

DELL EMC VXRAIL™ APPLIANCE TECHBOOK

A Hyperconverged Infrastructure Appliance from
Dell EMC® and VMware®

ABSTRACT

This TechBook is a conceptual and architectural review of the Dell EMC VxRail Appliance, powered by VMware vSAN, and with Intel Inside®. The TechBook first describes how hyperconverged infrastructure drives digital transformation and then focuses on the VxRail Appliance as a leading hyperconverged technology solution.

May 2019

Contents

IT's new challenge: Leading digital transformation	6
Dell EMC digital transformation: Faster outcomes. Simpler IT	7
Accelerating IT transformation with Dell EMC hyperconverged infrastructure	7
Fully transforming to the hybrid cloud	8
Innovate rather than integrate	8
Hyperconverged infrastructure: A modern infrastructure for modern IT challenges	10
Enabling technologies for HCI	11
Drivers for hyperconverged infrastructure adoption	11
Dell EMC hyperconverged infrastructure platforms	13
Dell EMC HCI delivers a turnkey customer experience	14
Dell EMC VxRail Appliances	15
VxRail Security and Compliance	17
VxRail hardware architecture	19
VxRail Appliance cluster	20
VxRail models and specifications (based on 14 th generation Dell EMC PowerEdge Servers)	20
VxRail node	21
Intel® Xeon® Scalable processor: Powerful processing for VxRail	23
VxRail node storage disk drives	23
VxRail hardware options	24
VxRail scaling	25
Upgradeable options	26
VxRail networking	26
1GbE network option	28
Dell EMC Open Networking & VxRail	29
Dell EMC SmartFabric Services (SFS)	29
VxRail software architecture	31
Appliance management	31
VxRail HCI System Software	31
VxRail Manager	34
Customer upgradeable software	36
vSphere and vSAN ordering information	37
VMware vSphere	39
VMware vCenter Server	39
vCenter services and interfaces	42
Enhanced Linked Mode	43
VMware vSphere ESXi	43
Communication between vCenter Server and ESXi hosts	44
Virtual machines	45
Virtual machine hardware	46
Virtual Machine Communication	46
Virtual networking	47

Virtual Distributed Switch.....	47
vMotion and Virtual Machine mobility	48
Enhanced vMotion Compatibility	50
Storage vMotion	50
vSphere Distributed Resource Scheduler	51
vSphere High Availability (HA)	53
vCenter Server Watchdog	55
vSphere Encryption	55
vSAN	56
Disk groups	57
Hybrid and All-Flash differences.....	58
Read cache: Basic function	58
Write cache: Basic function	58
Flash endurance	59
vSAN impact on flash endurance	59
Client cache	59
Objects and components	59
Witness	60
Replicas	60
Storage Policy Based Management (SPBM).....	61
Dynamic policy changes.....	61
Storage policy attributes	62
Sparse Swap	64
I/O paths and caching algorithms	65
Read caching.....	65
Anatomy of a hybrid read	66
Anatomy of an All-Flash read	67
Write caching	67
Anatomy of a write I/O—hybrid and All-Flash (FTM=mirroring)	68
Distributed caching considerations.....	69
vSAN high availability and fault domains.....	70
Fault domain overview.....	70
Fault domains and rack-level failures	71
Cautions when deploying a minimum cluster configuration	72
vSAN Stretched Cluster.....	72
Stretched Cluster with Local Protection	73
Site locality.....	74
Networking.....	74
Stretched cluster heartbeats and site bias	74
vSphere HA settings for stretched cluster	74
2-Node Configuration.....	75
Snapshots	75
Storage efficiency using deduplication and compression.....	76
Deduplication and compression overhead	78
Erasure coding.....	79
Enabling Erasure Coding.....	80
Erasure coding overhead	80
vSAN Encryption.....	81

VxRail integrated software	82
VM Replication.....	82
VMware vSphere Replication	82
Dell EMC RecoverPoint for Virtual Machines	83
VxRail replication use case.....	85
Support for external network storage	86
iSCSI with VxRail.....	86
NFS with VxRail.....	87
VxRail solutions and ecosystem	88
VMware Validated Design with VxRail	88
VMware Cloud Foundation on VxRail.....	90
Pivotal Ready Architecture (PRA)	91
Flexible consumption options	92
Cloud Flex for HCI - a cloud-like consumption option	92
VDI Complete	93
VMware Horizon	93
VMware Horizon with VxRail	94
VMware vSphere Platinum	95
IsilonSD Edge	95
SAP HANA Certification with VxRail.....	96
Reference Architecture for Splunk.....	97
Additional Product information	99
Dell EMC ProSupport for Enterprise.....	99
Dell EMC ProDeploy Services for VxRail Appliances	99

Preface

The Dell EMC TechBook is a conceptual and architectural review of the Dell EMC VxRail™ Appliance, powered by VMware vSAN with Intel Inside. The TechBook describes how hyperconverged infrastructure drives digital transformation and focuses on the VxRail Appliance as a leading hyperconverged technology solution.

Audience

This TechBook is intended for Dell EMC field personnel, partners, and customers involved in designing, acquiring, managing, or operating a VxRail Appliance solution.

Related resources and documentation

Refer to the following items for related, supplemental documentation, technical papers, and websites.

Dell EMC VxRail Network Guide: <https://vxrail.is/networkplanning>

VxRail Planning Guide for Virtual SAN Stretched Cluster:
<https://vxrail.is/stretchedclusterplanning>

An overview of VMware vSAN Caching Algorithms at
<https://www.vmware.com/files/pdf/products/vsan/vmware-virtual-san-caching-whitepaper.pdf>

VMware vSAN 6.2 Space Efficient Technologies Technical White Paper at
<http://www.vmware.com/files/pdf/products/vsan/vmware-vsan-62-space-efficiency-technologies.pdf>

VxRail Stretched Cluster at: <https://www.dell.com/resources/en-us/asset/offering-overview-documents/products/converged-infrastructure/vxrail-stretch-cluster-so.pdf>

vSphere Virtual Machine Administration Guide at
<https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-6-pubs.html>

vSphere Resource Management at <http://pubs.vmware.com/vsphere-65/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-65-resource-management-guide.pdf>

Dell EMC Hyperconverged infrastructure at <http://www.dell.com/en-us/work/shop/category/hyper-converged-infrastructure>

VxRail vCenter Server Planning Guide: <https://vxrail.is/vcenterplanning>

IT's new challenge: Leading digital transformation

In the digital economy, applications are both the face and the backbone of the modern enterprise.

For the digital customer, user experience trumps all. Customer-facing applications must be available anytime, anywhere and on any device, and must provide real-time updates and intelligent interactions. For the business, the insights gleaned from the data collected from these interactions inform and drive future development needs.

Applications and the underlying infrastructure are strategic to the business. Businesses that can efficiently leverage modern datacenter technologies to rapidly deliver innovative capabilities to customers are positioned for real success.

The importance of applications in the modern enterprise presents a huge opportunity for IT organizations. No longer simply a back-office function, IT can lead a digital transformation that positions the business for success moving forward. IT can become an active enabler of the business.

Traditional IT teams are faced with a massive amount of complexity when building, configuring, maintaining and scaling applications. Organizations need to successfully deploy and operate an environment that takes full advantage of the innovation taking place across the industry – without the complexity of piecing together and supporting a wide range of patchwork tools.

The challenge is how to go about this transformation. Dell EMC surveyed¹ over 1,000 executives across multiple industries about the state of their digital transformation efforts. Survey questions focused on:

Modern datacenter technology utilization, such as the use of All-Flash arrays, scale-out architectures, converged and hyper converged platforms, and software-defined solutions across networking and storage domains.

Automated IT processes, as measured by the progress the organization has made in terms of running IT more like a public cloud provider (enabling self-service infrastructure provisioning; rapid scalability; and usage-based tracking and chargeback).

Transformed business and IT relationships: enabled by consistent communication between IT and business stakeholders and continuous inspection of IT outcomes by line of business (LOB) leadership.

The findings show that progress has been, at best, mixed. Some companies have barely started their digital transformation. Many have taken a piecemeal approach. Only a small minority have almost completed their digital transformation. Why is it taking so long?

The bottom line is that IT transformation is difficult. It requires a great deal of planning, evaluation, re-organization and modernization of infrastructure technologies and applications. Multiple factors including costs, skill sets, governance, the drive to innovate and willingness to transform influence whether a business moves beyond the traditional three-tier datacenter structure.

1

How IT Transformation Maturity Drives IT Agility, Innovation, and Improved Business Outcomes, Enterprise Strategy Group, April 2017. <https://www.emc.com/collateral/analyst-reports/esg-dellemc-it-transformation-maturity-report.pdf>

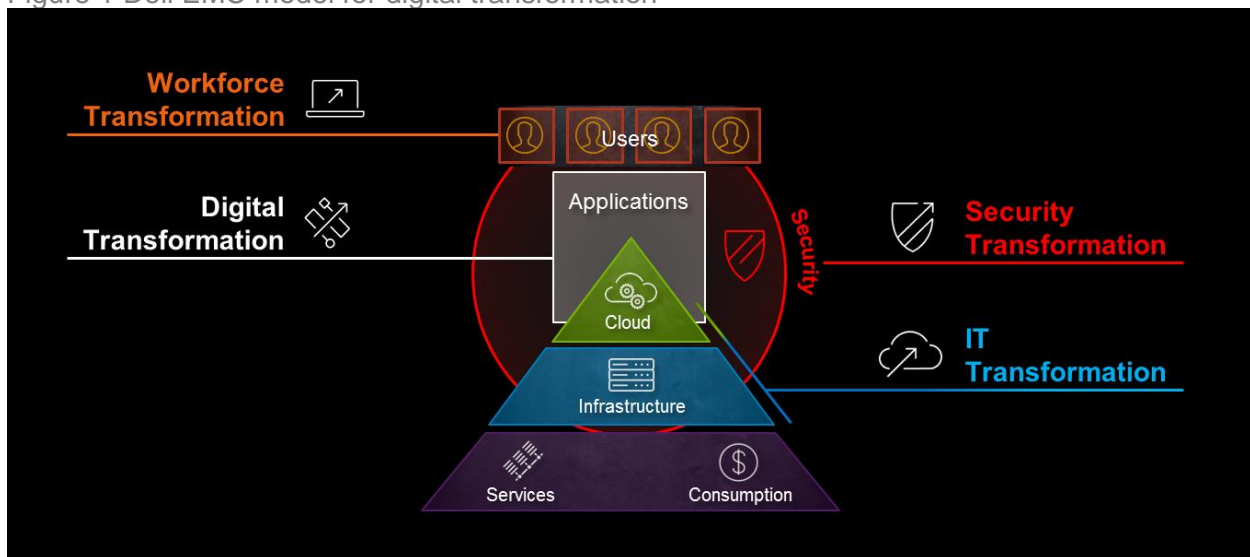
Every business approaches IT transformation at a different pace and has different goals for that transformation. Not every business wants or needs to go to a full cloud service delivery model. What is needed is an approach that enables businesses to transform to a place that provides the desired transformation benefits and at a pace that makes sense for their business model.

Dell EMC digital transformation: Faster outcomes. Simpler IT.

To have a complete and strategic view of transformation, you must start with an application centric point of view. Ensuring that applications have the right strategy for service level agreements whether on premises or off premises entails having a hybrid cloud strategy. Dell EMC leverages hyperconverged infrastructure (HCI) to deliver completely integrated and supported HCI solutions and hybrid cloud platforms that modernize, automate, and transform the enterprise datacenter and IT Transformation.

The following figure shows the Dell EMC model for digital transformation.

Figure 1 Dell EMC model for digital transformation



Dell EMC delivers fully engineered turnkey hyperconverged infrastructure solutions that enable businesses to innovate faster and accelerate IT operations.

Dell EMC converged and hyperconverged infrastructure delivers application-focused solutions built on best-of-breed hardware and software that provide real business value, while dramatically reducing the risk and cost of deploying mission critical, general purpose, and cloud native applications.

And finally, Dell EMC offers a full range of services and flexible consumption models to help make it faster and easier to consume these solutions.

Accelerating IT transformation with Dell EMC hyperconverged infrastructure

One of the first steps a business can take in their transformation journey is to simplify infrastructure deployment and management by introducing hyperconverged infrastructure (HCI) into the environment. HCI systems essentially collapse the traditional three-tier server, network, and storage model so that the infrastructure itself is much easier to manage.

Adopting hyperconverged infrastructure solutions that natively integrate compute, storage, virtualization, management and data services significantly reduces IT administrative tasks and create the foundation for a modern IT infrastructure. HCI solutions are optimal for reducing

infrastructure costs and simplifying management, regardless of workload deployment and extent of implementation.

Fully transforming to the hybrid cloud

Many businesses would ultimately like to automate IT service delivery through a self-service catalog via a hybrid cloud. The hybrid cloud delivers the following benefits:

- A single control point for on- and off-premises resources
- Automation streamlines delivery of IT resources, delivering them in a consistent and repeatable manner aligned with business best practices
- Metering allows the IT team to communicate the value of services while providing visibility to the business on resource cost and consumption
- Self-service empowers application owners and business users to access the resources they need, when they need them
- Capacity management allows the IT team to better manage resources across the hybrid cloud
- Monitoring and reporting provides visibility to the capacity, performance and health of the environment
- Built-in security protects enterprise workloads
- Service-level choice aligns workloads to service levels and cost objectives
- Ability to meet the service level agreements with application level granularity

The vision of hybrid clouds is not new. Businesses have tried to deploy hybrid clouds using traditional infrastructure based on scale-up storage accessed over a storage network that is deployed and scaled in big chunks. While it is possible to build cloud capabilities on traditional three-tier infrastructure with scale-up storage, this is not the optimal solution.

If businesses want full IT transformation to the cloud support their application environment, Dell EMC can modernize, automate, and transform IT operations with complete turnkey, hybrid cloud platforms built on hyperconverged infrastructure.

Innovate rather than integrate

Businesses do have the option of building a completely customized solution. Integrating storage, networking, compute, data protection, monitoring and reporting, and then figuring out how to get all of them to work together can be time consuming, but provides the most flexibility for an organization that may want prescribed vendor components as a part of their solution. Planning, designing and building a custom solution is a complex project that often takes months or years to come to fruition—too long if a business needs to roll out a solution to address immediate business needs and it can be costly to maintain or upgrade over the long term.

The challenge for IT is that complexity exists at each of these layers, so building and maintaining a functional, resilient cloud can be very difficult. Many companies find that doing it themselves requires more than 70% of their IT resources and budget, leaving few resources to focus on innovation and projects that add real value to the business.

For most businesses, the best way to consume HCI solutions is to buy them fully integrated with lifecycle management and single source of support. Buying versus building delivers the

best time-to-value, operational simplicity, and 5-year total cost of ownership savings of 619% over a traditional three-tier, build-your-own approach.²

2

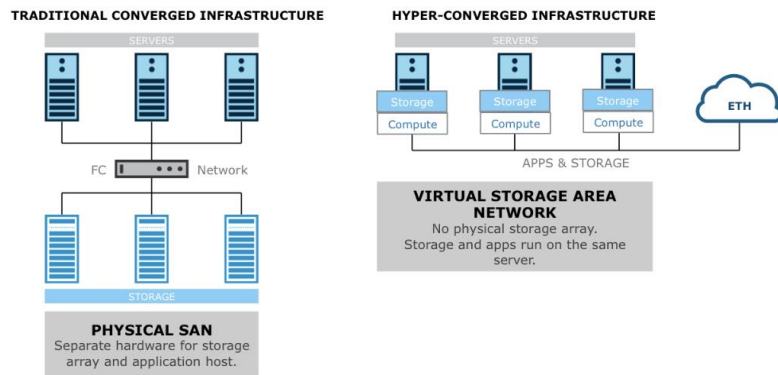
Source: IDC Oct 2017

Hyperconverged infrastructure: A modern infrastructure for modern IT challenges

Converged infrastructure platforms are fully pre-integrated server, traditional storage arrays, and networking hardware “stacks”. *Hyperconverged infrastructure* platforms are solutions that deliver compute, software-defined storage, and networking infrastructure services in a cluster of industry-standard servers.

Hyperconverged infrastructure extends the converged infrastructure model by incorporating the virtualization capabilities of software-defined storage (SDS). Hyperconverged infrastructure collapses the core components of traditional datacenter – compute and storage – into a server, effectively eliminating expensive and complex SAN environments. See the figures below.

Figure 2 CI and HCI



Because HCI is software-defined – which means the infrastructure operations are logically separated from the physical hardware – the integration between components is much tighter than with CI. HCI manages everything as a single system through a common toolset.

Hyperconverged infrastructure is particularly valuable because it lets you scale up quickly without a ton of added expense. That is not the case in traditional settings: customers either must buy more resources than they need in anticipation of scaling up, or wait until current workloads exhaust the allocated resources, then add infrastructure after the fact. Buying at the inopportune time means that resources are not optimally allocated and can even slow down your business from expanding.

HCI enables a pay-as-you-grow approach – start with what is needed today and expand incrementally rather than purchasing large amount of compute and storage up front. It also addresses the typical over-

HCI deployment models

Dell EMC identifies four deployment models for IT transformation, defined by the business' desired end state and operational readiness to realize that end state.

Build your own

Businesses wanting the benefits of HCI but prefer to maintain control & flexibility regarding server vendor and configuration choice.

HCI appliances

Appliances are a more turnkey option, including the server, software stack, and lifecycle management. They are fully engineered and supported as a single product.

Rack-scale HCI

Rack scale HCI solutions extend what is provided by HCI appliances, and additionally, tightly integrates physical and software defined networking along with software defined storage, compute, virtualization, data protection, and management software to fully automate complete infrastructure configuration and provisioning through a software-defined management and orchestration layer. Like HCI appliances they are engineered and supported as a single product.

Turnkey HCI-based hybrid cloud

Combine either appliance or rack-scale HCI with additional cloud management, automation & orchestration software and pre-engineered services & workflows to deliver an agile, elastic, automated, self-sustaining private cloud that can integrate with off-premises public cloud providers.

provisioning and over-purchasing that occurs when technology is intended to last for multiple year cycles.

Enabling technologies for HCI

The following table lists the confluence of technologies that has spurred the growth and development of hyperconverged infrastructure.

Table 1. Enabling technologies for HCI

Technology	Description
Software defined storage	<p>Abstracts the storage intelligence from the underlying storage infrastructure.</p> <p>Virtualizes direct-attach storage into a shared pool.</p> <p>Automates provisioning and load balancing.</p> <p>Allows a business to increase available storage resources, both capacity and processing power, by adding entire nodes (e.g., a server with storage software and media) to a cluster. The resulting cluster of nodes in turn acts as a single pool of storage capacity.</p>
Virtualization	<p>Abstracts compute and network functions.</p> <p>Enables physical resources to be shared.</p> <p>Improves utilization, mobility and security.</p>
X86 servers	<p>High performance processors, large memory.</p> <p>Flash media delivers consistent, predictable performance.</p>
Solid state storage	<p>Uses solid-state drives (most frequently various types of flash memory) to store data. This storage can reside in a storage controller or in a server, but for this assessment we are considering use cases limited to tiered and All-Flash storage arrays.</p> <p>In hybrid arrays, a portion of the drives in the array are solid-state and house the most active data on the array.</p> <p>In All-Flash arrays, all drives in the array are solid-state.</p>
High-speed networks	<p>Connects nodes together to create cluster.</p> <p>Enables HCI to deliver IOPS and reduced latencies.</p> <p>Connect applications to users</p>

Drivers for hyperconverged infrastructure adoption

Customers that have transitioned or plan to transition to HCI state cost reduction, accelerated deployment, improved ability to scale, improved operational efficiencies and reduction in infrastructure tasks as top benefits they expect to realize when implementing HCI.

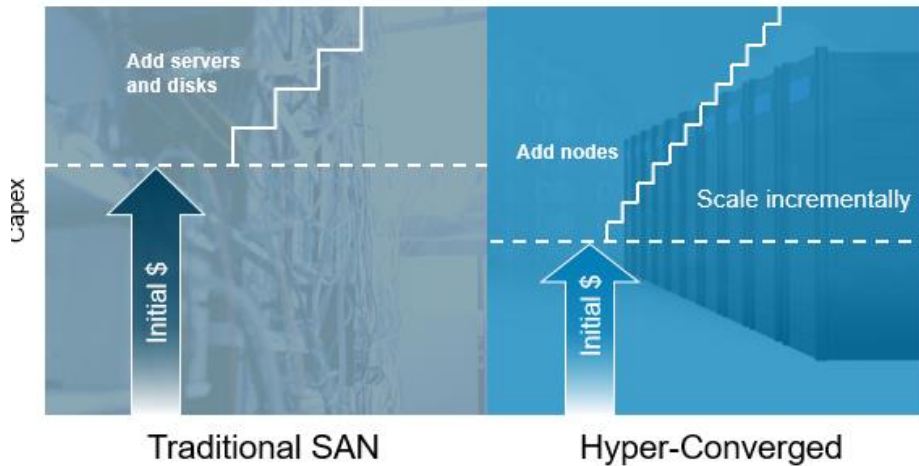
HCI delivers a compelling story for both CapEx and OpEx.

Savings in initial investments are lower, and operational expenses are also lower when compared to traditional three-tier architectures. Cost savings include power and cooling, ongoing system administration, and the elimination of disruptive upgrades and data migrations.

Rather than buying monolithic SAN-based infrastructure, a business can buy infrastructure that targeted for specific workloads. A main contributor to lower TCO and the increased agility of hyperconverged solutions is the ability start smaller and scale incrementally.

Not only is initial CapEx investment lower, but you can also scale them more incrementally, adding smaller amounts of compute or even expanding just the storage capacity as required, as shown in the figure below.

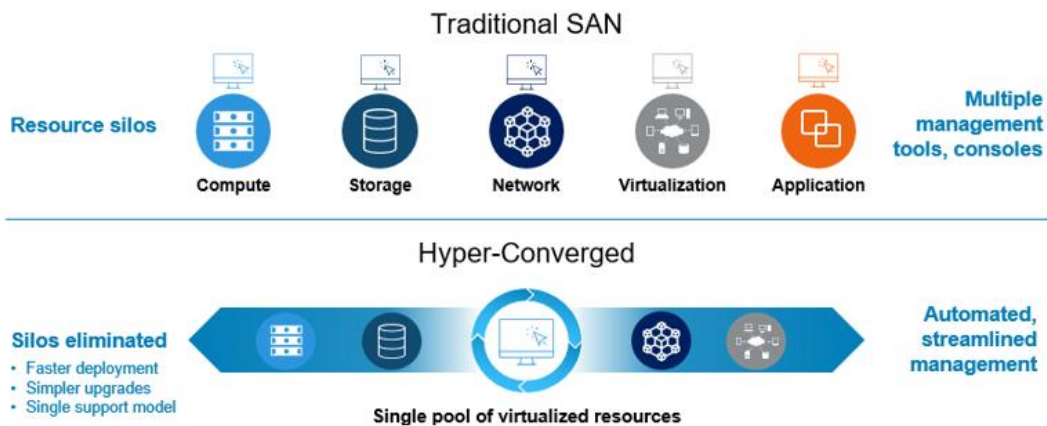
Figure 3 HCI: Buy what you need, pay as you go



Further, some HCI appliances support multiple nodes in a single chassis. So, if space is available, you can simply plug a node into an existing chassis in a matter of minutes to scale your infrastructure. Deploying HCI reduces time, risk and complexity.

The following figure shows how HCI simplifies IT management and operations as compared to a traditional SAN environment.

Figure 4 HCI simplifies IT management and operations



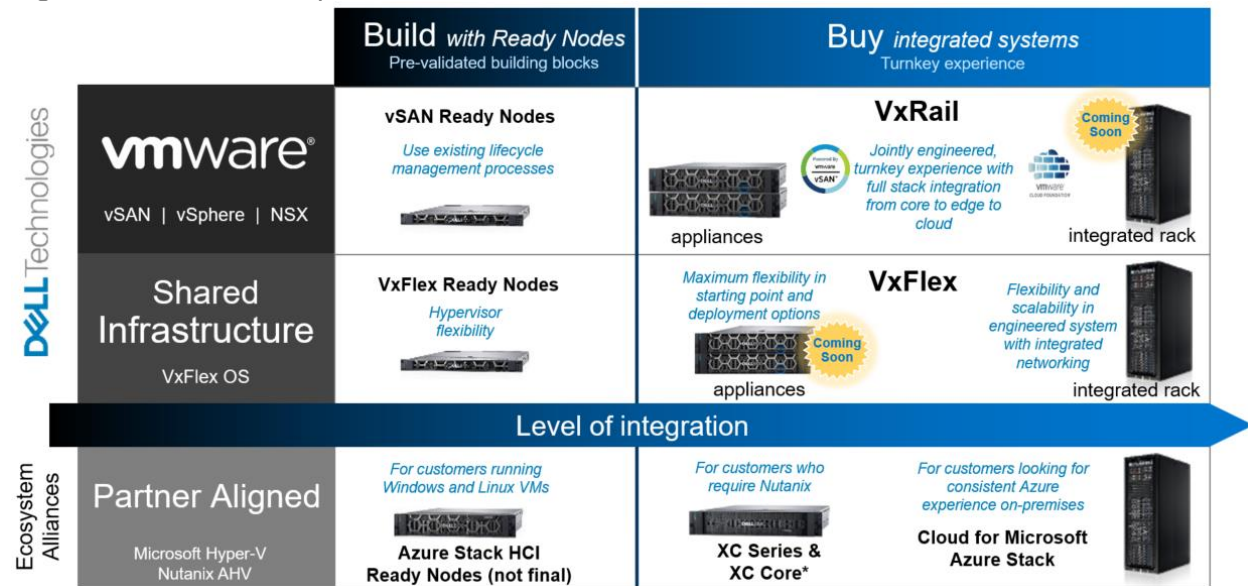
The most compelling HCI solutions leverage management frameworks that drastically improve operational tasks, reduce the burdens of lifecycle management, and improve the level of responsiveness of IT staff to meet the needs of their business. Ideally, existing investments in management and orchestration stacks can be leveraged as well.

Dell EMC hyperconverged infrastructure platforms

Dell EMC provides organizations the flexibility to choose HCI solutions that best fit their current state of IT transformation while ensuring IT certainty, continuous innovation and predictable evolution as they move toward cloud implementations. Whether a business is modernizing existing applications or deploying turnkey engineered solutions, the Dell EMC HCI portfolio delivers the power, simplicity, and certainty a business needs for the next phase of digital transformation.

The following figure shows the Dell EMC HCI portfolio, powered by Intel®.

Figure 5 Dell EMC HCI portfolio



The Dell EMC hyperconverged infrastructure portfolio includes both appliance and rack-scale offerings and differentiates with both fully-integrated VMware-based solutions or opportunities to get turnkey outcomes through hypervisor choice, bare metal or multi-hypervisor options.

Appliances accelerate the transformation of both the compute and storage layers for customers' datacenters by delivering turnkey outcomes on all-flash, software-defined, scale-out architectures. Rack-scale offerings offer additional transformation for those customers who are ready to fully modernize their datacenter by adopting software defined networking as well as compute and storage in a fully integrated, turnkey fashion.

Building upon the bedrock of the modern datacenter, Dell EMC delivers the #1 hyperconverged infrastructure portfolio purpose-built for HCI with the newest 14th-generation Dell EMC PowerEdge server platform. This portfolio delivers tailor-made performance and reliability powerful enough for any workload, combined with a polished approach to intelligent deployment and operations that simplifies and accelerates IT. Dell EMC HCI on latest generation PowerEdge servers are powerful, purposeful, and polished hyperconverged platforms that provide the ideal foundation for software-defined datacenter initiatives.

Customers demand more performance and reliability as HCI moves into the core datacenter where it must run a wider range of applications and workloads. With improved performance (double the IOPS), better economics (up to 3x more VDI users per node), and flexibility (more

configurability to meet more use cases), Dell EMC HCI on 14th Generation PowerEdge servers has enough power for any workload.

Dell EMC PowerEdge servers are purpose-built for HCI. With up to 150 customer HCI requirements built-in, PowerEdge servers are designed specifically for and tailored to HCI workloads that depend on the tenets of both servers and storage. This results in a more consistent, predictable, and reliable high-performing HCI that can meet any use case. With our comprehensive portfolio, Dell EMC can deliver the best fit for your customers' specific HCI needs – from workload requirements, to customer environment/standardization, to deployment preferences.

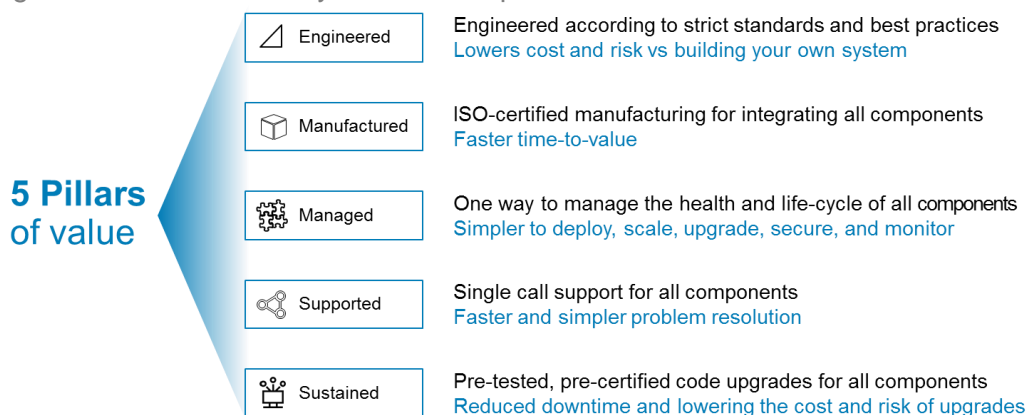
Dell EMC leads in hyperconverged sales with over 28% market share according to IDC³. More customers are choosing Dell EMC HCI over all others. Dell EMC PowerEdge is the world's best-selling server. Industry-leading Dell EMC HCI built on industry-leading PowerEdge, coupled with a single point of support and full lifecycle management for the entire system, makes for a winning solution.

Dell EMC HCI delivers a turnkey customer experience

Dell EMC has invested thousands of engineering hours designing; integrating and testing its hyperconverged solutions to make sure all components are hardened, work together and are sustained as one. As a result, Dell EMC HCI solutions can be implemented in weeks rather than months.

The following figure identifies the key aspects of the Dell EMC HCI turnkey experience.

Figure 6 Dell EMC turnkey customer experience



Dell EMC is constantly innovating and delivering new capabilities to ensure its HCI solution continues to evolve to meet new customer requirements. New capabilities are added as part of the solution release cycles with Dell EMC testing the entire solution end-to-end, including upgrades from previous versions. In addition, major upgrades are supported by Dell EMC Global Services across the entire solution stack eliminating interoperability management concerns for IT Ops teams.

³ Based on IDC Converged Tracker Q4 2018

Dell EMC VxRail Appliances

VxRail Appliances are jointly developed by Dell EMC and VMware and are the only fully integrated, preconfigured, and tested HCI appliance powered by VMware vSAN technology for software-defined storage. Managed through the ubiquitous VMware vCenter Server interface, VxRail provides a familiar vSphere experience that enables streamlined deployment and the ability to extend the use of existing IT tools and processes.

VxRail Appliances are fully loaded with integrated, mission-critical data services from Dell EMC and VMware including compression, deduplication, replication, and backup. VxRail delivers resiliency and centralized-management functionality enabling faster, better, and simpler



VxRail essentials

Fully integrated, preconfigured, and tested hyperconverged infrastructure appliance simplifies and extends VMware environments

Seamlessly integrates with existing VMware eco-system management solutions for streamlined deployment and management in VMware environments.

Start small, with a few as three nodes. Single node scaling, storage capacity expansion, and vSphere license independence enable growth that meets business demands.

Backup distributed applications or workloads with integrated data protection options, including RecoverPoint for VMs.

Single point of global 24x7 support for both the hardware and software

management of consolidated workloads, virtual desktops, business-critical applications, and remote-office infrastructure. As the exclusive hyperconverged infrastructure appliance from Dell EMC and VMware, VxRail is the easiest and fastest way to stand up a fully virtualized VMware environment.

VxRail is the only HCI appliance on the market that fully integrates Intel-based Dell EMC PowerEdge Servers with VMware vSphere, and vSAN. VxRail is jointly engineered with VMware and supported as a single product, delivered by Dell EMC. VxRail seamlessly integrates with existing (and optional) VMware eco-system and cloud management solutions, including vRealize, NSX, Horizon, Platinum and any solution that is a part of the vast and robust vSphere ecosystem.

VxRail provides an entry point to the software defined datacenter (SDDC) for most workloads. Customers of all sizes and types can benefit from VxRail, including small- and medium-sized environments, remote and branch offices (ROBO), and edge departments, as well as providing a solid infrastructure foundation for larger datacenters.

Small-shop IT personnel benefit from the simplicity of the appliance model to expedite the application-deployment process while still taking advantage of data services only typically available in high-end systems.

Larger datacenters benefit by rapid deployment where a complete vSphere environment can be installed and be ready to deploy applications within few hours of the system arriving on site. VxRail

allows businesses to start small and scale non-disruptively. Storage is configured to meet appropriate application capacity and performance requirements.

In addition, nodes are available with different compute power, memory, and cache configurations to closely match the requirements of new and expanding use cases. As

requirements grow, the system easily scales out and scales up in granular increments. Finally, because the VxRail is jointly engineered, integrated, and tested, organizations can leverage a single source of support and remote services from Dell EMC.

VxRail environments are configured as a cluster consisting of a minimum of two server nodes, with each node containing internal storage drives. VxRail systems are delivered with the software loaded, ready to attach to a customer-provided network. While most environments use 10Gb Ethernet for internal and external communications, 25Gb or 1Gb Ethernet connectivity is also available. Using a simple wizard at the time of install, the system can be configured to match unique site and networking requirements.

VxRail Appliances enable organizations to start small and scale out as the IT organization transforms and adapts to managing converged infrastructure versus silos. With a rich set of data services, including data protection, tiering to the cloud, and active-active datacenter support, VxRail can be the foundational infrastructure for IT. Best of all, you can simply add new appliances into existing clusters (and decommission aging appliances) to provide an evergreen HCI environment, never having to worry about costly SAN data migrations ever again. As organizations continue to transform to a cloud model, integration with the VMware vRealize Suite enables full cloud automation and service delivery capabilities.

Dell EMC VxRail Appliances offer a choice of Dell EMC PowerEdge servers, powered by new Intel® Scalable® processors, variable RAM, and storage capacity, allowing customers to buy what they need now. The VxRail Appliance uses a modular, distributed system architecture that starts with as few as two nodes and scales near linearly up to 64 nodes. Single-node scaling and storage capacity expansion provide a predictable, “pay-as-you-grow” approach for future scale up and out as business and user requirements evolve.

The VxRail software layers use VMware technology for server virtualization and software-defined storage. VxRail nodes are configured as ESXi hosts, and VMs and services communicate using the virtual switches for logical networking. VMware vSAN technology, implemented at the ESXi-kernel level, pools storage resources. This highly efficient SDS layer consumes minimal system resources, making more resources available to support user workloads. The kernel-level integration also dramatically reduces the complexities involved in infrastructure management. vSAN presents a familiar datastore to the nodes in the cluster and Storage Policy Based Management provides the flexibility to easily configure the appropriate level of service for each VM.

VxRail HCI System Software, the VxRail management platform, is a strategic advantage for VxRail and further reduces operational complexity. VxRail HCI System Software provides out-of-the-box automation and orchestration for day 0 to day 2 appliance-based operational tasks, which reduces the overall IT OpEx required to manage the stack. No build-it-yourself HCI solution provides this level of lifecycle management, automation, and operational simplicity. With VxRail HCI System Software, upgrades are simple and automated with a single-click. You can sit back and relax knowing you are going from one known good state to the next, inclusive of all the managed software and hardware component firmware. No longer do you need to verify hardware compatibility lists, run test and development scenarios, sequence and trial upgrades, and so on. The heavy lifting of sustaining and lifecycle management is already done for you.

Within the VxRail HCI System Software, the VxRail Manager plugin presents a simple integrated dashboard interface on vCenter Server for infrastructure monitoring and automation of lifecycle management tasks such as software upgrades. Since VxRail nodes function as ESXi hosts, vCenter Server is used for VM-related management, automation, monitoring, and security.

VxRail Appliances are powered by VMware vSAN software, which is fully integrated in the kernel of vSphere and provides full-featured and cost-effective software-defined storage. vSAN implements an efficient architecture, built directly into hypervisor. This distinguishes vSAN from solutions that typically install a virtual storage appliance (VSA) that runs as a guest VM on each host. Embedding vSAN into the ESXi kernel layer has advantages in performance and memory requirements. It has little impact on CPU utilization (less than 10 percent) and self-balances based on workload and resource availability. It presents storage as a familiar data store construct and works seamlessly with other vSphere features such as VMware vSphere vMotion.

vSphere is a well-established virtualization platform—a familiar usable entity in most datacenters. Dell EMC leverages vSphere for ESXi-based virtualization and VM networking in multiple product offerings, and they support a common set of VMware and Dell EMC services. This enables a VxRail implementation to integrate smoothly into VMware-centric datacenters and to operate in concert with Dell EMC converged, hyperconverged, and traditional storage offerings. NSX for SDN can optionally be added to the VxRail solution. NSX for SDN allows all datacenter assets to be maintained using a single administrative platform. Monitoring, upgrading, and diagnostics activities are performed efficiently and reliably.

Additional data services integrated into VxRail include RecoverPoint for VM replication and Dell EMC Remote Secure Services (ESRS).

To summarize, VxRail creates IT certainty. It is built on the enterprise-proven VMware hypervisor, powered by vSAN technology, delivers simplified lifecycle and daily operational management, and supports the dynamic and robust vSphere ecosystem. The system architecture is designed to predictably evolve with the business. The system scales by deploying preloaded appliances to effortlessly extend the environment, rather than by building and manually integrating servers.

VxRail, powered by vSAN and featuring Intel-based Dell EMC PowerEdge Servers, continues to evolve so business can thrive without worrying about the IT infrastructure. VxRail is a fully integrated, preconfigured, and pre-tested VMware hyperconverged infrastructure appliance family. Based on VMware's vSphere and powered by vSAN and Dell EMC software, the VxRail Appliance provides the IT infrastructure foundation for digital transformation that empowers organizations to continuously innovate.

VxRail Security and Compliance

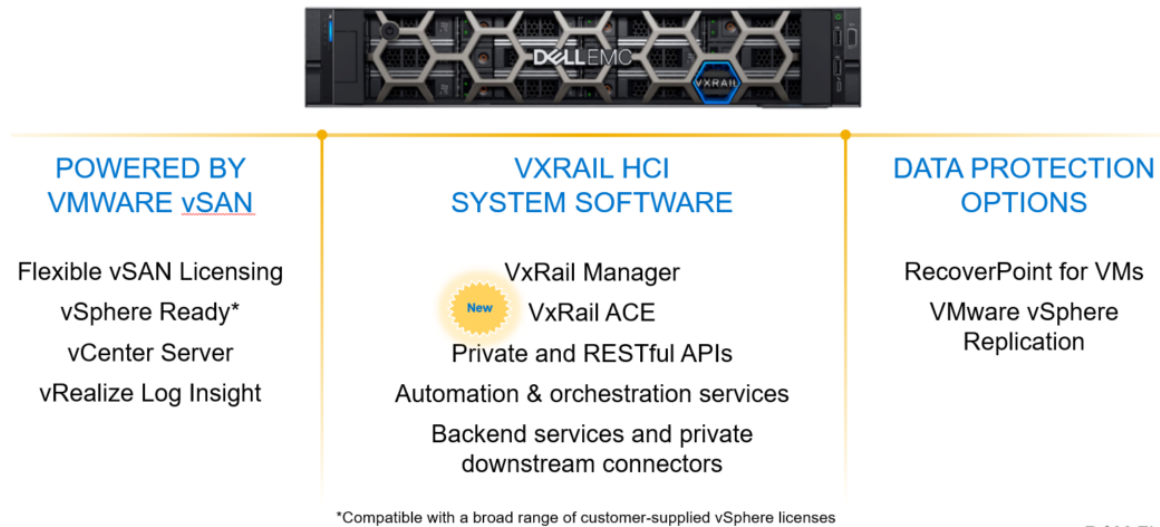
Dell EMC VxRail Appliance is a resilient, secure, and modern hyperconverged infrastructure system that directly addresses the challenges of security and compliance in modern day environments.

VxRail Appliance is engineered, built, configured, and maintained following the Dell EMC Secure Development Lifecycle, which follows a rigorous approach to secure product development, including executive-level risk management before products are shipped to market. Additionally, VMware vSphere—a significant part of VxRail hyperconverged infrastructure—has also been developed using a similar Security Development Lifecycle.

Everything that comprises VxRail is secure and can be seen in the figure below. Each component has security built in, with corporate security processes, unique security features, and supply chain control, so you can feel confident that VxRail can fit into your secure IT infrastructure design. The hardware is comprised of Dell EMC PowerEdge servers and Intel processors. The virtualization and software layers are comprised of vSphere and vSAN which is integrated into the kernel of vSphere. The integrated software and management included with

VxRail is comprised of software from VMware, the vRealize Log Insight and vCenter, and software from Dell EMC, RP4VM, ESRS, and the VxRail HCI System Software for all the lifecycle management of everything in this stack. (RP4VM excluded from LCM.) All of this is jointly engineered with Dell EMC and VMware, and delivered by and supported exclusively by Dell EMC as a single product—VxRail.

Figure 7 What comes with VxRail



VxRail is designed to a number of standards, has attained the Common Criteria EAL2+certificate, USGv6 certification making it IPv6 Ready, and provides a VxRail Product Security Configuration Guide to further harden VxRail deployments. Additionally, customers can leverage the VxRail STIG Compliance Guide and automated scripts to further harden their environments.

To learn more about VxRail’s Comprehensive Security by Design, please download the whitepaper: <https://www.emc.com/collateral/white-papers/vxrail-comprehensive-security-design.pdf>

VxRail hardware architecture

The Dell EMC VxRail Appliance family is the standard in hyperconverged infrastructure, providing extreme flexibility to granularly add capacity and performance on demand and enabling customers to easily extend use cases across the VMware virtualized environment. The appliance-based design allows IT centers to scale capacity and performance non-disruptively, so they can start small and grow incrementally with minimal up-front planning. VxRail environments can be designed to support a small number of virtual machines and scale to thousands.

The VxRail architecture enables a predictable pay-as-you-grow approach that aligns to changing business goals and user demand. Dell EMC and VMware are continuously innovating, and VxRail introduced new Dell EMC PowerEdge-based models that offer extreme configuration flexibility. This flexibility allows customers to choose performance, graphics, and capacity as required for VMware environments, and supports more use cases.

The Dell EMC VxRail family of appliances offers a range of platforms:

G Series—general-purpose and compute dense, multi-node form factor, ideal for widely deployed general purpose applications and VDI workloads (not requiring GPU cards)

E Series—everywhere from datacenter core to edge deployments, the combination of density, drive groups, and balance of resources in a low profile 1U form factor enables it to be deployed for a wide range of use cases.

P Series—performance optimized for high-end use cases with business-critical, performance-intensive applications and/or in-memory databases.

V Series—VDI optimized for specialized use cases with graphics intensive applications such as high-end 2D/3D visualization applications; the only series that supports GPU cards.

S Series—storage-dense configurations targeted at specialized use cases that require higher storage capacity at the server level such as Big Data, analytics, or collaboration applications.

The E, P, V, and S Series are single-node appliances based on Dell EMC PowerEdge server technology, the number-one selling X86 server platform, with greater storage capacity, larger memory, and more powerful CPU options. The G Series a four-node appliance in a compact 2U chassis, providing a compute dense footprint.

VxRail Appliances are built using a distributed-cluster architecture consisting of modular blocks that scale linearly as the system grows from as small as three nodes to as large as 64 nodes. Nodes are available with different form factors, with single-node appliances for use cases: low-profile systems; performance optimized; VDI optimized with GPU; and storage-optimized configurations supporting high-capacity HDD drives.

Extensive compute, memory, and storage options are designed to fit multiple use cases. Customers can choose from a range of next-gen Intel processors, variable memory sizes, storage, and cache capacity to provide the right balance of compute, memory, and storage. Single-node scaling and a low-cost entry point let customers buy just the right amount of storage and compute for today's requirements and effortlessly scale to accommodate tomorrow's growth. Systems are available with all-flash storage configurations that deliver the industry's most powerful HCI for applications that demand maximum performance and low latency.

VxRail Appliance cluster

VxRail nodes are enclosed in a one-node, single server appliance, with each node having one or two multi-core processors and either all-flash solid state disks (SSDs) or a hybrid mix of flash SSDs and hard disk drives (HDDs). The nodes form a networked cluster with a minimum of three nodes and a maximum of 64. Nodes within a cluster must be of the same storage configuration, either all hybrid or all-flash. The flexibility to mix nodes within a cluster is supported. The first three nodes must have the same compute, memory, and storage configuration, and mixing 1GbE, 10GbE, and 25GbE is not supported. From the minimum configuration to the maximum, the VxRail cluster is easily expanded one node at-a-time.

Appliance models support either 25GbE, 10GbE or 1GbE network. 10Gb and 25Gb Ethernet networks are required for all-flash configurations and environments that will scale to more than eight nodes. Additional ports are available, allowing the customer to expand VM-network traffic.

VxRail models and specifications (based on 14th generation Dell EMC PowerEdge Servers)

VxRail Appliances built on the new 14th Generation Dell EMC PowerEdge server platform deliver the performance and reliability your customers need for the widest range of workloads, all with full lifecycle management from a single point of support. In short, VxRail is the fastest and easiest way to transform infrastructure. It takes a lot of work and expertise to engineer a high performance and reliable HCI solution, and the work does not stop after the initial deployment. Continuous validation is needed to keep it running smoothly through software upgrades and node additions. As a turnkey, pre-integrated, tested, and validated HCI solution, VxRail can be quickly deployed, easily distributed, and counted on to increase the predictability, availability, and performance of your IT environment.

VxRail Appliances on next generation servers include multiple purpose-built platforms with build-to-order configurations that support a wide range of customer use cases, including graphics-intensive VDI, big data and analytics, high performance computing, remote office, and more. With more processor options, new SATA SSDs, more additional network connectivity options, and more GPU expansion, you can now more closely match a VxRail to your workload requirements. No over provisioning here: buy what is needed, when it is needed.

VxRail Appliance models are available to meet the requirements of a wide set of use cases. For smaller workloads, there is a low-profile system space efficient configuration that uses 1U single-node appliances. A performance-optimized and a VDI-optimized configuration is available in both all-flash and hybrid configurations. For use cases requiring even greater storage, a hybrid storage-dense configuration that uses larger-capacity 3.5-inch drives is available. All models have a wide range of available memory, SSD cache, capacity storage configuration options and can start as with as few as two nodes.

The VxRail on 14th generation PowerEdge servers is now more powerful and predictable than the previous generation with utmost flexibility to meet any use case and more demanding workloads in VMware environments.

Double the IOPS, half the response times – VxRail is powerful enough for anything






Sustained response time less than 1ms – VxRail delivers highly predictable performance

Dell EMC offers the world’s most configurable HCI appliances – VxRail perfectly matches any HCI requirements



The following figure shows the range of platforms designed to support multiple use cases.

Figure 8 VxRail based on 14th generation Dell EMC PowerEdge servers

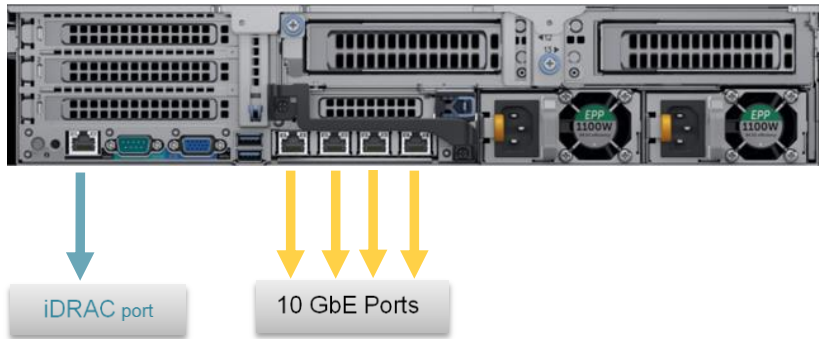
G Series Nodes	E Series Nodes	P Series Nodes	V Series Nodes	S Series Nodes
				
Compute dense	Low profile	Performance optimized	VDI optimized	Storage dense
G560/F	E560/F	P570/F	V570/F	S570
Full platform refresh to Dell EMC 14G PowerEdge servers based on Intel® Xeon® Processor Scalable Family Higher core counts, faster clock frequency, more memory channels, faster memory, higher endurance and redundant boot devices				
2000W or 2400W PSU Single or Dual socket 2x10GbE SFP+ onboard NVMe cache support	1100W PSU 4x10GbE onboard NVMe cache support 25GbE support*	1100W or 1600W PSU 20 capacity drives 4x10GbE onboard NVMe cache support 25GbE support	2000W PSU Up to 3 GPUs 8 more capacity drives 4x10GbE onboard 25GbE support	1100W PSU 4x10GbE onboard 25GbE support

VxRail node

The VxRail Appliance is assembled with proven server-node hardware that has been integrated, tested, and validated as a complete solution by Dell EMC. All the nodes in the current generation of VxRail use Intel Xeon Scalable family processors. The Intel Xeon Scalable family processors feature a multi-threaded, multi-core CPU designed to handle diverse workloads for cloud services, high-performance computing, and networking. The number of cores and memory capacity differ for each VxRail Appliance model.

The figure below shows a physical view of a node server with its processors, memory and supporting components. All VxRail models have similar components but may be physically laid out differently.

Figure 9 VxRail P Series node server: back view



Each node server includes the following technology:

One or two Intel Xeon Scalable processors with up to 28 cores per processor

Up to 24 DDR4 DIMMs, providing memory capacity ranging from 64GB to 3072GB per node, depending on model

A PCIe SAS disk-drive controller supporting 12GB SAS speeds

A mirrored pair of BOSS 240GB SATA M.2 cards used to boot ESXi on the node

Four port 10GbE Network Daughter Card (auto-negotiable to 1GbE)

iDRAC port

Intel® Xeon® Scalable processor: Powerful processing for VxRail

Intel® Xeon® Scalable platforms are powerful infrastructure that represents an evolutionary leap forward in agility and scalability. Disruptive by design, it sets a new benchmark in platform convergence and capabilities across compute, storage, memory, network and security. An innovative approach to platform design in Intel® Xeon® Scalable processors unlocks the power of scalable performance for today's datacenters and communications networks—from the smallest workloads to your most mission-critical applications.

With up to 28 cores delivering highly enhanced per core performance, and significant increases in memory bandwidth (six memory channels) and I/O bandwidth (48 PCIe lanes), your most data-hungry, latency-sensitive applications such as in-memory databases and high-performance computing will see notable improvements enabled by denser compute and faster access to large data volumes. And the latest generation processors designated with an 'M' can support denser memory, with up to 1536GB per processor.

The convergence of compute, memory, network, and storage performance combined with software ecosystem optimizations make Intel® Xeon® Scalable platforms ideal for fully virtualized, software-defined datacenters that dynamically self-provision resources—on-premises, through the network, and in the public cloud—based on workload needs.

VxRail node storage disk drives

Storage capacity for the VxRail Appliance is provided by disk drives that have been integrated, tested, and validated by Dell EMC. Most VxRail configurations use 2.5" form-factor SSDs and mechanical HDDs, and a configuration that uses 3.5" form-factor drives is also available for dense-storage requirements. Disk drives are logically organized into disk groups. Disk groups are configured in two ways:



Intel Inside®. Trusted clouds outside.

Intel innovation is driving the modernization and hybrid cloud transformation of the traditional enterprise datacenter.

Migrating to the newest generation of high-performing and energy-efficient Intel-based hardware tunes a datacenter for highly optimized performance across a broad set of enterprise workloads while lowering costs and improving resource utilization.

Over time, evolving to a software-defined infrastructure (SDI) across all the critical domains of the datacenter (compute/storage/network) will deliver critical automation, orchestration and telemetry capabilities to help businesses unlock the full capabilities of multi-cloud computing.

With modern, industry-standard Intel® servers and technologies that run on software-defined infrastructure, you can seamlessly manage an environment that supports development and delivery of cloud-native applications and mission-critical workloads on secure private clouds, while also integrating with public clouds, many of which already run on Intel® architecture.

Hybrid configurations, which contain a single SSD flash-based disk for caching (the cache tier) and multiple HDD disks for capacity (the capacity tier)

All-flash configurations, which contain all SAS SSD or NVMe drives for both cache and SAS or SATA SSD for capacity. NVMe is supported in dual processor configurations only.

The flash drives used for caching and capacity have different endurance levels. Endurance level refers to the number of times that an entire flash disk can be written every day for a five-year period before it has to be replaced. A higher-endurance SSD is used for write caching, and capacity-optimized SSDs are used for capacity. Currently, the 400GB, 800GB, and 1600GB SAS SSD or 1600 GB NVMe are used for caching, and for capacity either 1.92 or 3.84TB flash SSDs, 1.2, 1.8, and 2.4 TB 10K HDDs, 2TB (7.2K) HDDs, and 4TB 7.2K (3.5" form-factor) are used. All VxRail disk configurations use a carefully designed cache-to-capacity ratio to ensure consistent performance. Capacity SSDs are offered in both higher endurance SAS and SATA. The SATA SSDs are a lower cost option, up to 30% per drive, and great for read or moderately intensive workloads. The figure below depicts the wide-ranging set of components available across the VxRail Appliance line-up.

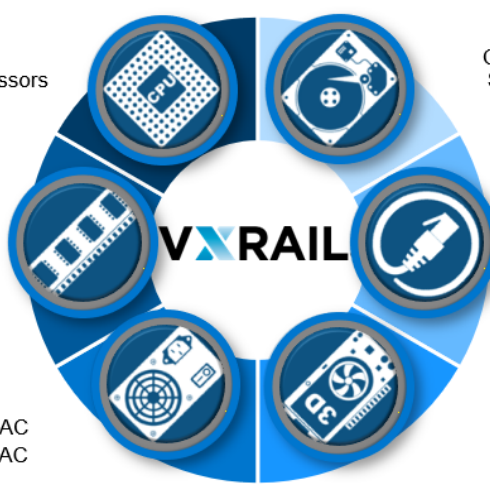
VxRail hardware options

VxRail nodes can be configured with choice of processor, memory, storage (cache and capacity drives), based networking, power supply, and GPU (in the V Series only). The figure below shows the comprehensive set of options available across the family. Customers can be assured their VxRail is configured to best match their workload requirements in a very prescriptive manner, with millions of possible configuration combinations in the VxRail Appliance Series. Combine this with the numerous ways to scale on demand, and it is clear that VxRail provides the agility demanded by today's modern IT. Some of the upgrade options for VxRail include, memory, GPU, NIC cards, cache SSD, and capacity drives.

Figure 10 A representative set of component options available across the VxRail series

VxRail Configuration Flexibility for Your Workload

G, E, P, S, V Series based on the latest Dell EMC PowerEdge servers



Processor
Choice of 40 Intel® Scalable® processors
4 to 56 cores per system

RAM
24 DIMM slots
16GB RDIMM
32GB RDIMM
64GB LRDIMM
128GB LRDIMM

Power supply

1100W	100-240V AC
1600W, 2000W, 2400W	200-240V AC
1100W	48V DC

Storage

NVMe Cache Drives: 1600GB
Cache SSDs: 400GB, 800GB, 1600GB
SSDs (SAS & SATA): 1.92TB, 3.84TB
HDDs: 1.2TB, 1.8TB, 2.4TB (10K)
2.0TB 4.0TB (7.2K)

Base networking

SFP28, SFP+, RJ45

2x 25GbE
4x 10GbE
2x 10GbE
4x 1GbE (4x 10GbE auto-negotiate)
Optional add-on NICs

GPUs

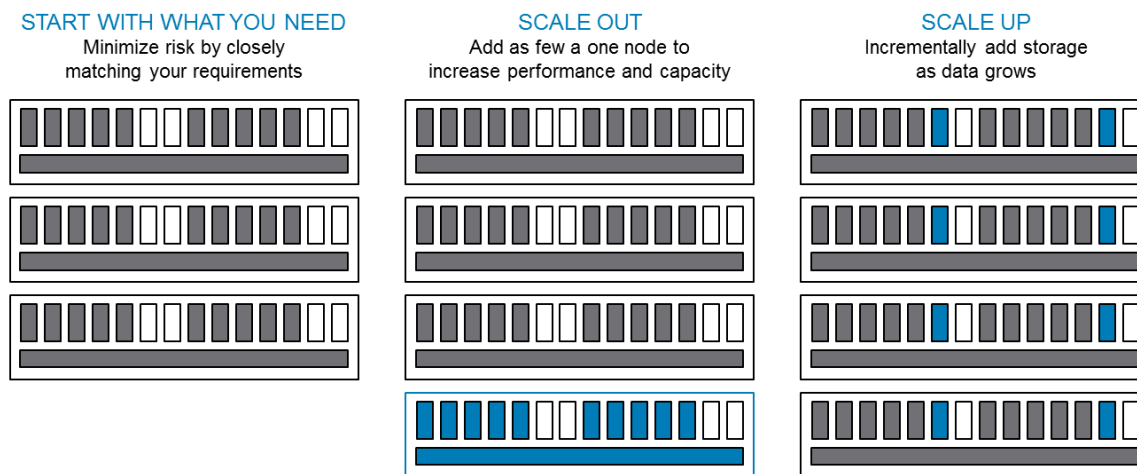
NVIDIA

Note: GPU SW & drivers sold separately

VxRail scaling

VxRail clusters start with as few as three nodes and can grow in one-node increments up to 64 nodes, providing performance and capacity to meet a wide range of use cases. Two node VxRail clusters are supported but cannot be expanded at this time. New appliances can be added non-disruptively and different models can be mixed within a VxRail cluster. Flexible storage options also allow a node to start with a few drives and add drives as capacity requirements grow, as shown in the following figure. Single node upgrades and drive scalability protect an optimized initial investment by allowing customers to start with what they need and expand the VxRail cluster by adding nodes and/or drives to increase performance and capacity as needed. Consult your Dell EMC representative for assistance.

Figure 11 VxRail scale on demand



A few basic rules regarding scaling are worth considering for planning:

1. Balance

- The first three nodes in a cluster must be the same processor and drive configuration. (2-Node vSAN configurations are not supported.)
- All nodes must be running the same version of software.
- Cannot mix hybrid and all flash nodes in the same cluster.
- 1GbE, 10GbE, and 25GbE base networking cannot be mixed in the same cluster.
- 1GbE must be hybrid and single processor node type.
- For G Series, all nodes in a chassis must be identical.

2. Flexibility

- Appliances in a cluster can be different models or series and can have different numbers of nodes.
- A cluster can have a varied number of drives, CPU, memory, and model types.
- A cluster can have between 3-64, but only a max of 8 if 1GbE networking is used.
- For G Series, a chassis can be partially populated.

Upgradeable options

With VxRail, nodes can upgrade or add memory, NIC cards, cache drives, and capacity drives. GPU can be upgraded or added in the V Series. It is not possible to upgrade from a single processor to a dual processor VxRail node. Please refer to the following table for information on which components are customer installable (replaceable).

Figure 12 VxRail customer and field replaceable parts

Hardware Component	Customer Replaceable Unit (CRU)	Field Replaceable Unit (FRU)
System Memory	Y	
Hard Drive	Y	
Solid State Drive (cache and capacity)	Y	
NVMe Cache Drive		Y
PCIe Network Interface Cards	Y	
Graphical Processing Unit (GPU)	Y	
Micro SDHC Card	Y	
Power Supply	Y	
Processors		Y
System Motherboard		Y
Host Bus Adapter (HBA330)	Y	
BOSS controller card and M.2 SATA disk		Y
Network Daughter Card (NDC)		Y

*The above table is a non-exhaustive list of FRUs that reflects common top level assembly parts.

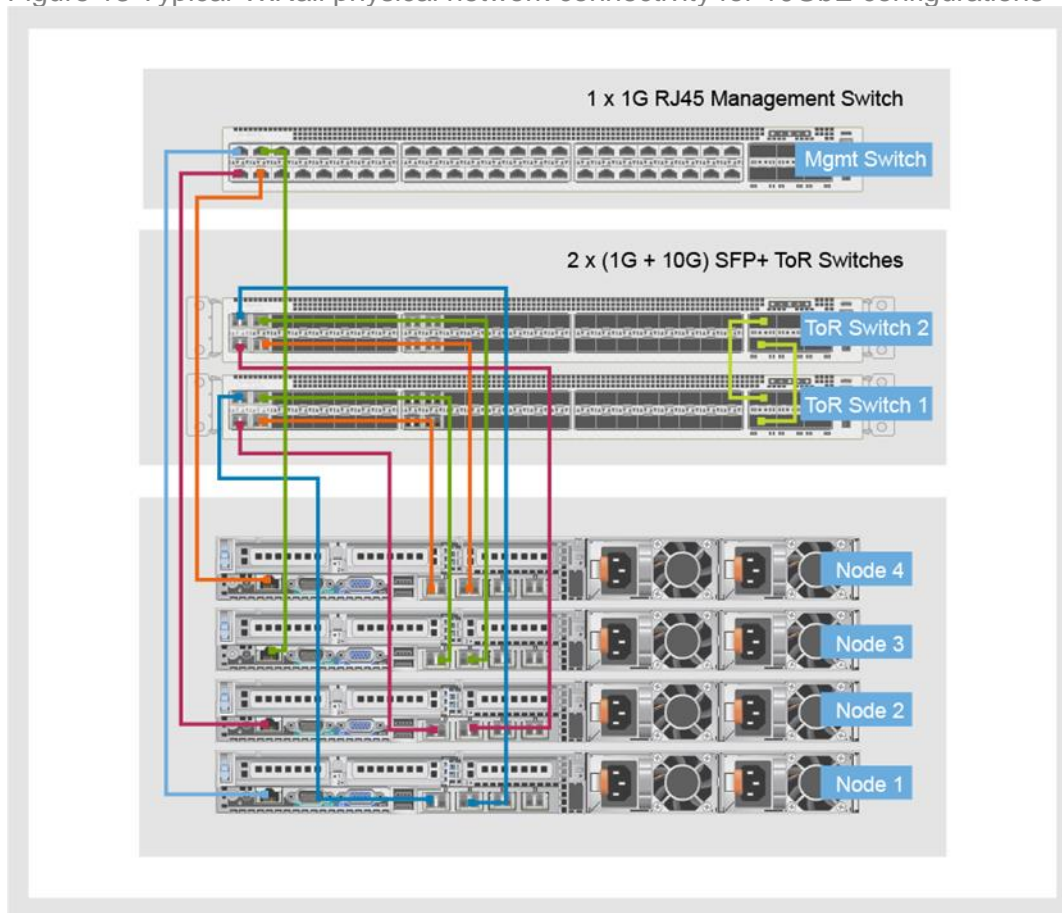
VxRail networking

The VxRail Appliance is a self-contained environment with compute, storage, server virtualization, and management services that make up a hyperconverged infrastructure. The distributed cluster architecture allows independent nodes to work together as a single system. Each node both contributes to and consumes system resources. This close coupling between nodes is accomplished through IP networking connectivity. IP networking also provides access to virtual machines and the services they provide.

While VxRail is a self-contained infrastructure; it is not a stand-alone environment. It is intended to connect and integrate with the customer's existing datacenter network. A typical implementation uses one or more customer-provided 10GbE Top of Rack (ToR) switches to connect each node in the VxRail cluster. For smaller environments, an option to use 1GbE switches is available, but these lower-bandwidth networks limit performance and scale. While the network switches are typically customer provided, Dell EMC offers an Ethernet switch, S4048, which can be included with the system.

The figure below shows typical network connectivity using two switches for redundancy. Single-switch implementations are also supported.

Figure 13 Typical VxRail physical network connectivity for 10GbE configurations



The number of Ethernet switch ports required depends on the VxRail model and whether it is configured for hybrid storage or for all flash. The all-flash system requires two or four 10GbE ports, and hybrid systems use either two 10GbE ports per node or four 1GbE ports per node. For 1GbE networks, the 10GbE ports auto-negotiate down to 1GbE. A two port 25GbE SFP28 is also an option. Additional network connectivity can be accomplished by adding additional NIC cards. The additional PCIe NICs are not configured by VxRail management but can be used by the customer to support non-VxRail traffic, primarily VM traffic. The additional ports are managed through vCenter.

Network traffic is segregated using switch-based VLAN technology and vSphere Network I/O Control (NIOC). Four types of network traffic exist in a VxRail cluster:

Management. Management traffic is used for connecting to VMware vCenter web client, VxRail Manager, and other management interfaces and for communications between the management components and the ESXi nodes in the cluster. Either the default VLAN or a specific management VLAN is used for management traffic.

vSAN. Data access for read and write activity as well as for optimization and data rebuild is performed over the vSAN network. Low network latency is critical for this traffic and a specific VLAN isolates this traffic.

vMotion. VMware vMotion™ allows virtual-machine mobility between nodes. A separate VLAN is used to isolate this traffic.

Virtual Machine. Users access virtual machines and the service provided over the VM network(s). At least one VM VLAN is configured when the system is initially configured, and others may be defined as required.

Pre-installation planning includes verifying that enough physical switch ports are available and that the ports are configured for the appropriate VLANs. VLANs along with IP addresses and other network-configuration information are used when the system is configured during installation. Detailed planning and configuration information is included in the [VxRail Network Guide](#).

When the system is initialized during installation, the configuration wizard automatically configures the required uplinks following VxRail standards and best practices. The wizard asks for the NIC configuration and accepts two options:

4X1GbE. Only valid for systems with hybrid-storage configuration with a single processor. The four (4) 10GbE ports auto-negotiate down to 1GbE. Management, vSAN, vMotion, and VM traffic is associated with these ports with the appropriate network teaming policy and NIOC settings.

2X10GbE. Management, vSAN, vMotion, and VM traffic is associated with these ports with the appropriate network teaming policy and NIOC settings. Note the additional two ports are available to use for additional VM traffic.

4X10GbE. Management, vSAN, vMotion, and VM traffic is associated with these ports with the appropriate network teaming policy and NIOC settings.

2X25GbE. Management, vSAN, vMotion, and VM traffic is associated with these ports with the appropriate network teaming policy and NIOC settings.

If nodes have additional physical NIC ports, they can be configured after installation using standard vSphere procedures.

1GbE network option

The Ethernet network not only provides connectivity to the VMs and services, it also provides the backplane for the nodes in a hyperconverged infrastructure to aggregate and share system resources. Therefore, network bandwidth is critical to system scale and performance. Today, most datacenters are built with 10Gb Ethernet connectivity, but 1GbE still exists in many environments. To support these environments, Dell EMC also offers 1GbE VxRail models for smaller, less-demanding workloads. The following are considerations for using the 1GbE connectivity option:

1GbE is only supported for hybrid-storage configurations, as it does not provide the bandwidth necessary for the ultra-high performance required by an all-flash system.

The maximum supported node count is eight nodes per cluster, because vSAN traffic increases with the number of nodes.

Only nodes with single-socket CPUs are supported.

A minimum of four ports are required per node. This increases the total number of switch ports required.

Dell EMC Open Networking & VxRail

As hyperconverged clusters scale, the network fabric becomes the critical piece of a successful deployment. Dell EMC fabrics deliver:

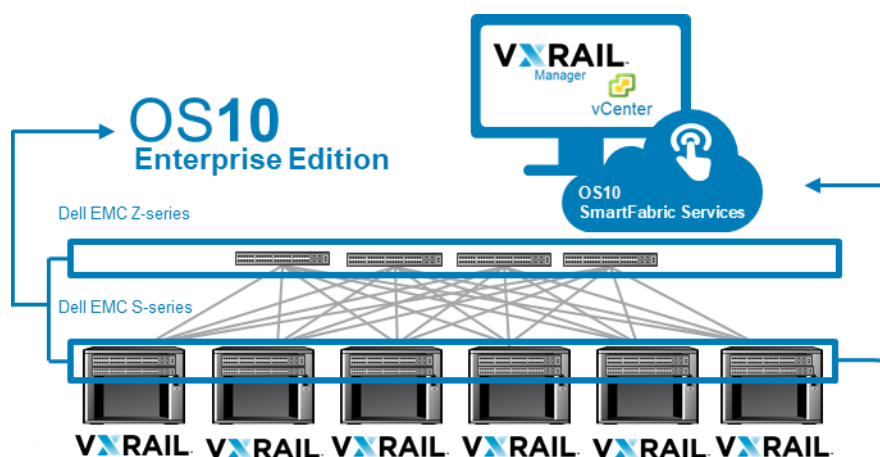
- On-demand scalability: to remain competitive, the modern datacenter requires the capability to dynamically grow and shrink based on business requirements. Together, Dell EMC Networking switching fabrics and Dell EMC VxRail provide an intelligent and capable architecture that scales on demand and increases the efficiency of the datacenter.
- Increased availability at scale: robust and redundant fabrics and storage are an absolute necessity for today's datacenter. A single failure should not cause a full-service interruption.

Dell EMC Fabric solutions provide a proven, fully-integrated and adaptable infrastructure. Our solutions leverage components of a hyperconverged architecture and purpose-built networking infrastructure elements, created solely to address the requirements of hyper-convergence and efficient data consumption.

Dell EMC SmartFabric Services (SFS)

Dell EMC Networking and VxRail are driving products and solutions by offering a well-engineered solution for the software defined enterprise that can deliver both operational and infrastructure efficiencies that were previously unavailable. Dell EMC is the only vendor with a complete set of hardware and software products capable of providing the necessary tools for the digital transformation. With the introduction of VxRail 4.7 and Dell EMC Networking OS10 Enterprise Edition SmartFabric Services, the conversation around automated dynamic infrastructure deployment can finally take place. Dell EMC, with VxRail and SmartFabric, is the first and only solution to deliver fully automated network awareness and configuration during set up, cluster expansion, and during day-to-day management to help create IT certainty.

Figure 14 SmartFabric Services



SmartFabric Services is a “one-of-a-kind” feature, part of the Dell EMC OS10 Enterprise Edition flagship networking operating system. Its introduction creates a fully integrated solution between the fabric and a hyperconverged cluster infrastructure such as VxRail. With SmartFabric Services, customers can quickly and easily deploy and automate datacenter networking fabrics. This enables faster time to production for hyperconverged and private cloud environments at any scale while being fully interoperable with existing datacenter infrastructure.

VxRail software architecture

The prior section introduced the flexible VxRail hardware architecture based on proven server technology. While the VxRail hardware helps differentiate the VxRail from other HCI solutions, VxRail is a complete appliance that includes software that enables a software-defined datacenter. These sections on software architecture provide a comprehensive examination of all the VxRail software components and their relationships and co-dependencies.

The VxRail Appliance is architected with software stack for appliance management, virtualization, and VM management. The stack comes pre-installed and simply requires running a configuration wizard on site to integrate the appliance into an existing network environment. VxRail HCI System Software is included for appliance management, operations, and automation. The VMware virtualization and virtual infrastructure management software includes:

- VMware vCenter Server
- vSphere ESXi
- vSAN (software-defined storage)
- VMware vRealize™ Log Insight™

Additional Dell EMC software includes:

RecoverPoint for VMs—5 VM licenses per node (for single node appliances), 15 VM per chassis for the G Series

VxRail provides a unique and tightly integrated architecture for VMware environments. VxRail deeply integrates VMware virtualization software. Specifically, VMware vSAN is integrated at the kernel level and is managed with VMware vSphere, which enables higher performance for the VxRail Appliance as well as automated scaling and wizard-based upgrades.

Appliance management

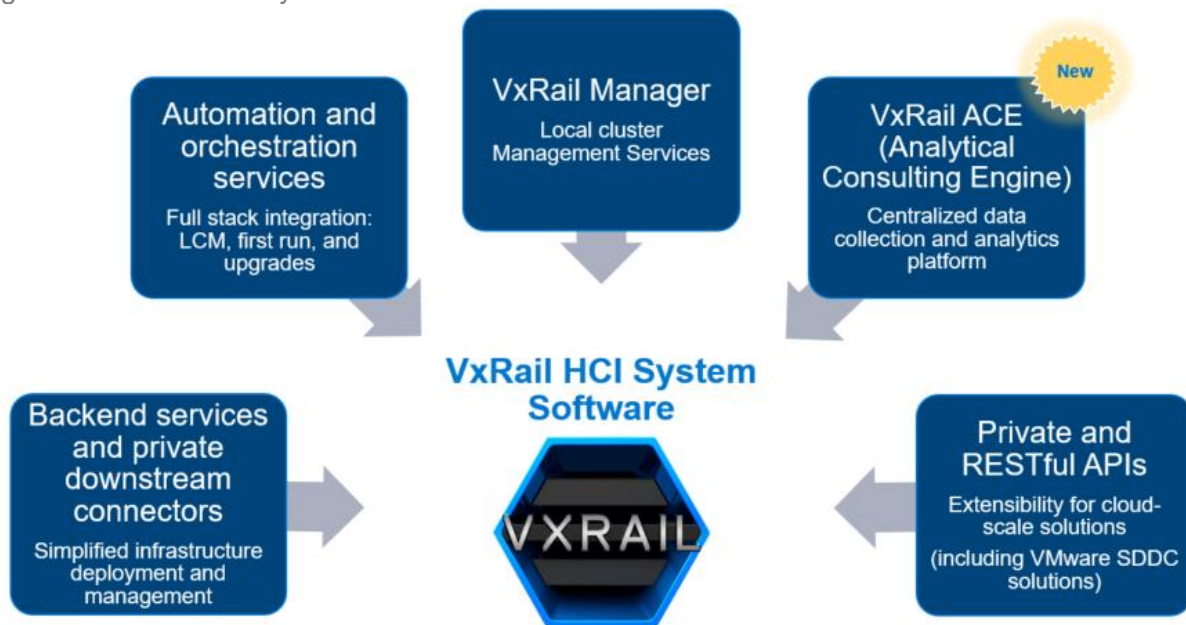
The introduction section of this TechBook addresses the complexity of the software-defined datacenter and the challenges of managing and maintaining an SDDC environment, and VxRail Manager responds to this challenge directly.

VxRail HCI System Software

VxRail HCI System Software, the VxRail management platform, is the appliance hardware lifecycle management and serviceability interface for VxRail clusters. It is a strategic advantage for VxRail and further reduces operational complexity. VxRail Manager provides out-of-the-box automation and orchestration for day 0 to day 2 appliance-based operational tasks, which reduces the overall IT OpEx required to manage the stack. No build-it-yourself HCI solution provides this level of lifecycle management, automation, and operational simplicity.

All virtualization management is performed using vCenter.
VxRail HCI System Software does not do any virtualization management.

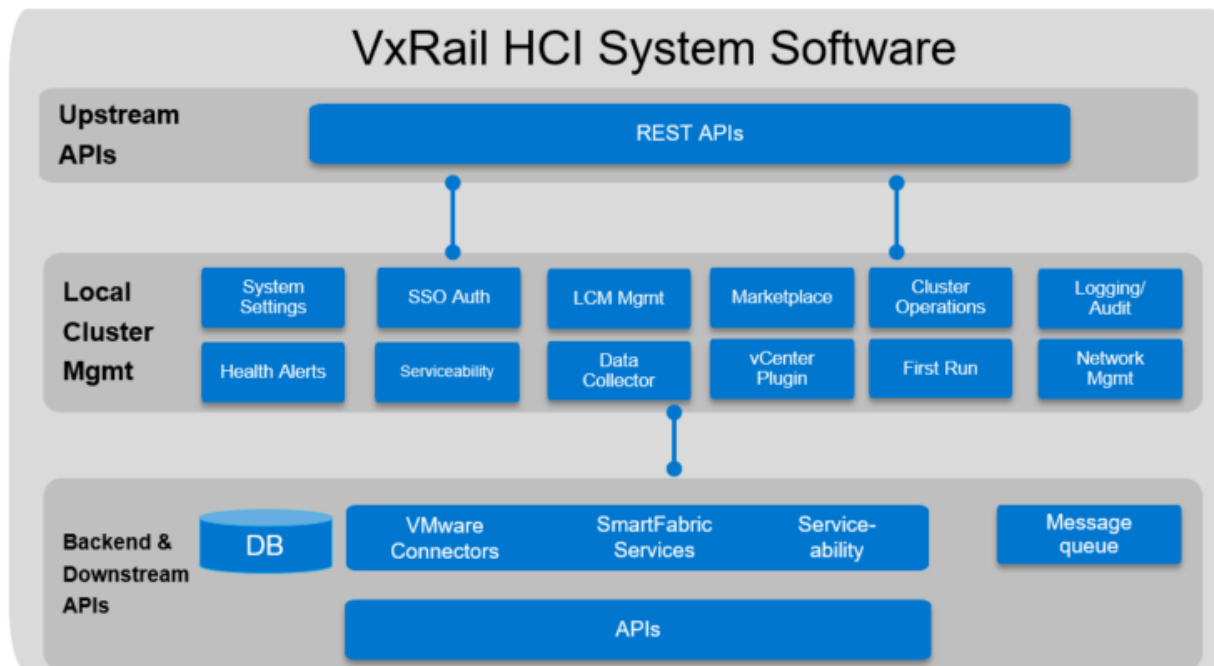
Figure 15 VxRail HCI System Software



To provide this level of differentiated value to customers, VxRail HCI System Software relies on a set of backend services to gather information and provide event coordination with the underlying infrastructure components including the vSAN cluster (i.e. ESXi, vSAN, vCenter Server), Dell PowerEdge server, and serviceability platforms (i.e. ESRS and eServices). This integration allows VxRail to automate and orchestrate infrastructure processes into critical services highly valued by HCI customers (i.e. lifecycle management, single package hardware/software upgrades, automated deployment capabilities). Users can access these cluster management services from their vCenter Server with the VxRail Manager plugin, or via a set of RESTful APIs. By providing public APIs, the value of VxRail for HCI can extend to SDDC solutions such as VMware Cloud Foundation, custom cloud solutions (i.e. Puppet, Ansible), or scripted solutions if customers, like service providers, wish to deploy and manage VxRail clusters at scale.

New to the VxRail management platform is another critical factor to customers' journey in IT transformation. While operational simplicity heavily impacts OpEx, the rise in infrastructure machine learning is fast becoming an essential component to aid in IT transformation. VxRail Analytical Consulting Engine (ACE) is the newly introduced analytical platform that leverages the data collection from VxRail clusters and best practices to deliver operational intelligence to customers about their HCI environment.

Figure 16 Architecture of VxRail System Software



The architecture diagram above expounds on VxRail HCI System Software. From the bottom up, the backend services include APIs that connect downstream to various members in the infrastructure layer. The VMware connectors provide communication to the components that make up a vSAN cluster. SmartFabric Services provides automated network provisioning capabilities to the VxRail management platform. More details about SmartFabric Services are described in this [section](#). To provide serviceability features in VxRail, communication with ESRS and eServices need to be established. For all the different types of information VxRail gathers, a database is used to store metrics and a message queue is used to facilitate the necessary sequence of events/transactions for automation and orchestration of processes.

The local cluster management layer is the set of proprietary services built to provide customers the benefit of operational simplicity and intelligence to more effectively manage their VxRail cluster.

- System settings – hardware enclosure and components status and information are propagated to the VxRail
- SSO Authentication – integration with the vCenter Server single sign-on service
- Lifecycle management – end-to-end upgrade of hardware and software components of VxRail
- Marketplace – access to download VxRail ecosystem software from a single area
- Cluster operations – cluster expansion or node removal services
- Logging/audit – system logging and auditing services that has the capability to send data to vRealize Log Insight

- Health alerts – notifications of component health
- Serviceability – access to ESRS for customer support and eServices for product knowledge
- Data collector – collects cluster metrics and sends to VxRail ACE for infrastructure machine learning
- vCenter plugin – VxRail Manager interface for local cluster management
- First Run – automated deployment wizard for Professional Services to deliver VxRail on Day 0
- Network management – automated network provisioning services enabled by SmartFabric Services

Customers looking to leverage these services to manage their VxRail clusters locally can use the VxRail Manager plugin on their vCenter Server. There is also upstream APIs that provide a subset of these capabilities, such as lifecycle management and cluster operations services, to cloud solution developers looking to orchestrate the provisioning of HCI for their cloud service delivery solutions. These APIs are also beneficial to customers looking to script their own solutions to manage VxRail clusters at scale.

VxRail Manager

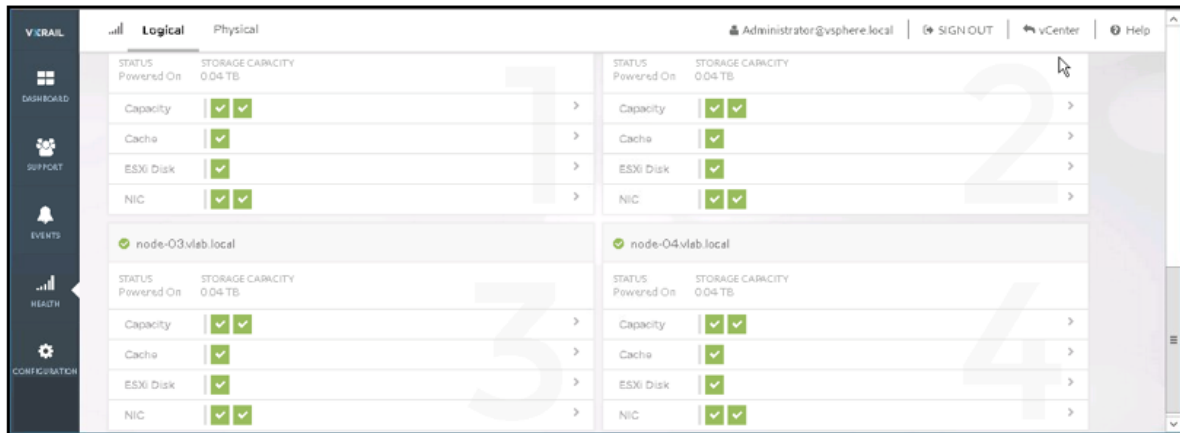
VxRail Manager features user-friendly workflows for automating VxRail deployment and configuration and monitoring the health of individual appliances and individual nodes in the entire cluster. It also incorporates functionality for hardware serviceability and appliance platform lifecycle management. For instance, it guides system administrators through adding new appliances to an existing cluster, and it automatically detects new appliances when they come online. VxRail Manager is also used to replace failed disk drives without disrupting availability, to generate and download diagnostic log bundles, and to apply VMware updates or software patches non-disruptively across VxRail nodes.

File-based backups of VxRail HCI System Software help to ensure business continuity in the rare event the VxRail VM needs to be rebuilt.

VxRail HCI System Software is preinstalled on the VxRail appliance as a single virtual machine. With VxRail Manager plugin for vCenter Server, all VxRail Manager features are integrated with and accessible from the vCenter Server so that users can benefit from these valuable capabilities on a familiar management interface. With the VxRail Manager plugin, the vCenter Server displays storage, CPU, and memory utilization at the cluster, appliance, or individual-node level.

Administrators can access additional real-time system health details for both logical and physical resources and can view and analyze resource operability, performance, and utilization data (as shown in the following figure).

Figure 17 VxRail Manager Health Tab for Logical Resources




VxRail also leverages VMware vRealize Log Insight to monitor system events and provide ongoing holistic notifications about the state of virtual environment and appliance hardware. It delivers real-time automated log management for the VxRail Appliance with log monitoring, intelligent grouping, and analytics to provide better troubleshooting at scale across VxRail physical, virtual, and cloud environments. Furthermore, VxRail HCI System Software simplifies appliance platform lifecycle management by delivering patch software and update notifications that can be automatically installed without interruption or downtime.


Dell EMC Software Remote Services (ESRS), also accessible from within VxRail Manager plugin or REST API, provide enterprise-class support and services. ESRS includes online chat support and Dell EMC field-service assistance (as seen in the figure below).

Figure 18 VxRail Manager ESRS details



- Proactive two-way remote connection for VxRail that is secure, high speed and **operates 24x7.**
- **Heartrate ensures continuous monitoring, 5%**
- Fast remote diagnosis and repair of potential problems before impact to business resulting in **5X faster service event resolution.**
- Receive support updates seamlessly and automatically.



CUSTOMER ENVIRONMENT

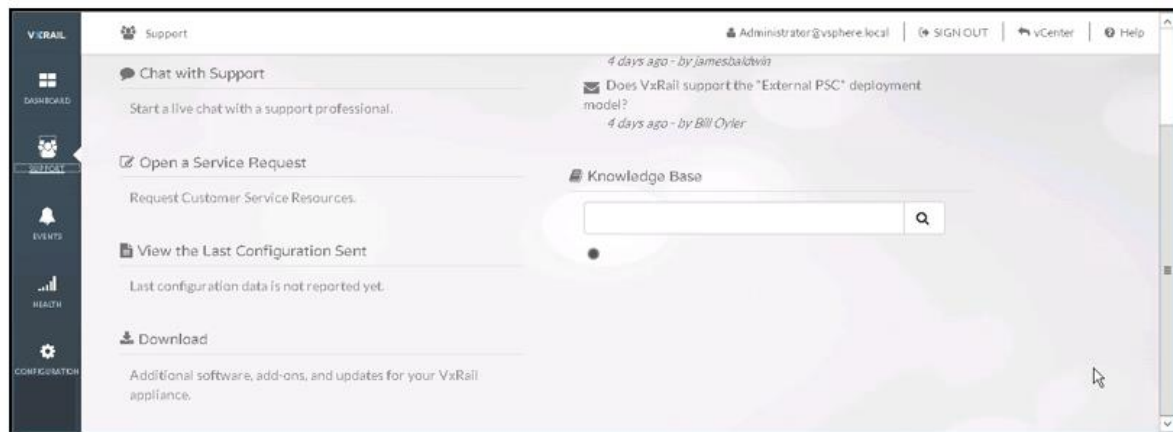


DELL EMC GLOBAL SUPPORT

In addition to ESRS-specific support, the VxRail Support page on vCenter Server links to VxRail Community pages for Dell EMC Knowledge Base articles, user forums for FAQ information and VxRail best practices. The figure below is an example of the support view.

Figure 19 VxRail Manager Support Tab



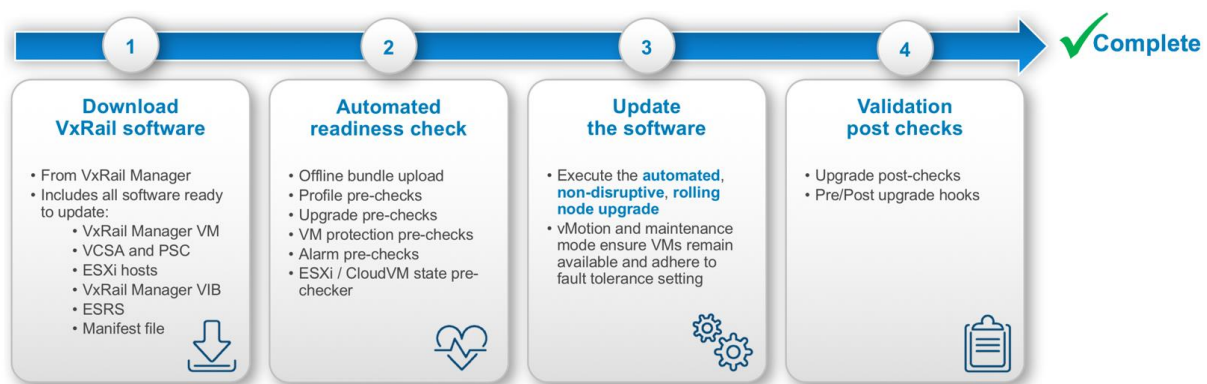
VxRail Manager plugin provides access to a digital market for finding and downloading qualified software packages such as VMware Horizon Cloud, Data Domain Virtual Edition, RecoverPoint for VM, vSphere Data Protection and other software options for VxRail Appliances.

Customer upgradeable software

VxRail Appliance software is customer upgradeable via a fully automated and validated process. The software upgrade is initiated from VxRail Manager plugin or REST API, and it automatically downloads all software ready to be updated including VxRail HCI System Software, vCenter Server and PSC, ESXi hosts, and ESRS. The automated process consists of four steps including download of the VxRail software, a readiness check, the actual update of the software, and finally, validation and upgrade post checks. The final validation step ensures the upgrade was successful, and the VxRail Appliance is fully functional at the new, upgraded version of software.

The figure below shows the four automated steps of a customer executed VxRail Appliance software upgrade.

Figure 20 Automated process steps for customer-executable VxRail appliance software upgrade



Step 3 is performed one node at a time, where the ESXi host is placed in maintenance mode, and using vMotion, the VMs are moved to other nodes making the upgrade process non-disruptive.

vSphere and vSAN ordering information

VxRail Appliance allows customers to use any existing eligible vSphere licenses with their VxRail, or the licenses can be purchased with a VxRail. This VxRail vSphere license independent model (also called “bring your own” or BYO vSphere License model) allows customers to leverage a wide variety of vSphere licenses they may have already purchased.

Several vSphere license editions are supported with VxRail including Enterprise+, Standard, and ROBO editions (vSphere Enterprise is also supported, but is no longer available from VMware). Also supported are vSphere licenses from Horizon bundles or add-ons when the appliance is dedicated to VDI.

If vSphere licenses need to be purchased, they should be ordered through Dell EMC, the customer’s preferred VMware channel partner, or from VMware directly. Licenses acquired through VMware ELA, VMware partners, or Dell EMC will receive single-call support from Dell EMC.

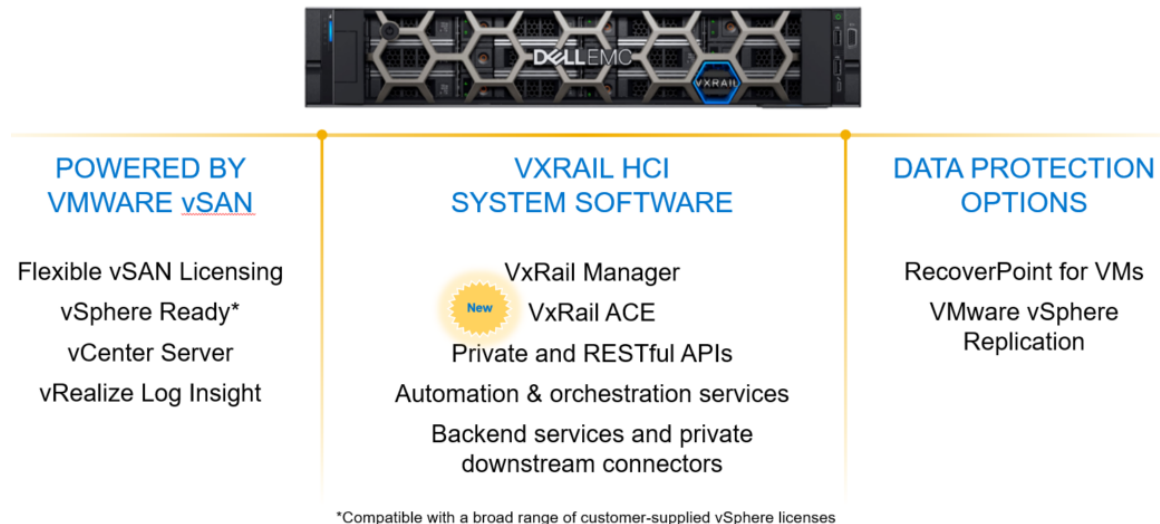
When determining the best vSphere license to use with the VxRail Appliance, a key consideration is the effect of VxRail functionality. DRS, a significant vSphere feature described earlier in this TechBook, provides the greatest amount of functional variance to VxRail clusters. Customers should consider the degree of automation that DRS provides to determine if the vSphere license they desire includes this functionality.

VxRail supports flexible vSAN licensing options and requires vSAN to be ordered with VxRail or applied via a vSAN ELA from VMware. VxRail supports all license editions of vSAN including Standard, Advanced, and Enterprise. For details on the differences between the vSAN versions please consult the VMware vSAN Comparison Guide here:

<https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsan/vmware-vsan-67-licensing-guide.pdf>

The figure below lists the software included with VxRail.

Figure 21 VxRail included software



Note: Check with your Dell EMC representative to verify the latest software versions levels supported.

Note: The vCenter license included with VxRail is for vCenter deployed on VxRail and is not transferrable to another external server. All other software integrated with VxRail is also non-transferrable.

Use the information in the following figures as a guide.

Figure 22 General vSphere options



 vSphere Enterprise Plus	 vSphere Standard
<p>Dramatically increases administrator productivity</p> <ul style="list-style-type: none"> • Automated workload rebalancing and affinity rules • Automated maintenance mode • Streamlined drive replacement • One-click software updates • vGPU support for VxRail V Series 	<p>Lower upfront costs; manual administration</p> <ul style="list-style-type: none"> • Manual workload balancing • Manual maintenance mode • Multi-step drive replacement • One-click software updates • Does not support vGPU

Figure 23 vSphere editions

VSPHERE EDITIONS & PACKAGES	VSPHERE EDITION FUNCTIONAL LEVEL	VSPHERE LICENSE KEY INPUT UNIT
vSphere Enterprise Plus	ENT+	CPU
vSphere Enterprise	ENT+	CPU
vSphere Standard	STD	CPU
vSphere ROBO Advanced	STD* (Additional Features, no DRS)	VM (Qty 25)
vSphere ROBO Standard	STD	VM (Qty 25)
vSphere Desktop	ENT+, virtual desktop only	User (Qty 100)
vSphere Platinum	ENT+	CPU + subscription
<hr/>		
vSphere with Operations Management Enterprise Plus (VSOM)	ENT+	CPU
<hr/>		
Horizon Enterprise	ENT+, virtual desktop only	User (Qty 10 or 100)
Horizon Advanced	ENT+, virtual desktop only	User (Qty 10 or 100)
Horizon Standard	ENT+, virtual desktop only	User (Qty 10 or 100)
vCloud Suite	ENT+	CPU

VMware vSphere

The VMware vSphere software suite delivers an industry-leading virtualization platform to provide application virtualization within a highly available, resilient, efficient on-demand infrastructure—making it the ideal software foundation for VxRail Appliances. ESXi and vCenter are components of the vSphere software suite. ESXi is a hypervisor installed directly onto a physical VxRail server node, enabling it to be partitioned into multiple logical servers or virtual machines. Virtual machines are configured on top of the ESXi server. VMware vCenter server is a centralized management application that is used to manage the ESXi hosts and VMs.

The following sections provide an in-depth examination of the VMware software components as implemented in the VxRail software architecture.

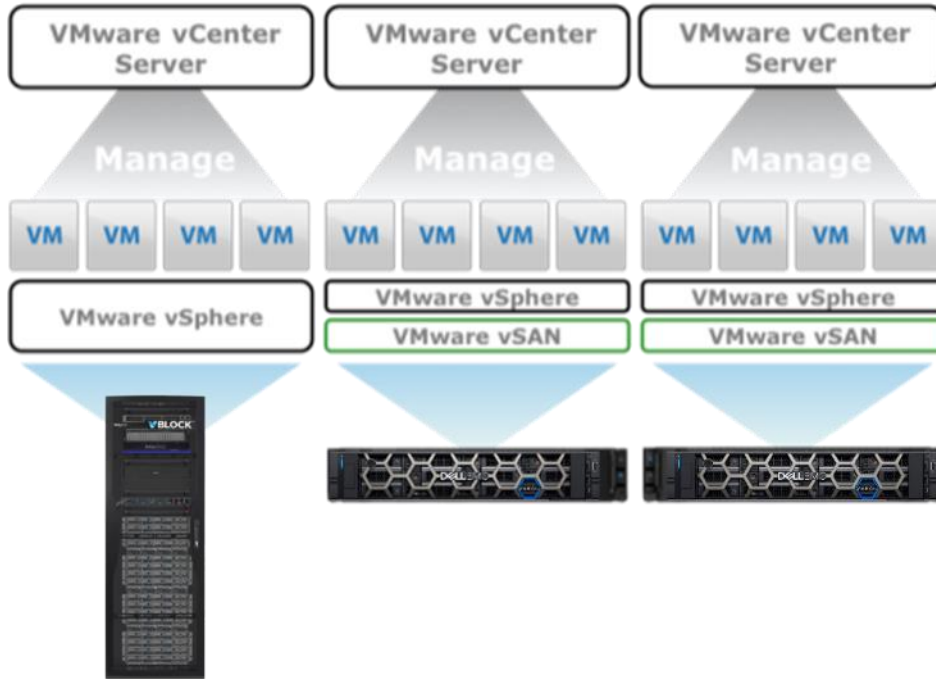
VMware vCenter Server

vCenter Server is the centralized platform for managing a VMware environment. VxRail includes a license to run vCenter hosted on VxRail (the license is not transferrable to run vCenter on an external server). It is the primary point of management for both server virtualization and vSAN and is the enabling technology for advanced capabilities such as vMotion, Distributed Resource Scheduler (DRS), and high availability (HA). vCenter scales to enterprise levels where a single vCenter can support up to 1,000 hosts (VxRail nodes) and 10,000 virtual machines. vCenter supports a logical hierarchy of datacenters, clusters, and hosts, which allows resources to be segregated by use cases or lines of business and allows resources to move dynamically as needed. This is all done from a single interface.

As part of a VxRail deployment, a standalone vCenter instance is optionally configured on VxRail and provides the primary point of management for both the virtual machine and vSAN environments. The vCenter instance will be configured with an external Platform Services Controller (PSC), where the vCenter server and the PSC are configured as separate virtual machines.

The figure below is an example of three separate vCenter environments.

Figure 24 Three separate vCenter environments

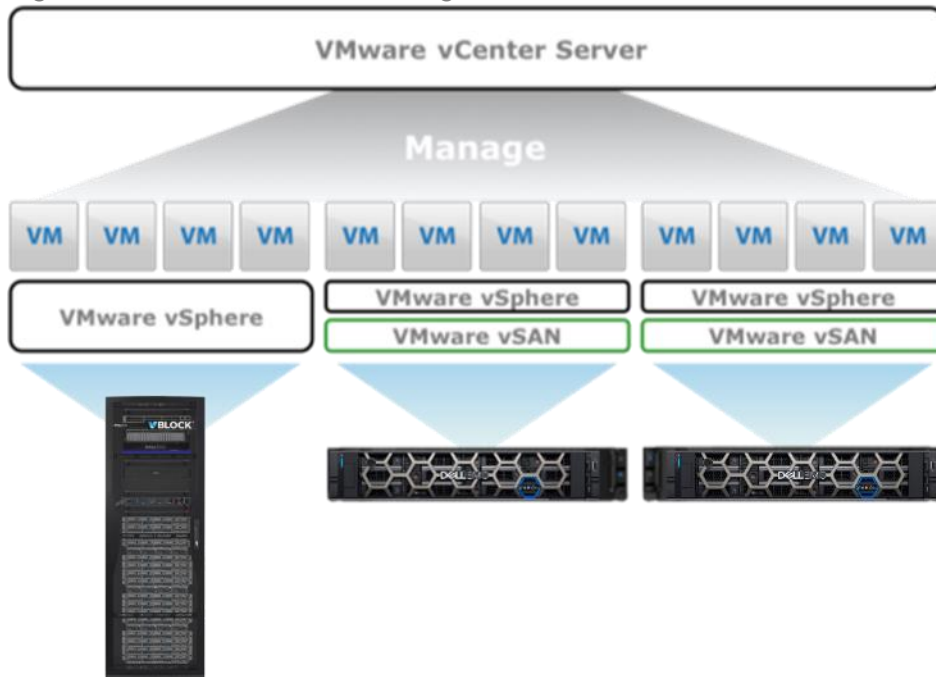


The VxBlock and two VxRail environments are managed as separate entities. This is a simple environment, and in some cases, the planning and deployment may be easier as there are no interactions.

A VxRail Appliance can optionally join an existing externally hosted vCenter Server environment during the initial configuration (if the versions are compatible with the VxRail software). This allows for a central vCenter Server to manage multiple VxRail Appliances from a single pane of glass. Each VxRail cluster appears within vCenter.

The figure below shows an example where multiple VxRail clusters are part of an existing vCenter environment. Each VxRail is a unique and independent vSAN data store within vCenter.

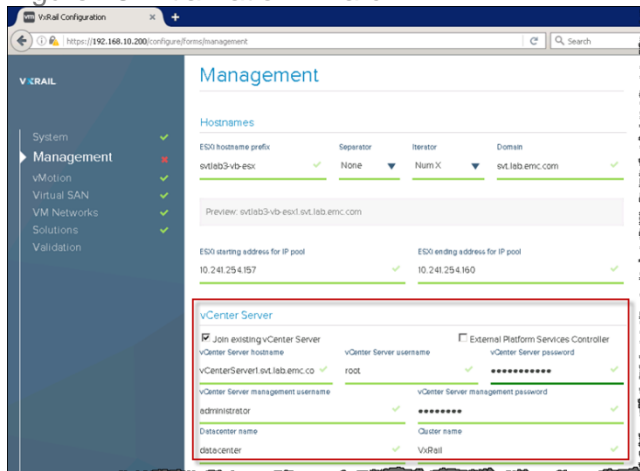
Figure 25 VxRail within an existing vCenter environment



The external vCenter server can be a physical server or a virtual server running as either a VCSA or in a Windows virtual machine. The Platform Services Controller (PSC) can be either embedded or non-embedded. As part of the initial configuration, the management page of the VxRail wizard presents the option to join an existing vCenter. If selected, specify the hostname of the vCenter Server and administrator password, identify the datacenter to add the VxRail environment, and supply the name of the cluster.

The figure below shows an example of the VxRail Initialization Wizard dialog to specify joining an existing vCenter Server.

Figure 26 Initialization wizard



The datacenter must already exist within vCenter, and the cluster will be created as part of the installation process. See the [VxRail vCenter Server Planning Guide](#) for more information on configuration options.

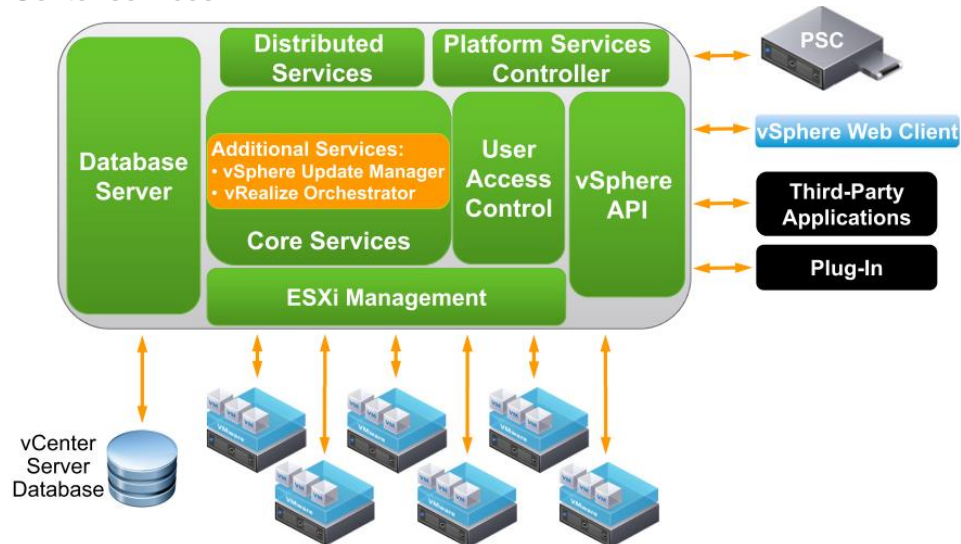
vCenter services and interfaces

vCenter provides a number of services and interfaces, including:

- Core VM and resource services such as an inventory service, task scheduling, statistics logging, alarm and event management, and VM provisioning and configuration
- Distributed services such as vSphere vMotion, vSphere DRS, and vSphere HA
- vCenter Server database interface

The figure below clarifies the organization of vCenter services within the vSphere environment.

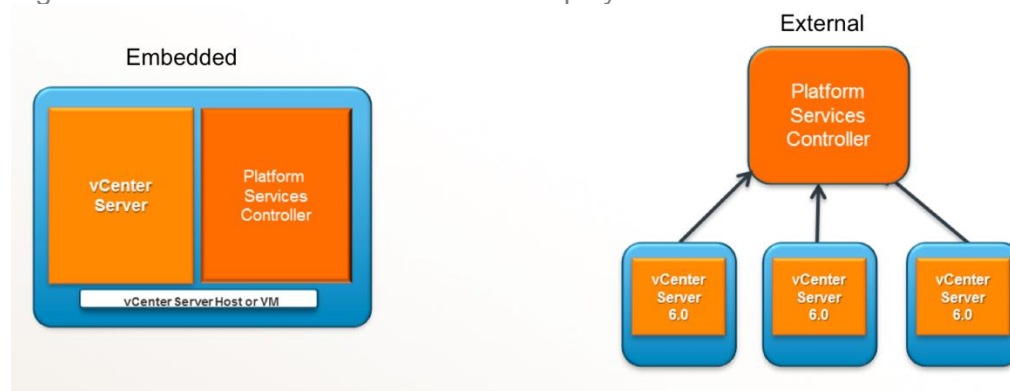
Figure 27 vCenter services



PSC deployment options

The Platform Services Controller (PSC) can be deployed as either embedded (in externally hosted vCenter) or external, as shown below. When VxRail-hosts vCenter the PSC is deployed external as a separate VM. See the figure below.

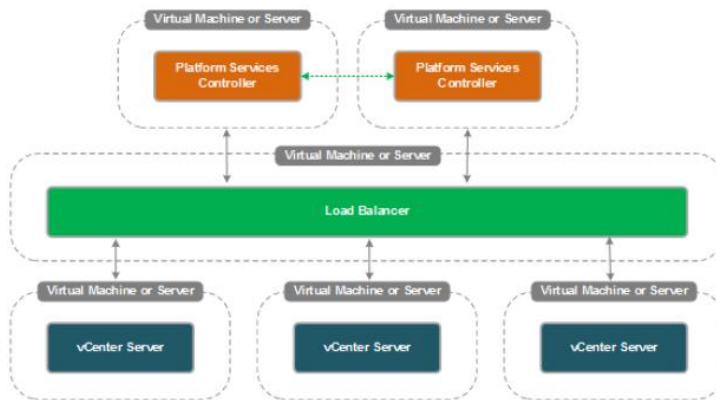
Figure 28 Embedded and external PCS deployments



Embedded PSC is ideal for small environments, or if simplicity and reduced resource utilization are key factors for the environment. The vCenter Server is bundled with an embedded PSC, and all the PSC services reside on the same host machine as vCenter Server.

External PSC (see the figure below) is ideal for larger environments, where there are multiple vCenter servers, but you want a single pane-of-glass for the site. The services bundled with the PSC and vCenter Server are deployed on different virtual machines or even different physical servers. When utilizing an externally hosted vCenter with external PSC, Enhanced Linked Mode (ELM) is supported.

Figure 29 External PSCs configured for high availability



Enhanced Linked Mode

Multiple vCenter Servers can connect to the same external Platform Service Controller. Enhanced Linked Mode enables a consolidated management view of multiple vCenter Servers configured to use the Platform Services Controller domain. This includes a common inventory where an administrator can search for objects across vCenter Servers. Roles, permissions, licenses, and other key data across systems are also replicated across vCenter instances.

This single pane-of-glass view provides enterprise-level scale and works particularly well in large, multiple VxRail cluster environments or when the VxRail appliance joins an existing large vSphere environment. In addition, Enhanced Link Mode also enables capabilities such as Cross vCenter vMotion where VMs can be moved between vCenters.

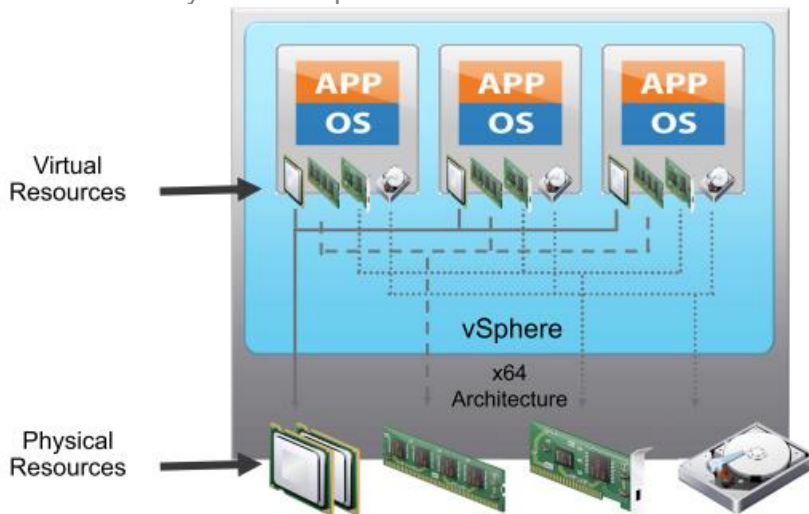
Enhanced Linked Mode is supported with VxRail when running an externally hosted vCenter with VxRail and utilizing external PSC.

VMware vSphere ESXi

vSphere is the core operational software in the VxRail Appliance. vSphere aggregates a comprehensive set of features that efficiently pools and manages the resources available under the ESXi hosts. Keep in mind that this TechBook focuses on vSphere technology specifically as it pertains to the VxRail Appliance. Features included in other vSphere implementations may not apply to VxRail and features included in VxRail may not apply to other implementations.

VMware ESXi is an enterprise-class hypervisor that deploys and services virtual machines. The following figure illustrates its basic architecture.

Figure 30 Birds-eye view: vSphere ESXi architecture



ESXi partitions a physical server into multiple secure and portable VMs that can run side by side on the same physical server. Each VM represents a complete system—with processors, memory, networking, storage, and BIOS—so any operating system (guest OS) and software applications can be installed and run in the virtual machine without any modification.

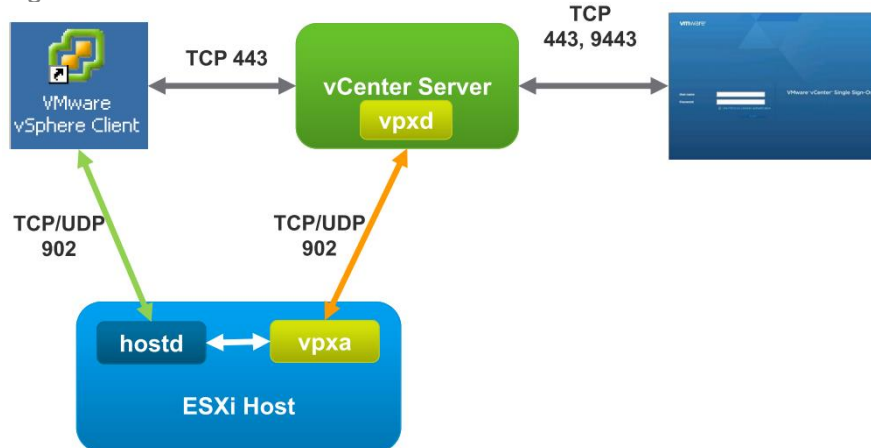
The hypervisor provides physical-hardware resources dynamically to virtual machines (VMs) as needed to support the operation of the VMs. The hypervisor enables virtual machines to operate with a degree of independence from the underlying physical hardware. For example, a virtual machine can be moved from one physical host to another. Also, the VM's virtual disks can be moved from one type of storage to another without affecting the functioning of the virtual machine.

ESXi also isolates VMs from one another, so when a guest operating system running in one VM fails, other VMs on the same physical host are unaffected and continue to run. Virtual machines share access to CPUs, and the hypervisor is responsible for CPU scheduling. In addition, ESXi assigns VMs a region of usable memory and provides shared access to the physical network cards and disk controllers associated with the physical host. Different virtual machines can run different operating systems and applications on the same physical computer.

Communication between vCenter Server and ESXi hosts

vCenter Server communicates with the ESXi host through a vCenter Server agent, also referred to as vpxa or the vmware-vpxa service, which is started on the ESXi host when it is added to the vCenter Server inventory. See the figure below.

Figure 31 Communication between vCenter and ESXi hosts



Specifically, the vCenter vpxd daemon communicates through the vpxa service to the ESXi host daemon known as the hostd process. The vpxa process acts as an intermediary between the vpxd process that runs on vCenter Server and the hostd process that runs on the ESXi host, relaying the tasks to perform on the host. The hostd process runs directly on the ESXi host and is responsible for managing most of the operations on the ESXi host including creating VMs, migrating VMs, and powering on VMs.

Virtual machines

A virtual machine consists of a core set of the following related files, or a set of objects, as shown in the figure below.

Figure 32 Virtual machine files

Name	Type	Size
.dvsData	Folder	
.sdd.sf	Folder	
TestVM01-572b3eef.v...	File	1,060,449.20 KB
TestVM01.nvram	Non-volatile Memory File	8.48 KB
TestVM01.vmdk	Virtual Disk	6,711,296.00 KB
TestVM01.vmsd	File	0.00 KB
TestVM01.vmx	Virtual Machine	3.44 KB
vmware-1.log	VM Log File	374.14 KB
vmware-2.log	VM Log File	221.98 KB
vmware-3.log	VM Log File	219.72 KB
vmware.log	VM Log File	218.44 KB

Except for the log files, the name of each file starts with the virtual machine's name (VM_name). These files include:

A configuration file (.vmx) and/or a virtual-machine template-configuration file (.vmtx)

One or more virtual disk files (.vmdk)

A file containing the virtual machine's BIOS settings (.nvram)

A virtual machine's current log file (.log) and a set of files used to archive old log entries (-#.log)

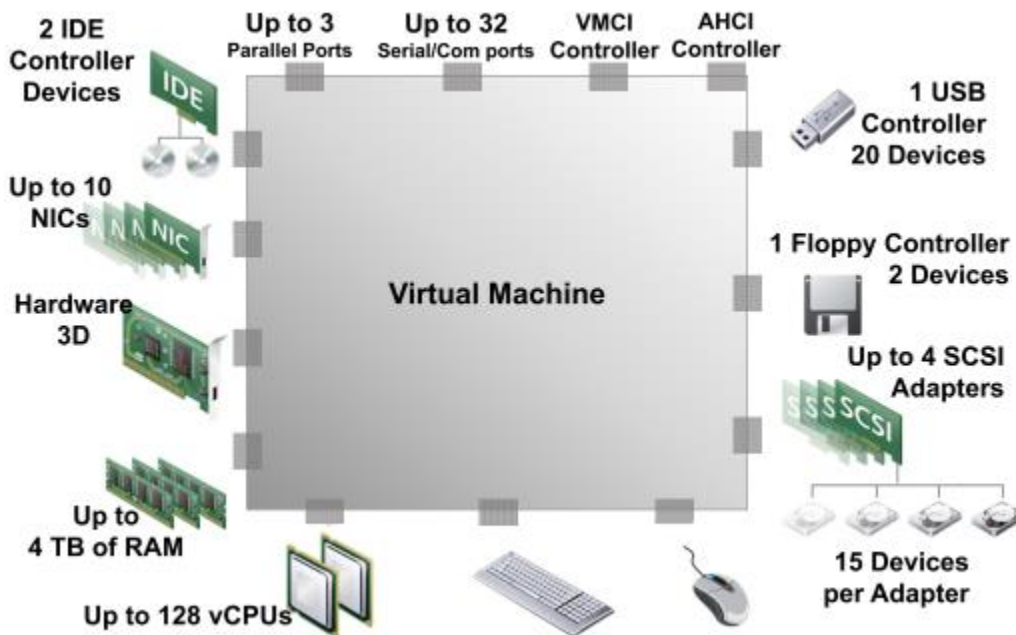
Swap files (.vswp), used to reclaim memory during periods of contention

A snapshot description file (.vmsd), which is empty if the virtual machine has no snapshots

Virtual machine hardware

A virtual machine uses virtual hardware. Each guest operating system sees ordinary hardware devices and does not know that these devices are virtual. Hardware resources are shown in the figure below.

Figure 33 Hardware resources for VMs



All virtual machines have uniform hardware, except for a few variations that the system administrator can apply. Uniform hardware makes virtual machines portable across VMware virtualization platforms. vSphere supports many of the latest CPU features, including virtual CPU performance counters. It is possible to add virtual hard disks and NICs, and configure virtual hardware, such as CD/DVD drives, floppy drives, SCSI devices, USB devices, and up to 16 PCI vSphere DirectPath I/O devices.

Virtual Machine Communication

The Virtual Machine Communication Interface (VMCI) provides a high-speed communication channel between a virtual machine and the hypervisor. VMCI devices cannot be added or removed. The SATA controller provides access to virtual disks and DVD/CD-ROM devices. The SATA virtual controller appears to a virtual machine as an AHCI SATA controller. Without VMCI, virtual machines would communicate with the host using the network layer, which adds overhead to the communication. With VMCI, communication overhead is minimal, and tasks requiring that communication can be optimized. An internal network can transmit an average of slightly over 2Gbps using VMXNET3. VMCI can go up to nearly 10Gbps with twelve 8k-sized queue pairs.

VMCI provides socket APIs that are very similar to the APIs already used for TCP/UDP applications.

For more information about the virtual hardware, see *vSphere Virtual Machine Administration Guide* at <https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-6-pubs.html>.

Virtual networking

VMware vSphere provides a rich set of networking capabilities that integrate well with sophisticated enterprise networks. These networking capabilities are provided by the ESXi server and managed by vCenter. Virtual networking provides the ability to network virtual machines in the same way physical machines are networked. Virtual networks can be built within a single ESXi server host or across multiple ESXi server hosts. VxRail ESXi hosts use a virtual switch for communication among virtual machines in the VxRail cluster using the same protocols that would be used over physical switches, without the need for additional networking hardware. The virtual switch also supports VLANs that are compatible with standard VLAN implementations from switch vendors. A virtual switch, like a physical Ethernet switch, forwards frames at the data-link layer.

Virtual Ethernet adapters are the key vSphere components for virtual networking. A virtual machine can be configured with one or more virtual Ethernet adapters, each of which has its own IP address and MAC address. As a result, virtual machines have the same properties as physical machines from a networking standpoint. In addition, virtual networks enable functionality not possible with physical networks today. Virtual Ethernet adapters are used by individual virtual machines and the virtual switches that connect VMs to each other and connect both virtual machines and the ESX Server service console to external networks.

The virtual switch links to the external network through outbound Ethernet adapters called vmnics, and the virtual switch can bind multiple vmnics together (much like NIC teaming on a traditional server), extending availability and bandwidth to the virtual machines it services.

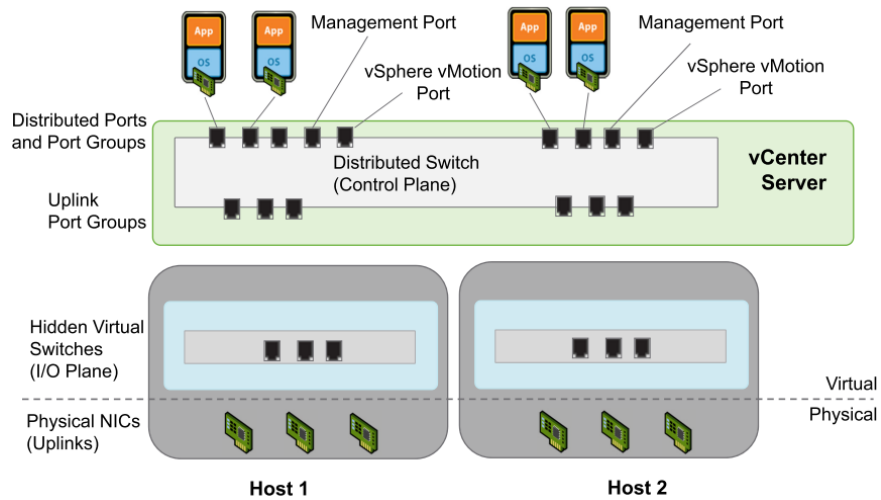
Virtual switches are similar to their physical-switch counterparts. Like a physical network device, each virtual switch is isolated for security and has its own forwarding table. An entry in one table cannot point to another port on another virtual switch. The switch looks up only destinations that match the ports on the virtual switch where the frame originated. This feature stops potential hackers from breaking virtual switch isolation. Virtual switches also support VLAN segmentation at the port level, so each port can be configured either as an access port to a single VLAN or as a trunk port to multiple VLANs.

Virtual Distributed Switch

VxRail clusters use the VMware Virtual Distributed Switch (VDS), which functions as a single switch that spans across multiple nodes in the same cluster. This switch enables virtual machines to maintain consistent network configuration as they migrate across multiple hosts. A distributed switch is configured in vCenter Server at the datacenter level and makes the configuration consistent across all hosts. vCenter Server stores the state of distributed ports in the vCenter Server database. Networking statistics and policies migrate with virtual machines when the virtual machines are moved from host to host. As discussed in upcoming sections, vSAN relies on VDS for its storage-virtualization capabilities, and the VxRail Appliance uses VDS for appliance traffic.

The following figure provides an overview of VDS.

Figure 34 Virtual Distributed Switch



vMotion and Virtual Machine mobility

VMware vMotion enables the live migration of running virtual machines from one physical server to another with zero downtime, continuous service availability, and complete transaction integrity. vMotion is a key enabling technology for creating the dynamic, automated, and self-optimizing datacenter. vMotion continuously and automatically allocates virtual machines within resource pools. It also improves availability by conducting maintenance without disrupting business operations.

The advanced capability for migrating workloads without disruption is one of the features that distinguish the VxRail solution from other HCI options. In the vSphere virtual infrastructure, migration refers to moving a virtual machine from one host, datastore, or vCenter Server system to another host, datastore, or vCenter Server system. Different types of migrations exist including:

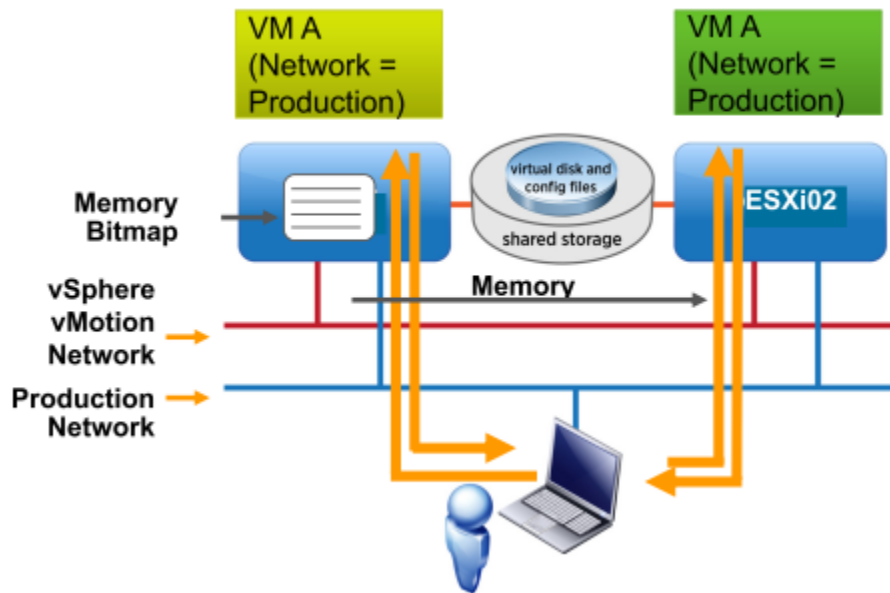
Cold, which is migrating a powered-off VM to a new host or datastore

Suspended, which is migrating a suspended VM to a new host or datastore

Live, which uses vSphere vMotion to migrate a “live,” powered-on VM to a new host and/or uses vSphere Storage vMotion to migrate the files of a live, powered-on VM to a new datastore

vMotion allows for live migration of virtual machines between ESXi hosts without disruption or downtime. The process is summarized in the figure below.

Figure 35 vMotion migration



With vMotion, while the entire state of the virtual machine is migrated, the data remains in the same datastore. The state information includes the current memory content and all the information that defines and identifies the virtual machine. The memory content consists of transaction data and whatever bits of the operating system and applications in memory. The definition and identification information stored in the state includes all the data that maps to the virtual machine hardware elements, including BIOS, devices, CPU, and MAC addresses for the Ethernet cards.

A vMotion migration consists of the following steps:

1. The VM memory state is copied over the vMotion network from the source host to the target host. Users continue to access the VM and, potentially, update pages in memory. A list of modified pages in memory is kept in a memory bitmap on the source host.
2. After most of the VM memory is copied from the source host to the target host, the VM is quiesced. No additional activity occurs on the VM. During the quiesce period, vMotion transfers the VM-device state and memory bitmap to the destination host.
3. Immediately after the VM is quiesced on the source host, the VM is initialized and starts running on the target host. A Gratuitous Address Resolution Protocol (GARP) request notifies the subnet that the MAC address for the VM is now on a new switch port.
4. Users access the VM on the target host instead of the source host. The memory pages used by the VM on the source host are marked as free.

Enhanced vMotion Compatibility

Enhanced vMotion Compatibility (EVC) is a cluster feature that prevents vMotion migrations from failing because of incompatible CPUs. EVC ensures that all hosts in a cluster present the same CPU feature set to virtual machines, even if the actual CPUs on the hosts differ. It prevents migration failures due to CPU incompatibility. This is on by default in VxRail Appliances.

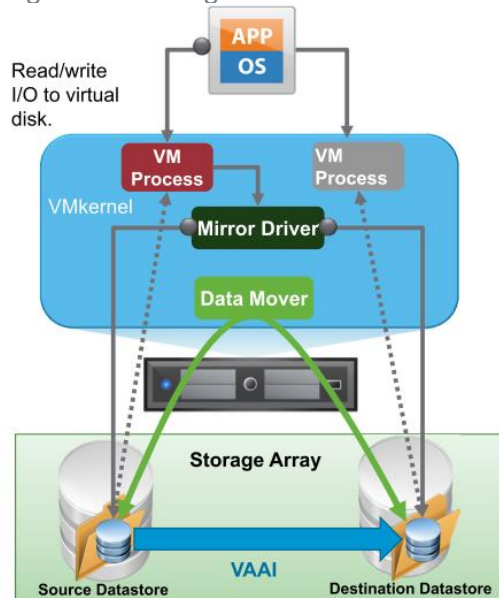
Storage vMotion

Storage vMotion uses an I/O-mirroring architecture to copy disk blocks between source and destination:

1. Initiate storage migration.
2. Use the VMkernel data mover and provide vSphere Storage APIs for Array Integration (VAAI) to copy data.
3. Start a new VM process.
4. Mirror I/O calls to file blocks that have already been copied to virtual disk on the target datastore. Switch to the target-VM process to begin accessing the virtual-disk copy.

The figure below illustrates the process

Figure 36 Storage vMotion



The storage-migration process copies the disk just once, and the mirror driver synchronizes the source and target blocks with no need for recursive passes. In other words, if the source block changes after it migrates, the mirror driver writes to both disks simultaneously which maintains transactional integrity. The mirroring architecture of Storage vMotion produces more predictable results, shorter migration times, and fewer I/O operations than more conventional storage-migration options. It's fast enough to be unnoticeable to the end user. It also guarantees migration success even when using a slow destination disk.

vSphere supports the following Storage vMotion migrations:

Between clusters

Between datastores (including non-vSAN to vSAN and vice versa)

Between networks

Between vCenter Server instances for vCenter Servers configured in Enhanced Link Mode with hosts that are time-synchronized

Over long distances (up to 150ms round-trip time)

Note that for VxRail clusters, Storage vMotion can only be used for migration into or out of the vSAN datastore.

vSphere Distributed Resource Scheduler

VMware Distributed Resource Scheduler (DRS) is a key feature included with vSphere Enterprise Plus and vSphere with Operations Management Enterprise Plus. DRS balances computing capacity across a collection of VxRail server resources that have been aggregated into logical pools. It continuously balances and optimizes compute resource allocation among the VMs.

When a VM experiences an increased workload, DRS evaluates the VM priority against user-defined resource-allocation rules and policies. If justified, DRS allocates additional resources. It can also be configured to dedicate consistent resources to the VMs of particular business-unit applications to meet SLAs and business requirements.

DRS allocates resources to the VM either by migrating the VM to another server with more available resources or by making more “resources” for the VM on the same server by migrating other VMs off the server. In the VxRail Appliance, all ESXi hosts are part of a vMotion network. The live migration of VMs to different node servers is completely transparent to end users through vMotion (see the figures below). DRS adds tremendous value to the VxRail cluster by automating VM placement and ensuring consistent and predictable application-workload performance.

Figure 37 DRS movement of VMs across node servers

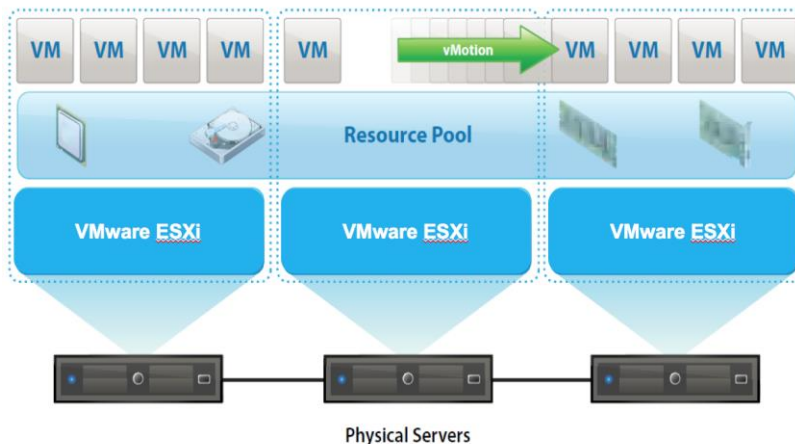
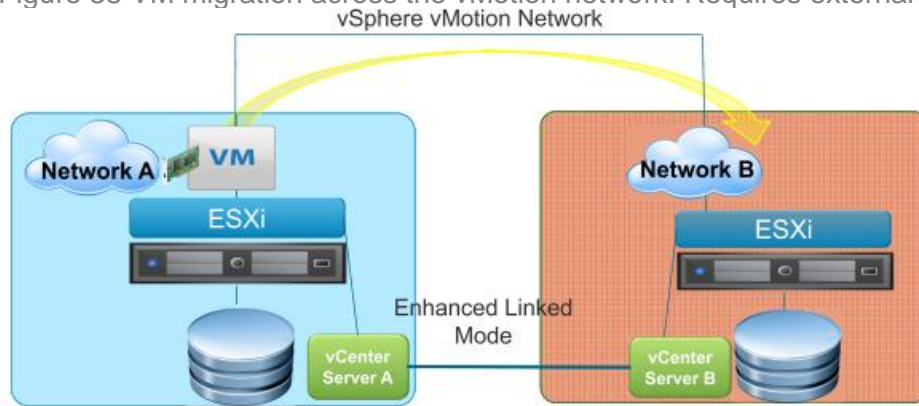


Figure 38 VM migration across the vMotion network. Requires external vCenter.

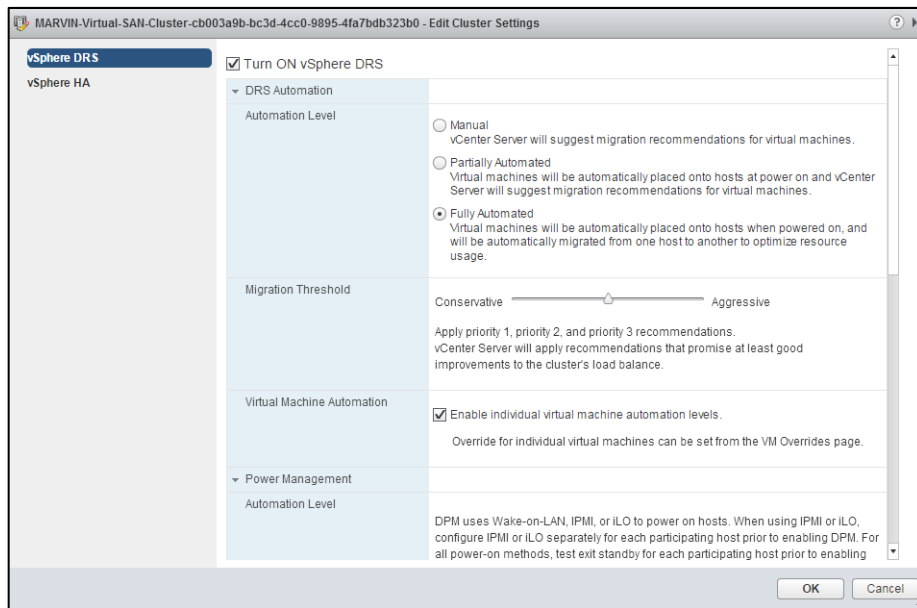


DRS offers a considerable advantage to VxRail users during maintenance situations, because it automates the tasks normally involved in manually moving live machines during upgrades or repairs. DRS facilitates maintenance automation, providing transparent, continuous operations by dynamically migrating all VMs to other physical servers. That way, servers can be attended to for maintenance, or new node servers can be added to a resource pool, all while DRS automatically redistributes the VMs among the available servers as the physical resources change.

In other words, DRS dynamically balances VMs as soon as additional resources become available when new server is added or when an existing server has finished its maintenance cycle. DRS allocates only CPU and memory resources for the VMs and uses vSAN for shared storage.

The following figure shows the settings for configuring DRS.

Figure 39 Configuring DRS settings



Some conditions and business operations warrant a more aggressive DRS migration strategy than others. Adjustable cluster parameters establish the thresholds that trigger DRS migrations (as shown in the screen shot above.) For example, a Level-2 threshold only applies specified migration recommendations to make a significant impact on the cluster's load balance, whereas a Level-5 threshold applies all the recommendations to even slightly improve the cluster's load balance.

DRS applies only to VxRail virtual machines. (vSAN uses a single datastore and handles placement and balancing internally. vSAN does not currently support Storage DRS or Storage I/O Control.)

vSphere High Availability (HA)

vSphere provides several solutions to ensure a high level of availability, during both planned and unplanned downtime scenarios. vSphere depends on the following technologies to make sure that virtual machines running in the environment remain available:

Virtual machine migration

Multiple I/O adapter paths

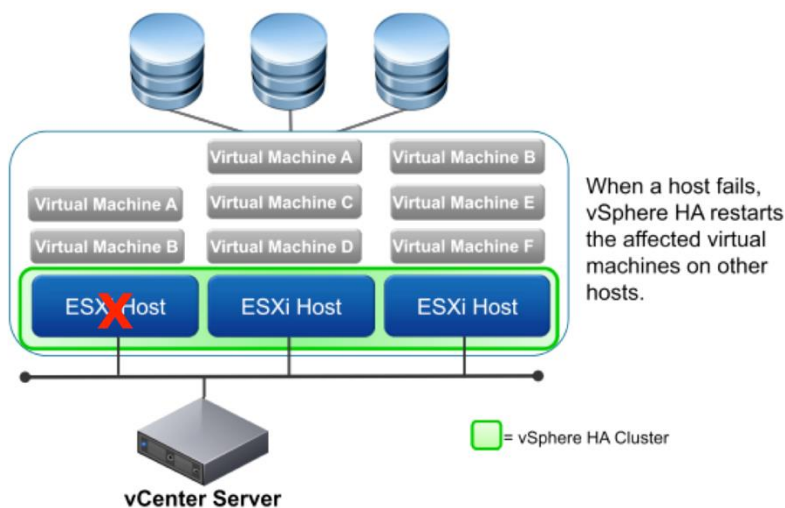
Virtual machine load balancing

Failure tolerance

Disaster recovery

Together with vSAN, vSphere HA produces a resilient, highly available solution for VxRail virtual machine workloads. vSphere HA protects virtual machines by restarting them in the event of a host failure. (See the figure below.) It leverages the ESXi cluster configuration to ensure rapid recovery from outages, providing cost-effective high availability for applications running in virtual machines. When a host joins a cluster, its resources become part of the cluster resources. The cluster manages the resources of all hosts within it. In a vSphere environment, ESXi clusters are responsible for vSphere HA, DRS, and the vSAN technology that provides VxRail software-defined storage capabilities. See the figure below.

Figure 40 vSphere HA



vSphere HA provides several points of protection for applications:

It circumvents any server failure by restarting the virtual machines on other hosts within the cluster.

It continuously monitors virtual machines and resets any detected VM failures.

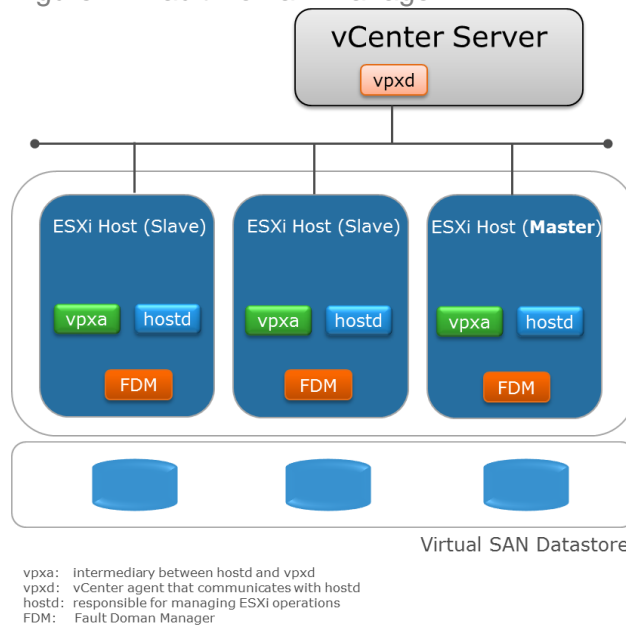
It protects against datastore accessibility failures and provides automated recovery for affected virtual machines. With Virtual Machine Component Protection (VMCP), the affected VMs are restarted on other hosts that still have access to the datastores.

It protects virtual machines against network isolation by restarting them if their host becomes isolated on the management or VMware vSAN network. This protection is provided even if the network has become partitioned.

Once vSphere HA is configured, all workloads are protected. No actions are required to protect new virtual machines and no special software needs to exist within the application or virtual machine.

Included in the failover capabilities in vSphere HA is a service called the Fault Domain Manager (FDM) that runs on the member hosts (shown in the figure below). After the FDM agents have started, the cluster hosts become part a fault domain, and a host can exist in only one fault domain at a time. Hosts cannot participate in a fault domain if they are in maintenance mode, standby mode, or disconnected from vCenter Server.

Figure 41 Fault Domain Manager



FDM uses a master-slave operational model (see the figure above). An automatically designated master host manages the fault domain, and the remaining hosts are slaves. FDM agents on slave hosts communicate with the FDM service on the master host using a secure TCP connection. In VxRail clusters, vSphere HA is enabled only after the vSAN cluster has been configured. Once vSphere HA has started, vCenter Server contacts the master host agent and sends it a list of cluster-member hosts along with the cluster configuration. That information is saved to local storage on the master host and then pushed out to the slave hosts in the cluster. If additional hosts are added to the cluster during normal operation, the master agent sends an update to all hosts in the cluster.

The master host provides an interface to vCenter Server for querying and reporting on the state of the fault domain and virtual-machine availability. vCenter Server governs the vSphere HA agent, identifying the virtual machines to protect and maintaining a VM-to-host compatibility list.

The agent learns of state changes through hostd, and vCenter Server learns of them through vpxa. The master host monitors the health of the slaves and takes responsibility for virtual machines that had been running on a failed slave host. Meanwhile, the slave host monitors the health of its local virtual machines and sends state changes to the master host. A slave host also monitors the health of the master host.

vSphere HA is configured, managed, and monitored through vCenter Server. Cluster configuration data is maintained by the vCenter Server vpxd process. If vpxd reports any cluster configuration changes to the master agent, the master advertises a new copy of the cluster configuration information and then each slave fetches the updated copy and writes the new information to local storage. Each datastore includes a list of protected virtual machines. The list is updated after vCenter Server notices any user-initiated power-on (protected) or power-off (unprotected) operation.

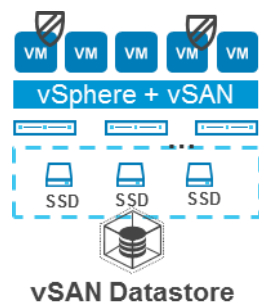
vCenter Server Watchdog

One method of providing vCenter Server availability is to use the Watchdog feature in a vSphere HA cluster. Watchdog monitors and protects vCenter Server services. If any services fail, Watchdog attempts to restart them. If it cannot restart the service because of a host failure, vSphere HA restarts the virtual machine running the service on a new host. Watchdog can provide better availability by using vCenter Server processes (PID Watchdog) or the vCenter Server API (API Watchdog).

vSphere Encryption

vSphere encryption enables customers to encrypt data on a per VM level. This level of encryption is ideal for customers who are concerned about rogue admins sending a VM and all its data to a non-secure location. Which VMs should be encrypted, is up to the virtualization administrative team and can be selected on a per VM basis (as seen in the figure below). A KMIP-compliant Key Management Server like CloudLink or Hytrust is required.

Figure 42 Per VM-level encryption with vSphere Encryption



vSAN

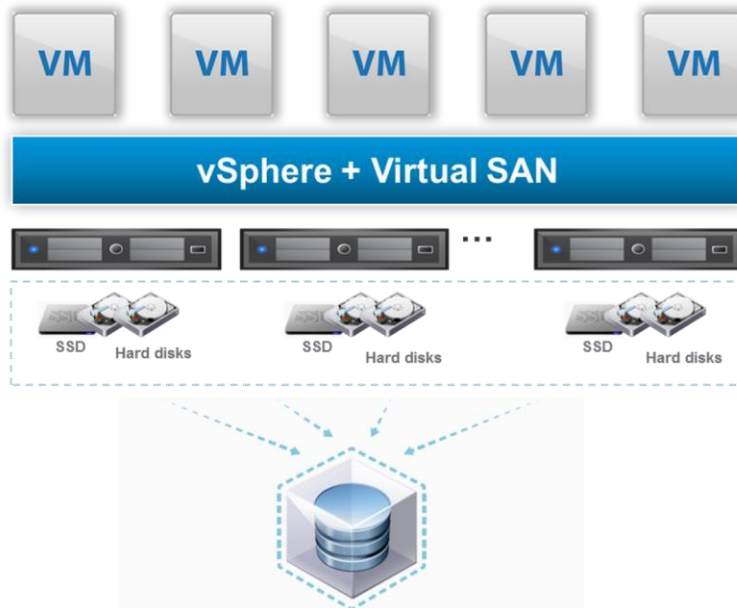
VxRail Appliances leverage VMware's vSAN software, which is fully integrated with vSphere and provides full-featured and cost-effective software-defined storage. vSAN implements a notably efficient architecture, built directly into hypervisor. This distinguishes vSAN from solutions that typically install a virtual storage appliance (VSA) that runs as a guest VM on each host. Embedding vSAN into the ESXi kernel layer has obvious advantages in performance and memory requirements. It has very little impact on CPU utilization (less than 10 percent) and self-balances based on workload and resource availability. It presents storage a familiar datastore construct and works seamlessly with other vSphere features such as vMotion.

vSAN aggregates locally attached disks of hosts in a vSphere cluster to create a pool of distributed shared storage. Capacity is easily scaled up by adding additional disks to hosts in the cluster and scaled out by adding additional ESXi hosts. This provides the flexibility to start with a very small environment and scale it over time. Storage characteristics are configured using Storage Policy Based Management (SPBM), which allows object-level policies to be set and modified on the fly to control storage provisioning and day-to-day management of storage service-level agreements (SLAs). vSphere and vSAN provide the foundation for VxRail performance and scale, and enable enterprise-class capabilities for hosted virtual machines including mobility using vMotion, High Availability, and Dynamic Resource Scheduler (DRS).

vSAN is preconfigured when the VxRail system is first initialized and managed through vCenter. The VxRail appliance-initialization process discovers locally attached storage disks from each ESXi node in the cluster to create a distributed, shared-storage datastore. The amount of storage in the vSAN datastore is an aggregate of all of the capacity drives in the cluster.

The figure below shows an example of a hybrid configuration where each node contributes storage capacity to the vSAN datastore. The SSD drive provides caching to optimize performance and hard disk drives (HDD) for capacity.

Figure 43 vSAN datastore



vSAN enables rapid storage provisioning within vCenter as part of the VM-creation and deployment operations. vSAN is policy driven and designed to simplify storage provisioning and

management. It automatically and dynamically matches requirements with underlying storage resources based on VM-level storage policies. VxRail provides two different vSAN node-storage configuration options: a hybrid configuration that leverages both flash SSDs and mechanical HDDs, and an all-flash SSD configuration. The hybrid configuration uses flash SSDs at the cache tier and mechanical HDDs for capacity and persistent data storage. The all-flash configuration uses flash SSDs for both the caching tier and capacity tier.

Disk groups

Disk drives in VxRail hosts are organized into disk groups, and disk groups contribute storage to the vSAN cluster. In a VxRail appliance, a disk group contains a maximum of one flash-cache drive and six capacity devices. Depending on the model, each VxRail node can be configured with up to four disk groups. The following figure shows the number of disk groups per node and the number of drives per disk group for each VxRail model.

Figure 44 Disk group configurations for VxRail based on 14th generation servers

Model	Disk Slots	Fixed Cache Disk Slots	Maximum Disk Groups (DG)	Max Capacity Disks Per DG
E Series	10 (2.5-inch)	8,9	2	4
P Series, V Series	24 (2.5-inch)	20,21,22,23	4	5
S Series	12 (3.5-inch) + 2 (2.5-inch)	12,13 (2.5-inch)	2	6

The disk group layout, based on capacity, performance, and availability, is determined when the system is designed and sized. Below are the considerations for designing disk-group layout:

The number of SSD cache drives. Each disk group requires one and only one high-endurance SSD flash-cache drive. More disk groups require more SSD flash-cache drives. Currently, the VxRail Appliance is offered with 400GB, 800GB, and 1600GB cache flash devices. For VxRail, the SSD cache drive must be installed in designated slots. See product documentation for specific slot locations.

The total number of capacity drives available. Each disk group requires at least one capacity drive. Only capacity drive counts when determining the total capacity available for a vSAN Cluster.

Cache requirements. Cache improves read and write performance in a hybrid storage configuration and write performance and endurance in an all-flash configuration. The optimal amount of cache depends on the active working set size. Workloads that have a larger working set size may require more cache, and this can be configured by either using larger SSD cache drives or configuring multiple disk groups with each disk group having one cache drive.

Note: All-flash configurations support a maximum of 600GB of write cache per disk group. Consider multiple disk groups for larger write cache per node.

Performance. For the same total capacity, a larger number of smaller drives will provide more IOPs than fewer larger drives. Depending on the model, configuring more drives may require multiple disk groups.

Fault Domains and recoverability. If the SSD cache drive fails, the full disk group goes offline and must be recovered. Larger disk groups take longer to recover. A node with only a single disk group will not be contributing any IOPs to the cluster during recovery, and this can reduce cluster performance.

Hybrid and All-Flash differences

Cache is used differently in hybrid and all-flash configurations. In hybrid disk-group configurations (which use mechanical HDDs for capacity and flash SSD devices for the caching), the caching algorithm attempts to maximize both read and write performance. The flash SSD device serves two purposes: a read cache and a write buffer. 70 percent of the available cache is allocated for storing frequently read disk blocks, minimizing accesses to the slower mechanical disks. The remaining 30 percent of available cache is allocated to writes. Multiple writes are coalesced and written sequentially if possible, again maximizing mechanical HDD performance.

In all-flash configurations, one designated SSD flash device is used for the cache, while additional SSD flash devices are used for the capacity. In all-flash disk-group configurations, there are two types of flash SSDs: a very fast and durable flash device that functions as write cache and more cost-effective SSD devices that function as capacity. Here, the cache-tier SSD is 100 percent allocated for writes. None of the flash cache is used for reads; read performance from capacity-tier flash SSDs is more than sufficient for high performance. Many more writes can be held by the cache SSD in an all-flash configuration, and writes are only written to capacity when needed, which extends the life of the capacity-tier SSD.

While both configurations dramatically improve the performance of VMs running on vSAN, all-flash configurations provide the most predictable and uniform performance regardless of workload.

Read cache: Basic function

The read cache, which only exists in hybrid configurations, keeps a collection of recently read disk blocks. This reduces the I/O read latency in the event of a cache hit; that is, the disk block can be fetched from cache rather than mechanical disk. For a given VM data block, vSAN always reads from the same replica/mirror. When there are multiple replicas (to tolerate failures), vSAN divides up the caching of the data blocks evenly between the replica copies.

If the data block being read from the first replica is not in cache, the directory service is referenced to discover whether or not the data block exists in the cache of another mirror (on another host) in the cluster. If the data block is found there, the data is retrieved. If the data block is not in cache on the other host, then there is a read-cache miss. In that case, the data is retrieved directly from the mechanical HDD.

Write cache: Basic function

The write cache, found in both hybrid and all-flash configurations, behaves as a non-volatile write buffer. This greatly improves performance in both hybrid and all-flash configurations and also extends the life of flash capacity devices in all-flash configurations. When writes are written to cache, vSAN ensures that a copy of the data is written elsewhere in the cluster. All VMs deployed with vSAN are set with a default availability policy that ensures at least one additional copy of the VM data is available. This includes making sure that writes end up in multiple write caches in the cluster.

Once an application running inside the guest OS initiates a write, it is duplicated to the write cache on the hosts that include replicas of the storage objects. This means that, in the event of a host failure, a copy of the data is in cache and no data loss occurs. The VM simply uses the replicated copy of the cache data.

Flash endurance

Flash endurance is related to the number of write/erase cycles that the cache-tier flash SSD can tolerate over its lifespan. It is important to choose the right level of endurance for each application or use case, according to the level of write activity that is expected.

For vSAN and VxRail configurations, the endurance specification uses Terabytes Written per day (TBW) as a metric to determine endurance.

For all-flash vSAN deployments, the specification for the cache device is 4TBW per day, which is an appropriate endurance for write intensive workloads.

vSAN impact on flash endurance

There are two commonly used approaches to improve NAND flash endurance: improve wear leveling and minimize write activity. A distributed storage implementation that focuses on localizing data on the same node where the VMs reside prevents the distribution of the writes across all the drives in the cluster. This localization inevitably increases drive usage, leading to early drive replacement.

In contrast, vSAN distributes the objects and components of a VM across all the disk groups in the VxRail cluster. This distribution significantly improves wear leveling and reduces write activity by deferring writes. vSAN also reduces writes by employing data-reduction techniques such as de-duplication and compression.

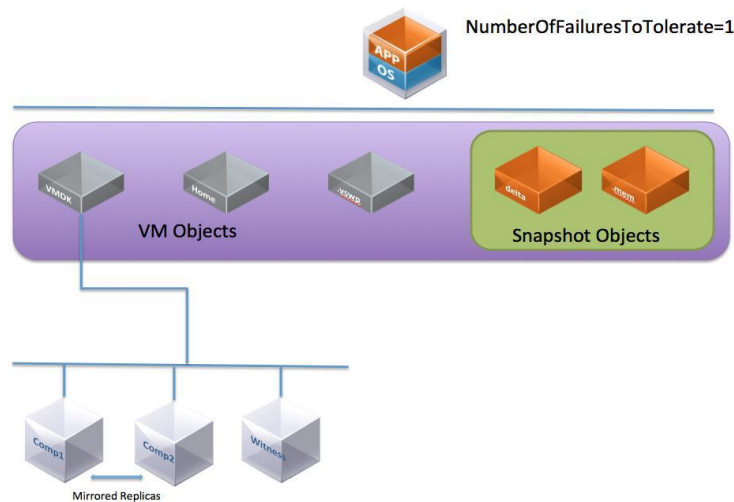
Client cache

The client cache is used on both hybrid and all-flash configurations. It leverages local DRAM server memory (client cache) within the node to accelerate read performance. The amount of memory allocated is .4% -1GB per host. vSAN first tries to fulfill the read request from the local client cache, so the VM can avoid crossing the network to complete the read. If the data is unavailable in the client cache, the cache-tier SSD is queried to fulfill the read request. The client cache benefits read cache-friendly workloads.

Objects and components

VxRail virtual machines are made up of a set of objects. For example, a VMDK is an object, a snapshot is an object, VM swap space is an object, and the VM home namespace (where the .vmx file, log files, and so on are stored) is also an object. See the figure below.

Figure 45 vSAN objects and components



Virtual-machine objects are split into multiple components based on performance and availability requirements defined in the storage policy applied to the objects of the VM. For example, if the VM is deployed with a policy to tolerate a single failure, the objects have two replica components. Distributed storage uses a disk-striping process to distribute data blocks across multiple devices. The stripe itself refers to a slice of data; the striped device is the individual drive that holds the stripe. If the policy contains a stripe width, the object is striped across multiple devices in the capacity layer, and each stripe is an object component.

Each vSAN host has a maximum of 9,000 components. The largest component size is 255GB. For objects greater than 255GB, vSAN automatically divides them into multiple components. For example, a VMDK of 62TB generates more than 500 x 255GB components. The figure above illustrates how components that make up VM objects are spread across drive on nodes based on Failure to tolerate policy.

Witness

In vSAN, witnesses are an integral component of every storage object when the object is configured to tolerate at least one failure and when using mirroring as the Failure Tolerance Method (FTM). Witnesses are components that contain no data, only metadata. Their purpose is to serve as tiebreakers when availability decisions are made to meet the Failures to tolerate (FTT) policy setting, and they're used when determining if a quorum of components exist in the cluster.

In vSAN, storage components can be distributed in such a way that they can guarantee availability without relying on a witness. In this case, each component has a number of votes—at least one or more. Quorum is calculated based on the rule that requires "more than 50 percent of votes."

Replicas

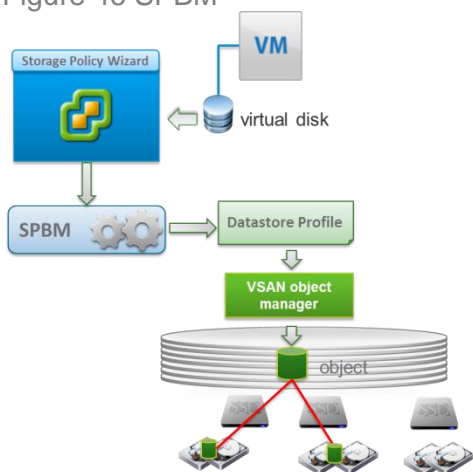
Replicas make up the virtual machine's storage objects. Replicas are instantiated when an availability policy (FTT) is specified for the virtual machine. The availability policy dictates how many replicas are created and lets virtual machines continue running with a full complement of data even when host, network, or disk failures occur in the cluster.

Storage Policy Based Management (SPBM)

vSAN policies define virtual-machine storage requirements for performance and availability. These policies determine how storage objects are provisioned and allocated within the datastore to guarantee the required level of service.

vSAN implements Storage Policy Based Management, and each virtual machine deployed in a vSAN datastore has at least one assigned policy. When the VM is created and assigned a storage policy, the policy requirements are pushed to the vSAN layer. See the figure below.

Figure 46 SPBM



Storage policies are a set of rules and are assigned to VMs either manually or a default policy is automatically assigned. A system may have multiple storage policies. For instance, all virtual machines that include PROD-SQL in their name or resource group might be set at RAID 1 and a five-percent read-cache reservation, and TEST-WEB would be automatically set to RAID 0.

Dynamic policy changes

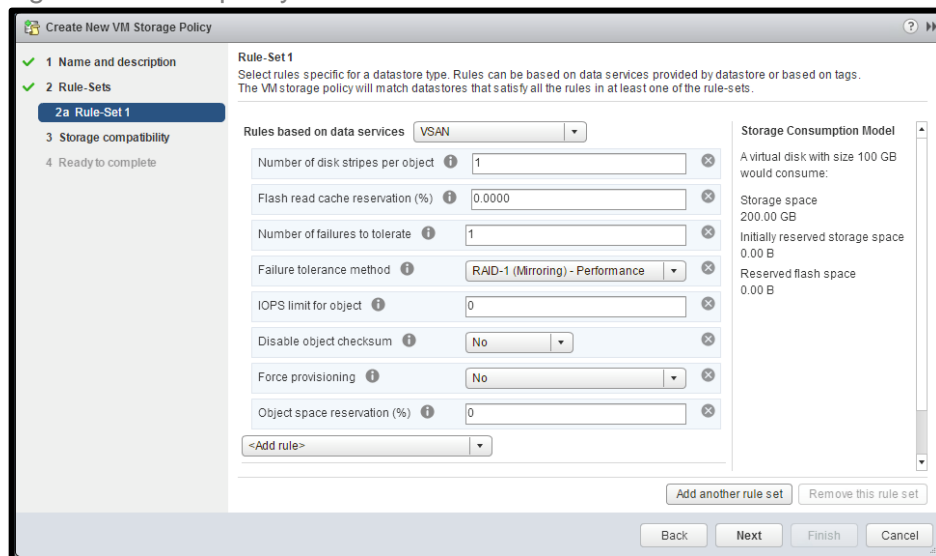
Administrators can dynamically change a VM storage policy. When changing attributes such as the number of Failures to tolerate, vSAN attempts to find a new placement for a replica with the new configuration. In some cases, existing parts of the current configuration can be reused, and the configuration just needs to be updated or extended. For example, if an object currently uses Failures to tolerate=1, and the user asks for Failures to tolerate=2, vSAN can simply add another mirror (and witness).

In other cases, such as changing the stripe width from one to two, vSAN cannot reuse existing replicas, and it creates a brand new replica (or replicas) without impacting the existing objects. A storage object has a status of *Compliant* if the attributes match the policy. When changing a policy online, the object may show as “*not compliant*” while vSAN reconfigures the object.

Storage policy attributes

The figure below displays an example of the rule set for a storage policy.

Figure 47 vSAN policy attributes



Number of disk stripes per object

This rule establishes the minimum number of capacity devices used for striping each virtual machine replica. A value higher than 1 might result in better performance, but it also results in higher resource consumption. The default value is the minimum, 1; the maximum value is 12. The stripe size is 1MB.

vSAN may decide that an object needs to be striped across multiple disks without any stripe-width policy requirement. The reason for this can vary, but typically it occurs when a VMDK is too large to fit on a single physical drive. However, when a particular stripe width is required, then it should not exceed the number of disks available to the cluster.

Flash read cache reservation

Flash cache reservation refers to flash capacity reserved as read cache for the virtual machine object, and it applies to hybrid configurations only. By default, vSAN dynamically allocates read cache to storage objects based on demand. As a result, no need typically exists to change the default 0 value for this parameter.

In very specific cases, when a small increase in the read cache for a single VM can provide a significant change in performance, it is an option. It should be used with caution to avoid wasting resources or taking resources from other VMs.

The default value is 0 percent. Maximum value is 100 percent.

Failures to tolerate

This FTT option generally defines the number of host and device failures that a virtual machine object can tolerate. For n failures tolerated, $n+1$ copies of the VM object area created and $2n+1$ hosts with storage are required.

The default value is 1. Maximum value is 3.

When erasure coding is enabled for a cluster (by setting FTM=Capacity), RAID 5 is applied if the number of Failures to tolerate is set to 1, and RAID 6 is applied if the number of Failures to tolerate is set to 2. Note a vSAN cluster requires a minimum of four nodes for RAID 5 and six nodes for RAID 6.

Failure tolerance method

Failure tolerance method (FTM) specifies whether the data replication method optimizes for performance or capacity. The RAID 1 failure tolerance method provides better performance and consumes less memory and network resources but uses more disk space. RAID 5/6 erasure coding provides more usable capacity but consumes more CPU and network resources. (An upcoming section on erasure coding section provides additional information.)

IOPS limit for object (QoS)

This attribute establishes Quality of Service (QoS) for an object by defining an upper limit on the number IOPS a VM/VMDK can perform. The default behavior is that all objects are not limited, and low priority applications could potentially consume resources in a manner that could impact more important workloads. This rule allows different limits for different applications and can be used to keep workloads from impacting each other (the noisy-neighbor issue) or to supply service-level agreements (SLAs) for different workloads and maintain performance for tier-1 applications.

IOPS is calculated based on all of the operations on this VM/VMDK (read & write) including snapshots.

A few notes regarding IOPS limits for objects:

When calculating IOPS, read and write are considered equivalent, but keep in mind that cache-hit ratio and sequentially are not considered.

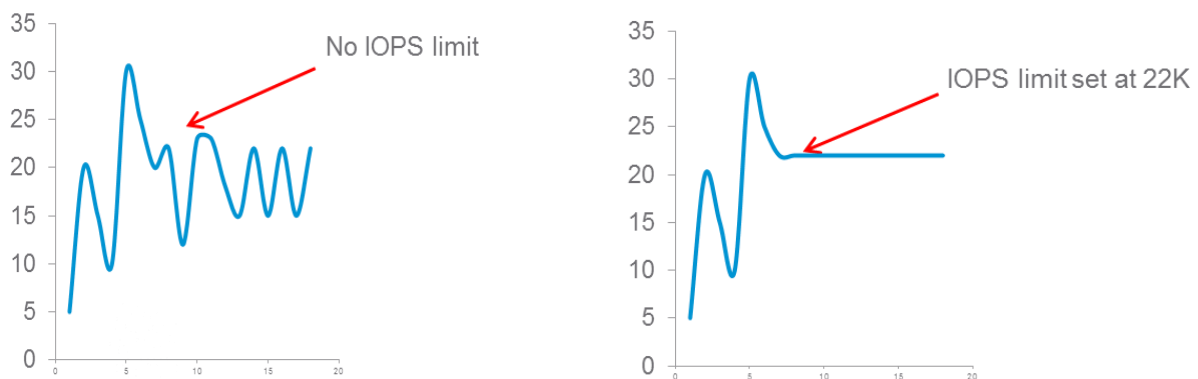
When an object exceeds its disk IOPS limit, I/O operations are throttled.

If the IOPS limit for object is set to 0, IOPS limits are not enforced.

vSAN allows the object to double the IOPS-limit rate during the first second of operation or after a period of inactivity.

The figure below illustrates how IOs are throttled when the IOPS limit policy is set.

Figure 48 IOPs limits establishes QoS



Disable object checksum

vSAN uses end-to-end checksum to ensure data integrity by confirming that each copy of an object's 4K chunk of data is exactly the same. The checksum is five bytes in length and persists with the data. When data is written, the checksum is verified to ensure no corruption occurred when the data transverses the network. When data is read, checksum verifies that the data read matches what was written. If an error is detected; vSAN uses a replica to recover the data and logs an error. Checksum calculation and error-correction are transparent to the user.

The default setting for all objects in the cluster is On, which means that checksum is enabled. The best practice is not to change this setting because detecting data corruption is a critical and valuable feature of vSAN and should not be disabled.

Force provisioning

If this option is set to Yes, the object is provisioned even if the rules specified in the storage policy cannot be satisfied by the datastore.

This parameter is sometimes used during an outage when resources are limited and normal provisioning policy cannot be satisfied.

The default is No and will not allow an object to be created if the policy rules cannot be satisfied. This is appropriate for most production environments. If set to Yes, an object can be created even not enough resources are available to satisfy the policy rules. In this case, the object will be displayed as Not Compliant.

Object space reservation

Object space reservation is specified as a percentage of the total object size. It reflects the reserved, thick-provisioned space required for deploying virtual machines.

The default value is 0%. Maximum value is 100%.

The value should be set either to 0% or 100% when using RAID-5/6 in combination with deduplication and compression.

Sparse Swap

Sparse Swap is a vSAN feature worth highlighting for its space efficiency, and it is available for both all-flash and hybrid environments. By default, vSAN swap files are thick provisioned—created with 100 percent space reservation. (For example, if the VM has 4GB of RAM, then the swap file also has 4GB of RAM.) While this guarantees sufficient capacity for the VM, it can consume too much memory—especially in large clusters with a lot of virtual machines. When Sparse Swap is enabled as an advanced host setting, the VM reserves less than 100 percent of memory space for the swap objects.

Sparse Swap is established by enabling the *Swap thick provision disabled* setting. Sparse swap files have an FTT setting of 1 and a FTM setting of RAID1 (Mirroring).

Leveraging Sparse Swap is an effective way to reduce memory overhead in environments that are not already over-committing memory, especially for all-flash VxRail environments where swap files represent a sizeable portion of the total required datastore capacity.

I/O paths and caching algorithms⁴

This section elaborates on some of the vSAN concepts that have been introduced so far with additional, general information about vSAN caching algorithms. The next paragraphs briefly describe how vSAN leverages flash, memory, and rotating disks. They also illustrate the I/O paths between the guest OS and the persistent storage areas.

Read caching

Each disk group contains an SSD drive used as a cache tier. On a hybrid system, 70 percent of the capacity is used by default for read cache (RC). The most active data is maintained in the cache tier and improves performance by minimizing the latency impact of reading from mechanical disk.

The RC is organized in terms of cache lines. They represent the unit of data management in RC, and the current size is 1MB. Data is fetched into the RC and evicted at cache-line granularity. In addition to the SSD read cache, vSAN also maintains a small in-memory (RAM) read cache that holds the most-recently accessed cache lines from the RC. The in-memory cache is dynamically sized based on the available memory in the system.

vSAN maintains in-memory metadata that tracks the state of the RC (both SSD and in memory), including the logical addresses of cache lines, valid and invalid regions in each cache line, aging information, etc. These data structures are designed to compress for efficiencies, using memory space without imposing a substantial CPU overhead on regular operations. No need exists to swap RC metadata in or out of persistent storage. (This is one area where VMware holds important IP.)

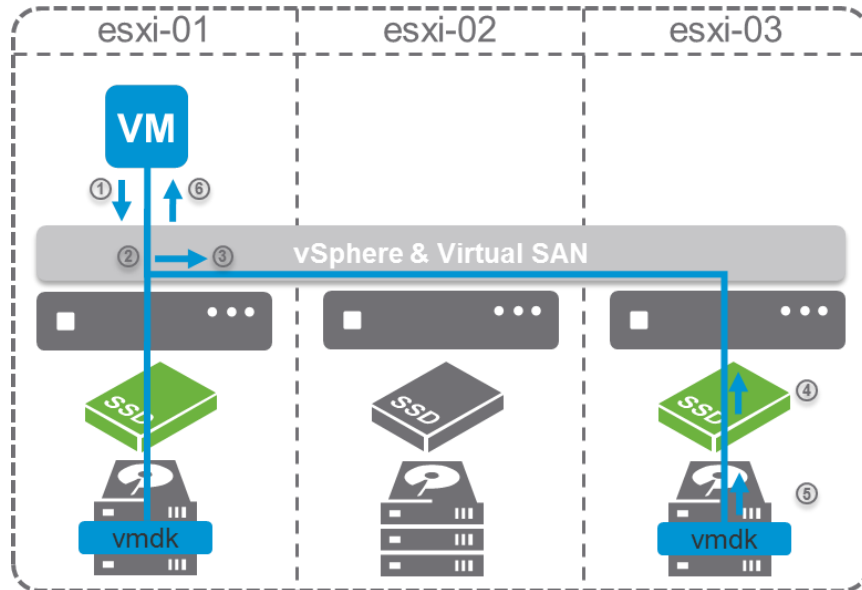
Read-cache contents are not tracked across power-cycle operations of the host. If power is lost and recovered, then the RC is re-populated (warmed) from scratch. So, essentially RC is used as a fast storage tier, and its persistence is not required across power cycles. The rationale behind this approach is to avoid any overheads on the common data path that would be required if the RC metadata was persisted every time RC was modified—such as cache-line fetching and eviction, or when write operations invalidate a sub-region of a cache line.

⁴ Much of the content in this *specific* section has been extracted from an existing technical whitepaper: *An overview of VMware vSAN Caching Algorithms*.

Anatomy of a hybrid read

Read operations follow a defined procedure. To illustrate, the VMDK in the example below has two replicas on esxi1 and esxi3. See the figure below.

Figure 49 Hybrid read



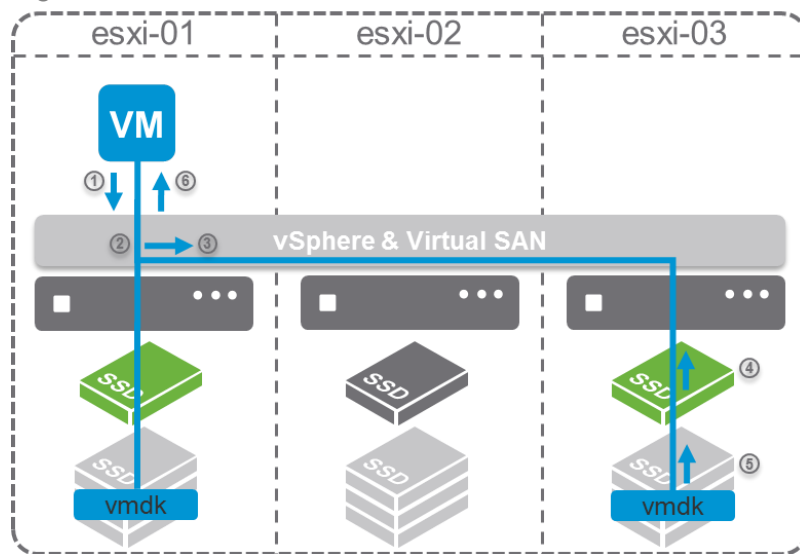
1. Guest OS issues a read on virtual disk
2. Owner chooses replica to read from
 - Load balance across replicas
 - Not necessarily local replica (if one)
 - A block always reads from same replica
3. At chosen replica (esxi-03): read data from flash write buffer, if present
4. At chosen replica (esxi-03): read data from flash read cache, if present
5. Otherwise, read from HDD and place data in flash read cache
 - Allocate a 1MB buffer for the missing cache line and replace “coldest” data (eviction of coldest data to make room for new read)
 - Each missing line is read from the HDD as multiples of 64KB chunks, starting with the chunks that contain the referenced data
6. Return data to owner
7. Complete read and return data to VM
8. Once the 1MB cache line is added to the in-line read cache, its population continues asynchronously. This occurs to explore both the spatial and temporal locality of reference, increasing the changes that the next reads will find in the read cache.

Anatomy of an All-Flash read

1. Guest OS issues a read on virtual disk
2. Owner chooses replica to read from
 - Load balance across replicas
 - Not necessarily local replica (if one)
3. At chosen replica (esxi-03): read data from flash write buffer, if present
4. Otherwise, read from capacity flash device
5. Return data to owner
6. Complete read and return data to VM

The operation is shown in the figure below.

Figure 50 All-flash read



The major difference is that read-cache misses cause no serious performance degradation. Reads from flash capacity devices should be almost as quick as reads from the cache SSD. Another significant difference is that no need exists to move the block from the capacity layer to the cache layer, as in hybrid configurations.

Write caching

In hybrid configurations write-back caching is done entirely for performance. The aggregate-storage workloads in virtualized infrastructures are almost always random, because of the statistical multiplexing of the many VMs and applications that share the infrastructure.

HDDs can perform only a small number of random I/O with a high latency compared to SSDs. So, sending the random write part of the workload directly to spinning disks can cause performance degradation. On the other hand, magnetic disks exhibit decent performance for sequential workloads. Modern HDDs may exhibit sequential-like behavior and performance even when the workload is not perfectly sequential. "Proximal I/O" suffices.

In hybrid disk groups, vSAN uses the write-buffer (WB) section of the SSD (by default, 30 percent of device capacity), as a write-back buffer that stages all the write operations. The key objective is to de-stage written data (not individual write operations) in a way that creates a benign, near-sequential (proximal) write workload for the HDDs that form the capacity tier.

In all-flash disk groups, vSAN utilizes the tier-1 SSD entirely as a write-back buffer (100 percent of the device capacity—up to a maximum of 600GB). The purpose of the WB is quite different in this case. It absorbs the highest rate of write operations in a high-endurance device and allows only a trickle of data to be written to the capacity flash. This approach allows low-endurance, larger-capacity SSDs for capacity.

Nevertheless, capacity SSDs are capable of serving very large numbers of read IOPS. Thus, no read caching occurs, except when the most-recent data referenced by a read operation still resides in the WB.

In both hybrid and All-flash, every write operation is handled through transactional processes: A record for the operation is persisted in the transaction log in the SSD.

The data (payload) of the operation is persisted in the WB.

Updated in-memory tables reflect the new data and its logical address space (for tracking) as well as its physical location in the capacity tier.

The write operation completes upstream after the transaction has committed successfully.

Under typical steady-state workloads, the log records of multiple write operations are coalesced before they are persisted in the log. This reduces the amount of metadata-write operations for the SSD. By definition, the log is a circular buffer, written and freed in a sequential fashion. Thus write amplification can be avoided (good for device endurance). The WB region allocates blocks in a round-robin fashion, keeping wear leveling in mind.

Even when a write operation overwrites existing WB data, vSAN never rewrites an existing SSD page in place. Instead, it allocates a new block and updates metadata to reflect that the old blocks are invalid. vSAN fills an entire SSD page before it moves to the next one. Eventually, entire pages are freed when all their data is invalid. (It is very rare to re-buffer data to allow SSD pages to be freed).

Also, because the device firmware does not have visibility into invalidated data, it sees no “holes” in pages. In effect, internal write leveling (by moving data around to fill holes in pages) is all but eliminated. This extends the overall endurance of a device. The vSAN design has gone to great lengths to minimize unnecessary writes to maximize cache SSD endurance. As a result, the life expectancy of SSDs implemented in vSAN may exceed the manufacturers’ specifications, which are developed with more generic workloads in mind.

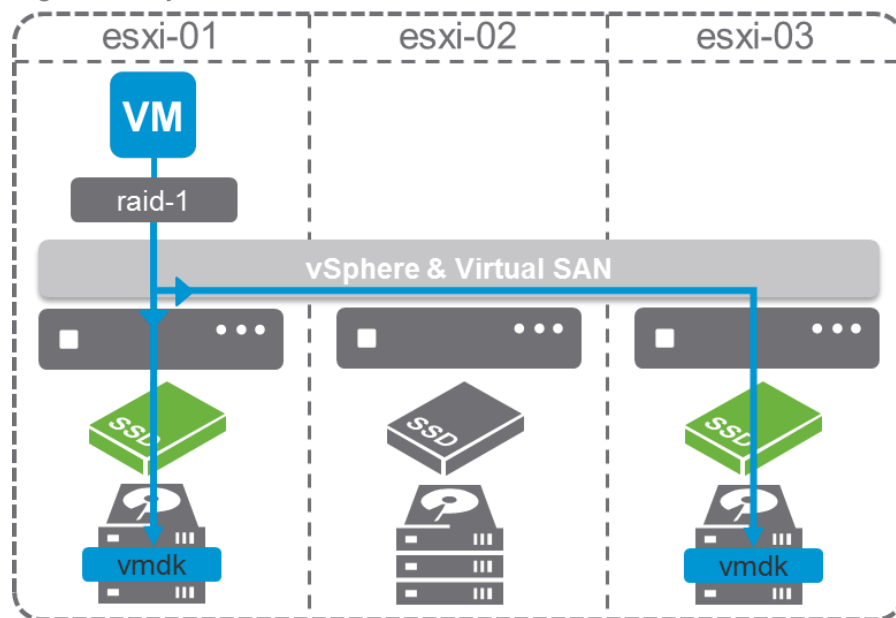
Anatomy of a write I/O—hybrid and All-Flash (FTM=mirroring)

1. VM running on host esxi-01.
2. esxi-01 is owner of virtual disk object.
 - *Failures to tolerate = 1*
3. Object has two (2) replicas on esxi-01 and esxi-03.
4. Guest OS issues write op to virtual disk.

5. Owner clones write operation.
 - In parallel: sends write op to esxi-01 (locally) and esxi-03
6. esxi-01, esxi-03 persist operation to flash (log).
7. esxi-01, esxi-03 ACK-write operation to owner.
8. Owner waits for ACK from both writes and completes I/O.
9. Later, backend hosts commit batch of writes.

The figure below illustrates the operation.

Figure 51 Hybrid and flash write I/O



Distributed caching considerations

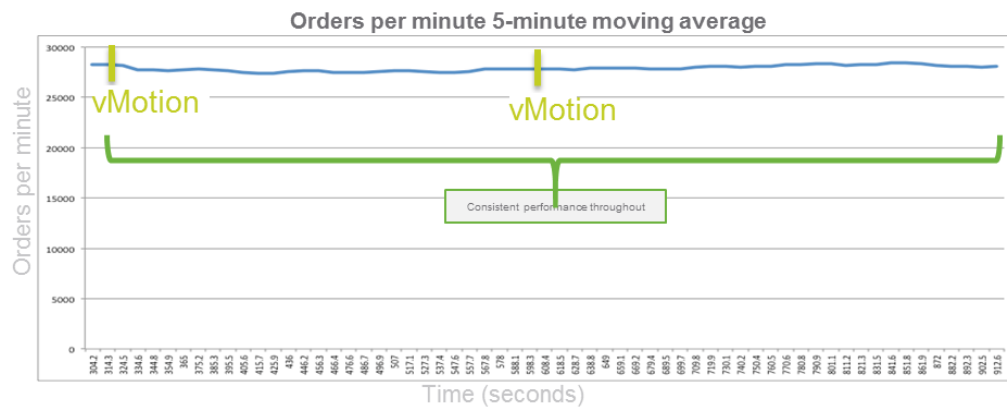
vSAN caching algorithms and data-locality techniques reflect a number of objectives and observations pertaining to distributed caching:

vSAN exploits temporal and spatial locality for caching.

vSAN implements a distributed, persistent cache on flash across the cluster. Caching is done in front of the disks where the data replicas live, not on the client side. A distributed-caching mechanism results in better overall flash-cache utilization.

Another benefit of distributed caching is during VM migrations, which can happen in some datacenters over ten times a day. With DRS and vMotion, VMs can move around from host to host in a cluster. Without a distributed cache, the migrations would have to move around a lot of data and rewarm caches every time a VM migrates. As the figure below illustrates, vSAN prevents any performance degradation after a VM migration.

Figure 52 vSAN prevents performance degradation after VM migration



The network introduces a small latency when accessing data on another host. Typical latencies in 10GbE networks range from 5 – 50 microseconds. Typical latencies of a flash drive, accessed through a SCSI layer, are near 1ms for small (4K) I/O blocks. So, for the majority of the I/O executed in the system, the network impact adds near 0.1 percent to the latency.

Few workloads are actually cache friendly, meaning that they don't take advantage of the way small increases in cache size can significantly increase the rate of I/O. These workloads can benefit from local cache, and enabling the Client Cache would be the right approach.

vSAN works with a View Accelerator (deduplicated, in-memory read cache), which is notably effective for VDI use cases.

vSAN features Client Cache that leverages DRAM memory local to the virtual machine to accelerate read performance. The amount of memory allocated is anywhere from .4 percent to 1GB per host.

vSAN high availability and fault domains

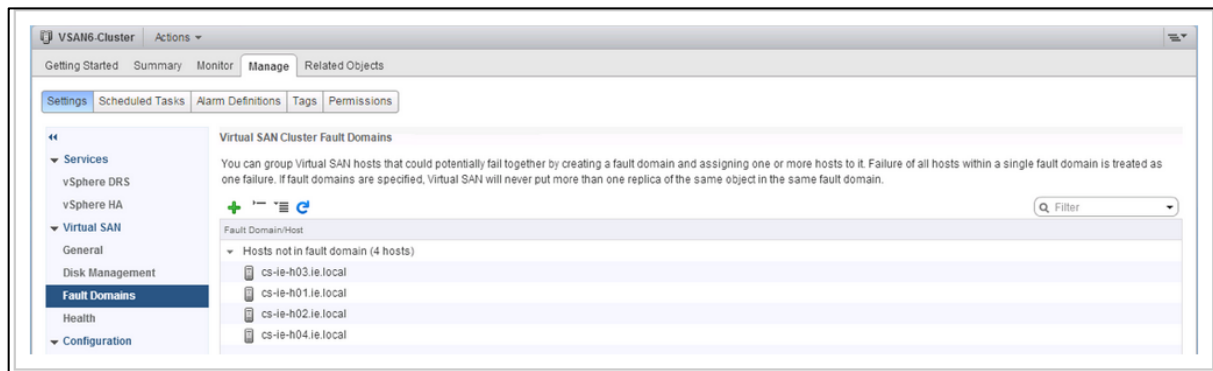
vSAN policy attributes establish parameters to protect against node failures, but they may not be the most effective or efficient way to build tolerance for events like rack failures. This section reviews the availability features for vSAN clusters on the VxRail Appliance. It starts out by looking at the availability implications on small VxRail deployments with fewer than four nodes.

Fault domain overview

vSAN and VxRail Appliances use fault domains as a way of configuring tolerance for rack and site failures. By default, a node is considered a fault domain. vSAN will spread components across fault domains, therefore, by default vSAN will spread components across nodes. Consider, for example, a cluster with four (4) four-node VxRail appliances, each VxRail appliance placed in a different rack. By explicitly defining each four-node appliance as separate fault domains, vSAN will spread redundancy components across the different racks.

In terms of implementation, any host that is not part of another fault domain is considered its own single-host fault domain. VxRail requires at least three fault domains, and each has at least one host. Fault domain definitions recognize the physical hardware constructs that represent the domain itself. Once the domain is enabled, vSAN applies the active virtual machine storage policy to the entire domain, instead of just to the individual hosts. The number of fault domains in a cluster is calculated based on the FTT attribute: (Number of fault domains) = 2 * (Failures to tolerate) + 1. Administrators can manage fault domains from the vSphere web client (as shown in the figure below.)

Figure 53 Managing fault domains



Fault domains and rack-level failures

The fault domain mechanism detects when the configuration is vulnerable. Consider a cluster that contains four server racks, each with two nodes. If the FTT is set to 1, and fault domains are not enabled, vSAN might store both replicas of an object with hosts in the same rack. In that case, applications are exposed to a potential rack-level failure. With fault domains enabled, vSAN ensures that each protection component (replicas and witnesses) is placed in a separate fault domain, making sure that the nodes cannot fail together.

The figure below illustrates a four-rack setup, each with two ESXi nodes (a subset of the available hosts in a VxRail Appliance). There are four defined fault domains:

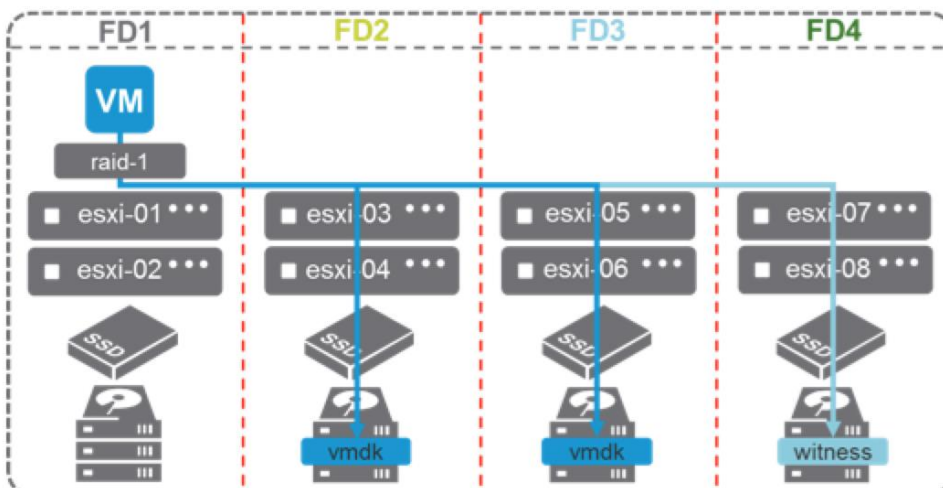
FD1 = esxi-01, esxi-02

FD2 = esxi-03, esxi-04

FD3 = esxi-05, esxi-06

FD4 = esxi-07, esxi-08

Figure 54 Fault domains for a four-rack VxRail configuration



This configuration guarantees that the replicas of an object are stored in hosts of different rack enclosures, ensuring availability and data protection in case of a rack-level failure.

Cautions when deploying a minimum cluster configuration

When deploying a cluster that just meets the minimum requirements, it is important to understand the high availability implications. Choosing a 3-node minimum configuration for RAID-1 protection or a 4-node minimum configuration for RAID-5 protection means that a cluster will not be able self-heal by rebuilding data on another host if one host fails. When a host is in maintenance mode such as a node upgrade, the data is exposed to a potential failure or inaccessibility if an additional failure occurs.

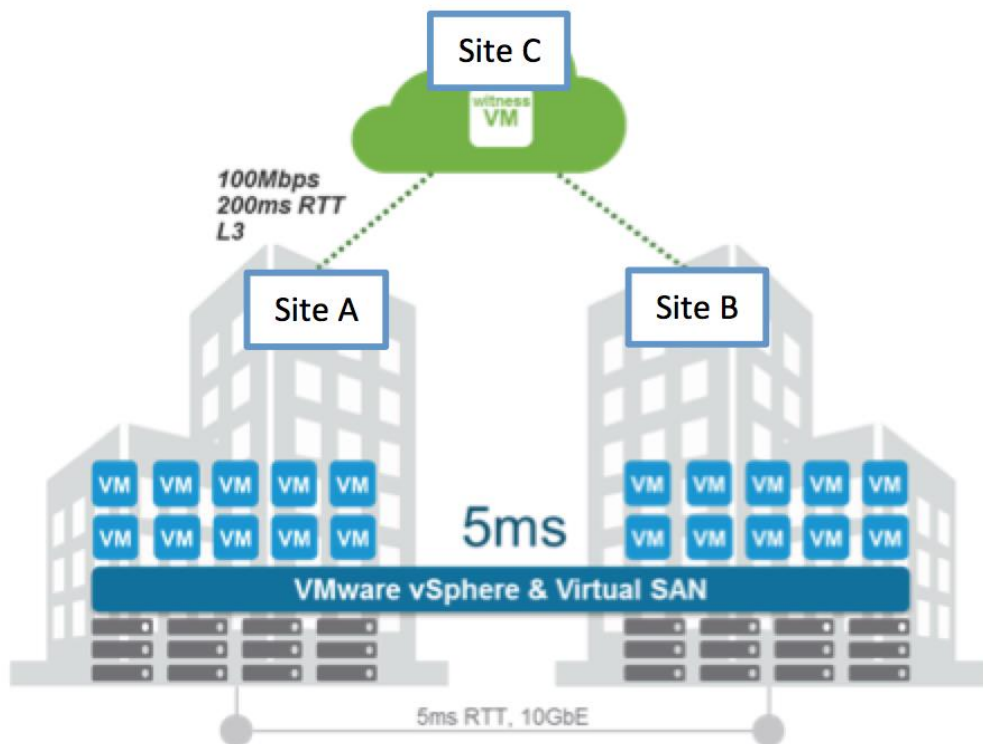
vSAN Stretched Cluster

The concept of a stretched cluster is a good example of vSAN's native integration with vSphere. With VxRail, stretch clustering extends availability of large enterprise datacenter. The stretched cluster is a specific configuration implemented in environments where the requirement for datacenter-level downtime avoidance is absolute. Similar to how fault domains enable "rack awareness" for rack failures; stretched clusters provide "datacenter awareness," maintaining virtual machine availability despite specific datacenter failure scenarios.

In a VxRail environment, stretched clusters with a witness host refers to a deployment where a vSAN cluster consists of two active/active sites with an identical number of ESXi hosts distributed evenly between them. The sites are connected via a high bandwidth/low latency networking.

In the figure below, each site is configured as a vSAN fault domain. The nomenclature used to describe the stretched cluster configuration is X+Y+Z, where X is the number of ESXi hosts at Site A, Y is the number of ESXi hosts at Site B, and Z is the number of witness hosts at site C.

Figure 55 Stretched VxRail cluster



A virtual machine deployed on a stretched cluster has one copy of its data on Site A, and another on Site B, as well as witness components placed on the host at Site C.

It is a singular configuration, achieved through a combination of fault domains, hosts and VM groups, and affinity rules. In the event of a complete site failure, the other site still has a full copy of virtual machine data and at least half of the resource components are available. That means all the VMs remain active and available on the vSAN datastore. The recovery point objective (RPO) is zero and the data recovery time objective (RTO) is zero. The application RTO is dependent upon the application recoverability.

The minimum configuration supported by VxRail is 3+3+1 (7 nodes); the maximum is 15+15+1 (31 nodes). Stretched clusters are supported by both hybrid and all-flash VxRail configurations. Additionally, customers will need to contact support to facilitate upgrades. Customer driven upgrades of VxRail Stretched Cluster implementations are not permitted, support should be contacted to perform the upgrade.

For more information, refer to *VxRail vSAN Stretched Clusters Planning Guide*:

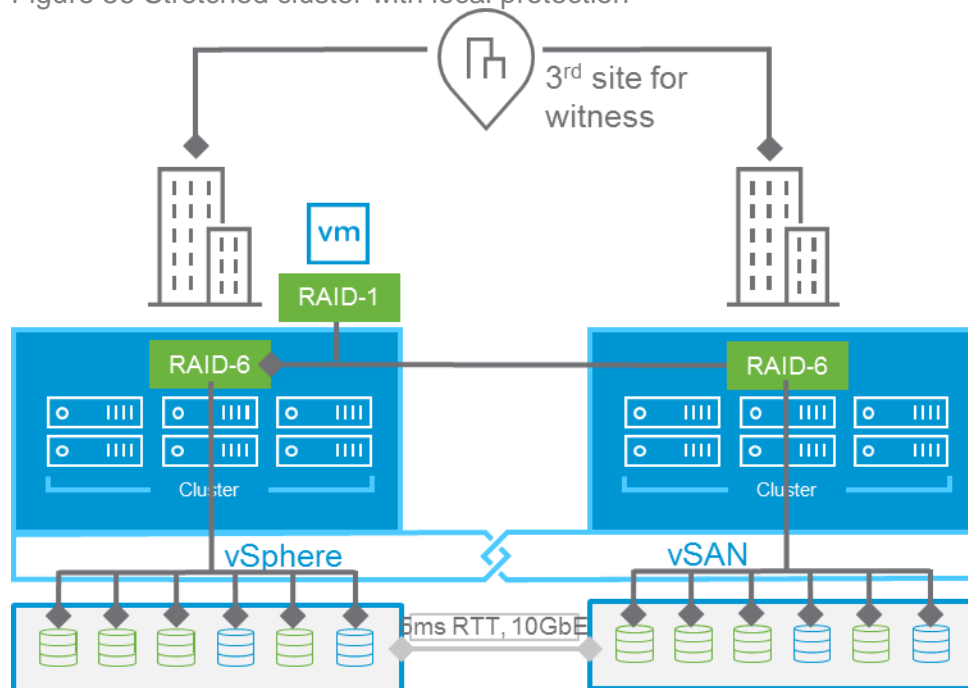
<https://vxrail.is/stretchedclusterplanning>

Stretched Cluster with Local Protection

VxRail Appliance software version 4.5 and vSAN 6.6 or above support Stretched Clusters with Local Protection. This feature mirrors data between sites, with each site applying local data protection for increased protection. The protection is specified using two parameters: Primary Failures To Tolerate (PFTT), and Secondary Failures To Tolerate (SFTT). PFTT refers to the protection between sites which is always RAID1 mirroring. SFTT is the local protection applied at each site. Hybrid configurations support SFTT of 0, 1, 2, or 3 with RAID1 (mirroring) Failure Tolerance Method (FTM). All-flash configurations support SFTT of 0, 1, 2, or 3 with RAID1 FTM or SFTT of 1 or 2 with Erasure Coding FTM.

Local protection for an all-flash stretched cluster configuration is shown in the figure below.

Figure 56 Stretched cluster with local protection



Site locality

In a conventional storage-cluster configuration, reads are distributed across replicas. In a stretched cluster configuration, the vSAN Distributed Object Manager (DOM) also takes into account the object's fault domain, and only reads from replicas in the same domain. That way, it avoids any lag time associated with using the inter-site network to perform reads.

Networking

Both Layer-2 (same subnet) and Layer-3 (routed) configurations are used for stretched cluster deployments. A Layer-2 connection should exist between data sites, and Layer-3 connection between the witness and the data sites.

The bandwidth between data sites depends on workloads, but Dell EMC requires a minimum of 10Gbps for VxRail Appliances in a stretched cluster configuration. The supported latency for witness hosts is up to 200ms RTT and a bandwidth of 2Mbps for every 1,000 vSAN objects. Also bear in mind that the latency between data sites should be no be greater than 5ms, generally estimated at 500km or about 310 miles.

Stretched cluster heartbeats and site bias

Stretched cluster configurations effectively have three fault domains. The first functions as the preferred data site, the second is the secondary data site, and the third is simply the witness host site.

The vSAN master node is placed on the preferred site and the vSAN backup node is placed on the secondary site. As long as nodes (ESXi hosts) are available in the preferred site, then a master is always selected from one of the nodes on this site—similarly for the secondary site, as long as nodes are available on the secondary site.

The master node and the backup node send heartbeats every second. If heartbeat communication is lost for five consecutive heartbeats (five seconds), the witness is deemed to have failed. If the witness has suffered a permanent failure, a new witness host can be configured and added to the cluster. Preferred sites gain ownership in case of a partition.

After a complete failure, both the master and the backup end up at the sole remaining live site. Once the failed site returns, it continues with its designated role as preferred or secondary, and the master and secondary migrate to their respective locations.

In terms of the communication with the witness, if the heartbeat pauses for five consecutive beats, the master assumes that the witness failed. If it's a permanent failure, a new witness host needs to be configured and added to the cluster, and preferred sites gain ownership in case of a partition.

vSphere HA settings for stretched cluster

A stretched cluster requires the following vSphere HA settings:

Host monitoring is enabled by default in all VxRail deployments, including of course stretched cluster configurations. This feature also uses network heartbeat to determine the status of hosts participating in the cluster. It indicates a possible need for remediation, such as restarting virtual machines on other cluster nodes.

Configuring admission control ensures that vSphere HA has sufficient available resources to restart virtual machines after a failure. This may be even more significant in a stretched cluster than it is in a single-site cluster, because it makes the entire, multi-site infrastructure resilient.

Workload availability is perhaps the primary motivation behind most stretched cluster implementations.

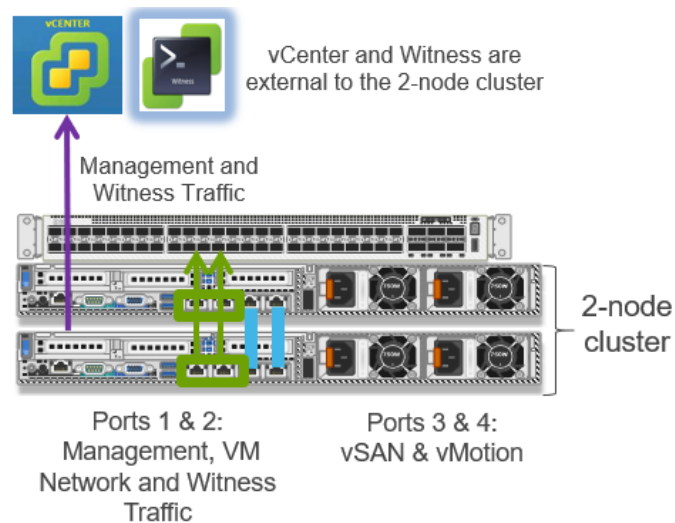
The deployment needs sufficient capacity to accommodate a full site failure. Since the stretched cluster equally divides the number of ESXi hosts between sites, Dell EMC recommends configuring the admission-control policy to 50 percent for both CPU and memory to ensure that all workloads can be restarted by vSphere HA.

2-Node Configuration

VxRail software version 4.7.100 introduces support for a fixed 2-node configuration that uses direct connection between the two nodes for cluster operations. With its small footprint, it can be an appropriate, cost-effective solution for locations with limited space and workload requirements. The support requires that the configuration to be a brand-new deployment which means existing clusters cannot utilize node removal to convert into a 2-node configuration. Cluster expansion is not supported therefore this solution should be targeted for specific use cases. Only the VxRail E560 and E560F are supported in a 2-node cluster. Users can still benefit from automated lifecycle management.

The configuration consists of the two nodes directly connected and a Witness to provide quorum for the cluster. The direct connection is for the vSAN and vMotion traffic. The Witness is a virtual appliance installed on an ESXi host which must reside outside of the 2-node cluster, i.e. in another datacenter or a physical host in the same rack/location. The Witness has individual connections to both nodes which requires VLANs to separate Witness management traffic from vSAN traffic. The configuration only supports mirroring (FTT=1). Witness host is used as the tiebreaker. Each node and the Witness is an individual fault domain for a total three in the cluster.

Figure 57 VxRail 2-node cluster



A special workflow in the First Run experience is used to deploy the 2-node cluster. The workflow includes the setup of the Witness appliance and Witness traffic separation. The configuration must use a customer-supplied vCenter for management. With only two data nodes in the cluster, users need be cognizant of the cluster load to prevent data unavailability in case of a node failure and a single node servicing the entire cluster workload.

The 2-node cluster supports per-socket vSAN Standard, Advanced, and Enterprise license editions. Refer to the [vSAN Licensing Guide](#) for more details.

For more information about this configuration, refer to the [VxRail Technical Deck](#).

Snapshots

Snapshots have been around for a while as a means of capturing the state of a data object at a particular point-in-time (PIT), so that it can be rolled back to that state if needed after a logical or physical failure. In the case of the VxRail solution, administrators can create, roll back, or delete VM snapshots using the Snapshot Manager in the vSphere web client. Each VM supports a chain of up to 32 snapshots.

A virtual machine snapshot generally includes the settings (.nvram and .vmx), the power state of all the VM's associated disks, and optionally, the memory state. Specifically, each snapshot includes:

Delta disk: The state of the virtual disk at the time the snapshot is taken is preserved. When this occurs, the guest OS is unable to write to its .vmdk file. Instead, changes are captured in an alternate file named VM_name-delta.Vmdk.

Memory-state file: VM_name-Snapshot#.Vlms, where # is the next number in the sequence, starting with 1. This file holds the memory state since the snapshot was taken. If memory is captured, the size of this file is the size of the virtual machine's maximum memory. If memory is not captured, the file is much smaller.

Disk-descriptor file: VM_name-00000#.vmdk, a small text file that contains information about the snapshot.

Snapshot-delta file: VM_name-00000#-delta.Vmdk, which contains the changes to the virtual disk's data at the time the snapshot was taken.

VM_name.Vmsd: This snapshot list file is created when virtual machine itself is deployed. It maintains VM snapshot information that goes into a snapshot list in the vSphere web client. This information includes the name of the snapshot .Vlms file and the name of the virtual-disk file.

The snapshot state uses a .Vlms extension and stores the requisite VM information at the time of the snapshot. Each new VM snapshot generates a new .vmsn file. The size of this file varies, based on the options selected during creation. For example, including the memory state of the virtual machine increases the size of the .vmsn file. It typically contains the name of the VMDK, the display name and description, and an identifier for each snapshot.

Other files might also exist. For example, a snapshot of a powered-on virtual machine has an associated snapshot_name_number.vmem file that contains the main memory of the guest OS, saved as part of the snapshot.

A quiesce option is available to maintain consistent point-in-time copies for powered-on VMs. VMware tools may use their own sync driver or use Microsoft's Volume Shadow Copy Service (VSS) to quiesce not only the guest OS files system, but also any Microsoft applications that understand VSS directives.

Storage efficiency using deduplication and compression

Storage capacity requirements continue to grow exponentially, and IT organizations are looking for ways to increase storage efficiency in order to meet their growing capacity requirements at the lowest cost. One way to do this is to use data deduplication and compression, which can result in more capacity at a lower cost. Many environments can achieve an effective capacity that is twice the raw capacity. These data reduction capabilities can be utilized on all-flash VxRail clusters.

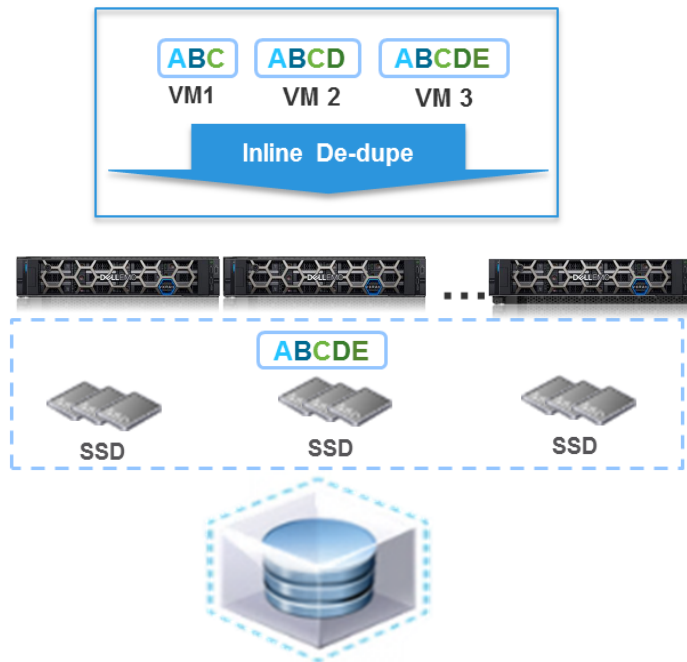
Compression and deduplication techniques have been in place for a number of years but have not been widely adopted because of the overhead and system resources required to implement. Today, VxRail all-flash models with many cores and lots of memory per processor are a

powerhouse! Along with the architectural efficiencies of vSAN, the space savings more than offset the slight overhead. A VxRail all-flash configuration often provides more effective capacity at a lower cost than a hybrid HDD solution.

With vSAN, deduplication and compression occurs inline when data is de-staged from the cache to the capacity drives. First data is deduplicated by removing redundant copies of blocks that contain the exact same data. This is done at the 4K block level.

The figure below shows a typical virtual machine environment.

Figure 58 Inline data deduplication



While all VMs are unique, they share some amount of common data. Rather than saving multiple copies of the same data, identical blocks are saved once on SSD, and references to the unique blocks are tracked using metadata that is maintained in the capacity tier.

The deduplication algorithm is applied at the disk-group level and results in only a single copy of each unique 4K block per disk group. While duplicated data blocks may exist across multiple disk groups, by limiting the deduplication domain to a disk group, a global lookup table is not required, which minimizes network overhead and CPU utilization.

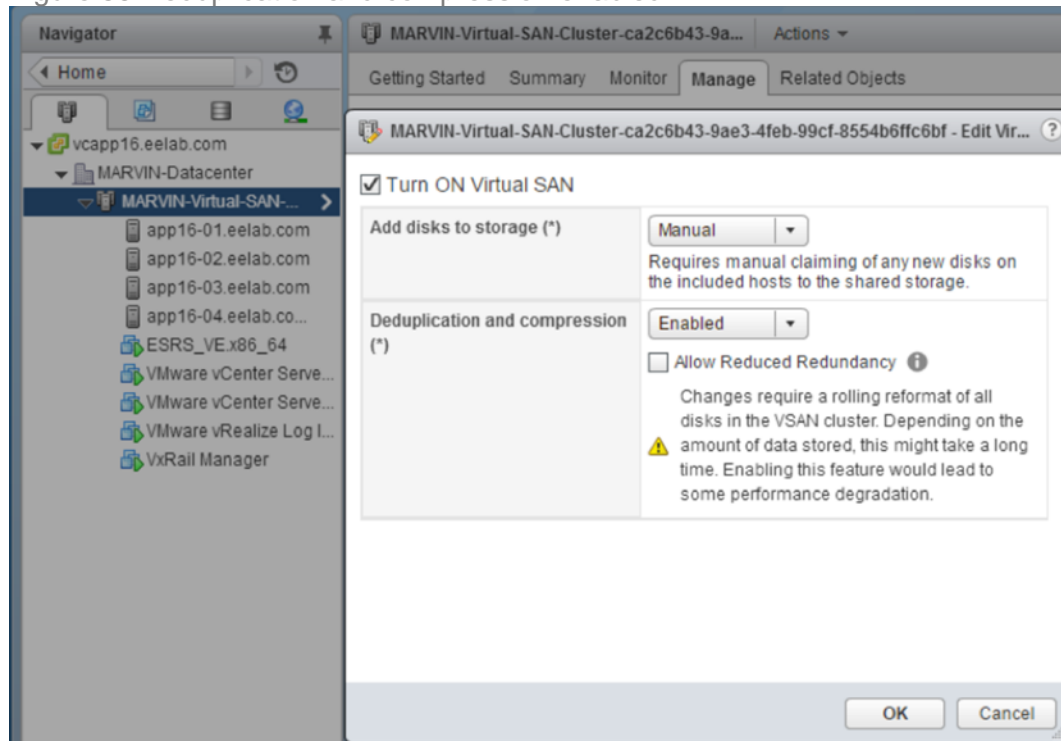
LZ4 compression is applied after the blocks are deduplicated and before being written to SSD. If the compression results in a block size of 2KB or less, the compressed version of the block is persistently saved on SSD. If the compression does not result in a block size of less than 2KB, the full 4K block is written to SSD.

Almost all workloads benefit some from deduplication. However typical virtual server workloads with highly redundant data such as full clone virtual desktops or homogenous server operating systems benefit most. Compression provides further data reduction. Text, bitmap, and program files are very compressible, and 2:1 is often possible. Other data types that are already compressed, such as certain graphics formats and video files or encrypted files, may yield little or no reduction.

Deduplication and compression are disabled by default and are enabled together at the cluster level. (See the figure below.) While it can be enabled at any time, enabling it when the system is

initially setup is recommended to avoid the overhead and potential performance impact of having to deduplicate and compress existing data through post processing rather than to do it inline.

Figure 59 Deduplication and compression enabled



Deduplication and compression overhead

Deduplication algorithms break data files into contiguous segments, or compute fingerprints, used to identify duplicate segments and reduce the data footprint. This is a basic deduplication concept. The specific approach varies among system vendors, but any deduplication method consumes CPU to compute the segment fingerprints or hash keys, and it executes I/O operations when performing lookups on the segment index tables.

vSAN computes the fingerprints and looks for duplicated segments only when the data is being de-staged from the cache to the capacity tier. This means that under normal operations, VM writes to the write buffer in the cache SSD should not incur any latency impact.

The cost of the deduplication occurs when data is de-staged from the cache to the capacity tier. It consumes a portion of CPU capabilities reserved for vSAN, and the disk operations generated by the index lookups consumes a portion of the backend I/O capabilities.

Because resource consumption varies according to I/O patterns, data types and so on, consult with an Dell EMC or VMware specialist before deciding whether deduplication is recommended for your application.

More information can be found in *Technical Whitepaper VMware VSAN 6.2 Space Efficiency Technologies* at <http://www.vmware.com/files/pdf/products/vsan/vmware-vsan-62-space-efficiency-technologies.pdf>.

Erasure coding

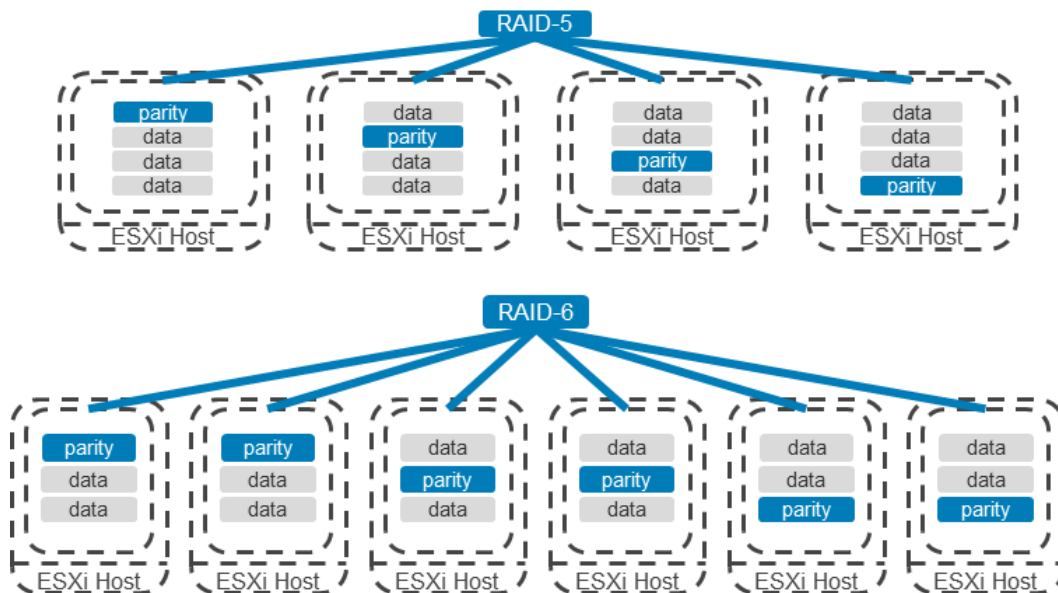
Erasure coding is another capacity-efficient solution for the failure tolerance method and data protection on all-flash VxRail configurations. As an alternative failure tolerance method to the data replication provided by RAID-1 mirroring, erasure codes can provide up to 50 percent more usable capacity than purely conventional RAID-1 mirroring, which drains storage space.

Erasure coding breaks up data into fragments and distributes redundant chunks of data across the system. It introduces redundancy by using data blocks and striping. To explain basically, data blocks are grouped in sets of n , and for each set of n data blocks, a set of p parity blocks exists. Together, these sets of $(n + p)$ blocks make up a stripe. The crux is that any of the n blocks in the $(n + p)$ stripe is enough to recover the entire data on the stripe.

In VxRail clusters, the data and parity blocks that belong to a single stripe are placed in different ESXi hosts in a cluster, providing a layer of failure tolerance for each stripe. Stripes don't follow a one-to-one distribution model. It is not a situation where the set of n data blocks sits on one host, and the parity set sits on another. Rather, the algorithm distributes individual blocks from the parity set among the ESXi hosts.

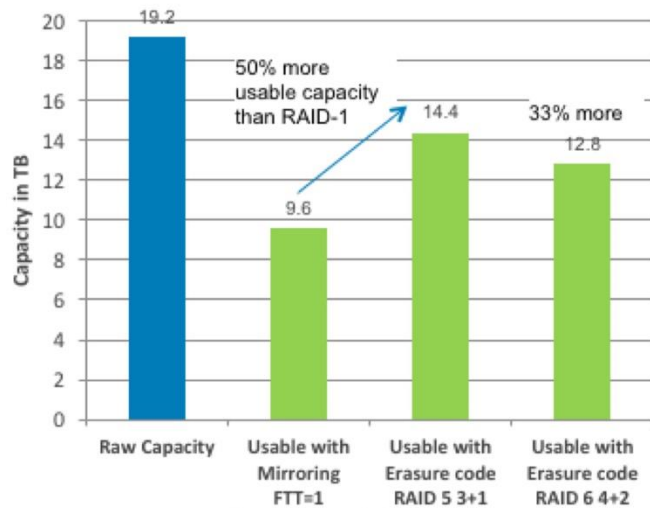
Erasure coding provides single-parity data protection (RAID-5) that can tolerate one failure (FTT=1) and double-parity data protection (RAID-6) that can tolerate two failures (FTT=2). The diagrams below illustrate the implementations. A single-parity stripe uses three data blocks and one parity block (3+1), and it requires a minimum of four hosts or four fault domains to ensure availability in case one of the hosts or disks fails (as shown below). It represents a 30 percent storage savings over RAID-1 mirroring. Dual parity saves as much as 50 percent capacity over RAID-1. It uses four data blocks plus two parity blocks (4+2) and requires a minimum of six nodes. See the figure below.

Figure 60 RAID-5 (FTT=1) requires a minimum of four nodes and RAID-6 (FTT=2) with 4+2 nodes



Look at the comparison of usable capacity in the figure below. The erasure-code protection method increases the usable capacity up to 50 percent compared to mirroring.

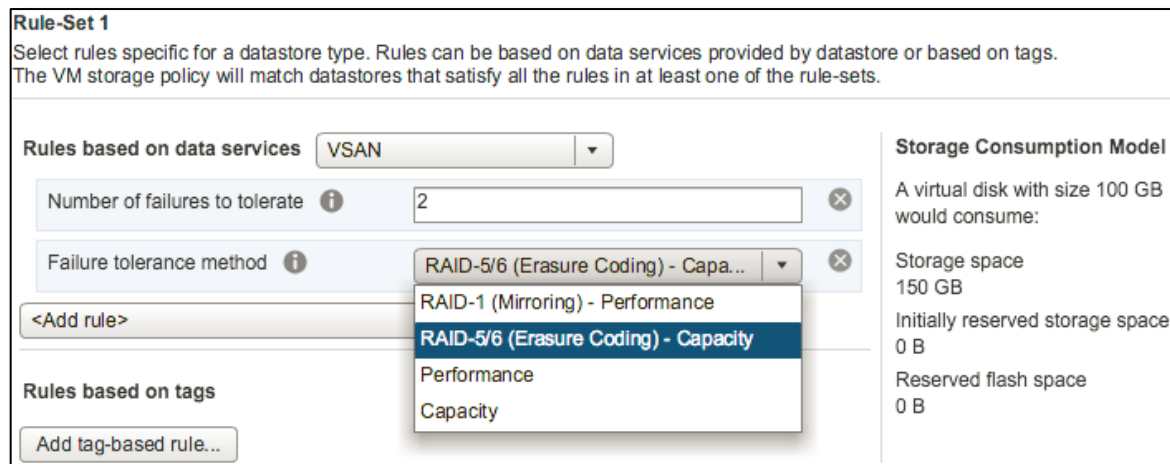
Figure 61 Erasure coding increases usable capacity up to 50 percent



Enabling Erasure Coding

The SPBM policy Failure Tolerance Method (FTM) lets administrators choose between RAID-1 (Mirroring) and RAID-5/6 (Erasure Coding). The FTT policy determines the number of parity blocks written by the erasure code. See the figure below.

Figure 62 FTT policy determines the number of parity blocks written by the erasure code



VxRail implements erasure coding at a very granular level, and it can be applied to VMDKs, making for a nuanced approach. Configurations for VMs with write-intensive workloads—a database log, for instance—can include a mirroring policy, while the data component can include an erasure coding.

Erasure coding overhead

Erasure coding saves space but increases backend overhead. Computing parity blocks consumes CPU cycles and adds overhead to the network and disks, as does distributing data slices across multiple hosts. This extra activity can affect latency and overall IOPS throughput.

The rebuild operation also adds overhead. In general, rebuild operations multiply the number of reads and network transfers used for replication. A formula is available here, too. If n refers to the number of blocks in a stripe, then the rebuild operations cost n times that of ordinary

replication. For a 3+1 stripe, that means three disk reads and three network transfers for every one of conventional data-replication. The rebuild operation can also be invoked to serve read requests for currently available data.

This additional I/O is the primary reason why only all-flash VxRail configurations use erasure coding. The rationale here is that the flash disks compensate for the extra I/O.

vSAN Encryption

In vSAN 6.6, encryption is a new datastore level setting. vSAN Encryption is a datastore, or cluster-wide, level setting applied to all VM components in the cluster. This feature solves the concern of media theft. An advantage of the cluster wide setting is that deduplication and compression are applied prior to the encryption. This provides space savings benefit over the vSphere 6.5 Encryption option. Like vSphere Encryption, a KMIP-compliant Key Management Server like CloudLink or Hytrust must be used in conjunction with vSAN Encryption. vSAN encryption is FIPS 140-2 Level 1, AES 256 compliant.

VxRail integrated software

VxRail has integrates software in two ways, both of which are fully engineered, tested, validated, manufactured and supported as a single offering from Dell EMC.

Products that are native to vSphere, including vSphere Replication

Dell EMC software products, including RecoverPoint for Virtual Machines

VM Replication

Several options are available for replicating virtual machines for data protection and disaster recovery in VxRail clusters. Among them are solutions integrated into the VMware software stack: VMware vSphere Replication (VR), and RecoverPoint for Virtual Machines (RP4VM), which is built on enterprise proven Dell EMC RecoverPoint technology.

When choosing the right solution, Recovery Point Objectives (RPO) is an important consideration. RPO defines the maximum acceptable age of data recovered from a replicated copy as a result of a data loss issue. For example, if a virtual machine is deleted and the RPO for the virtual machine is 24 hours, a recovered copy of the virtual machine should contain all data except for any changes that occurred in the last 24 hours. Depending on the technology, RPOs as low as a few seconds or minutes can be configured.

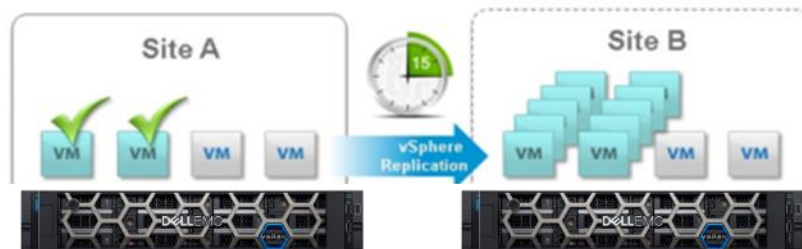
Another consideration is the number of recovery points to retain. When configuring replication for a virtual machine, an administrator has the option to enable the retention of multiple recovery points (point-in-time instances). This can be useful when an issue is discovered several hours, or even a few days, after it occurred. For example, a replicated virtual machine with a four-hour RPO contracts a virus, but the virus is not discovered until six hours after infestation. As a result, the virus has been replicated to the target location. With multiple recovery points, the virtual machine can be recovered and then reverted to a recovery point retained before the virus issue occurred.

VMware vSphere Replication

VMware vSphere Replication is a hypervisor-based, asynchronous replication solution that is fully integrated with VMware vCenter Server. It uses a proprietary replication engine developed by VMware and is included with VMware vSphere Essentials Plus Kit and higher editions of VMware vSphere. While VR works well with vSAN storage in a VxRail hyperconverged environment, it is completely independent of the underlying storage and allows for replication between heterogeneous storage types. This is useful when a VxRail Appliance is part of a larger virtualization environment that includes SAN or other storage types. VR provides data protection locally within the same vCenter environment as well as disaster recovery and avoidance to another vCenter site. It also supports replication to Service Provider clouds such as vCloud Air.

The figure below shows VR replication between two VxRail sites.

Figure 63 VMware vSphere Replication with VxRail



In this example, several VMs are replicated to a remote site with a RPO of 15 minutes. The remote site maintains multiple VM images, allowing roll back to different points-in-time.

The VR components that transmit replicated data are built into vSphere and use Secure Sockets Layer (SSL) connection. A best practice is to isolate network traffic to improve performance and security. VR also includes one or more prebuilt Linux-based vSphere Replication vApps. The first virtual appliance is referred to as the vSphere Replication Management server. It receives replicated data, manages authentication, and maintains mappings between the source virtual machines and the replicas at the target location. Each appliance requires 18GB of VMDK storage, 4GB of memory, and either two or four virtual CPUs.

VR configuration includes specifying the Recovery Point Objectives (RPO) within a range of 15 minutes to 24 hours as well as the number of point-in-time time images to maintain. This is all done from within the vSphere web client using a simple wizard. Once replication has been configured for a virtual machine, vSphere Replication begins the initial full synchronization of the source virtual machine to the target location. The time required to complete this initial synchronization can vary depending on the amount of data that must be replicated and the amount of available network bandwidth. After the initial full synchronization, only changed data is transmitted in each replication cycle, minimizing the network bandwidth requirements.

The replicated data is first written to a redo log. After all changes for the current replication cycle have been received and written to the redo log, the data in the redo log is merged into the base image. This process ensures a consistent and recoverable VM image is available at all times.

vSphere Replication delivers flexible, reliable and cost-efficient replication for data protection and disaster recovery. It is seamlessly integrated within the VMware product stack for simple deployment, configuration, and management.

Dell EMC RecoverPoint for Virtual Machines

RecoverPoint for Virtual Machines (RP4VM) is based on the RecoverPoint continuous data protection technologies that have been proven in enterprise environments for over ten years with over 350 million run hours and an entire exabyte of data protected. Like VR, RecoverPoint for VMs enables customers to replicate virtual machines simply and is configured and managed from within vSphere web client using the RecoverPoint plug-in.

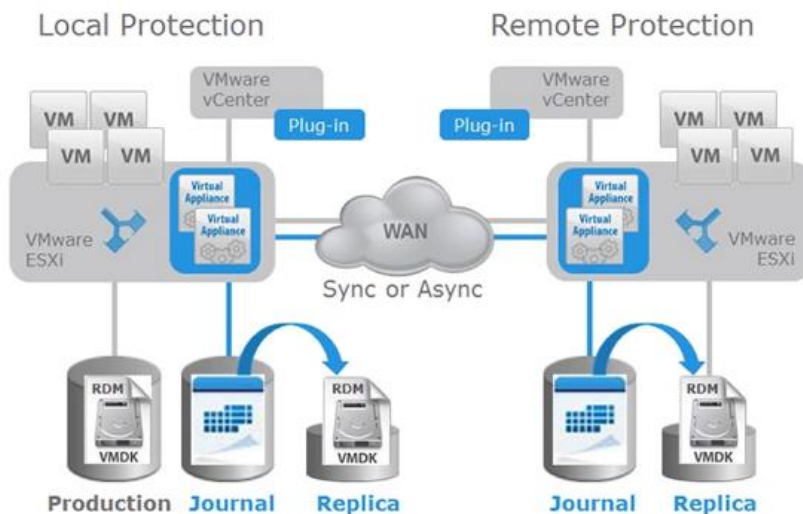
While similar in concept to VR, RP4VM has a number of unique capabilities including the ability to configure an RPO as low as zero seconds, as well as concurrent local and remote replication, and multi-site 2:1 and 1:2 configurations.

RecoverPoint for VM includes a RecoverPoint Virtual Appliance (vRPA) that manages all aspects of data replication. Two-to-eight vRPAs are grouped as a cluster. For local replication, a single cluster would be configured. For remote replication, a cluster is configured at both the local and remote sites. A RecoverPoint splitter is also installed on each ESXi server. The splitter takes each write and sends it to both the VMDK and to the vRPA. Management of RP4VM is all done from within vCenter using the RecoverPoint for VMs plugin.

The source of replication is the virtual machine, and associated application data and is referred to as the production copy. RP4VM performs continuous data replication of writes to the production copy. Each write is split and sent to both the VMDK and to a journal. The journal provides DVR-like roll back and is used to create a local copy that reflects any point in time. The local copy can be used to recover from logical errors and data corruption. Optionally, RP4VM can be configured to synchronously or asynchronously replicate to a remote vRPA cluster. The local and remote vRPA clusters communicate over a WAN connection and compression and de-duplication provides optimization and reduces bandwidth consumption.

The following figure shows an environment with both local and remote data protection and how writes are split, journaled, and used to create point-in-time images.

Figure 64 Continuous local and remote protection



Other RP4VM features include automated discovery, provisioning and orchestration of DR workflows, including test, and failover and failback of a single VM or multiple VMs using consistency groups.

A license Starter Pack is included with each VxRail Appliance. Additional licenses can be purchased as needed.

VxRail replication use case

Both VR and RPVM provide data protection and disaster-recovery capabilities. Both are software-only solutions that are embedded in the hypervisor, use vApps for control, and are managed from vCenter. Which one to use depends on the use case.

Local replication to other appliances, remote replication to other VxRail clusters, or non-VxRail vSphere clusters may be configured. Remote replication may be used to provide DR to ROBO sites. Other use cases include VM migration from other vCenter environments into and between VxRail environments.

For basic replication that is fully integrated and easy to manage using VMware vCenter, VR provides these capabilities at no additional cost. If an RPO of less than five minutes, only RP4VM can provide continuous data replication with DVR-like playback. RPVM also supports more flexible deployment options including concurrent local and remote replication and multiple sites in a 2:1 or 1:2 configuration.

Regardless of the replication choice, solution sizing should include the additional storage capacity required and the overhead of replication vApps.

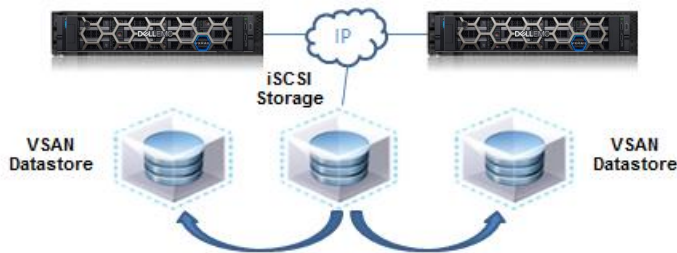
Support for external network storage

The vSAN presents a robust, secure, and efficient shared datastore to all nodes within a VxRail cluster. External SAN storage is typically not part of a VxRail environment. However, often a requirement exists to access external storage in order to move virtual machines and data into a VxRail environment or move data between environments. Because of the HBA requirement, Fibre Channel SAN connectivity is not possible. However, IP-based storage is ideal for this purpose, and VxRail Appliances support both iSCSI and NFS. An important distinction is that data in the iSCSI or NFS datastore is self-contained and is not distributed to the disk groups within the VxRail cluster.

iSCSI with VxRail

iSCSI can be used to provide mobility for VMs and associated data onto and between VxRail environments. The figure below shows a VxRail environment that includes iSCSI storage in addition to the vSAN datastore.

Figure 65 Data mobility into and between VxRail environments



Data on the iSCSI storage is easily moved into the VxRail vSAN environment or between VxRail environments.

Existing iSCSI storage can also be used to provide additional capacity to the VxRail environment. However with the VxRail scale-up and scale-out configuration flexibility, external storage is typically not used to meet capacity requirements.

iSCSI provides block-level storage using the SCSI protocol over an IP network. SCSI uses a client-server, initiator-target model where initiators issue read/write operations to target devices, and targets either return the requested read data or persistently save write data. iSCSI in a VMware environment is standard functionality. A software adapter using the NIC on an ESXi host is configured as an initiator, and targets on an external storage system present LUNs to the initiators. The external LUNs could be used by ESXi as raw device mapping (RDM) devices, however usually, the use case is for VxRail to configure them as VMFS datastores. (Refer to vSphere documentation for more information: *Using ESXi with iSCSI SAN.*)

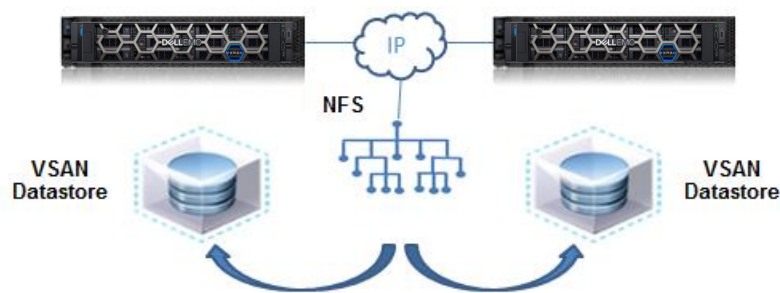
iSCSI configuration is performed using the vSphere web client. The steps involve creating a port group on the VDS, creating a VMkernel Network Adapter and associating it with the port group, and assigning an IP address. Then, from the vCenter Manage Storage Adapters view, the Add iSCSI Software Adapter dialog is used to create the software adapter. The last step is to bind the iSCSI software adapter with VMkernel adapter. Once this is complete, iSCSI targets and LUNs can be discovered and used to create new datastores and map them to the hosts in the cluster. (Refer to the VMware documentation for more details.)

iSCSI works best in a network environment that provides consistent and predictable performance, and a separate VLAN is usually implemented. iSCSI network requirements should be considered when planning the network requirements for VxRail environment to make sure connectivity to the external iSCSI storage system exists, and the additional network traffic will not impact other applications.

NFS with VxRail

NFS is a network filesystem that provides file-level storage using the NFS protocol over an IP network. It can work in use cases similar to iSCSI—the difference being that NFS devices are presented as file systems rather than block devices. The figure below shows an NFS file system that has been exported from a network-attached server and mounted by the ESXi nodes in the VxRail environment.

Figure 66 Network-attached file system with VxRail



This enables data mobility into and between VxRail environments as well as enabling additional storage capacity.

The external NFS server can be an open system host, typically Unix or Linux, or a specially built appliance. The NFS server takes physical storage and creates a file system. The file system is exported and client systems, in this example ESXi hosts in a VxRail Appliance, mount the file system and access it over the IP network.

Similar to iSCSI, NFS is a standard vSphere feature and is configured using the vCenter web client. This is done in the Hosts and Clusters view under Related Objects and the New Datastore dialog. Select NFS as datastore type, the NFS version, the name of the datastore, the IP address or hostname of the NFS server that exported the filesystem, and the host that will mount it. The NFS filesystem will appear like the vSAN datastore. VMs, templates, OVA files, and other storage objects can be easily moved between the NFS filesystem and the vSAN datastore using vMotion.

As with iSCSI, NFS works best in network environments that provide consistent and predictable performance. The network requirements for NFS should be considered when initially planning the network requirements for VxRail environment.

VxRail solutions and ecosystem

Dell EMC offers a full range of flexible consumption models that make it faster and easier for businesses to use VxRail to fuel digital transformation. These consumption models include both the technology itself and how businesses pay for this technology.

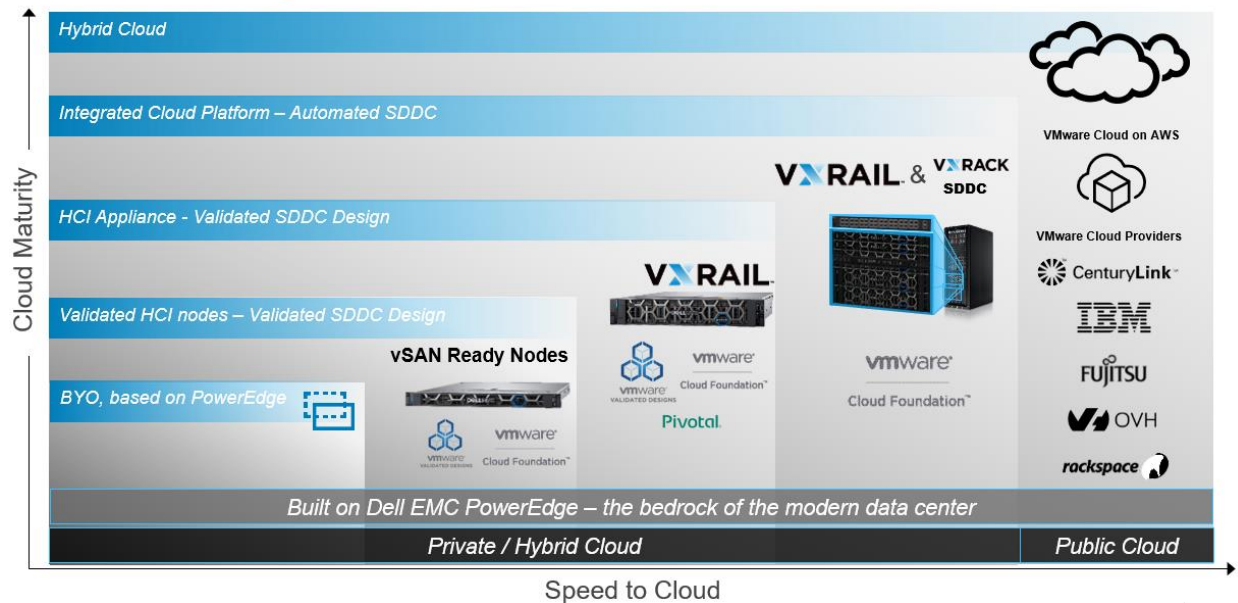
VMware Validated Design with VxRail

VMware Validated Design (VVD) is a family of solutions for datacenter designs that span compute, storage, networking, and management, serving as a blueprint for your Software-Defined Datacenter (SDDC) implementation. The VVD is a reference architecture for how to deploy, operate, and maintain a VMware SDDC. VxRail is supported starting with the VVD 4.2.

The VMware VVD provides a framework for complete NSX and vRealize capabilities on top of VxRail. It required end-to-end validation of HW and SW with interoperability and scalability testing. Further, it provides Day 2 guidance on how to monitor, backup, restore and failover management components. As such it creates a trusted implementation design that de-risks deployments, simplifies operations, and further drives IT agility for customers to create a private cloud and accelerate their transformation to a multi-cloud VMware environment. The figure below shows some of the different deployment models for modernizing vSphere and VMware environments, including the VVD using VxRail.

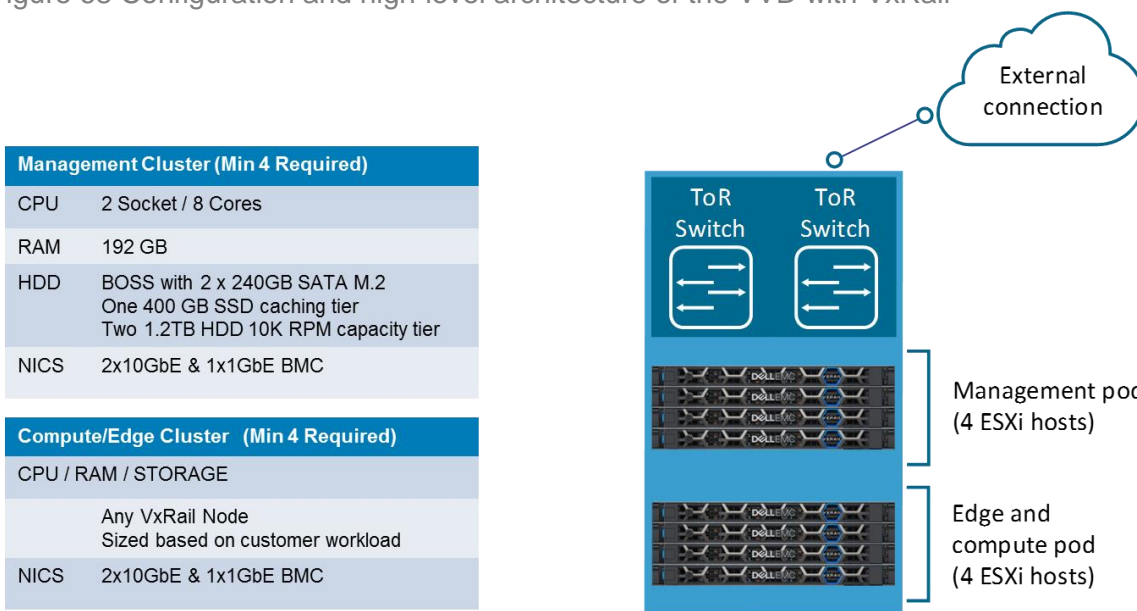
Figure 67 Paths to the VMware SDDC and multi-cloud IT model

Accelerating IT transformation to VMware multi-cloud



The VVD requires a management cluster with four nodes and a compute/edge cluster with a minimum of 4 nodes. The compute/edge cluster can be any VxRail node as long as it adheres to the cluster design rules described in the [VxRail Scaling section](#) of this paper. The Management Cluster has a prescribed set of minimum resources. Both cluster requirements are detailed in the figure below.

Figure 68 Configuration and high-level architecture of the VVD with VxRail



Leverage the VxRail model that **meets or exceeds** these requirements and the customer needs

The VVD provides guidance to the VMware SDDC and software bill of materials as seen in the following figure. VMware recommends upgrade licenses to VMware Cloud Foundation (VCF) for VxRail customers, such that titles already paid for titles and software are not “double” charged. Purchasing VCF licenses as a bundle will result in a lower overall license cost and will also provide a path to leveraging VCF and the SDDC Manager automation more broadly on VxRail in the future. (Note: Today, SDDC Manager is not supported on VxRail. NSX and vRealize licenses included in the VCF add-on for VxRail must be run as independent titles, not as a part of SDDC Manager.)

Figure 69 Example software Bill of Materials from the VVD 4.2

Cloud Component	Product Item	Version
Virtual Infrastructure	ESXi	6.5 u1
	vCenter Server Appliance (VIMISO)	6.5 u1e
	NSX for vSphere	6.3.4
Cloud Management	vRealize Automation Appliance	7.3.0
	vRealize Orchestrator	7.3.0
	vRealize Orchestrator Plug-in for NSX	1.0.4
	vRealize Business	7.3.1
Service Management	vRealize Operations Manager Appliance	6.6.1
	Management Pack for NSX for vSphere	3.5.1
	Management Pack for vRealize Log Insight	6.0
	Manager Management Pack for vRealize Automation	3.0
	Management Pack for Storage Devices	6.0.5
	vRealize Log Insight	4.5.0
Infrastructure	Windows	2012 R2
	SQL Server	2012 R2

For more information please visit:

The VMware VVD site:

<https://www.vmware.com/solutions/software-defined-datacenter/validated-designs.html>

Dell EMC Community Network, VMware Validated Design 4.2 on VxRail Deployment Guides:

<https://community.emc.com/docs/DOC-66332>

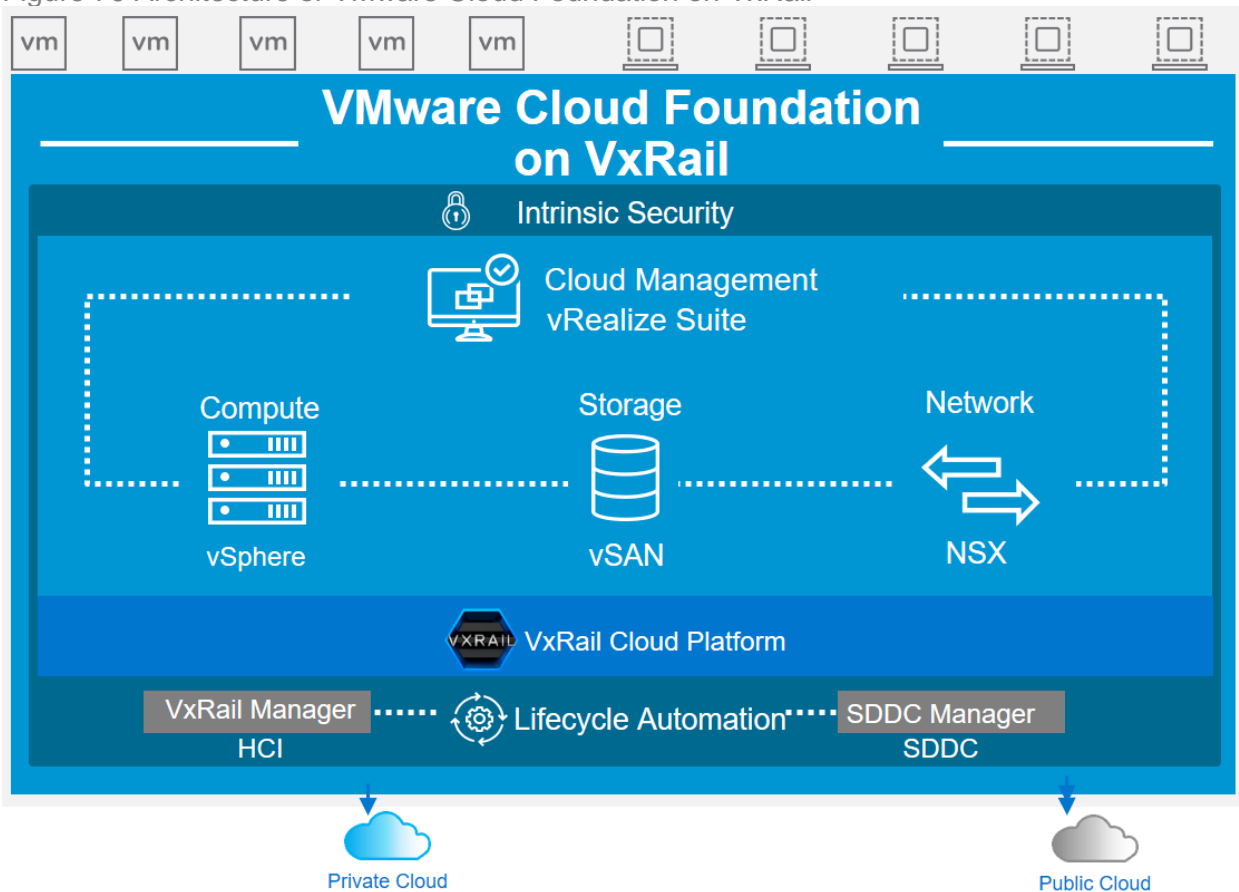
VMware Cloud Foundation on VxRail

VMware Cloud Foundation on VxRail is a Dell EMC and VMware jointly engineered integrated solution with features that simplify, streamline, and automate the operations of your entire SDDC from Day 0 through Day 2. The new platform delivers a set of software-defined services for compute (with vSphere and vCenter), storage (with vSAN), networking and security (with NSX), and cloud management (with vRealize Suite) in both private and public environments, making it the operational hub for your hybrid cloud.

VMware Cloud Foundation on VxRail provides the simplest path to the hybrid cloud through a fully integrated hybrid cloud platform that leverages native VxRail hardware and software capabilities and other VxRail unique integrations (such as vCenter plugins and Dell EMC networking) working together to deliver a new turnkey hybrid cloud user experience with full-stack integration. Full-stack integration means you get both the HCI infrastructure layer and cloud software stack in one completely automated lifecycle turnkey experience.

An important aspect of the offering is the introduction of a standardized architecture for how these SDDC components are deployed together with the introduction of Cloud Foundation, an integrated cloud software platform that is based on VVD. Having a standardized design incorporated as part of the platform provides you with a guarantee that these components have been certified with each other and are backed by Dell Technologies. You can then be assured that there is an automated and validated path forward to get from one known good state to the next across the end-to-end stack.

Figure 70 Architecture of VMware Cloud Foundation on VxRail



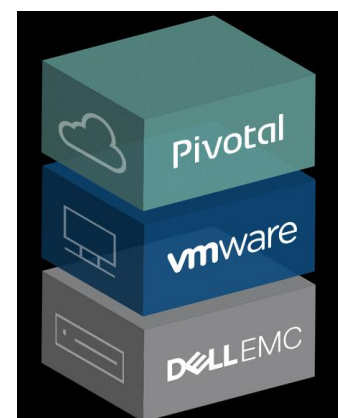
Pivotal Ready Architecture (PRA)

Pivotal Ready Architecture is a tested and validated reference architecture for deploying Pivotal Cloud Foundry on VxRail. With configurations for high availability, comprehensive product support, and options for object storage, Pivotal Ready Architecture is the best way to deploy Pivotal Cloud Foundry on-premises.

Cloud-native patterns are a modern approach to application architecture, development and delivery that has emerged as a natural response to the changes in business needs and infrastructure capabilities. This new model directly increases the speed and agility of application delivery for IT organizations and has proven its benefits for startups and established enterprises alike. Pivotal Ready Architecture is the fastest way to get Pivotal Cloud Foundry up and running in your datacenter. Accelerate your transformation with an "it just works" experience. PRA supports Pivotal Application Service (PAS) and Pivotal Container Service (PKS).

Business benefits derived from the PRA include:

- **Reliable Deployment.** PRA is a proven hardware and software solution.



- Ready Infrastructure. PRA is built on the only fully integrated, pre-configured, and pre-tested VMware hyperconverged infrastructure appliance family on the market.
- Resilient Architecture. PRA offers multi-site, multi-foundation, and multiple availability zone configuration options that deliver maximum uptime, geographic coverage, and resiliency.

PRA provides a tested, validated reference architecture on which to deploy a highly available enterprise-grade developer platform. Built on hyper-converged VxRail, PRA delivers automated lifecycle management of the infrastructure, a critical element in accelerating your transformation into a digital business.

- Pivotal Application Service (PAS) and Pivotal Container Service (PKS) reference architectures on VxRail
- Fully software defined infrastructure
- “Always on” highly available configurations
- A central management console
- Modular design that scales with you
- Integrated backup & disaster recovery options

For more information visit: <https://pivotal.io/pivotal-ready-architecture>

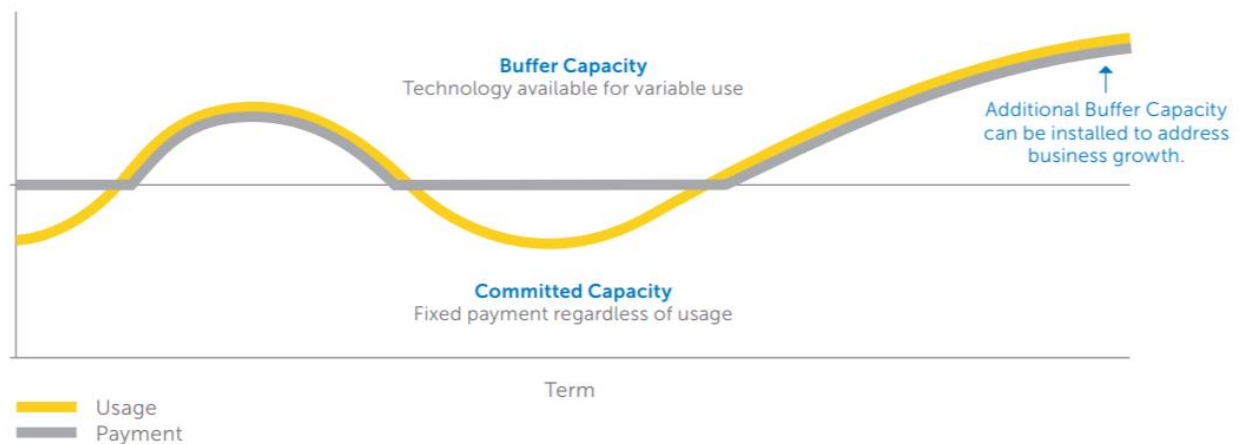
Flexible consumption options

Flex on Demand - a cloud-like consumption option

Flex on Demand by Dell Financial Services (DFS) allows you to acquire technology you need to support your changing business with payments that scale to match your actual usage. This model helps align your cost with usage and avoid paying for buffer capacity that is not used. It improves agility by providing instant deployment of capacity for usage when spikes occur in business operations. It improves budget agility and power by delivering better operational economics.

- DFS works with you to establish the “committed capacity” presently needed and the “buffer capacity” required in the future.
- Buffer capacity is measured using automated tools with your equipment. Each payment is comprised of the fixed committed capacity and variable buffer capacity amount.
- If your usage consistently consumes most of the buffer capacity, you have the option to receive additional buffer capacity. Once installed, your level of committed capacity and related payment will increase.

Figure 71 Relationship between technology usage and Flex on Demand payment



See <https://www.dell.com/en-us/flexibleconsumption/cloud-flex-for-hci.htm> for more information.

VDI Complete

VDI Complete is a series of end-to-end desktop and application virtualization solutions that feature a superior solution stack and exceptional total cost of ownership. The solutions are built on Dell EMC VxRail appliances and leverage VMware Horizon virtual desktops and applications.

VDI Complete is an end-to-end desktop and application virtualization solution from Dell EMC that includes everything you need to get started: the infrastructure appliances, the software, the storage and the endpoints.

VDI Complete is built with best-of-breed technology from Dell Technologies. It leverages proven and trusted infrastructure appliances and endpoints from Dell and Dell EMC. And it takes advantage of VMware Horizon, an industry leader desktop and application virtualization. VDI Complete is fully validated and tested, lowering risk and reducing complexity. It's a single go-to source for both purchase and complete solution support.

See [https://downloads.dell.com/solutions/general-solution-resources/White%20Papers/DellEMC.VxRail\(14G\).VMware.Horizon.RA.pdf](https://downloads.dell.com/solutions/general-solution-resources/White%20Papers/DellEMC.VxRail(14G).VMware.Horizon.RA.pdf) for more information.



95%

Faster initiation for 1st VxRail appliance



47%

Faster to provision 10 desktop VMs



96%

Faster to add a new appliance

VMware Horizon

VMware Horizon is VMware's VDI and desktop-management environment. Horizon provisions user desktops using a flexible and secure delivery model. The desktop environments are accessed by the user from almost any device, including mobile devices, with the security and resiliency of the datacenter. Because the application software and data components reside in the datacenter, traditional security, backup, and disaster recovery approaches may be applied.

If a user's device is lost or the hardware fails, the recovery is straight forward. The user simply restores the environment by logging in using another device. With no data saved on the user's device, if the device is lost or stolen, there is much less chance that critical data could be retrieved and compromised.

The following figure shows how Horizon View encapsulates the OS, applications, profiles, and user data into isolated layers and dynamically assembles desktops on demand to provide users with a personalized view of their individual environments.

Figure 72 Highly available and secure desktops



Availability and security, along with ease of management and support, are compelling reasons for moving from traditional physical desktops and laptops to VDI.

VMware Horizon is a comprehensive desktop management environment that runs in a vSphere environment. The environment is managed through vCenter centralized management and can leverage advanced capabilities including, Snapshots, vMotion, DRS, and vSAN storage.

The user's desktop environment runs as a View Desktop VM on an ESXi server, and is accessed via the View Client that uses either Remote Desktop Protocol (RDP) or PC over IP protocols. The View Client can be an application running on a physical desktop, laptop, mobile device, or a web browser using the View Portal. The user's desktop environment can be either a dedicated VM or a floating VM (a VM assigned from a pool when the user logs in). Using the optional View Composer, rather than full images, linked clones can reduce the disk space required. Horizon View includes additional components used to manage the connection, provisioning the environment, authenticate users, and other applications and services.

VMware Horizon with VxRail

The VxRail Appliance is a self-contained compute, storage, and vSphere virtualization, and management environment that is ideally suited for VMware Horizon. VxRail accelerates the Horizon infrastructure deployment, and an environment can be up in running in hours rather than days.

VxRail hyperconverged infrastructure is available in configurations that support hundreds to thousands of virtual desktops. The number of desktops supported is based on the user-workload profile.

Dell EMC has developed tools which provide the ability to model the number of VDI environments and the expected workload profiles to determine appropriate configuration that will meet the immediate and longer term requirements. As demand increases, VxRail non-disruptively scales-up by adding additional appliances and nodes while providing the users with expected performance and consistent user experience.

When deploying Horizon on VxRail Appliances, there are two general approaches: dedicating the VxRail environment to VDI or mixing VDI with other workloads. Horizon Editions or Horizon Add-on Editions are offered exclusively for use with VxRail. VMware or Dell EMC sales representatives can provide more details for the best customer-specific option.

In summary, VxRail with VMware Horizon allows an organization to quickly implement Desktops-as-a-Service (DaaS) and overcome the traditional capital expenditure (CAPEX) barriers of desktop virtualization. The environment can start small and easily scale up as needed. This lowers the initial startup investment. VxRail hyperconverged infrastructure is not only quick to setup, its integrated compute, storage, virtualization, and single-vendor support model eliminate the complexity of traditional infrastructure.

VMware vSphere Platinum

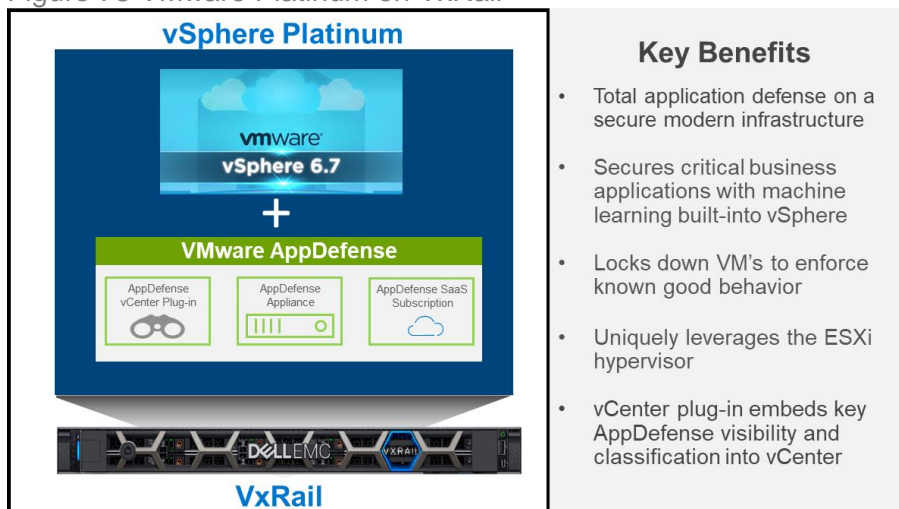
VMware vSphere® Platinum is a purpose-built security solution protecting enterprise applications, infrastructure, data, and access. It combines two proven products: vSphere, the industry-leading, efficient, and secure hybrid cloud platform for all workloads, and VMware AppDefense™, datacenter endpoint security powered by machine learning and embedding threat detection and response into the virtualization layer, to reduce security risk. While being operationally simple, vSphere Platinum ensures applications and virtual machines are running in their known-good states, with minimal overhead and performance impact. For more information about VMware vSphere Platinum, refer to

<https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vsphere/vmw-vsphere-platinum-solution-brief.pdf>

VMware vSphere Platinum can be run on VxRail to stand up total application defense on a secure modern infrastructure. Using Dell EMC VxRail as a platform for VMware vSphere Platinum ensures that optimized cyber resilience and security are built into every layer.

The following figure summarizes the benefits of running VMware vSphere Platinum on VxRail.

Figure 73 VMware Platinum on VxRail



IsilonSD Edge

The EMC IsilonSD product family combines the power of Isilon scale-out NAS with the economy of software-defined storage. IsilonSD Edge is purpose-built software that addresses the need in enterprise edge locations to store growing amounts of unstructured data. IsilonSD Edge allows you to quickly deploy a simple and efficient scale-out NAS solution in a VMware environment. It also extends the reach of the data lake from your core datacenter to your edge locations by economically supporting smaller capacity deployments in a virtualized infrastructure. The data lake enables you to improve storage utilization, eliminate islands of storage and lower your TCO.

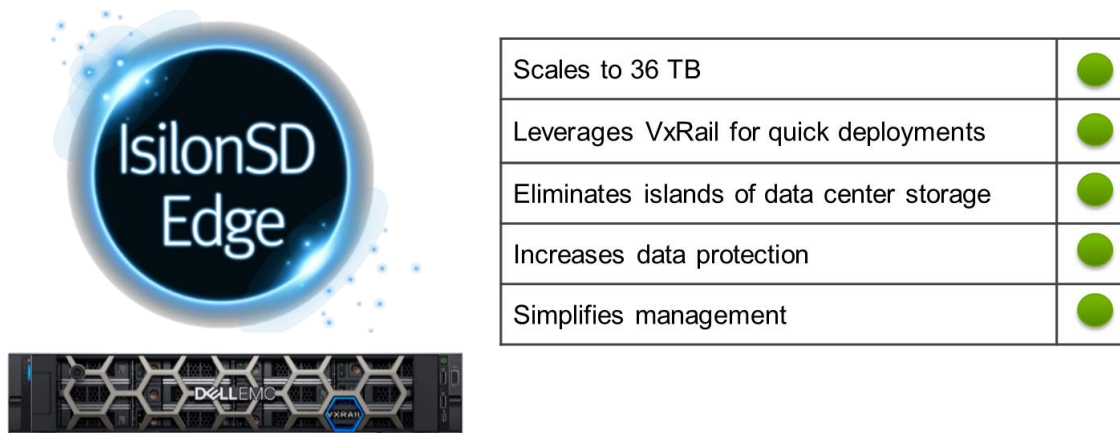
Running IsilonSD Edge on VxRail provides a simple, agile and cost-efficient platform to deliver file services from within a VxRail cluster. It is easy to manage with standard VMware tools. And it allows customers to consolidate and distribute data from and to remote locations. Best of all, it deploys in minutes. IsilonSD Edge includes all Isilon’s OneFS data services & protocols, including SMB, NFS, HDFS and OpenStack Swift.

IsilonSD Edge is tightly integrated with the VMware environment and runs on top of VMware ESXi 6.0 with VxRail 4.0 software or VMware ESXi 6.5 on VxRail 4.5 software. It leverages vCenter® with a management server that runs as a virtual image and can be used to install licenses, or add and remove nodes from a cluster. The IsilonSD Edge management server also installs a plug-in into vCenter that can be used to configure the cluster. The server and storage resources on VxRail do not need to be dedicated to IsilonSD Edge. If performance and capacity needs are met, other applications can run concurrently in the VxRail cluster. IsilonSD Edge with vSAN on VxRail is a validated and tested solution using VMFS or vSAN datastores.

The following figure summarizes the benefits of running IsilonSD Edge on VxRail.

Figure 74 IsilonSD Edge running on VxRail benefits

Delivering Enterprise Grade File Services for VxRail



Each IsilonSD Edge instance on your VxRail cluster can scale up to 36 TB, which is sufficient to handle the needs of many remote and branch offices. You do not have to dedicate your VxRail environment to your IsilonSD Edge cluster – you can run it alongside other workloads in the VxRail cluster.

SAP HANA Certification with VxRail

VxRail is among the first HCI platforms, and the first VMware-based HCI to achieve certification to run SAP HANA, SAP's in-memory database management system. SAP will leverage VxRail's persistent memory to support the application and its use cases.

Customers will benefit from running SAP HANA on VxRail because of the appliance's automation to get implementations up and running quickly, flexibility to offer the right mix of components to support the application from day one, and scalability to ensure future requirements are met. Start fast with automation and full lifecycle management to quickly and effectively support your HANA implementation using VxRail P Series nodes. VxRail is fully certified as a part of the Dell EMC Ready Solution for SAP v1.5 release.

VxRail is best for SAP HANA as it is fast, flexible, powerful, and scalable:

Fast — automation, ease of deployment / management ensure you're up and running quickly

Flexible — configure a system to meet specific needs with build-to-order VxRail on PowerEdge

Powerful — a rich mix of components deliver performance, density and power efficiency for both transactional process and analytics

Scalable — increase power and performance without rip-and-replace system upgrades

For more information and solution guides, please visit: <https://www.dell EMC.com/en-us/solutions/business-applications/sap/hana/index.htm>

Reference Architecture for Splunk

Splunk Enterprise is the industry-leading platform for analyzing machine-generated data. To gain valuable business insights, Splunk Enterprise uses its powerful Splunk Search Processing Language (SPL™) to extract meaningful information from machine data. The insights that are generated from analyzing machine data are called operational intelligence, which has many use cases, including:

IT Operations—Utilization, capacity growth

Security—Fraud detection, real-time detection of threats, forensics

Internet of Things (IoT)—Sensor data, machine-to-machine, human interactions.

Dell EMC and Splunk have partnered to provide jointly validated reference architectures that are optimized for maximum scalability and performance. Splunk software running on Dell EMC converged infrastructure delivers the operational intelligence that is required to drive an organization's digital transformation. When paired together, Dell EMC and Splunk combine the operational intelligence that is provided by the Splunk eco-system with the cost-effective, scalable, and flexible infrastructure of Dell EMC.

The primary benefits Dell EMC provides to your Splunk Enterprise environments include:

Optimized storage data tiering—Aligns storage to hot/warm, cold, and frozen data requirements with high retention and performance.

Cost-effective and flexible scale-out—Provides scale-out capacity and compute, independently or as a single, converged platform.

Powerful data services—Include secure encryption, compression and deduplication, and fast, efficient snapshots for protection.

A reference architecture using Dell EMC VxRail Appliance with Isilon™ for a virtualized Splunk Enterprise environment has been jointly tested and validated by Splunk and Dell EMC to meet

or exceed the performance of Splunk Enterprise running on Splunk's documented reference hardware. VxRail offers the performance and capacity required to meet the infrastructure requirements of a small or medium-sized enterprise Splunk deployment.

See <https://www.emc.com/collateral/service-overviews/h15699-splunk-vxrail-sg.pdf> for more information.

Additional Product information

For documentation, release notes, software updates, or for information about Dell EMC products, licensing, and service, go to the Dell EMC Online Support site (registration required) at: <https://support.emc.com>.

Dell EMC ProSupport for Enterprise

Enterprises need unwavering support for hardware and software and a smart way to manage the mix of vendors in the datacenter. Dell EMC offers a single source with the expertise, know-how and capabilities to help you support your business.

ProSupport offers highly trained experts around the clock and around the globe to address your IT needs, minimize disruptions and maintain a high level of productivity. With over 55,000+ Dell EMC & partner professionals, across 165 countries, speaking more than 55 languages, you can rest assured that with Dell EMC you will be able to:

1. Maximize productivity by leveraging Dell EMC scale and skill
2. Minimize disruptions with around the clock access to highly trained experts
3. Gain efficiency through a single source for all your support needs

Single source, 24X7 global support is provided for VxRail Appliance hardware and software via phone, chat, or instant message. Support also includes access to online support tools and documentation, rapid on-site parts delivery and replacement, access to new software versions, assistance with operating environment updates, and remote monitoring, diagnostics and repair with Dell EMC Secure Remote Services (ESRS).

Our 12 Centers of Excellence and Joint Solution Centers deliver in-house collaboration and industry-leading levels of support, leveraging Dell EMC's alliances with leading application providers such as Oracle and Microsoft. Our 87 technical support sites are comprised of 71 total Dell Tech Support Sites and 16 total EMC Customer Service Centers.

Dell EMC support is recognized with 94% customer satisfaction rating and has received multiple awards including Temkin Group CE Excellence, TSIA STAR awards, Microsoft Deployment Partner of the Year and many more.

The Dell EMC difference is clear, when it comes to your IT strategy we allow you to fearlessly adopt new technology giving you freedom to focus on your business. Having the same enterprise-class support from Dell EMC across your infrastructure gives you that freedom.

Dell EMC ProDeploy Services for VxRail Appliances

Dell EMC offers ProDeploy installation and implementation services to ensure smooth and rapid integration of VxRail Appliances into customer networks. The standard service, optimal for a single appliance, provides an expert on site to perform a pre-installation checklist with the data-center team, confirm the network and Top of Rack (TOR) switch settings, conduct site validation, rack and cable, configure, and initialize the appliance. Finally, an on-site Dell EMC service technician will configure EMC Secure Remote Services (ESRS) and conduct a brief functional overview on essential VxRail Appliance administrative tasks. A custom version of this installation and implementation service is available for larger-scale VxRail Appliance deployments, including those with multiple appliances or clustered environments. Also offered is VxRail Appliance extended service, which is delivered remotely and provides an expert service

technician to rapidly implement VxRail Appliance pre-loaded data services (RecoverPoint for Virtual Machines, and vSphere Data Protection).