Bright Computing

Cluster-as-a-Service for VMware

The easy button for high-performance clusters

March 2021

vmware

Table of Contents

Executive Summary

Overview

High-performance computing applications have historically been a finite set of modeling and simulation applications running on a cluster of bare metal servers, but that trend is changing. The reach of Linux clusters is expanding as organizations in every industry grapple with strategic new applications in areas like machine learning and data analytics, where time to value is of the essence in achieving competitive advantage.

These new applications that require HPC-like infrastructure are increasing the demand for HPC cluster resources—and those resources are often limited. Virtualization technologies, such as VMware vSphere, are closing the gap between virtual machines (VMs) and bare metal servers, making virtual clusters built from VMs viable for a growing set of high-performance applications.

Through advanced automation, Bright Cluster Manager eliminates the complexity of building and managing high-performance Linux clusters, enabling greater organizational efficiency and flexibility. And now, Bright Cluster Manager also supports clusters in VMware vSphere—and goes one step further to enable high-performance clusters as a service. Easily spin up high-performance clusters in vSphere to give engineers and data scientists the freedom to access the computational resources they need, when and how they need them.

Evolution of Clusters on Demand

HPC is undergoing massive changes that present both challenges and opportunities. Historically, HPC has been defined by a consistent and well-known set of applications for modeling and simulation running on bare metal servers in an organization's data center—but that is changing. In addition to the well-known HPC applications, we now have machine learning and data analytics applications vying for access to our trusted HPC systems, ultimately throttling important work in one place or another. And while the seemingly straight-forward answer to this dilemma is to increase the size of an existing system by adding more servers (or entirely new clusters), the cost and delay of procuring those servers creates a drag on innovation that can be far more costly than the servers themselves.

Enter virtual servers available from the public cloud, or from virtualization technologies such as VMware vSphere — virtual servers that can be combined to create a virtual cluster offer the distinct advantage of immediate availability and utility-based pricing that makes a lot of sense for the most urgent, computationally demanding applications. Performance and data security must also be considered; and similar to their physical cousins, virtual clusters must also be built and managed.

With hundreds (or thousands) of server and software elements spanning compute, networking, and storage, clusters are notoriously difficult to build and manage effectively.

Virtualization vs. Bare Metal—the Gap is Shrinking

HPC workloads have traditionally run on bare metal hardware, due to the performance penalty incurred by the virtualization software layer between the operating system and the hardware. Today, however, advances in the VMware vSphere hypervisor, coupled with improvements in virtualization support in x86 microprocessors, have yielded dramatic increases in performance for these highly computation-intensive workloads.

The numerous operational benefits of VMware vSphere are helping to shrink the gap between virtualization and bare metal—enabling virtualized HPC environments to become increasingly common.

Rising Difficulty, Complexity of High-performance Clusters

In the days of traditional HPC, teams built a static cluster and focused energy on keeping it up and running for its lifespan. As such, research institutions and commercial HPC practitioners alike were able to get by with building custom scripts to integrate a collection of different open-source tools to manage their clusters. Integrating tools for server provisioning, monitoring, alerts, and change management is difficult, labor-intensive, and creates an ongoing maintenance burden—but it is possible nonetheless, for organizations with the human resources and skill.

In the emerging new era of HPC, clusters are far from static and far more complex as a result. The need to (1) leverage new types of processors and accelerators, (2) implement servers from different manufacturers, (3) integrate with the cloud, (4) extend to the edge, (5) host machine learning and data analytics applications, and (6) offer end-users VMs and containers alongside bare metal servers raises the bar exponentially for organizations that contemplate a do-it-yourself approach to building a cluster management solution.

Bright Cluster Manager—the Easy Button for High-performance Clusters

Now more than ever before, there is an increasing need for a professional, supported cluster management software that spans hardware, software, and consumption models for the new era in HPC. Bright Cluster Manager is a perfect example of a commercial tool with the features and built-in know-how to build and manage heterogeneous high-performance virtualized clusters for HPC, machine learning, and analytics with ease.

Bright Cluster Manager automatically builds a cluster from bare metal—setting up networking, user directories, security, DNS, and more. Bright sits across an organization's HPC resources — whether on-premises, in the cloud, or at the edge — and manages them across workloads. Bright can also react to increasing demand for different types of applications and instantly re-assign resources within the cluster to service high-priority workloads based on the policies the business sets.

Intersect360 states, "Fundamentally, Bright Computing helps address the big question in HPC: how to match diverse resources to diverse workloads in a way that is both efficient today and future-proof for tomorrow."

Bright's Cluster-as-a-Service for VMware makes running virtualized HPC clusters in VMware fast and easy. Even users with no VMware expertise can create a Bright-managed cluster in minutes. Each virtual cluster is fully configured and ready to use upon creation, so users spend very little time setting up the environment before use. With Bright's Cluster-as-a-Service for VMware, organizations benefit from:

- Greater agility; authorized users can quickly provision the resources they need without waiting for IT
- Higher productivity, by enabling organizations to fully utilize resources
- Proven isolation, where each user is separated from all other users, projects, and data
- Organization control; users can control their own clusters, but the vSphere administrator controls the end-to-end virtualized environment
- Peace of mind, using a commercially supported, enterprise-ready solution.

Business Use Cases

Virtualized clusters as a service are well suited for a wide variety of business use cases, with the most common ones described below.

Resource Consolidation

Resource consolidation is a popular, cost-effective strategy organizations employ to use hardware resources more efficiently. Silos of under-utilized resources are consolidated into a single, shared cluster, or they are replaced with virtual servers that run on that cluster. The result is a cluster containing fewer servers that can do more work.

Using Bright's Cluster-as-a-Service for VMware, organizations can tap into this additional capacity to run virtualized HPC clusters in vSphere. Resources that are used during the day (e.g., providing virtual desktops for interactive users) can be used at night for number crunching as part of the HPC cluster. Similarly, the HPC cluster can be scaled down if/when resources are needed elsewhere.

Isolation Between Users and Groups

Some organizations want to share their infrastructure with groups of internal or external users (customers) that need to be kept totally isolated from one another. Assigning a unique virtualized cluster to each user/group is the perfect solution.

Creating isolated clusters using physical hardware would require sizing each cluster for the largest workload it will support. This would not only create an inflexible and unnecessarily large cluster, but also drive up costs. With Bright's Cluster-as-a-Service for VMware, however, the nodes/resources assigned to each virtual cluster can scale up or scale down, depending on user/workload demand.

Full Control for Power Users

Power users who need complete control over their software environment can be assigned their own virtualized Bright cluster. This way, power users enjoy ensured isolation from all other power users, and they benefit from the freedom to modify their environment as necessary without impacting anyone else.

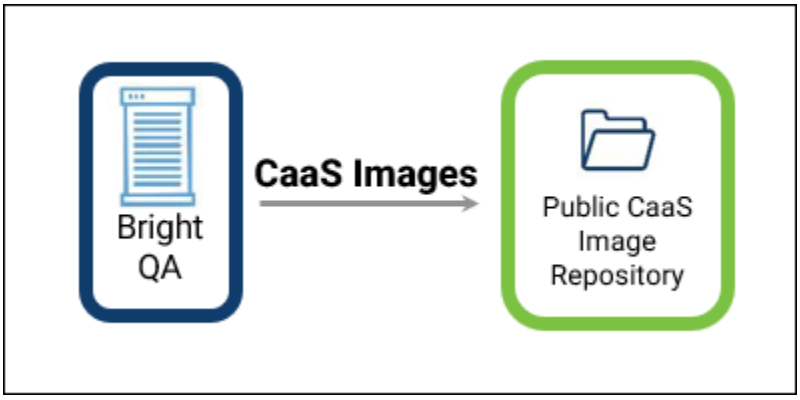
Application Development

Bright's Cluster-as-a-Service for VMware supports one of the best practices for today's DevOps teams—continuous integration/continuous delivery (CI/CD). This agile methodology enables software development teams to meet business requirements and ensure code quality and security by automating development processes. With CI/CD, developers can spin up short-lived virtualized Bright clusters for one specific purpose, such as to run an automated test on a distributed/parallel application or set of applications, and then release the cluster when the test is complete.

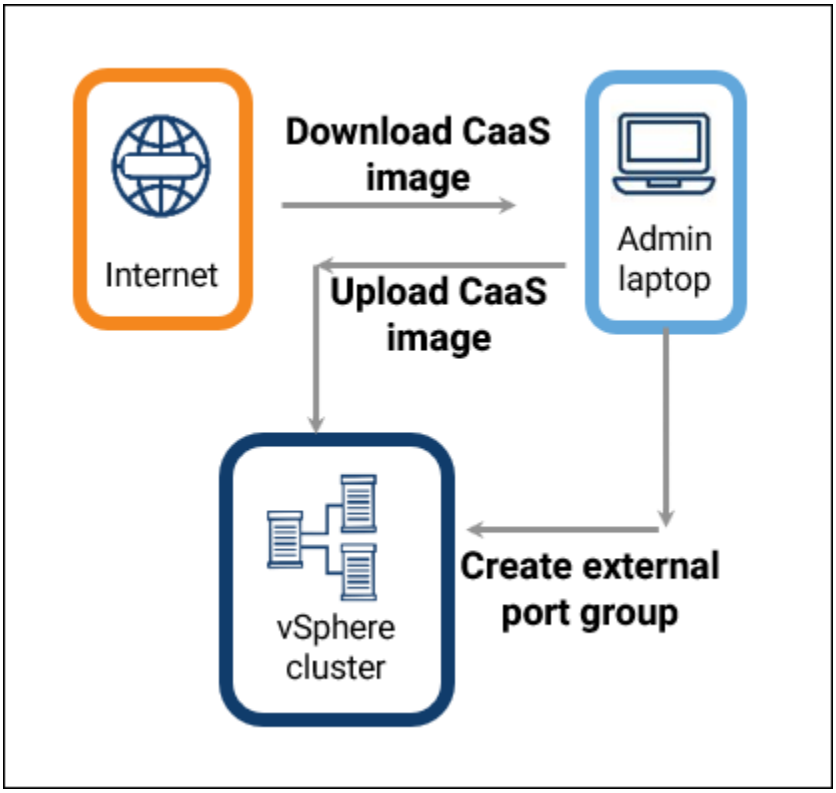
Solution Architecture Design

How it works

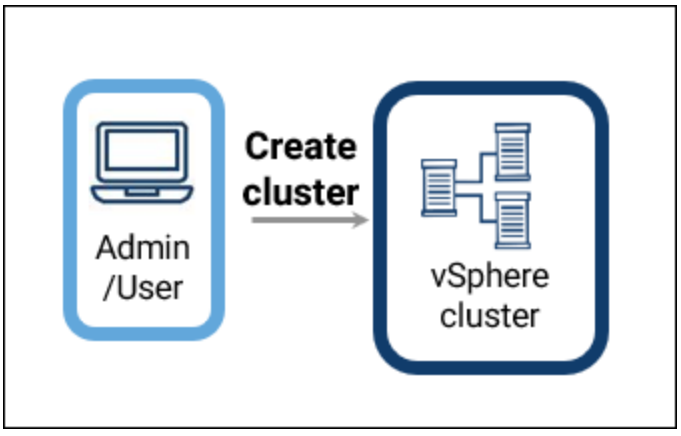
Bright develops the Cluster-as-a-Service container image and uploads it to Docker Hub.



The vSphere administrator downloads the image from Docker Hub and uploads it to vSphere. The administrator creates an external port group and authorizes user accounts to allow them to create clusters.



Authorized users can now create Bright-managed Linux clusters in vSphere. Clusters can be created by the vSphere administrator, and then handed off to the intended user. Or they can be created by a “customer administrator” on behalf of a group. Or they can be created by an authorized end user for his or her own use. In any case, the cluster creator has full control of the cluster.



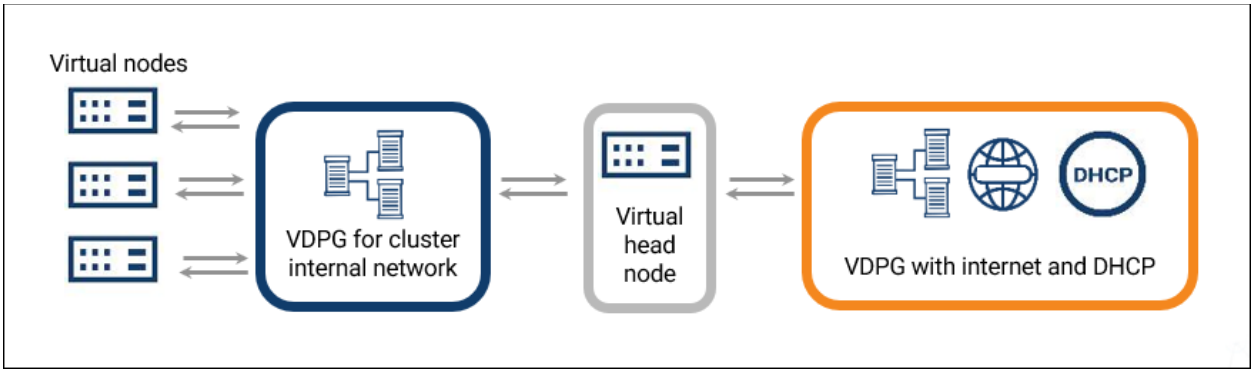
The cluster is created and ready for use in less than five minutes. This is a simple example, but the cluster create command provides a rich set of parameters that allows the user to tailor the cluster to specific needs.

```
[bob@laptop ~]$ cm-cod-vmware cluster create --distro centos8 --name democluster
16:30:49:     INFO: Please wait...
16:31:06:     INFO: ------------------------------------------------------------------
16:31:06:     INFO:         Cluster:  bob-democluster
16:31:06:     INFO: ------------------------------------------------------------------
16:31:06:     INFO:      Image name:  bcmh-centos8u1-trunk-170(centos8u1-trunk:170)
16:31:06:     INFO:      Node image:  bcmn-centos8u1-trunk-170(centos8u1-trunk:170)
16:31:06:     INFO:      Image date:  2020-09-06 22:18 (18h 12m ago)
16:31:06:     INFO:  Package groups:  none
16:31:06:     INFO:      Git hashes:  cmdaemon: n/a
16:31:06:     INFO:                   cm-setup: n/a
16:31:06:     INFO:                   cluster-tools: n/a
16:31:06:     INFO:         Version:  trunk
16:31:06:     INFO:          Distro:  centos8u1
16:31:06:     INFO: ------------------------------------------------------------------
16:31:06:     INFO: Press ENTER to continue and create the cluster.
16:31:06:     INFO: Press ctrl+c (or type 'a') to abort. Type 'i' for more info.

16:31:15:     INFO: Creating cluster bob-democluster
16:31:15:     INFO: Creating Portgroup bob-democluster with VLAN ID 400 on switch DSwitch01
16:31:17:     INFO: Deploying OVF
16:32:36:     INFO: Powering on head node
16:32:42:     INFO: Node was powered on. Waiting for network
16:33:42:     INFO: Waiting for sshd to start (ssh root@10.2.182.243).
16:33:42:     INFO: Waiting for CMDaemon to start.
16:34:16:     INFO: Waiting for CMDaemon to initialize.
16:35:17:     INFO: Cluster has been created successfully.
16:35:17:     INFO: Cluster name:    bob-democluster
16:35:17:     INFO:   SSH string:    ssh b1 (alias defined in ~/.ssh/config)
16:35:17:     INFO:           IP:    10.2.182.243
[bob@laptop ~]$
```

Figure 1: Example cluster create command

The cluster create command creates a fully functional Bright-managed cluster in just minutes resulting in the architecture shown below. Two vSphere distributed port groups (VDPG) are created — a private VDPG that connects the compute nodes to the head node, and public VDPG that connects the head node to the outside world. A head node is created and connected to both VDPGs, and the compute nodes are created and connected to the private VDPG.



Once a cluster is created, its owner can log in to create user accounts and prepare it for its intended use. Users can then log in and begin using the resource.

The users do not need to know anything about the deployment, or even that the cluster is running in vSphere. Furthermore, neither the cluster owner nor the cluster users will see any other clusters because every cluster is isolated from all others.

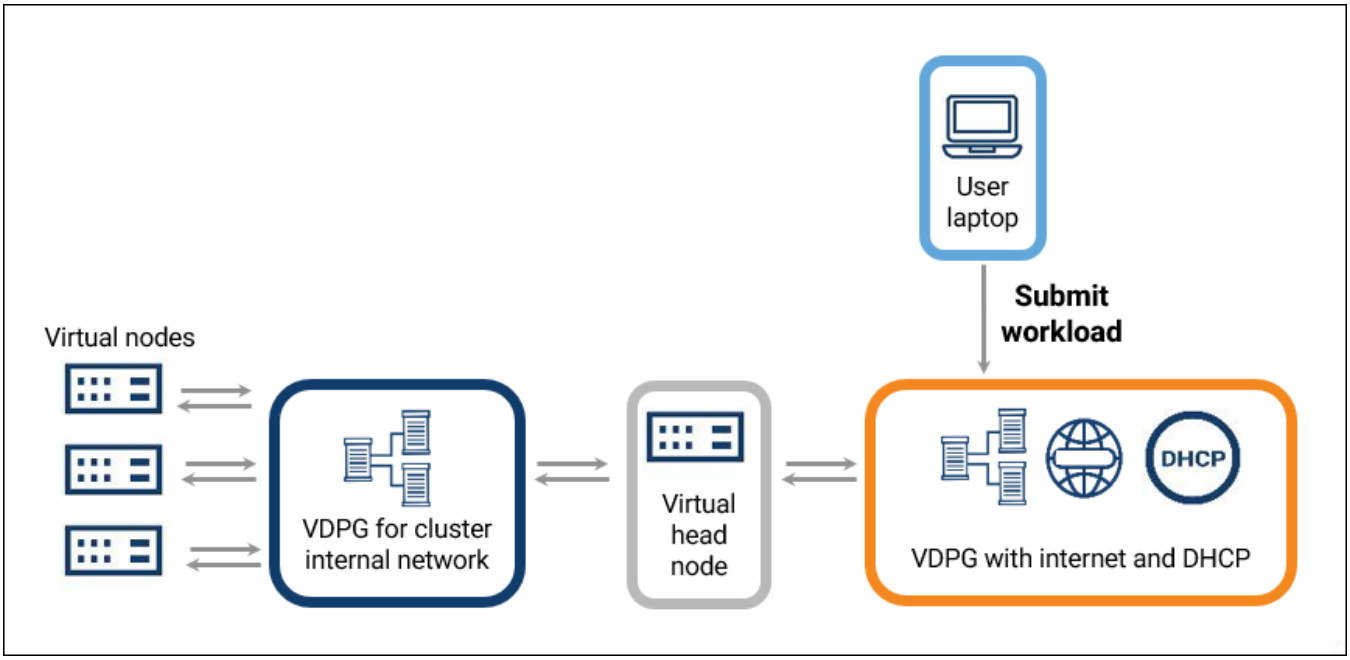


Figure 2: End users log in and use the cluster in the same as they would access any other HPC cluster

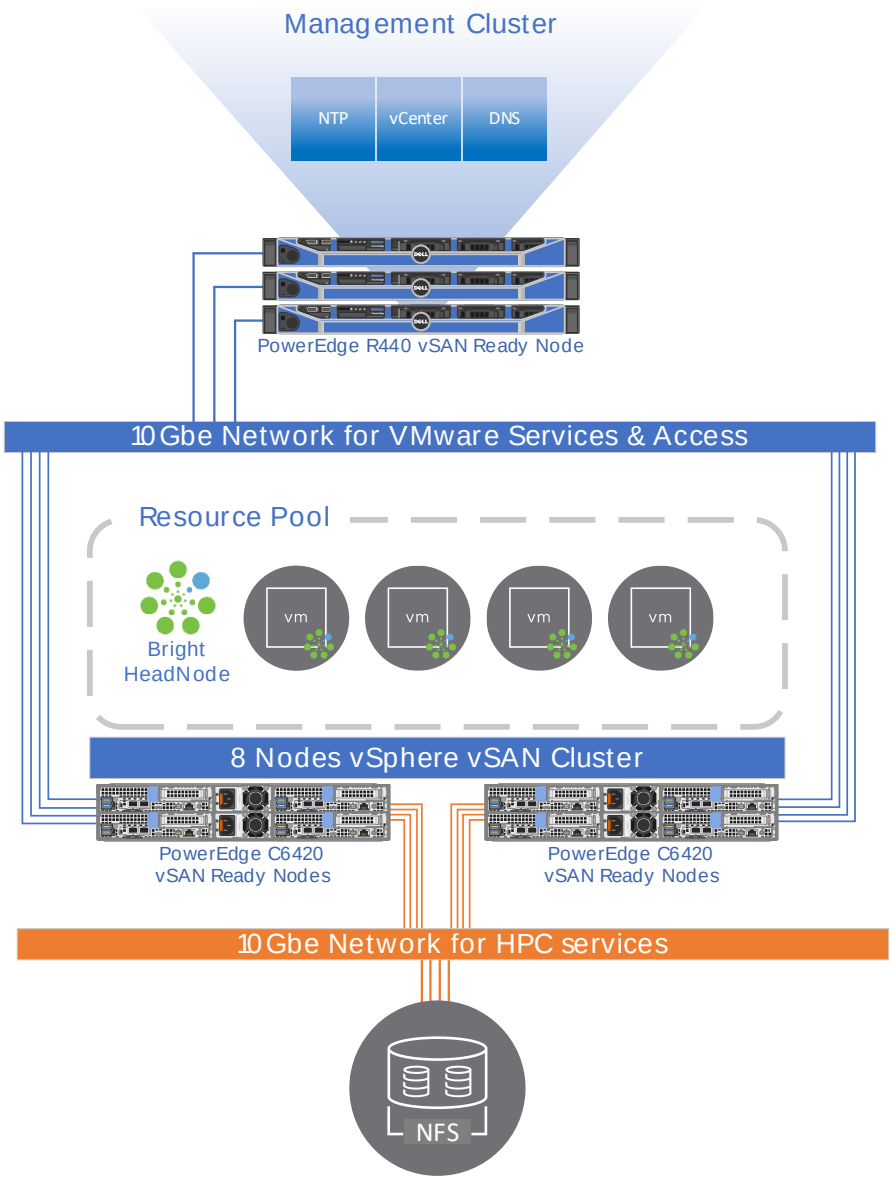
Network Design

The physical network topology was designed to segment the VMware services networks (e.g., vSAN, vMotion, etc.) from the HPC data plane to reduce network congestion and jitter when running latency-sensitive jobs. Customers requiring optimal performance for their HPC workloads can choose to implement a more robust HPC network fabric by considering higher bandwidth devices (25, 40, 100Gbe). They can also adopt RoCE and leverage the PVRDMA adapters that are supported natively on ESXi 6.7+.

In the diagram below, the management cluster is running on a dedicated fabric and the HPC traffic is also running on dedicated switches. The eight-node vSAN cluster is being used as the HPC compute cluster and enables an HPC admin to provision a Bright head node with Cluster-as-a-Service that can then be handed off to an end user. The head node is connected to both the VMware Services and the HPC networks. This allows the end user to access the head node on the corporate network and isolates the communication between the compute nodes to the HPC network. In addition, the NFS storage is connected to the HPC network so datasets and home directories can be mapped to the compute resources and head node if/as necessary.

The following section outlines the vSphere configuration to demonstrate how the logical networking is configured on the vSphere cluster using Distributed Virtual Switches (DVS).

Physical Topology



Logical Topology

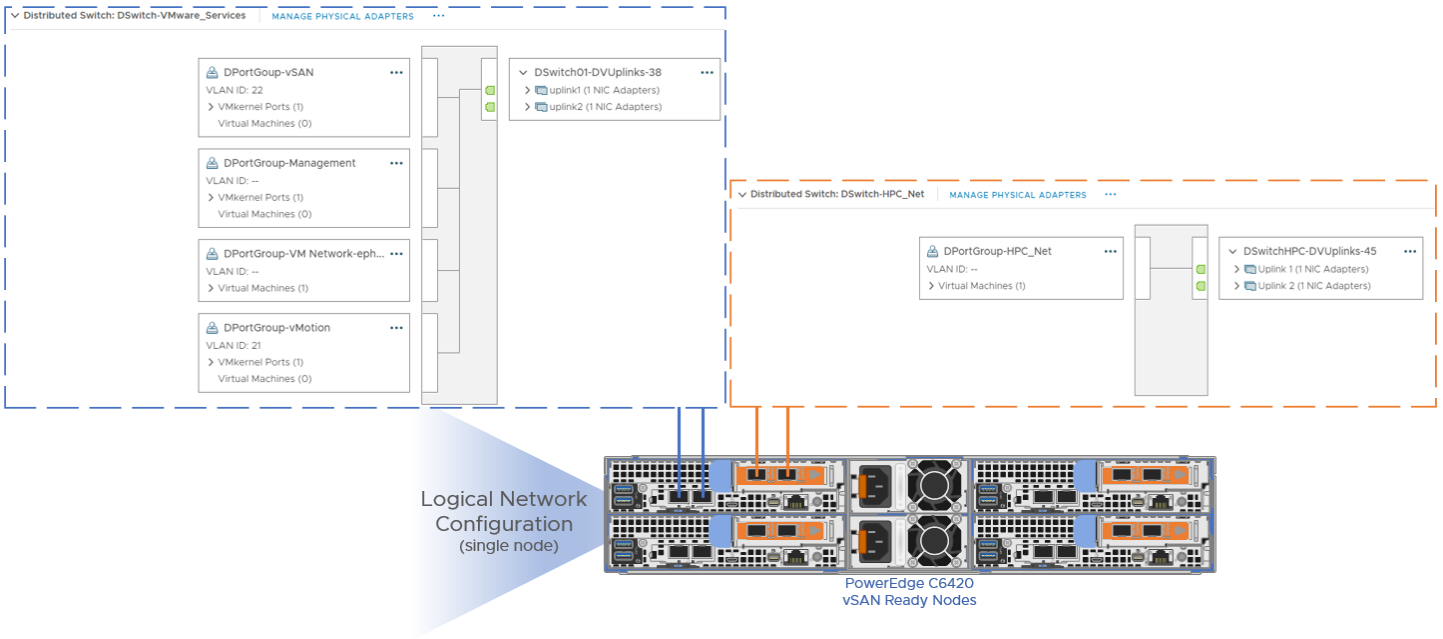
When deciding how to set up your ESXi host networking, it is recommended to use vSphere Distributed Switches (VDS). The VDS extends the features and capabilities of virtual networks while simplifying provisioning and the ongoing configuration, monitoring and management process.

vSphere network switches can be divided into two logical sections: the data plane and the management plane. The data plane implements the packet switching, filtering, tagging and so on. The management plane is the control structure used by the operator to configure data plane functionality. Each vSphere Standard Switch (VSS) contains both data and management planes, and the administrator configures and maintains each switch individually.

The VDS eases this management burden by treating the network as an aggregated resource. Individual host-level virtual switches are abstracted into one large VDS spanning multiple hosts at the data-center level. In this design, the data plane remains local to each VDS, but the management plane is centralized.

As shown below, a graphical representation of how each physical interface maps to its respective VDS and what networks are configured as port groups. In this example, the ESXi management interface and VM Network are using an untagged VLAN. All other port groups have been assigned a tagged VLAN to adhere to best practices. The VLAN has also been left untagged on the HPC network because there is only one VLAN used for this network for simplicity purposes. It should be noted that for networking resiliency, each physical connection should be connected to a ToR (top of rack) pair of switches that follows high availability (HA) best practices to ensure connectivity in case there are any failures at the top of rack (ToR) switch layer.

The following section dives into some of the consumption model use cases for how Cluster-as-a-Service can be implemented to better understand the personas involved, and how a flexible consumption model allows for several user experiences.



Consumption Use Cases

VMware administrator creates clusters

In this case, the VMware administrator may create the clusters on behalf of users, which may be individual users or a group of users, such as a business unit. In this case, none of the cluster users need VMware credentials. Once the cluster is created, the cluster is handed over to the user and the user has full control of the cluster.

Customer administrator creates clusters

Some companies want to share their vSphere infrastructure with several independent groups with isolation between them. Each group is a customer and designates a "customer administrator." The customer administrator creates and manages the cluster on behalf of his group. The customer administrator needs a VMware account, but the customer users do not. They access the cluster using a standard Linux account.

Users create their own clusters

In some cases, organizations may want to allow individual users to create clusters. In this case, each user needs a VMware account. The user has full control of the cluster subject to quotas established by the VMware administrator.

Technical Specification

The table below shows the technical specification of this solution validation.

Hardware

Solution Specification	Quantity
Dell EMC PowerEdge C6240 (Compute Nodes)	8
Dell EMC PowerEdge R440 (Management Cluster)	3
Dell EMC PowerSwitch s4048	4
Dell EMC PowerScale Isilon H500	4 nodes

Software

These are the versions of VMware software that were used during validation of this solution:

Components	Version	(Build#) {minimum}
ESXi	6.7U3 7.0.1	14320388 16850804
vSAN	6.7U3 7.0.1	14320388 16850804
vCenter	6.7U3 7.0	6.7.0.45000 7.0.0.10400

Conclusion

Bright's Cluster-as-a-Service for VMware allows organizations to *fully* utilize their vSphere infrastructure 24x7x365. This saves money and gives teams the ability to spin-up Clusters-as-a-Service and accelerate innovation for the business by increasing the capacity of high-performance infrastructure that drives innovation. Tapping into available vSphere capacity for high-performance needs will also reduce the need for organizations to use public cloud infrastructure, further reducing their costs, and eliminating the need to move data into the public cloud.

Additional Resources/References

- This solution is documented in the Bright Cloudbursting Manual (https://support.brightcomputing.com/manuals/9.1/cloudbursting-manual.pdf)
- Bright technical documentation (https://www.brightcomputing.com/documentation)
- Free online Bright training (https://training.brightcomputing.com/)
- vSphere Documentation
- vSphere Networking Guide
- vSAN Documentation
- Dell EMC PowerEdge C6420
- Dell EMC PowerEdge R440
- Dell EMC PowerSwitch S4048-ON Spec Sheet
- Dell EMC PowerScale Isilon H500

Authors

Chris Gully, Staff Solutions Architect, VMware OCTO

Robert Stober, Product Manager, Bright Computing

Available in Dell Technologies Customer Solution Centers