Validation Guide

Abstract
This validation guide describes the architecture and performance of the integration of VMware Horizon components for virtual desktop infrastructure (VDI) on Dell EMC vSAN Ready Nodes and Dell EMC VxRail.
CONTENTS

Chapter 1  Executive Summary  
Document purpose........................................................................................................ 6  
Audience..................................................................................................................... 6  
We value your feedback......................................................................................... 6

Chapter 2  Test Environment Configuration and Best Practices  
Validated hardware resources.................................................................................. 8  
Validated software resources............................................................................... 9  
Validated system version matrix.......................................................................... 9  
Virtual networking configuration.......................................................................... 9  
Management server infrastructure....................................................................... 10  
  NVIDIA GRID License Server........................................................................ 10  
  SQL Server databases..................................................................................... 10  
  DNS............................................................................................................. 11  
High availability.................................................................................................. 11  
VMware Horizon architecture............................................................................ 12

Chapter 3  Solution Performance and Testing  
Testing process........................................................................................................ 14  
  Resource monitoring...................................................................................... 14  
  Load generation......................................................................................... 15  
  Profiles and workloads............................................................................. 15  
  Desktop VM test configurations............................................................ 18  
Test results and analysis.................................................................................... 18  
  Density Optimized configuration........................................................... 20

Chapter 4  Conclusion  
Test results and density recommendations......................................................... 36  
Conclusions........................................................................................................... 36

Chapter 5  References  
Dell EMC documentation...................................................................................... 38  
VMware documentation....................................................................................... 38  
NVIDIA documentation....................................................................................... 38
CHAPTER 1

Executive Summary

This chapter presents the following topics:

- Document purpose ................................................................. 6
- Audience ........................................................................... 6
- We value your feedback .................................................. 6
Executive Summary

Document purpose

This validation guide details the architecture, components, testing methods, and test results for Dell EMC VxRail appliances and vSAN Ready Nodes with VMware Horizon 7. It includes the test environment configuration and best practices for systems that have undergone testing.

Audience

This guide is intended for architects, developers, and technical administrators of IT environments. It provides an in-depth explanation of the testing methodology and basis for VDI densities. It also validates the value of the Dell EMC Ready Architectures for VDI that deliver Microsoft Windows virtual desktops to users of VMware Horizon 7 VDI components on VxRail appliances or vSAN Ready Nodes.

We value your feedback

Dell EMC and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell EMC Solutions team by email or provide your comments by completing our documentation survey.

Authors: Dell EMC Ready Architectures for VDI Team.
CHAPTER 2

Test Environment Configuration and Best Practices

This chapter presents the following topics:

- Validated hardware resources ................................................................. 8
- Validated software resources ................................................................. 9
- Validated system version matrix ............................................................ 9
- Virtual networking configuration ......................................................... 9
- Management server infrastructure ....................................................... 10
- High availability ................................................................................. 11
- VMware Horizon architecture ......................................................... 12
Validated hardware resources

Dell EMC validated the solutions with the specific hardware resources listed in this section.

**Enterprise platforms**

We used the VxRail appliances and vSAN Ready Nodes hardware components listed in the following table.

**Table 1 Validated hardware configurations**

<table>
<thead>
<tr>
<th>Config</th>
<th>Enterprise platform</th>
<th>CPU</th>
<th>Memory</th>
<th>RAID ctrl</th>
<th>BOSS</th>
<th>HD config</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density-optimized V570F R740XD</td>
<td>Intel Xeon Gold 6248 (20 core 2.5 GHz)</td>
<td>768 GB @ 2,933 MT/s</td>
<td>HBA 330 adapter</td>
<td>2 x 120/240 GB M.2</td>
<td>2 x 800 GB SSD (Cache) 6 x 1.92 TB HDD (Capacity)</td>
<td>4 x Intel X710 rNDC</td>
<td></td>
</tr>
</tbody>
</table>

**Graphics hardware**

We used the NVIDIA T4 in our tests for graphics-intensive workloads. This card is a single-slot form-factor 6.6-inch PCI Express Gen 3.0 graphics card featuring a high-end NVIDIA T4 with a total of 16 GB GDDR6 memory per card.

**Network hardware**

The following network hardware was used in our test environment:

- Dell Networking S4048 (10 GbE ToR switch)—A high-density, ultra-low-latency ToR switch that features 48 x 10 Gbe SFP+ and 6 x 40 Gbe ports and up to 720 Gbps performance
- Dell Networking S5248 (25 GbE ToR switch)—A high-density, high performance, open networking ToR switch that features 48 x 25 Gbe SFP28, 4 x 100 Gbe QFSP28 ports, 2 x 100 Gbe QFSP28-DD ports and up to 2.0 Tbps switch fabric capacity
Validated software resources

Dell EMC validated this solution with the software components listed in the following table.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description/Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypervisor</td>
<td>ESXi 6.7</td>
</tr>
<tr>
<td>Broker technology</td>
<td>VMware Horizon 7 version 7.7</td>
</tr>
<tr>
<td>Broker database</td>
<td>Microsoft SQL Server 2016</td>
</tr>
<tr>
<td>Management VM OS</td>
<td>Microsoft Windows Server 2016 (Connection Server &amp; DB)</td>
</tr>
<tr>
<td>Virtual desktop OS</td>
<td>Microsoft Windows 10 Enterprise</td>
</tr>
<tr>
<td>Office application suite</td>
<td>Microsoft Office Professional 2016</td>
</tr>
<tr>
<td>Login VSI test suite</td>
<td>Version 4.1.32.1</td>
</tr>
<tr>
<td>Platform</td>
<td>VxRail v 4.7.100</td>
</tr>
<tr>
<td>NVIDIA GRID software (for graphics testing)</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Validated system version matrix

Dell EMC validated this solution using the system versions listed in the following table.

<table>
<thead>
<tr>
<th>Server configuration</th>
<th>Nvidia vGPU version</th>
<th>Hypervisor</th>
<th>Hypervisor version</th>
<th>Hypervisor build</th>
<th>Bios</th>
<th>VxRail version</th>
<th>Windows 10 version</th>
<th>Windows 10 patches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density-optimized</td>
<td>n/a</td>
<td>ESXi 6.7</td>
<td>11675023</td>
<td>2.1.3</td>
<td>4.7.100(^a)</td>
<td>1803 - 17134.523</td>
<td>KB4480966</td>
<td>KB4100347</td>
</tr>
<tr>
<td>Density-optimized + 6 x T4</td>
<td>7.1 (410.92)</td>
<td>6.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Although testing was conducted using version 4.7.100, we recommend using 4.7.110 or later. Using the latest version available does not affect user density.

Virtual networking configuration

The network configuration for the uses a 25 GbE converged infrastructure model. All required VLANs traverse two 25 GbE NICs configured in an active/active team. We used the following VLAN configurations for the compute and management hosts in our validation testing.

- **VLAN configuration:**
- Management VLAN: Configured for hypervisor infrastructure traffic—L3 routed via core switch
- VDI VLAN: Configured for VDI session traffic—L3 routed via core switch
- VMware vSAN VLAN: Configured for VMware vSAN traffic—L2 switched only via ToR switch
- vMotion VLAN: Configured for Live Migration traffic—L2 switched only, trunked from Core (HA only)
- VDI Management VLAN: Configured for VDI infrastructure traffic—L3 routed via core switch
- A VLAN iDRAC is configured for all hardware management traffic—L3 routed via core switch

Management server infrastructure

The management server component sizing recommendations are listed in the following table.

**Table 3** Sizing for VxRail appliance, RDSH, and NVIDIA GRID license server (optional)

<table>
<thead>
<tr>
<th>Component</th>
<th>vCPU</th>
<th>RAM (GB)</th>
<th>NIC</th>
<th>OS + data vDisk (GB)</th>
<th>Tier 2 volume (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMware vCenter Appliance</td>
<td>2</td>
<td>16</td>
<td>1</td>
<td>290</td>
<td></td>
</tr>
<tr>
<td>Platform Services Controller</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Horizon Connection Server</td>
<td>4</td>
<td>16</td>
<td>1</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>SQL Server</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>40</td>
<td>210 (VMDK)</td>
</tr>
<tr>
<td>File server</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>40</td>
<td>2048 (VMDK)</td>
</tr>
<tr>
<td>VxRail Appliance Manager</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Log Insight</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>530</td>
<td></td>
</tr>
<tr>
<td>RDSH VM</td>
<td>8</td>
<td>32</td>
<td>1</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>NVIDIA GRID License Server</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>40 + 5</td>
<td></td>
</tr>
</tbody>
</table>

**NVIDIA GRID License Server**

When using NVIDIA vGPU cards, graphics-enabled VMs must obtain a license from a GRID License Server on your network to be entitled for vGPU.

We installed the GRID License Server software on a system running a Windows 2016 operating system to test vGPU configurations.

We made the following changes to the GRID License Server to address licensing requirements:
- Used a reserved fixed IP address
- Configured a single MAC address
- Applied time synchronization to all hosts on the same network

**SQL Server databases**

During validation, a single dedicated SQL Server 2016 VM hosted the VMware databases in the management layer. We separated SQL data, logs, and tempdb into their respective volumes, and created a single database for Horizon Connection Server.
We adhered to VMware best practices for this testing, including alignment of disks to be used by SQL Server with a 1,024 KB offset and formatted with a 64 KB file allocation unit size (data, logs, and tempdb).

DNS

DNS is the basis for Microsoft Active Directory and also controls access to various software components for VMware services. All hosts, VMs, and consumable software components must have a presence in DNS. We used a dynamic namespace integrated with Active Directory and adhered to Microsoft best practices.

High availability

Although we did not enable high availability (HA) during the validation that is documented in this guide, we strongly recommend that HA be factored into any VDI design and deployment. This process involves following the N+1 model with redundancy at both the hardware and software layers. The design guide for this architecture provides additional recommendations for HA.
VMware Horizon architecture

The following figure shows the Horizon communication flow.

Figure 1 VMware Horizon architecture
CHAPTER 3
Solution Performance and Testing

This chapter presents the following topics:

- Testing process .............................................................. 14
- Test results and analysis .................................................... 18
Testing process

To ensure the optimal combination of end-user experience (EUE) and cost-per-user, we conducted performance analysis and characterization testing on this solution using the Login VSI load-generation tool. Login VSI is a carefully designed, holistic methodology that monitors both hardware resource utilization parameters and EUE during load-testing.

We tested each user load against four runs: a pilot run to validate that the infrastructure was functioning and valid data could be captured, and three subsequent runs to enable data correlation.

During testing, while the environment was under load, we logged into a session and completed tasks that correspond to the user workload. While this test is subjective, it helps to provide a better understanding of the EUE in the desktop sessions, particularly under high load. It also helps to ensure reliable data gathering.

Resource monitoring

To ensure that the user experience was not compromised, we monitored the following important resources:

- **Compute host servers**—VMware vCenter (for VMware vSphere-based solutions) or Microsoft Performance Monitor (for Hyper-V-based solutions) gathers key data (CPU, memory, disk, and network usage) from each of the compute hosts during each test run. This data is exported to .csv files for single hosts, and then consolidated to show data from all hosts. While the report does not include specific performance metrics for the management host servers, these servers are monitored during testing to ensure that they are performing at an expected level with no bottlenecks.

- **Hardware resources**—Resource contention, which occurs when hardware resources have been exhausted, can cause poor EUE. We monitored the relevant resource utilization parameters and applied relatively conservative thresholds, as shown in the following table. Thresholds are carefully selected to deliver an optimal combination of good EUE and cost-per-user while also providing burst capacity for seasonal or intermittent spikes in usage.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pass/fail threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical host CPU utilization</td>
<td>85% a</td>
</tr>
<tr>
<td>Physical host memory utilization</td>
<td>85%</td>
</tr>
<tr>
<td>Network throughput</td>
<td>85%</td>
</tr>
<tr>
<td>Storage I/O latency</td>
<td>20 ms</td>
</tr>
<tr>
<td>Login VSI Failed Session</td>
<td>2%</td>
</tr>
</tbody>
</table>

  a. The Ready Solutions for VDI team recommends that average CPU utilization not exceed 85% in a production environment. A 5% margin of error was allocated for this validation effort. Therefore, CPU utilization sometimes exceeds our recommended percentage. Because of the nature of LoginVSI testing, these exceptions are reasonable for determining our sizing guidance.

- **GPU resources**—vSphere Client monitoring collects data about the GPU resource use from a script that is run on ESXi 6.7 and later hosts. The script runs for the duration of the test and contains NVIDIA System Management Interface commands. The commands query each GPU and log the GPU processor, temperature, and memory use to a .csv file.
Load generation

Login VSI from Login VSI, Inc. is the industry-standard tool for testing VDI environments and Remote Desktop Session Host (RDSH) environments.

Login VSI installs a standard collection of desktop application software (for example, Microsoft Office, Adobe Acrobat Reader) on each VDI desktop. It then uses launcher systems to connect a specified number of users to available desktops within the environment. When the user is connected, a logon script starts the workload, configures the user environment, and starts the test script. Each launcher system can launch connections to a number of VDI desktops (target machines). A centralized management console configures and manages the launchers and the Login VSI environment.

In addition, we used the following login and boot paradigm:

- Users were logged in within a login timeframe of 1 hour, except when testing low-density solutions such as GPU/graphic-based configurations, in which users were logged in every 10 to 15 seconds.
- All desktops were started before users logged in.
- All desktops ran an industry-standard anti-virus solution. Windows 10 machines used Windows Defender.

Profiles and workloads

Machine profiles and user workloads determine the density numbers that the solution can support. Each profile and workload is bound by specific metrics and capabilities, with two targeted at graphics-intensive use cases.

Profiles and workloads are defined as follows:

- **Profile**—The configuration of the virtual desktop; the number of vCPUs and the amount of RAM that is configured on the desktop and available to the user
- **Workload**—The set of applications that is used

We load-tested two profiles by using a workload that is representative of the profile. The following table describes each use case.

<table>
<thead>
<tr>
<th>Profile name/workload</th>
<th>Workload description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task worker</td>
<td>The least intensive of the standard workloads. This workload primarily runs Microsoft Excel and Microsoft Internet Explorer, with some minimal Microsoft Word activity, as well as Microsoft Outlook, Adobe, and copy and zip actions. The applications are started and stopped infrequently, which results in lower CPU, memory, and disk I/O usage.</td>
</tr>
</tbody>
</table>
| Knowledge worker      | Designed for virtual machines with 2 vCPUs. This workload includes the following activities:  
  - Outlook—Browse messages.  
  - Internet Explorer—Browse websites and open a YouTube style video (480p movie trailer) three times in every loop.  
  - Word—Start one instance to measure response time and another to review and edit a document.  
  - Doro PDF Printer and Acrobat Reader—Print a Word document and export it to PDF.  
  - Excel—Open a large randomized sheet. |
### Table 5 Virtual desktop profiles and workloads (continued)

<table>
<thead>
<tr>
<th>Profile name/workload</th>
<th>Workload description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Power worker</strong></td>
<td>The most intensive of the standard workloads. The following activities are performed with this workload:</td>
</tr>
<tr>
<td></td>
<td>• Begin by opening four instances of Internet Explorer and two instances of Adobe Reader, which remain open throughout the workload.</td>
</tr>
<tr>
<td></td>
<td>• Perform more PDF printer actions than in the other workloads.</td>
</tr>
<tr>
<td></td>
<td>• Watch a 720p and a 1080p video.</td>
</tr>
<tr>
<td></td>
<td>• Reduce the idle time to two minutes.</td>
</tr>
<tr>
<td></td>
<td>• Perform various copy and zip actions.</td>
</tr>
<tr>
<td><strong>Graphics performance configuration/multimedia</strong></td>
<td>A workload that is designed to heavily stress the CPU when using software graphics acceleration. GPU-accelerated computing offloads the most compute-intensive sections of an application to the GPU while the CPU processes the remaining code. This modified workload uses the following applications for its GPU/CPU-intensive operations:</td>
</tr>
<tr>
<td></td>
<td>• Adobe Acrobat</td>
</tr>
<tr>
<td></td>
<td>• Google Chrome</td>
</tr>
<tr>
<td></td>
<td>• Google Earth</td>
</tr>
<tr>
<td></td>
<td>• Microsoft Excel</td>
</tr>
<tr>
<td></td>
<td>• HTML5 3D spinning balls</td>
</tr>
<tr>
<td></td>
<td>• Internet Explorer</td>
</tr>
<tr>
<td></td>
<td>• MP3</td>
</tr>
<tr>
<td></td>
<td>• Microsoft Outlook</td>
</tr>
<tr>
<td></td>
<td>• Microsoft PowerPoint</td>
</tr>
<tr>
<td></td>
<td>• Microsoft Word</td>
</tr>
<tr>
<td></td>
<td>• Streaming video</td>
</tr>
</tbody>
</table>
A comparison of linked clones and instant clones

Horizon supports two provisioning methods that deliver space-optimized virtual desktop pools: linked clones and instant clones. For this testing we have used instant clones to create virtual machines. The user density per host is not impacted by using one over the other. The differences in the test graphs for these two methods are a result of the following processes:

- For linked clones, all the VMs are re-booted before the test starts to simulate a boot storm. The CPU spike during the boot storm phase is due to the CPU being utilized by all VMs during powering on. Once the VMs are booted up, CPU utilization comes down to near zero as shown in the following figure. During the logon phase CPU utilization again increases and once all users have logged in CPU utilization remains constant as shown in the steady state phase in the figure. Once the steady state phase is over and users start to log off, CPU utilization decreases and dips to near zero when all users log off.

  Figure 2  Host CPU utilization for linked clones

- For instant clones, the VMs are re-booted after the session ends because when a user logs out of the instant clone, the clone is destroyed and re-created for the next user. CPU utilization gradually increases during the log on phase when users start logging in and then remains constant during the steady state phase when logins have been completed, as shown in the following figure. Once the steady state period is over and users starts to log off, CPU utilization again decreases and reaches near zero when all users have logged off. After user logoff the instant clone pool is re-created. During this phase there is a CPU spike that then drops to near zero when pool re-creation is complete.
Desktop VM test configurations

The following table summarizes the compute VM configurations for the profiles and workloads that we tested.

Table 6 Desktop VM specifications

<table>
<thead>
<tr>
<th>User profile</th>
<th>vCPUs</th>
<th>ESXi configured memory</th>
<th>ESXi reserved memory</th>
<th>Screen resolution</th>
<th>Operating system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task worker</td>
<td>2</td>
<td>3 GB</td>
<td>1.5 GB</td>
<td>1280 x 720</td>
<td>Windows 10 Enterprise 64-bit</td>
</tr>
<tr>
<td>Knowledge worker</td>
<td>2</td>
<td>4 GB</td>
<td>2 GB</td>
<td>1920 x 1080</td>
<td>Windows 10 Enterprise 64-bit</td>
</tr>
<tr>
<td>Power worker</td>
<td>4</td>
<td>8 GB</td>
<td>4 GB</td>
<td>1920 x 1080</td>
<td>Windows 10 Enterprise 64-bit</td>
</tr>
<tr>
<td>Multimedia</td>
<td>4</td>
<td>8 GB</td>
<td>8 GB</td>
<td>1920 x 1080</td>
<td>Windows 10 Enterprise 64-bit</td>
</tr>
</tbody>
</table>

a. Dell EMC has validated the LoginVSI Task worker workload with two vCPUs assigned per VM, although LoginVSI lists the typical VM vCPU profile for this workload as being a single vCPU. Dell EMC diverges from this definition to deliver virtual desktops with great user experience. Increasing the vCPU count to 2 in the vCPU profile associated with the Task worker workload does have a minor impact on densities but generates improved user experience in return.

Test results and analysis

We used the Login VSI test suite to simulate the user experience for several profile types under the typical workload for that type. The following table summarizes the test results for the compute hosts using the various workloads and configurations.


<table>
<thead>
<tr>
<th>Server configuration</th>
<th>Workload</th>
<th>User density</th>
<th>Protocol</th>
<th>Avg CPU</th>
<th>Avg memory consumed</th>
<th>Avg active memory</th>
<th>Avg IOPS / user</th>
<th>Avg net Mbps/user</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density Optimized</td>
<td>Knowledge worker</td>
<td>160</td>
<td>PCoIP</td>
<td>84%</td>
<td>640</td>
<td>151 GB</td>
<td>5.9</td>
<td>4.9</td>
</tr>
<tr>
<td>Density Optimized</td>
<td>Power worker</td>
<td>120</td>
<td>PCoIP</td>
<td>87%</td>
<td>731</td>
<td>198 GB</td>
<td>6.9</td>
<td>8.6</td>
</tr>
<tr>
<td>Density Optimized + 6 x T4</td>
<td>Multimedia (Virtual Workstation: T4-2Q)</td>
<td>48</td>
<td>Blast</td>
<td>48%</td>
<td>444</td>
<td>392 GB</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>Density Optimized + 6 x T4</td>
<td>Power worker (Virtual PC: T4-1B)</td>
<td>96</td>
<td>Blast</td>
<td>95%</td>
<td>451</td>
<td>488 GB</td>
<td>9.5</td>
<td>5.9</td>
</tr>
</tbody>
</table>

- The Ready Solutions for VDI team recommends that average CPU utilization not exceed 85% in a production environment. A 5% margin of error was allocated for this validation effort. Therefore, CPU utilization sometimes exceeds our recommended percentage. Because of the nature of LoginVSI testing, these exceptions are reasonable for determining our sizing guidance.
- Note that 96 users were achieved at 95% CPU utilization. The CPU utilization threshold of 85% is relaxed when testing with graphics cards. This test represents maximum utilization of the graphical resources available to the system as well as full user concurrency. Ideally, in a production environment, you decrease the user density slightly or use higher bin processors to bring the CPU utilization closer to the 85% threshold. All LoginVSI tests completed successfully without reaching VSI maximum, indicating that user experience was good.

The table headings are defined as follows:

- **User density**—The number of users per compute host that successfully completed the workload test within the acceptable resource limits for the host. For clusters, this number reflects the average of the density achieved for all compute hosts in the cluster.
- **Avg CPU**—The average CPU usage over the steady-state period. For clusters, this number represents the combined average CPU usage of all compute hosts. On the latest Intel processors, the ESXi host CPU metrics exceed the rated 100 percent for the host if Turbo Boost is enabled, which is the default setting. An additional 35 percent of CPU is available from the Turbo Boost feature, but this additional CPU headroom is not reflected in the VMware vSphere metrics where the performance data is gathered. Therefore, CPU usage for ESXi hosts is adjusted and each CPU graph includes a line indicating the potential performance headroom that is provided by Turbo boost.
- **Avg active memory**—For ESXi hosts, the amount of memory that is actively used, as estimated by the VMkernel based on recently touched memory pages. For clusters, this is the average amount of guest physical memory that is actively used across all compute hosts over the steady-state period.
- **Avg IOPS per user**—IOPS calculated from the average disk IOPS over the steady state period divided by the number of users.
Density Optimized configuration

We performed the following multimedia performance testing on the VxRail Ready Node Density Optimized configuration described in Validated hardware resources on page 8. These results also apply to the vSAN Density Optimized configuration, as the configurations are the same.

Knowledge Worker, 480 users, ESXi 6.7, Horizon 7.7

We ran the following tests for this workload.

CPU usage

Each compute host was populated with 160 virtual machines per host. With all user virtual machines powered on and before starting the test, the CPU usage was approximately 7 percent. The Compute A host also hosted the three VxRail management VMs.

The following figure shows the performance data for 160 user sessions per host. The CPU reaches a steady state average of 84 percent across the three compute hosts during the test cycle when 160 users are logged on to each host. The Compute A host's CPU usage was slightly higher due to the presence of the three VxRail management VMs.

Memory

With regard to memory consumption for the cluster, out of a total of 768 GB available memory per node, memory usage was not an issue. The compute hosts reached a maximum memory consumption of 665 GB (Compute A host) with active memory usage reaching a max of 530 GB during the recreation of the instant clones. There was no ballooning or swapping at any point during the test.
Network usage

Network bandwidth is not an issue on this test run with a steady state peak of approximately 1,237 Mbps. The busiest period for network traffic was at the end of the logon phase. The steady state average was 791 Mbps.
IOPS

The IOPS graphs clearly display the initial logon of the desktops, the steady state and logoff phases, and finally the recreation of the desktops after testing was complete. The graph displays the disk IOPS figure for the vSAN cluster.

The cluster reached a maximum total (read + write) of disk 34,371 IOPS during the instant clone recreation period after testing. The steady state total average was 2,850 disk IOPS. The steady state peak was 3,701 disk IOPS (read + write IOPS).

Disk I/O latency

Disk I/O latency was not an issue during the Login VSI testing period of this test run. Total latency (read + write) reached its maximum of approximately 3.1 ms during the recreation of the instant clones. This was well below the 20 ms threshold where latency becomes potentially troublesome. The average total latency during steady state was 0.9 ms.
**User experience**

The following figure shows that the user experience score did not reach the Login VSI maximum for this test.

---

**Power Worker, 360 users, ESXi 6.7, Horizon 7.7**

We ran the following tests for this workload.

**CPU**

Each compute host was populated with 120 virtual machines per host. With all user virtual machines powered on and before starting the test, the CPU usage was approximately 6 percent. The Compute A host also hosted the three VxRail management VMs. The following figure shows the performance data for user sessions per host. The CPU reaches a steady state average of 87 percent across the three compute hosts during the test cycle when 120 users are logged on to each host. The Compute A host’s CPU usage was slightly higher due to the presence of the three VxRail management VM’s.
Memory

With regard to memory consumption for the cluster, out of a total of 768 GB available memory per node, memory usage was quite high. The compute hosts reached a maximum memory consumption of 748 GB (Compute A host) with active memory usage reaching a maximum of 868 GB\(^1\). There was a small amount of memory swapping on all the hosts during the test. However, the maximum amount of swapping reached only .06 GB.

\(^1\) During the recreation of the instant clones, the Active Memory metric exceeded the physical installed memory. This is possible because vCenter uses sampling to statistically estimate the number of active memory pages that are touched, rather than actually measuring the number of pages. Measuring every touched memory page would generate too much resource overhead for the hypervisor.
Network usage

Network bandwidth was not an issue on this test run with a steady state peak of approximately 2,954 Mbps. The busiest period for network traffic was during the recreation of the instant clones after testing had completed. One of the Compute A host reached a peak of 4,618 Mbps during the deletion and recreation of the instant clones. The steady state average was 1,041 Mbps.

IOPS

The IOPS graphs clearly display the initial logon of the desktops, the steady state and logoff phases, and finally the recreation of the desktops after testing was complete. The graph displays the disk IOPS figure for the vSAN cluster.
The cluster reached a maximum of 25,918 disk IOPS (read + write IOPS) during the instant clone recreation period after testing and a steady state average of 2,481 (read + write IOPS). The steady state peak was 4,019 disk IOPS (read + write IOPS).

**Disk I/O latency**

Disk I/O latency was not an issue during the Login VSI testing period of this test run. Total latency (read + write) reached its maximum at approximately 2.7 ms during the logon phase. This was well below the 20 ms threshold where latency becomes potentially troublesome. The average total latency during steady state was 1 ms and the steady state peak was 1.1 ms (read + write latency).
**User experience**

As shown in the following figure, the Login VSI Max user experience score for this test was not reached, indicating that there was no deterioration in user experience at the number of users tested.

![User experience chart](image)

### Multimedia Workload, 48 vGPU users, ESXi 6.7, Horizon 7.7

We ran the following tests for this workload

**CPU usage**

The GPU-enabled Compute host was populated with 48 vGPU enabled virtual machines and used the NVIDIA T4-2Q profile. With all user virtual machines powered on and before starting the test, the CPU usage was approximately 8 percent on the GPU-enabled Compute host.

The following figure shows the performance data for 48 user sessions on the Management and GPU-enabled Compute hosts. The CPU reaches a steady state average of 48 percent during the test cycle when all users are logged on to the GPU-enabled Compute host. The Management host had no load during the test.

![CPU usage chart](image)

**GPU usage**

The GPU usage during the steady state period averaged approximately 35 % and reached a peak usage of 38 % with the Multimedia workload.
Memory

There were no constraints on the Management or GPU-enabled Compute hosts. Of a total of 768 GB available memory per node, the GPU Compute host reached a maximum memory consumption of 444 GB with active memory usage reaching a maximum of 392 GB during the steady state phase. There were no variations in memory usage throughout the test as all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on hosts.
Network usage

Network bandwidth was not an issue on this test run with a steady state peak of approximately 1,203 Mbps. The busiest period for network traffic was during the logoff phase. The Compute/GPU hosts reached a peak of 4897 Mbps at the end of logoff.

IOPS

The IOPS graphs and IOPS numbers are taken from vCenter metrics. They clearly display the initial logon of the desktops, the steady state, and the logoff phase.

The Cluster IOPS reached a maximum of 7,281 read IOPS during logoff and averaged 433 read and write IOPS during the steady state.
Cluster latency

The cluster latency reached a maximum read latency of 7,281 and write latency of 518 during the logoff.

User experience

The Login VSI Max user experience score for this test was not reached. When manually interacting with the sessions during steady state, the mouse and window movement was responsive and video playback was good.

The baseline performance of 810 indicates that the user experience for this test run was Good. The index average reached 1,003, which is well below the threshold of 1,810.
Power Workload, 96 users, ESXi 6.7, Horizon 7.7

We ran the following tests for this workload.

**CPU usage**

The GPU-enabled Compute Host was populated with 96 vGPU-enabled virtual machines and used the NVIDIA T4-1B profile. With all user virtual machines powered on and before starting test, the CPU usage was approximately 16 percent on the GPU-enabled Compute host.

The following figure shows the performance data for 96 user sessions on the Management and GPU-enabled Compute hosts. The CPU reaches a steady state average of 96 percent during the test cycle when all users are logged on to the GPU-enabled Compute host.

**GPU usage**

The GPU usage during the steady state period averaged approximately 34 percent and reached a peak usage of 38 percent with the Power Workload.
Memory

There were no constraints on the Compute hosts. Of a total of 768 GB available memory per node, the Compute host reached a maximum memory consumption of 451 GB with active memory usage reaching a maximum of 395 GB during the steady state phase. There were no variations in memory usage throughout the test as all vGPU-enabled VM memory was reserved. There was no memory ballooning or swapping on any host.
Network usage

Network bandwidth was not an issue on this test run with a steady state peak of approximately 585 Mbps. The busiest period for network traffic was during the recreate phase with a peak of 2,319. The steady state averaged at 294 Mbps.

IOPS

The IOPS graphs and IOPS numbers are taken from vCenter metrics. They clearly display the Initial logon of the desktops, the steady state, and the logoff phase.

The Cluster IOPS reached a maximum of 9,325 read IOPS during the Recreate Clones and averaged 306 disk IOPS during steady state.
Cluster latency

The cluster latency reached a maximum Latency of 232 read and 845 write latency during the logoff.

User experience

The Login VSI Max user experience score for this test was not reached. When manually interacting with the sessions during steady state, the mouse and window movement was responsive and video playback was good.

The baseline performance of 1,054 indicated that the user experience for this test run was Good. The index average reached 1,619, which is well below the threshold of 2,055.
CHAPTER 4

Conclusion

- Test results and density recommendations ................................................................. 36
- Conclusions .................................................................................................................. 36
Test results and density recommendations

Test results provided recommended user densities.

The user densities in the following table were achieved by following the VMware best practices of FTT = 1 and a reserved slack space of 30 percent. All configurations were tested with Microsoft Windows 10 and Microsoft Office 2016. We implemented all mitigations to patch the Spectre, Meltdown and L1TF vulnerabilities at the hardware, firmware, and software levels to ensure an improved performance impact, which is reflected in the achieved user densities.

Table 8 User density recommendations for VMware vSphere ESXi 6.7 with VMware Horizon

<table>
<thead>
<tr>
<th>Server configuration</th>
<th>Workload</th>
<th>User density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density Optimized</td>
<td>Knowledge worker</td>
<td>160</td>
</tr>
<tr>
<td>Density Optimized</td>
<td>Power worker</td>
<td>120</td>
</tr>
<tr>
<td>Density Optimized + 6 x T4</td>
<td>Multimedia (Virtual Workstation: T4-2Q)</td>
<td>48</td>
</tr>
<tr>
<td>Density Optimized + 6 x T4</td>
<td>Power worker (Virtual PC: T4-1B)</td>
<td>96(^a)</td>
</tr>
</tbody>
</table>

\(^a\) The user density of 96 users was achieved at 95% CPU utilization. The CPU utilization threshold of 85% is relaxed when testing with graphics cards. This test represents maximum utilization of the graphical resources available to the system as well as full user concurrency. Ideally, in a production environment, you would decrease the user density slightly or use higher bin processors to bring the CPU utilization closer to the 85% threshold. All LoginVSI tests completed successfully without reaching VSI maximum, indicating that user experience was good.

Conclusions

Both vSAN Ready Nodes and VxRail appliances reach similar user densities when configured similarly, as shown during this validation. Remember to use VxRail sizing tools to reserve resources for management tools when designing your VDI infrastructure.

The configurations for the VxRail appliances and vSAN Ready Nodes have been optimized for VDI. We selected the memory and CPU configurations that provide optimal performance. You can change these configurations to meet your own requirements, but keep in mind that changing the memory and CPU configurations from those that have been validated in this document affects the user density per host. We applied mitigations to the Spectre and L1TF vulnerabilities during the validation.

With the introduction of the six-channels-per-CPU requirement for Skylake, and now Cascade Lake, the Density-Optimized memory configuration recommendation has increased from the previous guidance of 512 GB to 768 GB. This change was necessary to ensure a balanced memory configuration and optimized performance for your VDI solution. The additional memory is advantageous, considering the resulting increase in operating system resource utilization and the enhanced experience for users when they have access to additional memory allocations.
CHAPTER 5

References

This chapter presents the following topics:

- Dell EMC documentation.......................................................................................................38
- VMware documentation........................................................................................................ 38
- NVIDIA documentation.......................................................................................................... 38
Dell EMC documentation

The following Dell EMC documentation provides additional and relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell EMC representative. Also see the Dell EMC VDI Information Hub for a complete list of VDI resources.

- Dell EMC VxRail Appliance documentation
- Dell EMC Virtual Desktop Infrastructure

This document is part of the documentation set for this architecture, which includes the following:

- Dell EMC Ready Architectures for VDI: Designs for VMware Horizon on VxRail and vSAN Ready Nodes Design Guide
- Dell EMC Ready Architectures for VDI: Designs for VMware Horizon on VxRail and vSAN Ready Nodes Validation Guide
- Dell EMC Ready Architectures for VDI: Designs for VMware Horizon on VxRail and vSAN Ready Nodes Deployment Guide

VMware documentation

The following VMware documentation provides additional and relevant information:

- VMware vSphere documentation
- VMware Horizon 7 documentation
- Best Practices for Published Application and Desktops in VMware Horizon Apps and VMware Horizon 7
- vSAN Ready Node Configurator
- VMware Compatibility Guide
- Horizon 7 Enterprise Edition Reference Architecture
- Horizon 7 Enterprise Edition Multi-Site Reference Architecture

NVIDIA documentation

The following NVIDIA documentation provides additional and relevant information:

- NVIDIA Virtual GPU Software Quick Start Guide