VMware Validated Design for Cloud Providers: Scale and Performance Guidelines

for vCloud Director 10 environments

Published December 2019
Contents

Introduction .................................................................................................................. 5

1.1 Audience ............................................................................................................ 5

1.2 Scope .................................................................................................................. 5

Interoperability Stack (Bill of Materials) ................................................................. 7

Scale and Performance ............................................................................................... 8

2.1 Performance Characteristics ............................................................................ 9

Sizing Guidelines ...................................................................................................... 11

3.1 vCloud Director Virtual Appliance ................................................................. 11

3.2 Management and Resource Component Sizing ............................................... 12

Appendix A – Test Environment and Benchmarking Methods ............................ 16

Test Environment .................................................................................................... 16
Management Cluster ............................................................................................... 16
Resource Cluster ..................................................................................................... 18
PostgreSQL Tuning ................................................................................................. 21
Test Driver ............................................................................................................... 21
Benchmarking Methods ......................................................................................... 21
Scale Test ................................................................................................................ 21
Performance Test .................................................................................................... 21
Uptime Tests .......................................................................................................... 22
List of Operations ................................................................................................... 22

Appendix B – FAQ ................................................................................................... 24
# List of Tables

Table 1. Bill of Materials ........................................................................................................... 7  
Table 2. Scale Profile B ............................................................................................................. 8  
Table 3. Performance and Throughput ....................................................................................... 9  
Table 4. API Operations Latency ............................................................................................. 9  
Table 5. OVF Upload and Download Times ............................................................................... 10  
Table 6. Time to Protect a VM ................................................................................................. 10  
Table 7. Network Latency Impact on Time to Protect (VM size = 100 GB) ............................ 11  
Table 8. Management and Resource Component Sizing ............................................................ 12  
Table 9. vCloud Director Operations (Part 1) .......................................................................... 22  
Table 10. vCloud Director Operations (Part 2) ....................................................................... 23

# List of Figures

Figure 1. Management Component Deployment ......................................................................... 17  
Figure 2. Management Cluster Networking ............................................................................. 18  
Figure 3. Resource Cluster Setup ............................................................................................. 19  
Figure 4. Resource Cluster Networking ..................................................................................... 20  
Figure 5. Component Architecture Overview ........................................................................... 24  
Figure 6. Management Component Configuration ................................................................... 25  
Figure 7. Resource Cluster Configuration ................................................................................. 26  
Figure 8. Network Configuration .............................................................................................. 27  
Figure 9. vCloud Director Operations ....................................................................................... 28  
Figure 10. Performance and Throughput ................................................................................... 29  
Figure 11. API Operations Latency ........................................................................................... 30  
Figure 12. OVF Upload and Download Times .......................................................................... 31  
Figure 13. Time to Protect a VM ............................................................................................... 32  
Figure 14. Network Latency Impact on Time to Protect (VM size = 100 GB) ....................... 33  
Figure 15. Management and Resource Component Sizing ....................................................... 34  
Figure 16. vCloud Director Operations (Part 1) ..................................................................... 35  
Figure 17. vCloud Director Operations (Part 2) ..................................................................... 36  
Figure 18. Component Architecture Overview ......................................................................... 37  
Figure 19. Management Component Configuration ................................................................. 38  
Figure 20. Resource Cluster Configuration .............................................................................. 39  
Figure 21. Network Configuration ............................................................................................ 40  
Figure 22. vCloud Director Operations ..................................................................................... 41  
Figure 23. Performance and Throughput .................................................................................. 42  
Figure 24. API Operations Latency .......................................................................................... 43  
Figure 25. OVF Upload and Download Times ......................................................................... 44  
Figure 26. Time to Protect a VM .............................................................................................. 45  
Figure 27. Network Latency Impact on Time to Protect (VM size = 100 GB) ..................... 46  
Figure 28. Management and Resource Component Sizing ...................................................... 47  
Figure 29. vCloud Director Operations (Part 1) .................................................................... 48  
Figure 30. vCloud Director Operations (Part 2) .................................................................... 49  
Figure 31. Component Architecture Overview ....................................................................... 50  
Figure 32. Management Component Configuration ............................................................... 51  
Figure 33. Resource Cluster Configuration .............................................................................. 52  
Figure 34. Network Configuration ........................................................................................... 53  
Figure 35. vCloud Director Operations .................................................................................... 54  
Figure 36. Performance and Throughput .................................................................................. 55  
Figure 37. API Operations Latency .......................................................................................... 56  
Figure 38. OVF Upload and Download Times ....................................................................... 57  
Figure 39. Time to Protect a VM ............................................................................................. 58  
Figure 40. Network Latency Impact on Time to Protect (VM size = 100 GB) ................. 59  
Figure 41. Management and Resource Component Sizing .................................................... 60  
Figure 42. vCloud Director Operations (Part 1) ................................................................... 61  
Figure 43. vCloud Director Operations (Part 2) ................................................................... 62  
Figure 44. Component Architecture Overview ...................................................................... 63  
Figure 45. Management Component Configuration .............................................................. 64  
Figure 46. Resource Cluster Configuration ............................................................................. 65  
Figure 47. Network Configuration .......................................................................................... 66  
Figure 48. vCloud Director Operations .................................................................................... 67  
Figure 49. Performance and Throughput .................................................................................. 68  
Figure 50. API Operations Latency .......................................................................................... 69  
Figure 51. OVF Upload and Download Times ....................................................................... 70  
Figure 52. Time to Protect a VM ............................................................................................ 71  
Figure 53. Network Latency Impact on Time to Protect (VM size = 100 GB) .............. 72  
Figure 54. Management and Resource Component Sizing .................................................... 73  
Figure 55. vCloud Director Operations (Part 1) ................................................................... 74  
Figure 56. vCloud Director Operations (Part 2) ................................................................... 75  
Figure 57. Component Architecture Overview ..................................................................... 76  
Figure 58. Management Component Configuration .............................................................. 77  
Figure 59. Resource Cluster Configuration ............................................................................. 78  
Figure 60. Network Configuration .......................................................................................... 79  
Figure 61. vCloud Director Operations .................................................................................... 80  
Figure 62. Performance and Throughput .................................................................................. 81  
Figure 63. API Operations Latency .......................................................................................... 82  
Figure 64. OVF Upload and Download Times ....................................................................... 83  
Figure 65. Time to Protect a VM ............................................................................................ 84  
Figure 66. Network Latency Impact on Time to Protect (VM size = 100 GB) ........ 85  
Figure 67. Management and Resource Component Sizing .................................................... 86  
Figure 68. vCloud Director Operations (Part 1) ................................................................... 87  
Figure 69. vCloud Director Operations (Part 2) ................................................................... 88  
Figure 70. Component Architecture Overview ...................................................................... 89  
Figure 71. Management Component Configuration .............................................................. 90  
Figure 72. Resource Cluster Configuration ............................................................................. 91  
Figure 73. Network Configuration .......................................................................................... 92  
Figure 74. vCloud Director Operations .................................................................................... 93  
Figure 75. Performance and Throughput .................................................................................. 94  
Figure 76. API Operations Latency .......................................................................................... 95  
Figure 77. OVF Upload and Download Times ....................................................................... 96  
Figure 78. Time to Protect a VM ............................................................................................ 97  
Figure 79. Network Latency Impact on Time to Protect (VM size = 100 GB) ........... 98  
Figure 80. Management and Resource Component Sizing .................................................... 99  
Figure 81. vCloud Director Operations (Part 1) ................................................................100 
Figure 82. vCloud Director Operations (Part 2) ................................................................101 
Figure 83. Component Architecture Overview ................................................................102
Introduction

The VMware Validated Designs (VVD) for Cloud Providers: Scale and Performance Guidelines (Scale and Performance) is an evolution of Certified Reference Design for VMware Cloud Providers™. It is a pre-validated set of software components that simplify the deployment of a VMware vCloud Director®-based multitenant cloud in a predictable and efficient manner. The intent of the Scale and Performance initiative is to document a verified stack and provide scale and performance benchmarking. It also helps reduce the complexity of figuring out dependencies between the VMware components required for a vCloud Director-based service. While this initiative does not yet involve software automation for software upgrades, it aims to present clearly what components are needed, which versions must be used, and what kind of scale and performance VMware Cloud Providers can expect.

VMware Cloud Providers get clarity and predictability about which version of each software component of the stack is recommended at a given time. Each Scale and Performance version also includes a predictable support timeframe for all underlying components, typically 12 – 18 months from the launch of the corresponding Scale and Performance release. This reduces the expense and time involved in determining what components to upgrade when and to which version, so that the entire software stack stays in support and incompatible combinations are avoided.

VMware Cloud Providers also benefit from clear guidelines for sizing hardware and software components to match their expected tenant load. While the Scale and Performance does not cover every cloud configuration and size, it provides a sizing recommendation for a “typical” cloud (a cloud size representative of a broad set of VMware Cloud Providers). Future versions of the Scale and Performance may address larger and less common environment configurations as well as more specialized use cases.

It is not the current intent of Scale and Performance to push vCloud Director to its absolute limits. For configuration maximums and limits, see VMware vCloud Director Configuration Maximums.

This document also includes the expected performance as observed by tenant users and VMware Cloud Provider administrators interacting with the vCloud Director user interface and API.

A vCloud Director-based platform can be properly sized by following the sizing guidelines for hardware and scale based on anticipated tenant demand.

1.1 Audience

This document is intended for VMware Cloud Provider architects and technical leads responsible for planning and executing the deployment and upgrades of a VMware-based cloud environment.

1.2 Scope

This document addresses the following aspects:

- Interoperability stack
  
  Provides a list of certified versions of all the component software comprising the software stack. Using the recommended versions guarantees known support life of the stack as well as performance characteristics.

- Sizing guidelines and software requirements

- Performance characteristics of the solution

The certified solution stack provides known performance and scale characteristics and includes recommendations and guidelines for hardware and scale based on anticipated tenant demand.
See the complimentary documents that are part of the *VMware vCloud® Architecture Toolkit™ for Service Providers*:

- *Architecting a VMware vCloud Director Solution for VMware Cloud Providers*
- *Architecting Tenant Networking with VMware NSX® in VMware vCloud Director*
- *Developing a Hyper-Converged Storage Strategy for VMware vCloud Director with VMware vSAN™*
- *VMware vCloud Director Configuration Maximums*

The [VMware Product Interoperability Matrices](#) is the authoritative resource for interoperability between the VMware software components.

A compliant solution must comply with all relevant security guidelines outlined in the product-specific documentation as well as security recommendations in the *VMware vCloud Architecture Toolkit for Service Providers* document.
Interoperability Stack (Bill of Materials)

The Bill of Materials table lists the pre-validated set of software components for Cloud Providers at the time of the Scale and Performance launch. While VMware Cloud Providers are free to choose and pick other versions or different combinations of VMware Cloud Provider Program software products, the specified stack guarantees a known predictable support time and specific performance and scaling characteristics. Performance and scaling information is provided later in this document. Products marked “Core” are required to officially achieve VVD for Cloud Providers compliance.

Table 1. Bill of Materials

<table>
<thead>
<tr>
<th>Component</th>
<th>Version and Build</th>
<th>Core/ Optional</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMware vCenter Server®</td>
<td>6.7 Update 3</td>
<td>Core</td>
<td>See Table 8 for patch level tested.</td>
</tr>
<tr>
<td>VMware ESXi™</td>
<td>6.7 Update 3</td>
<td>Core</td>
<td>See Table 8 for patch level tested.</td>
</tr>
<tr>
<td>VMware NSX-V</td>
<td>6.4.6</td>
<td>Core</td>
<td>See Table 8 for patch level tested.</td>
</tr>
<tr>
<td>VMware vSAN</td>
<td>6.7 Update 3</td>
<td>Core</td>
<td>See Note 2</td>
</tr>
<tr>
<td>vCloud Director</td>
<td>10</td>
<td>Core</td>
<td>10.0.0.1 Virtual Appliance with an embedded database.</td>
</tr>
<tr>
<td>vCloud Availability</td>
<td>3.5</td>
<td>Optional</td>
<td></td>
</tr>
<tr>
<td>Container Service</td>
<td>2.5.1</td>
<td>Optional</td>
<td>Container Services Extension</td>
</tr>
<tr>
<td>VMware vRealize® Log Insight™</td>
<td>8.0</td>
<td>Optional</td>
<td></td>
</tr>
<tr>
<td>VMware vRealize® Network Insight™</td>
<td>5.0</td>
<td>Optional</td>
<td>Apply the latest available patch.</td>
</tr>
<tr>
<td>VMware vRealize® Orchestrator™</td>
<td>7.6</td>
<td>Optional</td>
<td>Apply the latest available patch.</td>
</tr>
<tr>
<td>VMware vCloud Usage Meter</td>
<td>3.6.1 Hot Patch 3</td>
<td>Core</td>
<td></td>
</tr>
<tr>
<td>VMware vRealize Operations™</td>
<td>8.0</td>
<td>Optional</td>
<td>Apply the latest available patch.</td>
</tr>
</tbody>
</table>

Note 1. These are the recommended set of products, but this is not a full interoperability matrix. For example, vCloud Director 10 is supported with multiple versions of NSX but in the current benchmarking we used a specific NSX version. Test results generally apply to all patches within
the specified major version of each component. See the VMware Product Interoperability Matrix for full vCloud Director interoperability information.

2. vSAN-based storage must be deployed in at least one cluster (either management or capacity).

**Scale and Performance**

The Scale Profile B table represents a common environment similar to the environments of approximately 60% of all VMware Cloud Providers. While vCloud Director is capable of a larger scale, the following profile is what is validated and benchmarked in the current Scale and Performance.

**Table 2. Scale Profile B**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tenants (Organizations in vCloud Director)</td>
<td>400</td>
</tr>
<tr>
<td>Number of powered-on tenant virtual machines (with an OS installed)</td>
<td>10,000</td>
</tr>
<tr>
<td>Number of data centers</td>
<td>1</td>
</tr>
<tr>
<td>Number of vCloud Director cells</td>
<td>4</td>
</tr>
<tr>
<td>Number of vCenter Server instances managed by vCloud Director</td>
<td>1 vCenter Server for management cluster</td>
</tr>
<tr>
<td></td>
<td>3 vCenter Server instances for resource capacity</td>
</tr>
<tr>
<td>Number of hosts and clusters</td>
<td>3 resource capacity clusters, 33 hosts total</td>
</tr>
<tr>
<td></td>
<td>1 management cluster; 5 hosts</td>
</tr>
<tr>
<td>Maximum network latency from vCloud Director to VMware vCenter Server, VMware NSX Manager™, and ESXi hosts</td>
<td>Network RTT latency up to 150 ms</td>
</tr>
<tr>
<td>Concurrent API operations</td>
<td>Up to 128 concurrent users executing operations against the vCloud Director API</td>
</tr>
<tr>
<td>Concurrent virtual machine migrations to vCloud Director from tenant environments by vCloud Availability</td>
<td>100</td>
</tr>
</tbody>
</table>
2.1 Performance Characteristics

2.2.1 Environment Setup

The multitenant cloud environment is set up based on Scale Profile B. Testing is performed at different levels of network latency from vCloud Director cells to vCenter Server and NSX Manager to measure the impact of network latency on performance.

2.2.2 Performance and Throughput

The test throughput is measured as the number of operations executed over 30 minutes. The test was run with different test concurrency (32, 64, and 128) and network latency (0.3 ms, 40 ms, and 150 ms). During this test, a representative random sample of operations from the List of Operations is used.

Table 3. Performance and Throughput

<table>
<thead>
<tr>
<th>Concurrency (Number of concurrent users)</th>
<th>Throughput at RTT = 0.3 ms (Successfully completed operations per minute)</th>
<th>Throughput at RTT = 40 ms</th>
<th>Throughput at RTT = 150 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>124</td>
<td>101</td>
<td>72</td>
</tr>
<tr>
<td>64</td>
<td>213</td>
<td>170</td>
<td>137</td>
</tr>
<tr>
<td>128</td>
<td>319</td>
<td>289</td>
<td>229</td>
</tr>
</tbody>
</table>

2.2.3 API Latency

The API Operations Latency table shows average user observed latency (in seconds) for a selection of API operations at RTT = 0.3 ms. See the

Results

Total Operations completed: 1341318
Average test throughput over 5 days: 186 Ops/min
Total failures over 5 days: 59
Failure rate over 5 days: 0.004%
Success rate: 99.996%

List of Operations for the full list of operations invoked during this test.

Table 4. API Operations Latency
2.2.4 Upload and Download Performance

The OVF upload and download times observed in the test environment vary depending on the different network latencies.

### Table 5. OVF Upload and Download Times

<table>
<thead>
<tr>
<th></th>
<th>RTT = 0.3 ms</th>
<th>RTT = 40 ms</th>
<th>RTT = 150 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>OVF upload time in</td>
<td>359</td>
<td>388</td>
<td>392</td>
</tr>
<tr>
<td>seconds (4 GB)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVF download time</td>
<td>232</td>
<td>235</td>
<td>245</td>
</tr>
<tr>
<td>in seconds (4 GB)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2.5 vCloud Availability

The Time to Protect a virtual machine in vCloud Availability represents the time to establish replication of virtual machines of various sizes between a vCloud Director and a disaster recovery target vCenter Server environment using vCloud Availability. 10 GB uplinks were configured between vCenter Server and vCloud Director. Network throughput was stable around 710 Mbps.

### Table 6. Time to Protect a VM

<table>
<thead>
<tr>
<th>VM size</th>
<th>Time to Protect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 GB</td>
<td>29 sec</td>
</tr>
<tr>
<td>10 GB</td>
<td>2 min</td>
</tr>
<tr>
<td>100 GB</td>
<td>27 min</td>
</tr>
<tr>
<td>500 GB</td>
<td>1 hr: 58 min</td>
</tr>
<tr>
<td>1 TB</td>
<td>3 hr: 55 min</td>
</tr>
</tbody>
</table>
The Network Latency Impact on Migration Performance table displays how the network latency between vCenter Server and vCloud Director impacts cold migration for virtual machines of 100 GB size.

Table 7. Network Latency Impact on Time to Protect (VM size = 100 GB)

<table>
<thead>
<tr>
<th>RTT Latency between vCenter Server and vCloud Director</th>
<th>Time to Protect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3 ms</td>
<td>27 min</td>
</tr>
<tr>
<td>40 ms</td>
<td>29 min</td>
</tr>
<tr>
<td>150 ms</td>
<td>33 min</td>
</tr>
</tbody>
</table>

**Sizing Guidelines**

Many environment variables influence the number of hosts and CPU and memory resources required to run a cloud service based on VVD for Cloud Providers. It is impossible to give a precise formula for how much of each resource is required. The current effort focuses on demonstrating how our deployment setup behaves in terms of scale and performance under the defined test load. The Scale Profile B captures the parameters of the load on the system in terms of number of tenants, organizations, VMs, network latency, and cloud management operations load. The results in terms of average response time, throughput, and uptime under these controlled conditions provide a starting point for Cloud Providers to estimate how much capacity is needed for their use cases. We strongly recommend that Cloud Providers extensively test each environment prior to production use to ensure that the performance meets the business SLAs.

**3.1 vCloud Director Virtual Appliance**

The current VVD for Cloud Providers recommends the use of a vCloud Director 10 virtual appliance. The virtual appliance includes an embedded and fully managed PostgreSQL database and built-in replication for maintaining consistency between cells. A properly configured virtual appliance cluster is also resilient to failures of individual cells.

The recommended vCloud Director virtual appliance deployment includes 1 primary cell, 2 stand-by cells and 0 or more application cells. See the vCloud Director Installation, Configuration, and Upgrade guide for more details about the deployment, configuration, and operation of the virtual appliance.

For the purposes of VVD scale and performance benchmarking, the following setup was used: 1 primary cell, 2 standby cells, 1 application cell. The Table 8 Management and Resource Component Sizing gives the details of memory and CPU capacity on each vCloud Director appliance node.

The embedded PostgreSQL database is automatically configured. However, some post-deployment tuning might be needed for best performance. See the section PostgreSQL Tuning for the database parameters used in this benchmarking.
3.2 Management and Resource Component Sizing

The following table summarizes sizing choices made for various management and resource components.

Table 8. Management and Resource Component Sizing

<table>
<thead>
<tr>
<th>Component</th>
<th>Version</th>
<th>Size</th>
<th>Resources</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management vCenter Server</td>
<td>6.7 Update 3</td>
<td>Tiny</td>
<td>RAM: 10 GB</td>
<td>1 management vCenter Server</td>
</tr>
<tr>
<td>(vCenter Server Appliance with an</td>
<td></td>
<td></td>
<td>CPU: 2</td>
<td></td>
</tr>
<tr>
<td>embedded Platform Services</td>
<td></td>
<td></td>
<td>Storage: 250 GB</td>
<td></td>
</tr>
<tr>
<td>Controller)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource vCenter Server</td>
<td>6.7 Update 3</td>
<td>Medium</td>
<td>RAM: 24 GB</td>
<td>3 resource vCenter Server instances</td>
</tr>
<tr>
<td>(vCenter Server Appliance with an</td>
<td></td>
<td></td>
<td>CPU: 8</td>
<td></td>
</tr>
<tr>
<td>embedded Platform Services</td>
<td></td>
<td></td>
<td>Storage: 400 GB</td>
<td></td>
</tr>
<tr>
<td>Controller)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESXi</td>
<td>6.7 Update 3</td>
<td></td>
<td>Cisco UCSC-C240-M5SX servers</td>
<td>33 hosts for resource cluster, 5 hosts for management cluster</td>
</tr>
<tr>
<td>vSAN (deployed in management</td>
<td>6.7 Update 3</td>
<td></td>
<td>RAM: 16 GB</td>
<td></td>
</tr>
<tr>
<td>cluster)</td>
<td></td>
<td></td>
<td>CPU: 4</td>
<td></td>
</tr>
<tr>
<td>NSX-V for vSphere</td>
<td>6.4.6</td>
<td></td>
<td>Storage: 60 GB</td>
<td></td>
</tr>
<tr>
<td>Component</td>
<td>Version</td>
<td>Size</td>
<td>Resources</td>
<td>Notes</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>vCloud Director</td>
<td>10.0.0.1</td>
<td>Virtual</td>
<td><strong>1 Primary, 2 Standby cells</strong>&lt;br&gt;RAM: 32 GB&lt;br&gt;CPU: 24&lt;br&gt;Storage: 170 GB + 500 GB NFS&lt;br&gt;<strong>1 application cell</strong>&lt;br&gt;RAM: 8 GB&lt;br&gt;CPU: 8&lt;br&gt;Storage: 170 GB</td>
<td>170 GB = 120 GB for database and 50 GB for logs per cell and 500 GB of shared NFS storage for the vCloud Director transfer service.</td>
</tr>
<tr>
<td>vCloud Availability</td>
<td>3.5</td>
<td></td>
<td>RAM: 4 GB&lt;br&gt;CPU: 2&lt;br&gt;Storage: 10 GB</td>
<td></td>
</tr>
<tr>
<td>vCloud Availability C4 Appliance</td>
<td></td>
<td></td>
<td>RAM: 6 GB&lt;br&gt;CPU: 4&lt;br&gt;Storage: 10 GB</td>
<td></td>
</tr>
<tr>
<td>vCloud Availability Replicator</td>
<td></td>
<td></td>
<td>RAM: 2 GB&lt;br&gt;CPU: 2&lt;br&gt;Storage: 10 GB</td>
<td></td>
</tr>
<tr>
<td>vCloud Availability Tunnel Appliance</td>
<td></td>
<td></td>
<td>RAM: 2 GB&lt;br&gt;CPU: 2&lt;br&gt;Storage: 10 GB</td>
<td></td>
</tr>
<tr>
<td>vCloud Director Database</td>
<td>PostgreSQL 10.5</td>
<td></td>
<td>RAM: 32 GB&lt;br&gt;CPU: 16&lt;br&gt;Storage: 300 GB</td>
<td>Database is embedded with the vCloud Director virtual appliance</td>
</tr>
<tr>
<td>Container Service Extension</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vCloud Director AMQP</td>
<td>RabbitMQ 3.7.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vCloud Director Metrics Database</td>
<td>Cassandra 3.11.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component</td>
<td>Version</td>
<td>Size</td>
<td>Resources</td>
<td>Notes</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>-------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>vRealize Log Insight deployment</td>
<td>8.0</td>
<td>Medium</td>
<td>RAM: 16 GB, CPU: 8, Storage: 500 GB</td>
<td>Use the vRealize Log Insight sizing calculator: <a href="http://www.vmware.com/go/loginsight/calculator">http://www.vmware.com/go/loginsight/calculator</a></td>
</tr>
<tr>
<td>vRealize Network Insight deployment</td>
<td>5.0</td>
<td>Large</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vRealize Orchestrator</td>
<td>7.6</td>
<td></td>
<td>RAM: 6 GB, CPU: 2, Storage: 20 GB</td>
<td>vRealize Orchestrator 7.6 was used as opposed to 8.0 (in the BOM) because of an issue affecting vRealize Orchestrator 8.0.</td>
</tr>
<tr>
<td>vRealize Orchestrator plug-in for vCloud Director</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vRealize Operations Manager</td>
<td>8.0</td>
<td>Medium</td>
<td>RAM: 32 GB x 3, CPU: 8 x 3, Storage: 300 GB x 3</td>
<td>1 Master, 1 Master Replica, 1 Data node Use the vRealize Operations sizing guidelines: <a href="https://kb.vmware.com/s/article/75162">https://kb.vmware.com/s/article/75162</a></td>
</tr>
</tbody>
</table>
## Component List

<table>
<thead>
<tr>
<th>Component</th>
<th>Version</th>
<th>Size</th>
<th>Resources</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Pack for NSX for vSphere</td>
<td>3.5.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Pack for vSphere</td>
<td>8.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Pack for vRealize Log Insight</td>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Pack for vCloud Director</td>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Pack for vSAN</td>
<td>8.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vRealize Operations Tenant App for vCloud Director</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix A – Test Environment and Benchmarking Methods

Test Environment

The test environment is broadly divided into three main setups:

- Management cluster
- Resource cluster (30% of workloads on vSAN, 70% on iSCSI storage)
- Test driver

Management Cluster

This is where all the management components were deployed.

- Management components
  - 1 x Management vCenter Server (Tiny)
  - 4 x vCloud Director virtual appliance cells (1 primary, 2 stand-by, 1 application)
  - 3 x Resource vCenter Server (Medium)
  - 3 x NSX Manager
  - 1 x Management NSX vCloud Director edge
  - 1 x vRealize Log Insight (Medium)
  - 1 x vRealize Operations (Large)
  - 1 x vCloud Usage Meter (Standard)

- Management Cluster Resources
  - 5 physical servers with 192 GB RAM and 28 cores, each with vSAN supported SSDs
  - 10 TB vSAN, 10 TB iSCSI
Figure 1. Management Component Deployment
Resource Cluster

This is where Tenant Organizations and workload virtual machines were created.

- Resource cluster resources
  - 33 physical servers with 192 GB RAM and 28 cores, each with vSAN supported SSDs
  - 30 TB vSAN, 10 TB iSCSI
Figure 3. Resource Cluster Setup
Figure 4. Resource Cluster Networking
PostgreSQL Tuning
PostgreSQL database parameters were set as follows:

shared_buffers = 8GB
effective_cache_size = 24GB
work_mem = 8MB
maintenance_work_mem = 512MB
max_parallel_workers_per_gather = 0
max_worker_processes = 24

See [How to Modify PostgreSQL Configuration](#) from the vCloud Director documentation.

Test Driver
The test driver suite is executed from this environment.

- 4 CPU, 8 GB memory, CentOS 7.3

Benchmarking Methods
The testing process is focused primarily on verifying and measuring environment behavior for:

- Scale – Verify whether the environment meets the Scale Profile B requirement of 10,000 powered-on virtual machines.
- Performance – Measure operation latency and throughput when the environment is running at scale (10,000 powered-on virtual machines).
- Uptime – Verify that the environment can operate at scale with reasonable performance for a long time.

The remainder of this section details the exact methods used for test execution and measurement.

Scale Test
Scale was carried out with a mix of manual operations and JMeter test tool-based script operations by using the following steps:

1. Create 400 Tenant Organizations in vCloud Director.
2. Create and power on 10,000 virtual machines across these 400 Tenant Organizations. All virtual machines were running Ubuntu OS with 2 GB disk, 1 GB memory
3. A sample of vCloud Director operations were carried out to verify that system behaves normally at this scale.

Performance Test
Performance tests were done by executing a well-known distribution of vCloud Director operations with the help of an internal test tool. For the complete operation list, see [List of Operations](#).

The following were the key steps in execution and measurement of the operations:

1. Scaled up the environment as outlined in the previous section.
2. After the environment was at scale, executed continuous stream of operations for 30 minutes with following distribution:
   - 35-40% vApp operations such as instantiate, deploy, edit, clone, and delete.
- 25% storage-centric operations such as create, attach, detach, and delete disk.
- 15% networking-related operations, such as create and delete gateway, routed networks and firewall configurations.
- 5% create and delete Orgs, users, catalogs, and virtual data centers.

3. Operations were executed using vCloud Director local users of different roles (vApp Author, Org Admin, System Admin) with 10% admin roles and 90% user operation roles.

4. Given that most of the operations are asynchronous, the test tool monitors the task returned by vCloud Director to get completion status and execution time details.

5. Steps 2 to 4 were repeated with 32, 64, and 128 concurrent users to ascertain the ability of the system to deal with concurrent operation invocation.

6. Step 5 was repeated for following latency (between vCloud Director and vCenter Server) values (achieved by artificial latency injection with a tool):
   - 0.3 ms (default)
   - 40 ms
   - 150 ms

Uptime Tests

Uptime tests involved running the environment (based on Scale Profile B) for 5 days and executing a constant stream of API calls covering a representative set of operations. The purpose of the test is to establish the API call success rate and system uptime.

1. Tests ran continuously for 5 days.

2. API workflows were triggered by 100 concurrent clients, each client would invoke an operation roughly every 20 seconds. 10,000 powered on VMs

3. No artificial latency injection was done.

Results

Total Operations completed: **1341318**

Average test throughput over 5 days: **186 Ops/min**

Total failures over 5 days: **59**

Failure rate over 5 days: **0.004%**

Success rate: **99.996%**

List of Operations

For performance benchmarking, API test clients executed a predetermined distribution across different types of vCloud Director operations as described in the following tables.

**Table 9. vCloud Director Operations (Part 1)**

<table>
<thead>
<tr>
<th>vApp Operations</th>
<th>Network Operations</th>
<th>Management Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instantiate vApp</td>
<td>Deploy a fenced vApp</td>
<td>Create an org</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Deploy (power on)</td>
<td>Undeploy a fenced vApp</td>
<td>Create a user</td>
</tr>
<tr>
<td>Edit vApp</td>
<td>Create an isolated network</td>
<td>Create an Org VDC</td>
</tr>
<tr>
<td>Compose vApp</td>
<td>Delete an isolated network</td>
<td>Create a direct VDC network</td>
</tr>
<tr>
<td>Clone vApp</td>
<td>Create a gateway</td>
<td>Create a catalog</td>
</tr>
<tr>
<td>Power off vApp</td>
<td>Create a routed Org network</td>
<td>Delete a catalog</td>
</tr>
<tr>
<td>Delete vApp</td>
<td>Instantiate a vApp in that network</td>
<td>Delete a VDC network</td>
</tr>
<tr>
<td></td>
<td>Deploy a vApp</td>
<td>Delete an Org VDC</td>
</tr>
<tr>
<td></td>
<td>Undeploy a vApp</td>
<td>Delete a user</td>
</tr>
<tr>
<td></td>
<td>Delete a vApp</td>
<td>Delete an Org</td>
</tr>
<tr>
<td></td>
<td>Delete a routed Org network</td>
<td>Delete a gateway</td>
</tr>
</tbody>
</table>

Table 10. vCloud Director Operations (Part 2)

<table>
<thead>
<tr>
<th>NSX Management Operations</th>
<th>Datastore Operations</th>
<th>OVF Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convert edge to Advanced edge</td>
<td>Create a disk</td>
<td>OVF upload</td>
</tr>
<tr>
<td>Edge routing services</td>
<td>Instantiate a vApp</td>
<td>OVF download</td>
</tr>
<tr>
<td>Edge firewall services</td>
<td>Attach a disk to a vApp</td>
<td></td>
</tr>
<tr>
<td>Edge NAT services</td>
<td>Detach a disk from a vApp</td>
<td></td>
</tr>
<tr>
<td>Distributed firewall services</td>
<td>Delete a disk</td>
<td></td>
</tr>
<tr>
<td>Load balancer services</td>
<td>Delete a vApp</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B – FAQ

How frequently will the Scale and Performance be updated?
➢ We expect to release an updated Scale and Performance with every major vCloud Director release.

How is this document related to the VMware interoperability matrix?
➢ The benchmarked stack is a subset of the full interoperability matrix and reflects the exact components we validated and benchmarked in this exercise. The full interoperability includes many more products and versions than what is tested in this exercise.

How is the Scale and Performance related to VMware Cloud Foundation?
➢ VMware Cloud Foundation is not currently part of the benchmarking, however the stack we are testing is closely aligned with the current VCF BOM.

Is Scale and Performance suitable for greenfield environments or brownfield environments?
➢ Any environment can be made compliant by simply upgrading all its components to versions listed in the Scale and Performance Bill of Materials. There is no other qualification.

How can we provide input/recommendations for future versions of this doc?
➢ Contact the vCloud Director team at vcd-feedback@vmware.com or reach out to your VMware account team and pass your feedback through them.

What is the support model for an environment configured according to these guidelines?
➢ Each component of the Cloud Provider Platform stack is supported according to its support lifecycle. A cloud deployment compliant with the Bill of Materials is in support for at least 12 months after the Scale and Performance release date.