Table of Contents

Lab Overview - HOL-1851-10-ADV - Horizon 7.1: Graphics Acceleration for 3D Workloads and vGPU ................................................................. 2
  Lab Guidance ........................................................................................................ 3
Module 1 - 3D Options in Horizon 7 (15 minutes - Basic) ........................................... 5
  Introduction ........................................................................................................... 6
  3D Desktop Graphics in Horizon ........................................................................... 7
  Choosing a 3D Graphics Acceleration Technology ............................................... 12
  Infrastructure Considerations .............................................................................. 20
Module 2 - NVIDIA Graphic Processing Unit (GPU) Manager VIB Installation .......... 21
  Module Overview ................................................................................................. 22
  Hands-on Labs Interactive Simulation: NVIDIA Graphics Processing Unit (GPU)
  Manager VIB Installation .................................................................................... 23
Module 3 - Virtual Graphic Processing Unit (vGPU) Profiles ....................................... 24
  Module Overview ................................................................................................. 25
  Hands-on Labs Interactive Simulation: Virtual Graphic Processing Unit (vGPU)
  Profiles ............................................................................................................... 26
Module 4 - Creating a 3D Horizon Desktop Pool ....................................................... 27
  Module Overview or Introduction ....................................................................... 28
  Hands-on Labs Interactive Simulation: Name of iSIM ......................................... 29
Module 5 - Creating a Horizon Remote Application Pool ........................................... 30
  Module Overview or Introduction ....................................................................... 31
  Hands-on Labs Interactive Simulation: Name of iSIM ......................................... 32
Lab Overview - HOL-1851-10-ADV - Horizon 7.1: Graphics Acceleration for 3D Workloads and vGPU
Lab Guidance

The Table of Contents can be accessed in the upper right-hand corner of the Lab Manual.

In this Hands-on Labs Interactive Simulation, you will learn about the different 3D technologies that exist, and considerations for each one. You will then walk through various stages of the installation and configuration process of setting up vGPU in Horizon 7.1 environment utilizing instant clones, application pools, and our Blast Extreme Protocol.

Lab Module List:

**Module 1 - Horizon Cloud Hosted Infrastructure Overview** (15 Minutes - Basic) - Overview of 3D options in an Horizon 7 environment (informational only)

**Module 2 - NVIDIA Graphics Processing Until (GPU) Manager VIB Installation** (15 Minutes - Advanced) - Walk through installation of a vGPU card on a ESXi 6.5 host

**Module 3 - Selecting Virtual Graphic Unit (vGPU) Profile** (15 Minutes - Advanced) - Walk through setup and modification of a VM utilizing a vGPU

**Module 4 - Creating a 3D Horizon Desktop Pool** (45 Minutes - Advanced) Walk through process of creating a Horizon desktop pool configured with a vGPU

**Module 5 - Creating a 3D Horizon Remote Application Pool** (45 Minutes - Advanced) Prepare RDS host, and deploy an application pool.

**Captain:** Pamela Norris, Staff Technical Account Manager, Chicago, Illinois

This lab may be available in other languages. To set your language preference and have a localized manual deployed with your lab, you may utilize this document to help guide you through the process:

Location of the Main Console

1. The area in the RED box contains the Main Console. The Lab Manual is on the tab to the Right of the Main Console.
2. Your lab starts with 90 minutes on the timer. The lab can not be saved. All your work must be done during the lab session. But you can click the **EXTEND** to increase your time. If you are at a VMware event, you can extend your lab time twice, for up to 30 minutes. Each click gives you an additional 15 minutes. Outside of VMware events, you can extend your lab time up to 9 hours and 30 minutes. Each click gives you an additional hour.
3. All work in this HOL Interactive Simulation will take place in the manual.
Module 1 - 3D Options in Horizon 7 (15 minutes - Basic)
Introduction

This Module contains the following lessons:

• Lesson 1: An overview of 3D technologies supported in Horizon 7
• Lesson 2: Choosing a 3D Graphics Acceleration Technology
• Lesson 3: General considerations in Horizon 7 to support 3D

Horizon 7 is VMware’s easy to use, enterprise class solution that provides users access to applications and virtual desktops, through a fat or web client. When 3D graphics are utilized within a Horizon 7 environment, it allows your users to emerge themselves in a true 3D experience while your organization can take advantage of the benefits such as cost and security since their desktop or application session is contained in the data center.

You have several 3D options to choose from when looking to utilize 3D in Horizon 7. These options will be reviewed in detail further in this lab.

• The first option is Soft 3D.
  ◦ Soft 3D does not require a physical GPU, but instead uses the graphics driver that is automatically installed with VMware Tools to render the images.

• The second set of options are various types of 3D hardware acceleration leveraged from a specific set of graphics cards from AMD, Intel and NVIDIA installed in your vSphere hosts.
  ◦ Virtual Dedicated Graphics Acceleration (vDGA)
  ◦ Virtual Shared Graphics Acceleration (vSGA)
  ◦ Virtual Graphics Processing Unit (vGPU)

The correct rendering solution for your environment will depend on multiple factors including cost and use case. Understanding user requirements is the key to getting maximum performance from each virtual machine on your host.

More information on the supported physical GPU cards in an Horizon 7 environment can be found in the VMware Compatibility Guide.
3D Desktop Graphics in Horizon

VMware vSphere servers with Horizon 7 hosted in enterprise data centers enable users to access virtual desktops running 3D applications from a wide range of client devices.

Horizon 3D Overview

This solution provides users with graphics performance roughly equivalent to high-end graphics workstations, using lower-cost clients or re-purposed devices.

For detailed information on, please see the [VMware Horizon 3D Engineering Workloads Reference Architecture](http://bit.ly/3D-RA)
Designing for 3D Desktop Workload Use Cases

Horizon 7 offers four types of 3D graphics acceleration: software-based Soft 3D, and hardware-based vSGA, vGPU, and vDGA.

The picture above illustrates how the 3D technologies map to the main use case categories: task workers, knowledge workers, power users, and designers.

Use Case Categories

VMware's methodology defines 4 main categories of 3D use cases.
Task workers often require only Soft 3D, a software-based 3D renderer suitable for less graphics-intensive applications. They do not need, or realize a noticeable benefit from, hardware-based 3D acceleration. For that reason, the task worker use case is not considered in this paper. Soft 3D is a standard component of Horizon 7.

Knowledge Workers
Office workers and executives fall into this category, typically using applications such as Microsoft Office, Adobe Photoshop, and other non-specialized end-user applications. A vSGA solution can improve performance for this use case by providing high levels of consolidation of users across GPUs.

However, vSGA does not provide a wide range of graphics API support, so it is often worthwhile to consider a vGPU-based solution for knowledge workers.

**Power Users**

These users consume more complex visual data, but their requirements for manipulations of large datasets and specialized software are less intense than for designers. Their needs can typically be served more than adequately with access to a shared vGPU.
Designers and advanced engineering and scientific users often create and work with large, complex datasets and require graphics-intensive applications such as 3D design, molecular modeling, and medical diagnostics software from companies such as Dassault Systèmes, Enovia, Siemens NX, and Autodesk. These users typically require either a vGPU- or vDGA-based solution.
Choosing a 3D Graphics Acceleration Technology

There are three types of hardware-based graphics acceleration available for View virtual desktops in Horizon 7

Choosing a 3D Graphics Acceleration Technology

<table>
<thead>
<tr>
<th>vDGA</th>
<th>vGPU</th>
<th>vSGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU dedicated to one user</td>
<td>GPUs shared among users but can be dedicated</td>
<td>GPUs shared among users</td>
</tr>
<tr>
<td>1:1 consolidation ratio (1 user per physical GPU)</td>
<td>Good consolidation ratio (8 users per physical GPU)</td>
<td>High consolidation ratio (limited by video memory on graphics card)</td>
</tr>
<tr>
<td>Workstation-level performance</td>
<td>Entry-level workstation performance under load</td>
<td>Solid performance for lightweight applications, but no driver certification</td>
</tr>
<tr>
<td>Maximum compatibility with all 3D GPU rendering and computation applications</td>
<td>Full compatibility with all 3D and GPU rendering applications; requires certification</td>
<td>Compatibility limited by API support and virtual machine video RAM capacity</td>
</tr>
<tr>
<td>DirectX 9, 10, or 11</td>
<td>DirectX 9, 10, or 11</td>
<td>DirectX 9.0 SM3 only</td>
</tr>
<tr>
<td>OpenGL 2.1, 3.x, or 4.x</td>
<td>OpenGL 2.1, 3.x, or 4.x</td>
<td>OpenGL 2.1 only</td>
</tr>
<tr>
<td>Hardware video playback</td>
<td>Hardware video playback</td>
<td>Software video playback only</td>
</tr>
<tr>
<td>Compute APIs with CUDA or OpenCL</td>
<td>Does not support compute APIs, CUDA, or OpenCL</td>
<td>Does not support compute APIs, CUDA, or OpenCL</td>
</tr>
<tr>
<td>Not compatible with VMware vSphere vMotion® and vSphere High Availability</td>
<td>Not compatible with vSphere vMotion and HA</td>
<td>vSphere vMotion, HA, and VMware vSphere Distributed Resource Scheduler™ compatible—automatically falls back to software renderer as needed</td>
</tr>
</tbody>
</table>

The table compares the main features of the GPU options. As you can see there is a trade-off between performance and compatibility.
Virtual Dedicated Graphics Acceleration (vDGA)

This technology provides a user with unrestricted, fully dedicated access to a single vGPU. Although consolidation and management trade-offs are associated with dedicated access, vDGA offers the highest level of performance for users with the most intensive graphics computing needs. It enables the use of applications that run OpenGL 4.4, Microsoft DirectX 9, 10, or 11, and NVIDIA CUDA 5.0.
With vDGA, the hypervisor passes the GPUs directly to guest virtual machines, so the technology is also known as GPU pass-through. No special drivers are required in the hypervisor. However, to enable graphics acceleration, the appropriate NVIDIA driver needs to be installed on the guest virtual machines. The installation procedures are the same as for physical machines.

Because the GPU is passed through to the guest OS, which uses native graphics drivers, vDGA fully supports everything the chosen driver can do natively, including but not limited to all versions of DirectX, OpenGL, and CUDA.

<table>
<thead>
<tr>
<th>BENEFITS</th>
<th>PROS</th>
<th>CONS</th>
</tr>
</thead>
</table>
| • Enables dedicated access to physical GPU hardware for 3D and high-performance graphical workloads. | • Outstanding performance  
• Performance equivalent to dedicated GPU in physical desktop  
• Supports the entire API stack  
• Direct driver support for GPU  
• Vendor certification—technology is direct pass-through  
• Could be a true workstation replacement option | • 1:1 consolidation ratio                                           |
Horizon 7 and vSphere 6.0 include vGPU support. Like vDGA, vGPU brings the benefit of wide API support and native NVIDIA drivers but with greater scalability.
vGPU is essentially vDGA with multiple users instead of one user. As with vDGA, a user or administrator needs to install the appropriate NVIDIA driver on the guest virtual machine, and all graphics commands are passed directly to the GPU without having to be translated by the hypervisor. Up to eight virtual machines can share a GPU. Calculating the exact number of desktops or users per GPU depends on application requirements, screen resolution, number of displays, and frame rate measured in frames per second (FPS).

The vGPU technology provides better performance than vSGA and higher consolidation ratios than vDGA. It is a good technology to use for low-, mid-, or even advanced-level engineers and designers as well as for power users with 3D application requirements. One drawback of vGPU, however, is that it might require applications be re-certified in order to be supported.
Choosing a vGPU Profile

<table>
<thead>
<tr>
<th>Virtual GPU Type</th>
<th>Physical Board</th>
<th>Physical GPUs</th>
<th>FB Per Virtual GPU</th>
<th>Display Heads</th>
<th>Maximum Resolution</th>
<th>Maximum Virtual GPUs Per Physical GPU</th>
<th>Maximum Virtual GPUs Per Physical Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRID M60-0q</td>
<td>GRID M60</td>
<td>two</td>
<td>512M</td>
<td>2</td>
<td>2560x1600</td>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>GRID M60-1q</td>
<td>GRID M60</td>
<td>two</td>
<td>1G</td>
<td>2</td>
<td>2560x1600</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>GRID M60-2q</td>
<td>GRID M60</td>
<td>two</td>
<td>2G</td>
<td>4</td>
<td>2560x1600</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>GRID M60-4q</td>
<td>GRID M60</td>
<td>two</td>
<td>4G</td>
<td>4</td>
<td>3840x2160</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>GRID M60-8q</td>
<td>GRID M60</td>
<td>two</td>
<td>8G</td>
<td>4</td>
<td>3840x2160</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Each physical GPU can support several virtual GPU types, or profiles. Each vGPU profile has a fixed amount of frame buffer memory, number of supported display heads, and maximum resolutions, and is targeted at different classes of workload.

<table>
<thead>
<tr>
<th>Virtual GPU Type</th>
<th>Physical Board</th>
<th>Physical GPUs</th>
<th>FB Per Virtual GPU</th>
<th>Display Heads</th>
<th>Maximum Resolution</th>
<th>Maximum Virtual GPUs Per Physical GPU</th>
<th>Maximum Virtual GPUs Per Physical Board</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRID M6-0q</td>
<td>GRID M6</td>
<td>one</td>
<td>512M</td>
<td>2</td>
<td>2560x1600</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>GRID M6-1q</td>
<td>GRID M6</td>
<td>one</td>
<td>1G</td>
<td>2</td>
<td>2560x1600</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>GRID M6-2q</td>
<td>GRID M6</td>
<td>one</td>
<td>2G</td>
<td>4</td>
<td>2560x1600</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>GRID M6-4q</td>
<td>GRID M6</td>
<td>one</td>
<td>4G</td>
<td>4</td>
<td>3840x2160</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>GRID M6-8q</td>
<td>GRID M6</td>
<td>one</td>
<td>8G</td>
<td>4</td>
<td>3840x2160</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The GPU profiles (ending in Q, as shown in the images) undergo the same application certification process as the NVIDIA Quadro workstation-class processors.
Virtual Shared Graphics Acceleration (vSGA)

This technology allows a GPU to be shared across multiple virtual desktops. It is an attractive solution for users who require the full potential of the GPU’s capability during brief periods. However, vSGA can create bottlenecks, depending on which applications are used and resources needed from the GPU. vSGA is generally used for knowledge workers and occasionally for power users, but it is restricted in its support for OpenGL and DirectX versions.
With vSGA, the physical GPUs in the host are virtualized and shared across multiple guest virtual machines. An NVIDIA driver needs to be installed in the hypervisor. Each guest virtual machine uses a proprietary VMware vSGA 3D driver that communicates with the NVIDIA driver in vSphere.

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Enables shared access to physical GPU hardware for 3D and high-performance graphical workloads</td>
<td>• Mature technology with appropriate number of ISVs supporting this type of configuration</td>
<td>• Not suitable for high-end or compute-intensive workloads</td>
</tr>
<tr>
<td>• Desktops still see abstracted VMware SVGA device for maximum virtual machine compatibility and portability</td>
<td>• Scales well and provides good performance</td>
<td>• Shared environment, shared problems</td>
</tr>
<tr>
<td>• Cost effective with multiple virtual machines sharing single GPU resource</td>
<td>• Full compatibility with hosts lacking physical GPUs (for vSphere vMotion, DRS, and so on)</td>
<td>• Limited API support (Microsoft)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Limited maximum video RAM of 512 MB</td>
</tr>
</tbody>
</table>

With vSGA, the physical GPUs in the host are virtualized and shared across multiple guest virtual machines. An NVIDIA driver needs to be installed in the hypervisor. Each guest virtual machine uses a proprietary VMware vSGA 3D driver that communicates with the NVIDIA driver in vSphere.
Infrastructure Considerations

As is the case with many workloads in a Virtual Desktop Infrastructure (VDI), it is essential to have a solid understanding of each of your user’s needs, in order to plan for the size and required throughput required for a positive user experience. Design it for the user experience.

Local SSD, VMware vSAN, or all-flash-based storage arrays are a good choice for workloads typical in a 3D environment. Storage suffers the biggest performance impact when bandwidth is shared directly by all virtual machines. It is essential to distribute your virtual machine workload to achieve maximum performance.

If using linked or instant clones, consider the impact of applications that load large data sets from the network and store locally. You may want to look at using App Volumes so that your applications can be directed to a separate storage tier. (Please refer to HOL-1851-03-ADV for more information on App Volumes)

Determining how your application utilizes CPU, is a critical component for a 3D. Operations, such as update, clash detection, drawing, and weight analysis are especially CPU-intensive. Because processes, including CAD operations, are mono-threaded, a higher CPU clock frequency increases performance more than an increase in CPU cores (or vCPUs). However, multi-threaded applications can benefit from running application threads across multiple CPU cores (or vCPUs).

3D requires more memory than a virtual desktop that is running traditional graphics. Consider using the highest frequency memory available in the host to produce the best results.

You will want to confirm there is sufficient bandwidth between the location of the graphics card and the end users device along with sufficient processing power on the client device itself. 3D applications typically have fast-changing graphics that can require considerable bandwidth that flows between the two devices.

Once configured, a user can access their virtual desktop from Workspace One or the Horizon Client. Providing access to their virtual desktop on anything from a smart phone to the hardiest MAC or PC, and everything in between. That being said, it is important not to underestimate the amount of CPU and memory that the end user device needs to decode the display protocol. It is possible that a lower end client may not have the CPU processing power decode the data fast enough for a positive user experience. The amount of end user device resource will be dependent on configuration, application workload, and how the virtual desktop is being accessed. Testing will be required in order to validate end user experience in your environment.
Module Overview

In this module you will walk through the process of installing the NVIDIA Grid Manager VIB installation for the NVIDIA Grid Tesla M60 card on a vSphere 6.5 host.

This Module contains the following lessons:

• Lesson 1: Installation of NVIDIA Grid Manager VIB Installation
• Lesson 2: Verification of VIB Installation
Hands-on Labs Interactive Simulation: NVIDIA Graphics Processing Unit (GPU) Manager VIB Installation

The interactive simulation will allow you to experience steps which are too time-consuming or resource intensive to do live in the lab environment. We recommend you use Chrome to take the iSIM.

1. Click here to open the interactive simulation. It will open in a new browser window or tab.
2. When finished, click the “Return to the lab” link to continue with this lab.
Module 3 - Virtual Graphic Processing Unit (vGPU) Profiles
Module Overview

In this module, you will review what a vGPU profile is, considerations when selecting a profile, and modifying a profile.

Lesson 1: Define a virtual graphics process unit (vGPU) Profile

Lesson 2: Set and Modify a vGPU Profile
Hands-on Labs Interactive Simulation: Virtual Graphic Processing Unit (vGPU) Profiles

The interactive simulation will allow you to experience steps which are too time-consuming or resource intensive to do live in the lab environment. We recommend you use Chrome to take the iSIM.

1. **Click here to open the interactive simulation.** It will open in a new browser window or tab.
2. When finished, click the “Return to the lab” link to continue with this lab.
Module Overview or Introduction

This module covers the following lessons:

• Prepare a vGPU parent virtual machine
• Deploy an Instant Clone desktop pool
• Access vGPU virtual desktop
Hands-on Labs Interactive Simulation: Name of iSIM

This part of the lab is presented as a **Hands-on Labs Interactive Simulation**. This will allow you to experience steps which are too time-consuming or resource intensive to do live in the lab environment. In this simulation, you can use the software interface as if you are interacting with a live environment. We recommend you use Chrome to take the iSIM.

1. [Click here to open the interactive simulation](#). It will open in a new browser window or tab.
2. When finished, click the “Return to the lab” link to continue with this lab.
Module 5 - Creating a Horizon Remote Application Pool
Module Overview or Introduction

This module covers the following lessons:

- Prepare and install RDS Host
- Install and Configure vGPU card
- Publish and RDSH Application
Hands-on Labs Interactive Simulation: Name of iSIM

The interactive simulation will allow you to experience steps which are too time-consuming or resource intensive to do live in the lab environment. We recommend you use Chrome to take the iSIM.

1. Click here to open the interactive simulation. It will open in a new browser window or tab.
2. When finished, click the “Return to the lab” link to continue with this lab.
Conclusion

Thank you for participating in the VMware Hands-on Labs. Be sure to visit http://hol.vmware.com/ to continue your lab experience online.

Lab SKU: HOL-1851-10-ADV

Version: 20170920-142740