# Table of Contents

1 Introduction ........................................................................................................................................... 4  
   1.1 About Attached Storage for vCenter Server ................................................................................. 4  
   1.2 Key Benefits .................................................................................................................................. 4  
2 Design .................................................................................................................................................. 5  
   2.1 Overview ....................................................................................................................................... 5  
   2.2 Physical Infrastructure ...................................................................................................................... 5  
      2.2.1 Physical Network Design ......................................................................................................... 6  
      2.2.2 Physical Storage Design ............................................................................................................ 7  
   2.3 Infrastructure Management ............................................................................................................ 8  
      2.3.1 Virtual Networking Design ..................................................................................................... 9  
      2.3.2 Storage Settings ....................................................................................................................... 10  
      2.3.3 Cluster Settings....................................................................................................................... 12
Summary of Changes
This section records the history of significant changes to this document. Only the most significant changes are described here.

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Author</th>
<th>Description of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>2016-12-15</td>
<td>IBM Cloud</td>
<td>Initial Release</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 About Attached Storage for vCenter Server

The purpose of this document is to define and describe the attached storage solution architecture for the vCenter Server offering deployed in the IBM Cloud. Specifically, it will detail the components of the solution and high-level configuration of each component in the design. This solution is considered to be an additional component and extension of the upcoming vCenter Server solution offering on IBM Cloud. As a result, this document will not cover the existing configuration of the vCenter Server solution on IBM Cloud. Therefore, it is highly recommended to review and understand the aforementioned architecture located on the IBM Architecture Center before reading this document.

![Diagram of vCenter Server on IBM Cloud]

Figure 1 vCenter Server on IBM Cloud

1.2 Key Benefits

While attached storage is not a prerequisite for VMware environments, its use as a shared storage device does provide many benefits to end-users as well as IT operations. These benefits are realized in the areas of high availability, distributed resource scheduler, and migration. Additionally, it allows for the efficient use of storage capacity, simplifies management, and enables the following functionality:

- **vSphere Distributed Resource Scheduler and Resource Pools** – feature that allows for abstraction and flexible management of resources via load balancing and virtual machine placement. Resource pools can be grouped into hierarchies and used to hierarchically partition available CPU and memory resources.
- **vSphere High Availability** - provides high availability for virtual machines by pooling them and the hosts they reside on into a cluster. Hosts in the cluster are monitored and in the event of a failure, the virtual machines on a failed host are restarted on alternate hosts.
- **vSphere Datastore Clusters** - a collection of datastores with shared resources and a shared management interface.
- **vSphere Fault Tolerance** - provides continuous availability to virtual machines, eliminating downtime and disruption—even in the event of a complete host failure.
2 Design

2.1 Overview

IBM Cloud for VMware Solutions provides VMware technology deployed in an automated fashion to this document’s prescribed architecture within IBM Cloud datacenters across the globe. Within the IBM Cloud solutions portfolio, the base VMware vCenter Server on IBM Cloud offering consists of up to twenty vSphere hosts, a single Platform Services Controller (PSC) and a vCenter Server Appliance capable of managing up to 100 hosts and 1,000 virtual machines. The architecture prescribed here complements the vCenter Server solution by adding attached storage as a shared storage device for the environment. The attached storage device is located within the same datacenter as the vCenter Server deployment and is comprised of a single NFS share or multiple NFS exports from the IBM Cloud.

![vCenter Server High Level Components](image)

2.2 Physical Infrastructure

The physical infrastructure consists of three main components including physical compute, physical storage and physical network. Since this architecture is considered an add-on component to the existing vCenter Server solution architecture, this document will only discuss the physical elements specific to the attached storage device. This includes the IBM Cloud services network as well as the physical storage consumed by the infrastructure.
2.2.1 Physical Network Design

Physical networking is handled by IBM Cloud. This section will describe the physical network provided by the IBM Cloud as it relates to attached storage.

2.2.1.1 IBM Cloud Network Overview

The physical network of IBM Cloud is separated into three distinct networks: Public, Private, and Management. The public, private, and management networks have already been described in the vCenter Server solution architecture, however, this document will discuss the services network that is part of the private network.

![Figure 3 IBM Cloud High-Level Network](image)

2.2.1.1.1 Private Services Network

The IBM Cloud contains a private services network that provides common services such as block storage, file storage, object storage, DNS resolvers, and NTP servers. This private network is separate from the customer private network, and it enables environments to seamlessly connect to services located in the IBM Cloud. The private network is multi-tiered in that servers and other infrastructure are connected to an aggregated backend customer switches (BCS). These aggregated switches are attached to a pair of separate routers (i.e., backend customer routers, BCR) for L3 networking. The private network also supports the ability to utilize jumbo frames (i.e., MTU 9000) for physical host connections.

2.2.1.2 VLANs

In this design, a new private VLAN with 2 additional subnets is created to segregate storage traffic. The VLAN contains 2 subnets, however, only the private portable is used. The primary subnet is reserved for IBM Cloud. Additionally, this VLAN is trunked to the existing switch ports so that both private VLAN A and private VLAN B are able to be used in the deployment.

<table>
<thead>
<tr>
<th>VLAN</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>Primary</td>
<td>Assigned to physical hosts for public network access. Not utilized upon initial deployment.</td>
</tr>
<tr>
<td>Private A</td>
<td>Primary</td>
<td>Single subnet assigned to physical hosts assigned by IBM Cloud. Used by the management interface for vSphere management traffic.</td>
</tr>
<tr>
<td>Private A</td>
<td>Portable</td>
<td>Single subnet assigned to virtual machines functioning as management components.</td>
</tr>
</tbody>
</table>
### VLAN and Subnet Summary

<table>
<thead>
<tr>
<th>VLAN</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private A</td>
<td>Portable</td>
<td>Single subnet assigned for vMotion traffic.</td>
</tr>
<tr>
<td>Private B</td>
<td>Primary</td>
<td>Single subnet assigned and reserved by IBM Cloud</td>
</tr>
<tr>
<td>Private B</td>
<td>Portable</td>
<td>Single subnet assigned for storage traffic</td>
</tr>
</tbody>
</table>

**Table 1 VLAN and Subnet Summary**

#### 2.2.2 Physical Storage Design

This section will discuss the configuration of the attached storage device that is present in the IBM Cloud. Note that this attached storage device complements the existing vCenter Server solution and as a result, locally attached disks that are internal to the physical hosts will not be discussed.

##### 2.2.2.1 Attached Storage

Performance and Endurance storage are IBM Cloud storage solutions that are designed to support high I/O applications requiring predictable levels of performance. This predictable performance is achieved through the allocation of protocol-level input/output operations per second (IOPS) to individual volumes. IOPS ranging from 100 through 48,000 can be provisioned with storage sizes that range from 20 GB to 12 TB. Performance and Endurance storage volumes are available for both block storage and file storage.

In this design, the attached storage used for the vCenter Server solution focuses on Endurance storage. As a result, the user will be able to select and attach (via automation) Endurance NFS exports ranging in size from 20GB to a maximum of 12TB. The IBM Cloud allows up to 64 vSphere ESXi hosts to connect to a single Endurance NFS export.

Endurance is available in three IOPS performance tiers to support varying application needs. Note that once an NFS share is provisioned, it cannot be resized or reconfigured to allow for more/less IOPS. The tiers are as follows:

- **0.25 IOPS per GB** is designed for workloads with low I/O intensity. These workloads are typically characterized by having a large percentage of data inactive at a given time. Example applications include storing mailboxes or departmental level file shares.

- **2 IOPS per GB** is designed for most general purpose usage. Example applications include hosting small databases backing web applications or virtual machine disk images for a hypervisor.

- **4 IOPS per GB** is designed for higher-intensity workloads. These workloads are typically characterized by having a high percentage of data active at a given time. Example applications include transactional and other performance-sensitive databases.

- **10 IOPS per GB** is designed for the most demanding workloads such as analytics and NoSQL Database workloads. This feature is only enabled in select datacenters: DAL09, SJC03, and WDC04.

In addition to the tiers of storage, IBM Cloud Endurance storage supports a wide selection of application needs, including snapshots and replication, and encryption at rest in the IBM Cloud datacenter locations.
2.3 Infrastructure Management

Infrastructure management refers to the VMware components that are managing the vSphere ESXi infrastructure. For this document, the architecture and configuration related to the vCenter Server constructs that allow for shared, attached storage is discussed.

![Diagram of Infrastructure Management with File Storage Network Diagram]
2.3.1 Virtual Networking Design

The network virtualization utilized in this design leverages the existing vSphere Distributed Switch (vDS) associated with the private network and specified in the vCenter Server solution architecture document.

2.3.1.1 vSphere Distributed Switch (vDS)

As stated previously, another VLAN is created within the vCenter Server solution and used to attach the NFS mount point to the ESXi hosts in the existing cluster. Since the vCenter Server solution already has a vDS associated with the private network, another port group is created and tagged with the additional VLAN number since this additional VLAN is not native. The settings of this new port group are described in Table 2 NFS Port Group Summary.

<table>
<thead>
<tr>
<th>Port Group Name</th>
<th>SDDC-DPG-NFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port binding</td>
<td>Static</td>
</tr>
<tr>
<td>VLAN type</td>
<td>&lt;Private VLAN B&gt;</td>
</tr>
<tr>
<td>Load balancing</td>
<td>Route based on originating virtual port</td>
</tr>
<tr>
<td>Active Uplinks</td>
<td>Uplink1 and uplink2</td>
</tr>
</tbody>
</table>

Table 2 NFS Port Group Summary

In addition to the creation of the vDS port group for NFS storage traffic, a VMkernel port is created on each vSphere ESXi host in the deployment and assigned to the SDDC-DPG-NFS port group. The VMkernel port is also assigned an IP address from the private portable subnet associated with the attached storage VLAN (i.e., Private VLAN B) and its MTU is set to 9000 to support jumbo frames.

![SDDC-Dswitch-Private](image)

Figure 5 Private vDS Portgroups and Uplinks

2.3.1.2 Sphere Host Static Routing

Although the vDS is configured with a new port group and a VMkernel port is assigned to the aforementioned port group, the solution creates a static route on each vSphere ESXi host in the deployment so that all NFS traffic traverses the VLAN and subnet for NFS. The static route is created in /etc/rc.local.d/local.sh so that it persists across host reboots.
2.3.2 Storage Settings

This design supports the attachment of shared storage via NFS v3 and NFS v4.1 (NFS v4 is not supported). Upon ordering of the vCenter Server solution, the user determines which NFS version to utilize for the environment. While NFS v4.1 introduces better performance and availability through load balancing and multipathing, it does restrict the utilization of vSphere features including Storage DRS, Storage I/O Control, and use of Site Recovery Manager. Table 3 NFS Protocols and vSphere Features below lists the NFS protocols and support for vSphere features when using vSphere 6.0.

<table>
<thead>
<tr>
<th>vSphere Feature</th>
<th>NFS v3</th>
<th>NFS v4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>vMotion and Storage vMotion</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>High Availability (HA)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fault Tolerance (FT)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Distributed Resource Scheduler (DRS)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Host Profiles</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Storage DRS</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Storage I/O Control (SIOC)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Site Recovery Manager</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Virtual Volumes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3 NFS Protocols and vSphere Features

Note that all attached storage for this design is limited to the IBM Cloud storage available in the same data center as the vCenter Server solution. Additionally, the NFS v3 and 4.1 datastores are mounted using the same NFS version on all hosts and all virtual disks stored to the datastore are thin provisioned by default.

2.3.2.1 NFS v3

When NFS v3 is chosen as the preferred method to attach the NFS share, the architecture is such that it utilizes a single IP address from IBM Cloud storage to connect to the share. In addition, the NFS share is attached to all hosts in the vCenter Server cluster and placed into a datastore cluster with storage DRS enabled.

2.3.2.1.1 vSphere Storage Distributed Resource Scheduler (Storage DRS)

Storage DRS allows users to manage the aggregated resources of a datastore cluster. When Storage DRS is enabled, it provides recommendations for virtual machine disk placement and migration to balance space and I/O resources across the datastores in the datastore cluster. The following features are available when storage DRS is turned on:

- Space load balancing among datastores within a datastore cluster
- I/O load balancing among datastores within a datastore cluster
- Initial place for virtual disks based on space and I/O workload.

In this design, Storage DRS is enabled with the automation level set to “Fully Automated”. As a result, files will be migrated automatically to optimize resource usage on the data cluster. Note that all other Storage DRS options are set to “Use cluster settings” since the cluster is fully automated.

2.3.2.1.2 Storage DRS Runtime Settings

The aggressiveness of Storage DRS is determined by specifying thresholds for space used and I/O latency. Storage DRS collects resource usage information for the datastores in a datastore cluster. vCenter Server uses this information to generate recommendations for placement of virtual disks on datastores.
When low aggressiveness level is set for a datastore cluster, Storage DRS recommends Storage vMotion migrations only when absolutely necessary, for example, if when I/O load, space utilization, or their imbalance is high. When high aggressiveness level is set for a datastore cluster, Storage DRS recommends migrations whenever the datastore cluster can benefit from space or I/O load balancing.

The following threshold categories are available in the datastore cluster.

- **Space Utilization:** Storage DRS generates recommendations or performs migrations when the percentage of space utilization on the datastore is greater than the threshold you set in the vSphere Web Client.
- **I/O Latency:** Storage DRS generates recommendations or performs migrations when the 90th percentile I/O latency measured over a day for the datastore is greater than the threshold.

In this design, the Storage DRS Runtime Settings are enabled and thresholds are kept to their respective default values. It is highly recommended to change these values based on the workload I/O requirements and latency sensitivity.

![Figure 6 Storage DRS Runtime Settings](image)

### 2.3.2.1.3 Storage I/O Control

When SIOC is enabled in the environment, it will change the device queue length for the single VM. The change to the device queue length reduces the storage array queue for all VMs to an equal share and throttle of the storage queue. SIOC engages only if resources are constrained and the storage I/O latency is above a defined threshold.

In order for SIOC to determine when a storage device is congested or constrained, it requires a defined threshold. The congestion threshold latency is different for different storage types, however, this design recommends and implements a threshold latency of 10 milliseconds.

Note that one can also limit individual virtual disks for individual VMs or grant them different shares with SIOC. The limiting of disks and granting different shares allows the end user to match and align the environment to the workload with the acquired file storage volume IOPS number. The limit is set by IOPS and it is possible to set a different weight or "Shares." Virtual disks shares set to High (2,000 shares) receive twice as much as I/O as a disk set to Normal (1,000 shares) and four times as much as one set to Low (500 shares). Normal is the default value for all the VMs, so users need to adjust the values above or below Normal for the VMs that actually require it.

### 2.3.2.1.4 Additional Storage

When the need arises to add additional storage to the environment due to insufficient space or high latency, the end user has the ability to order another NFS share from the console. Once the share is ordered, the automation attaches the export to the vSphere ESXi hosts in the cluster and places it into the aforementioned storage cluster. Placing the new NFS share in the storage cluster
effectively and seamlessly scales out the storage associated with the environment without burdening the end-user with storage-level migrations.

2.3.2.2 **NFV v4.1**

When NFS v4.1 is chosen as the preferred method to attach the NFS share, the architecture is such that it utilizes multiple IP addresses from IBM Cloud storage to connect to the share. While the share is attached to all hosts in the vCenter Server cluster, it is not placed into a datastore cluster due to the vSphere limitation of using Storage DRS with NFS v4.1. Also note that NFS 4.1 with vSphere 6.0 does not support hardware acceleration. This limitation does not allow the creation of thick virtual disks on NFS v4.1 datastores.

2.3.2.2.1 **Additional Storage**

When the need arises to add additional storage to the environment due to insufficient space or high latency, the end user has the ability to order another NFS share from the console. Once the share is ordered, the automation attaches the export to the vSphere ESXi hosts in the cluster. It is then up to the end user to migrate virtual machines and load balance the datastores so that space is properly utilized and latency requirements are met.

2.3.2.3 **Advanced Configuration Parameters**

Regardless of whether the user chooses NFS v3 or NFS v4.1, this design adds advanced configuration parameters that are recommended by IBM Cloud. As a result, the following parameters in Table 4 NFS Advanced Configuration Parameters are set on each vSphere ESXi host that is attached to the IBM Cloud NFS share.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net.TcpipHeapSize</td>
<td>32</td>
</tr>
<tr>
<td>Net.TcpipHeapMax</td>
<td>512</td>
</tr>
<tr>
<td>NFS.MaxVolumes</td>
<td>256</td>
</tr>
<tr>
<td>NFS.HeartbeatMaxFailures</td>
<td>10</td>
</tr>
<tr>
<td>NFS.HeartbeatFrequency</td>
<td>12</td>
</tr>
<tr>
<td>NFS.HeartbeatTimeout</td>
<td>5</td>
</tr>
<tr>
<td>NFS.MazQueueDepth</td>
<td>64</td>
</tr>
</tbody>
</table>

*Table 4 NFS Advanced Configuration Parameters*

2.3.3 **Cluster Settings**

Prior to the addition of attached storage, the vCenter Server solution did not enable advanced features such as vSphere Distributed Resource Scheduler (DRS) and vSphere High Availability (HA). With the addition of the NFS attached storage device, these features are enabled on the cluster with the settings listed in the sections that follow.

2.3.3.1 **vSphere Distributed Resource Scheduler**

There are two main features that are enabled when turning on vSphere DRS on a cluster: Load Balancing and Power Management.

2.3.3.1.1 **Load Balancing**

With load balancing the distribution and usage of CPU and memory resources for all hosts and virtual machines in the cluster are continuously monitored. DRS compares these metrics to an ideal resource utilization given the attributes of the cluster’s resource pools and virtual machines,
the current demand, and the imbalance target. It then performs (or recommends) virtual machine migrations accordingly. When a virtual machine is first powered on in the cluster, DRS attempts to maintain proper load balancing by either placing the virtual machine on an appropriate host or making a recommendation. The placement or recommendation settings are set in the DRS Automation section of the cluster settings.

In this design, the DRS Automation level is set to fully automated so when virtual machines are powered on they are automatically placed onto hosts with sufficient capacity. The virtual machines are also automatically migrated from one host to another to load balance the cluster. Additionally, the migration threshold of the DRS cluster is set at the midpoint between conservative and aggressive such that priority 1, priority 2, and priority 3 recommendations are applied. This means that vCenter Server will provide at least good improvements to the cluster’s load balance.

![DRS Automation Settings](image)

**Figure 7 DRS Automation Settings**

In addition to the automation level and migration threshold of the cluster, this design enables virtual machines automation so that overrides to individual virtual machines can be set by the end user. This allows more granular control of virtual machines and enables end users to further prioritize the load balancing of virtual machines.

### 2.3.3.1.2 Power Management

When the VMware Distributed Power Management feature is enabled, DRS compares cluster- and host-level capacity to the demands of the cluster’s virtual machines, including recent historical demand. It places (or recommends placing) hosts in standby power mode if sufficient excess capacity is found or powering on hosts if capacity is needed. Depending on the resulting host power state recommendations, virtual machines might need to be migrated to and from the hosts as well.

In this design, power management is disabled since there is no operational or financial benefit to powering on and off hosts in the cluster.

### 2.3.3.2 vSphere High Availability

vSphere provides high availability for virtual machines by pooling them and the hosts they reside on into a cluster. Hosts in the cluster are monitored and in the event of a failure, the virtual machines on a failed host are restarted on alternate hosts.

In this design, vSphere High Availability is enabled with host monitoring and virtual machine monitoring on the cluster.
2.3.3.2.1 Host Monitoring

Host monitoring allows hosts in the cluster to exchange network heartbeats and allows vSphere HA to take action when it detects failures. This feature is enabled in this design.

2.3.3.2.2 Virtual Machine Monitoring

The virtual machine monitoring feature uses the heartbeat information that VMware Tools captures as a proxy for guest operating system availability. This allows vSphere HA automatically to reset or restart individual virtual machines that have lost their ability to heartbeat. This design enables both VM and application monitoring.

2.3.3.2.2.1 Failure Conditions and VM response

The failure conditions define how the virtual machines fail and the response given to each failure. In this design, the VM restart priority is set to medium; it is highly recommended to review this value and adjust settings accordingly so the restart priority matches the importance of the workload. Additionally, the response for host isolation is set to “Power off and restart VMs” so that virtual machines are not effected by an isolated host in the cluster. The rest of the values for this setting are set to default.

![Figure 8 vSphere HA Failure Conditions and VM response](image-url)
2.3.3.2.2  Admission Control

vCenter Server uses admission control to ensure that sufficient resources are available in a cluster to provide failover protection and to ensure that virtual machine resource reservations are respected. In this design, the failover capacity is reserved by specifying a percentage of the cluster resources. The defined failover capacity is set to 25% CPU and 25% Memory.

2.3.3.2.2.3  Datastore for Heartbeating

vSphere HA uses datastore heartbeating to distinguish between hosts that have failed and hosts that reside on a network partition. Datastore heartbeating allows vSphere HA to monitor hosts when a management network partition occurs and to continue to respond to failures that occur. In this design, the heartbeat datastore selection policy is set to “Automatically select datastores accessible from the host”.

© Copyright IBM Corporation 2016