of almost 50%](#) in comparison to proprietary models.

To meet the needs of those who wish to simplify and accelerate the deployment of software-defined storage systems with minimal risk, Dell EMC has a portfolio of Ready Solutions that are certified reference systems for simplified configuration, sizing, quoting, deployment, financial packaging. Ready Nodes are PowerEdge servers pre-configured for specific workloads

Dell EMC has a portfolio of Ready Solutions that are certified reference systems for simplified configuration, sizing, quoting, deployment, financial packaging.

and, available with software and services. Ready Bundles combine software, servers, network, and storage optimized together for specific workloads. Ready Nodes and Ready Bundles evolve with new components and versions so check the corresponding Solution Overviews for the latest specifications.

This document is focused on Dell EMC vSAN Ready Nodes and Microsoft Storage Spaces Direct Ready Nodes — a practical networking guide.

Dell EMC vSAN Ready Nodes

- Reduce project risk
- Improve storage efficiency
- Scale quickly



Dell EMC Microsoft Storage Spaces Direct Ready Nodes

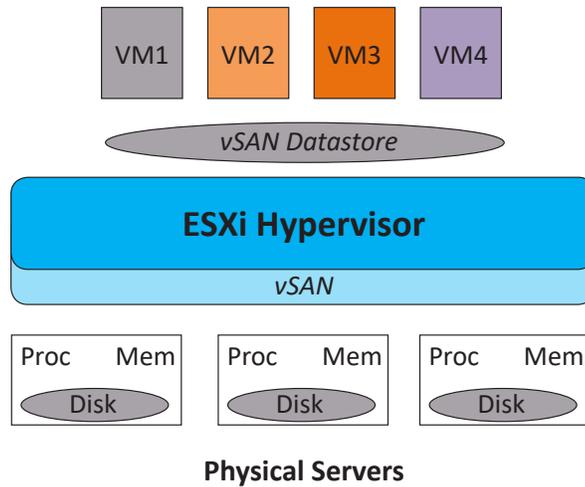
- Confidence
- Convenience
- Customer support



How do I network vSAN Ready Nodes?

Dell EMC vSAN Ready Nodes are built on Dell EMC PowerEdge servers that have been pre-configured, tested and certified to run VMware vSAN. Each Ready Node includes just the right amount of CPU, memory, network I/O controllers, HDDs and SSDs for VMware vSAN.

VMware vSAN is Software Defined Storage (SDS) that leverages a distributed control plane abstraction to create a pool of storage from disparate server-based disk hardware. That abstraction is comparable to the way the vSphere ESXi hypervisor converts a cluster of server hardware into a pool of compute resources (VMs). As an integrated feature of the ESXi kernel, vSAN exploits the same clustering capabilities to deliver a comparable virtualization paradigm to storage that has been applied to server CPU and memory for many years.



Compute and Storage (Disk) Virtualization

vSAN SDS is an extension of hypervisor-based virtualization

See [Architecting a Dell EMC Hyper-converged Solution with VMware vSAN](#)

Dell EMC vSAN ReadyNode Example

Dell EMC vSAN Ready Nodes

PRIMARY CUSTOMER

VMware vSAN customers looking to scale capacity, performance or both

CUSTOMER BENEFITS

- **Reduces project risk** with validated, tested, and certified configurations for vSAN deployment
- **Improves storage efficiency** with software-defined storage that leverages servers and flash
- **Scale quickly** with faster configuration, fewer update steps and reduced time for maintenance

DELL EMC DIFFERENTIATION²

- Dell EMC and VMware have collaborated on vSAN for years and **thousands of hours of testing**
- Patented failsafe memory fault isolation, automatic failover redundant hypervisors and Memory Page Retire help **protect uptime**
- Installed, implemented and supported by Dell EMC as a **single trusted source**

Reduce risks and lower total cost of ownership with a VMware vSAN building block that's fast and easy to scale

- | | |
|-----------------|--|
| VMware | VMware vSphere, vCenter, ESXi, and vSAN |
| DELL EMC | PowerEdge™ Servers including R440, R640, R740, R740xd, R6415, R7415, R7425, C6420, FC430 |
| DELL EMC | Dell EMC OpenManage software integration for VMware vCenter |
| DELL EMC | Dell EMC Professional Services, ProDeploy, ProSupport, Dell Financial Services |



DELL EMC

Each physical host in a [vSAN cluster](#), can consist of three to 64 hosts (as of vSAN 6.6), contributing direct attached storage to a [Disk Group](#) in the aggregate storage pool. A hybrid vSAN deployment consists of physical servers that have both spinning disks (HDD) and solid state drives (SSD/flash) to accommodate tiered storage deployments, with the SSD reserved for intense read and write operations. There are also all-flash deployments for particularly demanding workloads, where sequential read/write transactions are dense and associated HDD latency is intolerable.

vSAN follows the same operational model for provisioning storage that vSphere uses to provision virtual machines. vCenter Server acts as the common management point for both vSAN and ESXi. In a vSAN environment, when a virtual machine is created, the administrator is asked to assign a previously-configured storage policy that reflects an application's performance and availability requirements. The policy that is applied to the workload is all that is needed with regard to storage provisioning. There is no need to manage Logical Unit Numbers, or configure masking or zoning, or set RAID levels. All the details and mundane tasks typically required in a legacy, non-software defined storage deployment are relegated to the vSAN software, which provides centralized management and hardware abstractions that place the burden of executing the policy's requirements on the software and away from the hardware.

Dell TechCenter

The [Dell TechCenter](#) web site contains many detailed reference resources and best practices guides such as the [Leaf-Spine Deployment and Best Practices Guide for Greenfield Deployments](#) to enable a network administrator or engineer with traditional networking experience to deploy a layer 2 or layer 3 leaf-spine architecture using the examples provided.

DELL EMC NETWORKING FOR VSAN

Depending on the design requirements, the network fabric can be a leaf and spine (Clos) or a traditional hierarchical, three-tiered architecture (access, aggregation and core). The distributed nature of vSAN's virtualized data store lends itself to a network architecture that can be scaled out horizontally, can provide predictable performance and control jitter between end nodes, offers multiple paths of equal cost between the different layers of switching and lowers oversubscription. Such characteristics are typical of a leaf-and-spine (L/S) architecture, thereby making it the preferred architecture for VMware vSAN. Both a routed (L3) L/S and a switched (L2) network fabric can be deployed to support vSAN, each with its design considerations.

Network bandwidth capacity

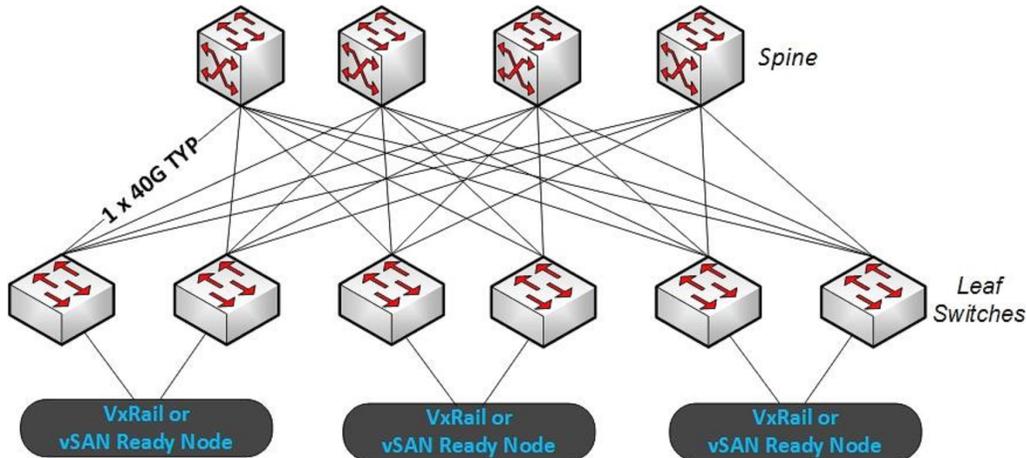
There are no "special" considerations that need to be made when [designing a network fabric that can support vSAN software-defined storage](#) other than total network bandwidth capacity. Like any other high-performance data center fabric, a vSAN network needs to be redundant, resilient and robust. The virtual constructs that are abstracted from the hardware are only as stable as the underlying infrastructure, so reliability and performance are of the utmost importance.

Do pay attention when using technologies such as Non-Volatile Memory express (NVMe) drives, capable of generating much higher volumes of network traffic.

The [VMware vSAN Network Guide](#) provides in-depth information and guidelines regarding network considerations. Starting with vSAN 6.6, multicast is not required on the physical switches. However, if some hosts in the vSAN cluster are running earlier versions of software, a multicast network is still required.

Typical Leaf and Spine Architecture (Clos)

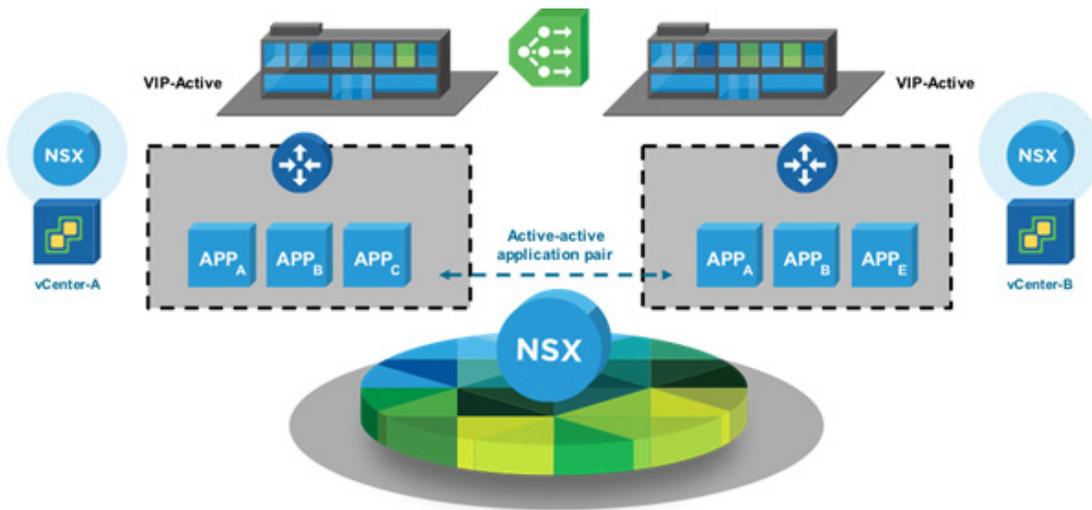
A preferred architecture for vSAN network fabrics



Typical leaf and Spine Architecture (Clos). Preferred architecture for vSAN network fabrics

See [Architecting a Dell EMC Hyper-converged Solution with VMware vSAN](#)

NETWORK SECURITY WITH VSAN



See [Designing the vSAN Network from VMware](#)

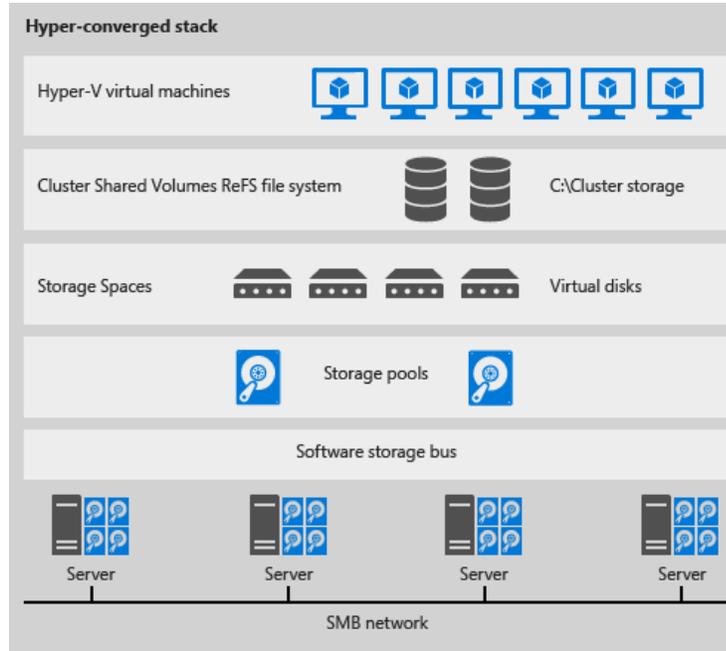
Alongside vSAN, [VMware NSX](#) is the VMware network virtualization and security platform that de-couples the network functions from the physical devices, in a way that is analogous to de-coupling virtual machines (VMs) from physical

servers. In order to de-couple the new virtual network from the traditional physical network, NSX natively re-creates the traditional network constructs in virtual space, which includes ports, switches, routers, and firewalls.

How do I network Dell EMC Microsoft Storage Spaces Direct Ready Nodes?

Dell EMC Microsoft Storage Spaces Direct Ready Nodes are built on Dell EMC PowerEdge servers, which provide the storage density and compute power to maximize the benefits of Storage Spaces Direct (S2D). They leverage the advanced feature sets in Windows Server 2016 Datacenter Edition to deploy a scalable hyper-converged infrastructure solution with Hyper-V and Storage Spaces Direct. For Microsoft server environments, S2D scales to 16 nodes in a cluster, and is a kernel-loadable module, (with no RDMA iWarp or RoCE needed) which is a low risk approach to implementing an S2D cluster. Dell EMC Microsoft Storage Spaces Direct Ready Nodes can be ordered with software installed with or as bare metal for customers with volume license agreements.

You can use Windows Server 2016 Datacenter Edition with the Server Core, or Server with Desktop Experience installation options, the steps in the “Configure the Network” and “Configure Storage Spaces Direct” sections are identical whether you are using Server with Desktop Experience or Server Core installations. The management system can be run inside of a virtual machine or on a physical machine, however, the management system needs to be joined to the same domain or a fully trusted domain. There are [several deployment options](#) available to connect the cluster nodes, and RDMA is not available for networking inside a virtual machine.



Hyper-V VMs run directly on the Storage Spaces Direct cluster that hosts the storage, as shown. Virtual machine files are stored on local CSVs. This allows for scaling Hyper-V compute clusters together with the storage it is using, reducing the number of clusters required.

See [Deploying Storage Spaces Direct from Microsoft](#), and see [Microsoft Storage Spaces Direct Ready Nodes - Sample Switch Configurations](#)

Dell EMC Microsoft Storage Spaces Direct Ready Node Example

Dell EMC Microsoft Storage Spaces Direct Ready Nodes

PRIMARY CUSTOMER

Customers or managed service providers invested in Microsoft Hyper-V®

CUSTOMER BENEFITS

- **Confidence** in preconfigured, tested and certified solutions
- **Convenience** of procuring, deploying with Dell EMC guidance that covers multiple scenarios
- **Customer support** that is streamlined, and collaborative support from initial contact

DELL EMC DIFFERENTIATION

- Takes the guesswork out of building **Storage Spaces Direct clusters**
- **Admin work reduced** by fewer interfaces and steps to complete tasks, with no specialized knowledge needed
- **Customer support:** Single point of contact for solution support

A pre-validated solution that offers new levels of flexibility, confidence and time to value



Microsoft Windows Server 2016 and Storage Spaces Direct software



PowerEdge™ Servers including R640, R740xd



Management integration



Dell EMC Professional Services, ProDeploy, ProSupport, Dell Financial Services



DELL EMC

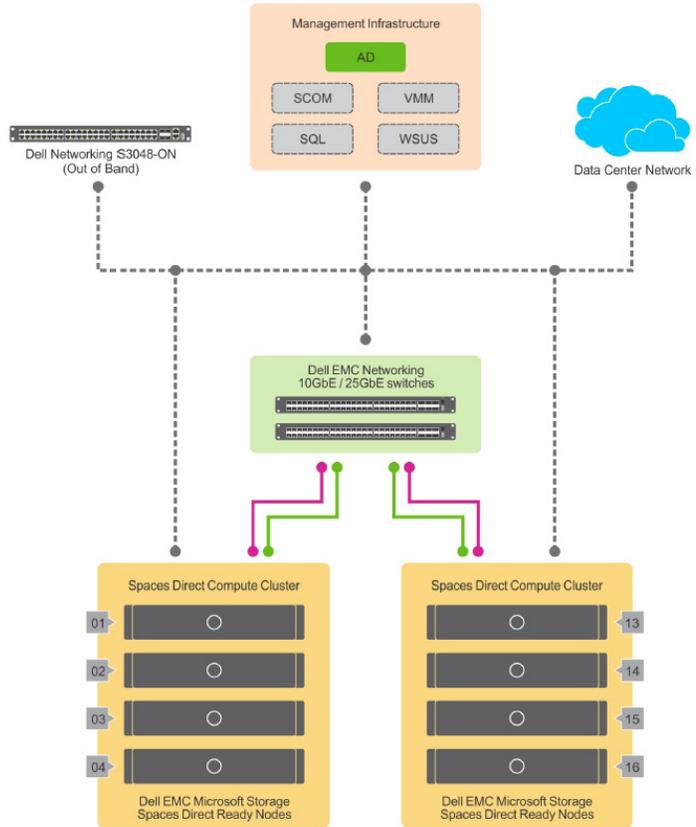
The disks intended for Storage Spaces Direct need to be empty and without partitions or other data. If a disk has partitions or other data, it will not be included in the Storage Spaces Direct system.

Storage Spaces Direct requires high bandwidth and low latency network connections between nodes. This network connectivity is important for both system performance and reliability. When using physical nodes instead of VMs, using at least two 10 Gb connections between the nodes is recommended, preferably with RDMA to increase throughput and reduce the CPU usage for network traffic.

There are two common versions of remote direct memory access (RDMA) network adapters — RDMA over Converged Ethernet (RoCE) and Internet Wide-area RDMA Protocol (iWARP). You can use either with Storage Spaces Direct as long as it has the Windows Server 2016 logo, but iWARP usually requires less configuration. Top of Rack switches and server configurations vary depending on the network adapters and switches. Configuring the server and switch correctly is important to ensure reliability and performance of Storage Spaces Direct.

Ensure that there is no data on any of the disks of the cluster before running configuration commands or scripts. It will remove any data on the disks that are not being used by the operating system.

Windows Server 2016 also introduces a new virtual switch that has network teaming built in called Switch Embedded Teaming (SET). This virtual switch allows the same physical NIC ports to be used for all network traffic while using RDMA. This reduces the number of physical NIC ports that would otherwise be required and allows network management through the Software Defined Network features of Windows Server.



If you’re deploying a hyper-converged cluster, the last step is to provision virtual machines on the Storage Spaces Direct cluster.

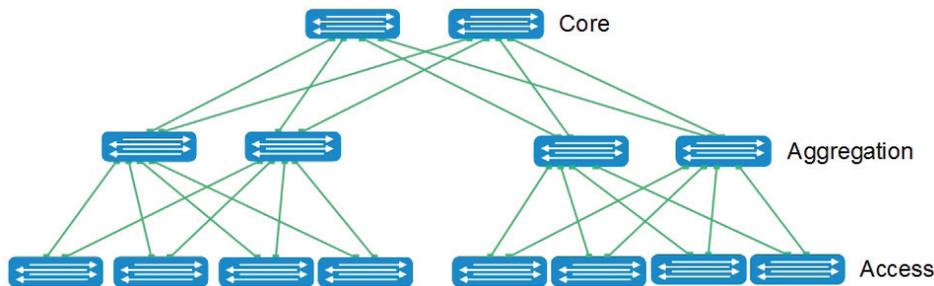
After deploying your clustered file server, we recommend testing the performance of your solution using test workloads before bringing up any production workloads. This lets you confirm that the solution is performing properly and work out any lingering issues before adding the complexity of production workloads. For more info, see [Test Storage Spaces Performance Using Synthetic Workloads](#).

Network switch configuration with Dell EMC Ready Nodes

Storage networks are constantly evolving. From traditional Fibre Channel to IP-based storage networks, each technology has its place in the data center. IP-based storage solutions have two main network topologies to choose from based on the technology and administration requirements. Dedicated storage network topology, shared leaf-spine network, software defined storage, and iSCSI SAN are all supported. Hybrid network architectures are common, but add to the complexity.

DEDICATED STORAGE NETWORKS

Traditional network architecture



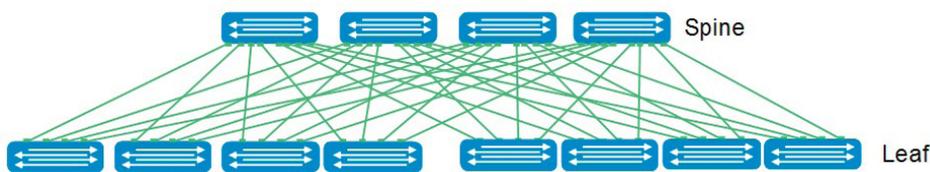
See [Dell EMC Switch Configuration Guide for iSCSI and Software Defined Storage](#)

Fibre Channel storage has imparted a traditional network philosophy and implementation methodology to IP-based storage. The dedicated storage network has a proven design that provides

performance, predictability, and manageability to storage deployments. The storage traffic is isolated from the application traffic to allow each network to be optimized for its own purpose.

DISAGGREGATION NETWORK ARCHITECTURE

Leaf-spine architecture



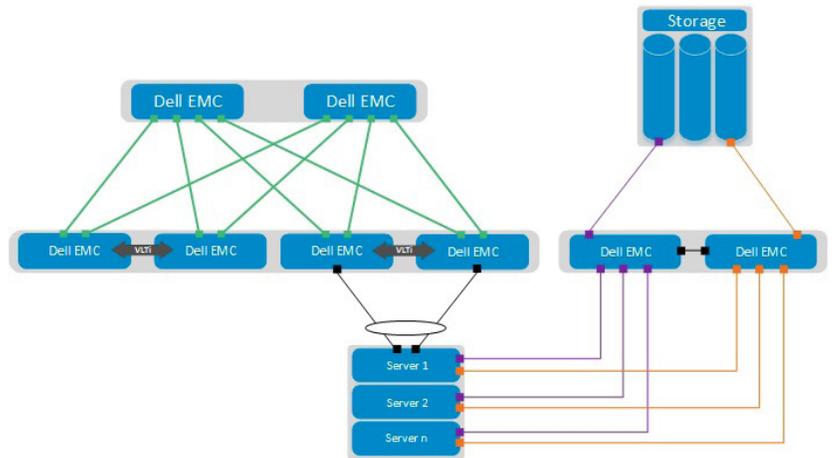
As a direct result of increasing east-west traffic within the data center (server-server, server-storage, etc.), an alternative to the traditional access-aggregation-core network model is becoming more widely used. This architecture is known as a Clos or leaf-spine network and is designed to minimize the number of hops between hosts, thus reducing network communication latency.

In a leaf-spine architecture, the access layer is referred to as the leaf layer. Servers and storage devices connect to leaf switches at this layer.

At the next level, the aggregation and core layers are condensed into a single spine layer. Every leaf switch connects to every spine switch to ensure that all leaf switches are no more than one hop away from one another. This minimizes latency and the likelihood of bottlenecks in the network. A leaf-spine architecture is highly scalable. As administrators add racks to the data center, a pair of leaf switches are added to each new rack. Spine switches may be added as bandwidth requirements increase.

HYBRID DEDICATED/LEAF-SPINE NETWORK ARCHITECTURE

A dedicated storage network has been a popular standard for iSCSI based storage for performance and administration reasons including low latency, predictability, lack of competition with application traffic, simplified management, and straight forward fault isolation. The following diagram shows the dedicated storage network on the right and a leaf-spine network for application traffic on the left, may need to coexist as hybrid networks to optimize the balance between north-south and east-west network traffic.



While these established approaches are generally well understood, the [configuration details](#) can be complex and change over time.

Where Ready Nodes are the optimal choice

The perennial question for IT asset acquisition is whether to build or consume, and there is no ‘one size fits all’ answer to the question. Individual components provide the maximum flexibility, but also require the maximum of self-support. Reference architectures provide the recipe and guidance, but the customer still has the responsibility for making the system work. At the other extreme of the product spectrum, engineered systems and hybrid cloud platforms provide a fully supported integrated offering, yet deliver limited flexibility based upon high level appliance-like building blocks and services.

By disaggregating networking system software, and by decoupling networking software from hardware

platforms, companies can free themselves from the rigid and proprietary environments of yesterday to embrace software-defined principles and unlock innovation at any scale, and based on open, industry standard hardware and software.

For more information visit the [Dell EMC TechCenter](#) for the latest in-depth guidance.

Dell EMC [Ready Solutions](#) offer the ultimate compromise between the flexibility of complete custom architectures and the restricted nature of appliance-like products and services, delivering the best of both worlds and supporting technology evolution with enterprise grade support.

“The network has become again the bottleneck of a system, mostly because of NVMe drives. Four NVMe drives, aggregated, are capable of generating around 11 gigabits per second of bandwidth, which tops a 100-gigabit connection. They may saturate and block I/O with just four drives, so we are looking to 25 gigabit, which becomes 50 gigabit because we’re using a spine-leaf approach. Every server has two lanes, so aggregate bandwidth is 50 gigabits per second. This helps us ensure that the network will not be too much of a storage bottleneck.”

– Antonio Cisternino, University of Pisa Chief Information Officer

For Dell EMC customer case studies, see <http://www.dellemc.com/poweredge-stories>

Intel® Technology Enables Scale-Out Storage Architectures

Intel® technologies are already helping data centers move to SDI and software-defined storage systems. Intel architecture-based solutions maximize infrastructure investments and include a portfolio of interoperable, scalable, and programmable products and technologies. This includes key software capabilities that significantly enhance the value of foundational Intel architecture in SDI hardware and software ecosystems. One example is the ability to speed replication with erasure coding, which is optimized for Ceph* software on the Intel® Xeon® processor family.

Key storage-related products and technologies include:

- The Intel Xeon processor family with storage workload optimization enables more efficient storage, smarter data protection, and exceptional system performance.
- The Intel® Solid-State Drive (Intel® SSD) Data Center Family with NVMe® delivers more performance with fewer resources to boost storage density, speed, and reliability.
- Intel® QuickAssist Technology, a workload acceleration tool in Intel Xeon processor E5 and E7 families for compression and encryption offload.
- Intel® Cache Acceleration Software (Intel® CAS) improves application performance by using an Intel SSD as a cache or frequently accessed data.
- Intel security technologies verify that virtual servers boot into “known good states” (Intel® Trusted Execution Technology, or Intel.®)

Source: [“Making the Case for Software Defined Storage”](#)