GTC EXPRESS WEBINAR: GETTING THE MOST OUT OF VMWARE HORIZON VIEW VDGA WITH NVIDIA GRID

Steve Harpster - Solution Architect NALA
AGENDA

- Overview
- Prerequisites
- Building vDGA-enabled VMs
- Basic perf tuning
- Support resources
- Q&A
World Leader in Visual Computing
NVIDIA Brands

GPU

Mobile

Cloud

GeForce®
Quadro®
Tesla®

Tegra®

NVIDIA GRID™

World Leader in Visual Computing
Visual Realism and Accuracy

Complex materials surfaces, reflections and shadows
Fast and Interactive Performance

Without RealView (without GPU)

With RealView (with GPU)
Workstation Unit Share

82%

PREFERRED, TRUSTED BRAND
ISV Certifications

Many times the ISV applications require an NV GPU for graphics

Examples: AutoCAD and SolidWorks require NV Graphics
Industries that use professional graphics:

- Aerospace
- Manufacturing & Automotive
- Construction
- Energy
- Engineering Services
- Hi Tech - Electronics
- Industrial Equipment
- Medical Equipment
What is a GPU?

CPU
Optimized for Serial Tasks

GPU
Optimized for Many Parallel Tasks

In a virtual world the CPU has been asked to be all things from vCPU to networking to storage, and GPU!
Segmenting the User Population

Tier 1 (e.g. design engineers)
Top rendering performance (dedicated GPU: vDGA)

Tier 2 (viewing/editing of 3D drawings)
Shared GPU: vSGA, or possibly vDGA

Tier 3 (typical knowledge workers)
Shared GPU: vSGA
importance of GPU

Becoming a Must Have

Must Have

3D Engineering & Design Apps
- Autodesk
- SolidWorks
- PTC

Office Productivity
- Microsoft Office
- Adobe PDF

PLM & Volume Design
- Teamcenter

Windows 7

Web
- HTML
- Flash
- HTML 5
### NVIDIA GRID K2

#### GPU
- 4 Kepler GPUs

#### CUDA Cores
- 768 (192/GPU)

#### Memory Size
- 16GB DDR3 (4GB/GPU)

#### Max Power
- 130 W

#### Equivalent Quadro with Pass-through
- Quadro K600 (entry)

### NVIDIA GRID K1

#### GPU
- 2 High End Kepler GPUs

#### CUDA Cores
- 3072 (1536/GPU)

#### Memory Size
- 8GB GDDR5 (4GB/GPU)

#### Max Power
- 225 W

#### Equivalent Quadro with Pass-through
- Quadro K5000 (high end)

---

1. Number of users depends on software solution, workload, and screen resolution.
GPU Pass-through Fundamentals

- GPU passthrough is 1:1 assignment of physical GPU to a guest VM
  - Brings the full power of a GPU to a guest VM
- The NVIDIA driver is resident in the guest
  - Must have for applications
- Can co-exist alongside vSGA and non GPU enabled guests
- First line of attack in diagnosing performance problems!
  - Treat like a physical PC, start with basics e.g. App issues, Task Manager!
  - Check for other issues in the virtualization configuration, network, storage etc.
VMware ESXi 5.5 and vDGA

- Supports all versions of DX, OpenGL and CUDA
- NO *.vib driver is necessary on ESX server host
  - Unless also doing vSGA
- PCI Pass thru must be enabled on the ESX server
- NVIDIA Win7 driver installed on the guests
GPU Virtualization - vDGA (pass-thru)

- **Virtual Driver**
- **Virtual Driver**
- **Virtual Driver**
- **Virtual Driver**
- **NVIDIA Driver**

**GUEST OS**

**VIRTUAL MACHINE**

- **vCPU**
- **vMemory**
- **vStorage**
- **vNetwork**
- **GPU**

**ESXi Hypervisor**

**Physical Server**

- **CPU**
- **Memory**
- **Storage**
- **Network**
- **GPU**

**NVIDIA**

- **Pass-Through**
- **Accelerated Capture**

**GRID K1**
**GRID K2**
**Quadro 2000-6000**
**Quadro K5000**

**Client**

**NVIDIA Driver**

**Physical Server**

**Virtual Driver**

**vCPU**

**vMemory**

**vStorage**

**vNetwork**

**GPU**
VMware ESXi 5.5 and vSGA

- GPU is gets “shimmed” to the VMs
- Intercept model
- VMware software adapter installed on guests
- The NVIDIA GRID *.vib driver installed on host
- API Support up to DirectX 9, OpenGL 2.1
- Support for vMotion
GPU Virtualization - vSGA

vSphere ESXi Hypervisor

Virtual Machine

- vCPU
- vMemory
- vStorage
- vNetwork
- GPU

Physical Server

- CPU
- Memory
- Storage
- Network
- GPU

Software

- Virtual Driver
- Virtual Driver
- Virtual Driver
- Virtual Driver
- Virtual Driver

Hardware

- NVIDIA GRID software from NVIDIA

View Guest

- VMware Software Adapter
- VMware ESXi

Virtualization

- NVIDIA VIB Driver

App

NVIDIA GRID K1, K2
Citrix XenServer and XenDesktop

- This is the currently available vGPU
- Supports all versions of DX, OpenGL
- NVIDIA Win7 driver is installed on the guests
- Driver manager needed in the host
- Citrix also supports vDGA - pass-thru
- VMware is releasing vGPU later this year
vDGA - Install
Installation / setup - prerequisites

Prerequisites:

- Server platform - Check the HCL!
  - Power, enablement kit
- GRID K1/K2 card(s)
  - Double wide, full sized, PCIe x16
- NVIDIA GRID Driver
- VMware ESXi vSphere 5.5 recommended, 5.1+ supported
- VMware Horizon View 5.3fp1/6.0
Getting started, getting help?

Demo’ing
- Solutions Centers
- POC’s, Pilots, and Trials

Getting help
- Certified server OEM
- VMware SEs
- NVIDIA Product Specialists and Solution Architects

Step by step
Installing vDGA

- Ensure BIOS is configured for virtualization
  - Intel Virtual Technology - enabled
  - Confirm via: `# esxcfg-module -l | grep vtddmar`
  - Hyper-threading - enabled
  - Above 4G Decoding - disabled

- Install and configure vSphere ESXi 5.5

- Prepare environment:
  - Active Directory
  - DNS/DHCP
  - vCenter
  - View Connection Manager
  - vDGA guest
  - vSGA guest
Installing vDGA

On ESXi host:

- Confirm available displays
  - At command prompt:
    - `# lspci | grep -i display`
    - VMware recommends a second inexpensive display for console
    - If only one display ESXi takes ownership for console access

- Confirm graphics card is installed and correct
  - At command prompt:
    - `# esxcli hardware pci list -c 0x0300 -m 0xff`
    - Result should identify the appropriate card
Installing vDGA - Enable Passthrough

- In vSphere Web Client, in “Hosts and Clusters”, right click on the host, select “Settings”
- In “Settings” section, select “PCI Devices” under “Hardware” then click “Edit”
- Select one GRID K2 GPU in “Edit PCI Device Availability” window -> OK
- Click “Refresh”, then click “Reboot This Host”, “Reboot Host” window pop -> OK
- Close vSphere web client, after host reboot, launch and log in vSphere web client
Installing vDGA - Create a VM

- In “Hosts and Clusters” section, right click on host and select “New Virtual Machine”
  1a Select a creation type, select “Create a new virtual machine” -> Next
  2a Select a name and folder, give VM a name -> Next
  2b Select a compute resource, keep default -> Next
  2c Select storage, keep default -> Next
  2d Select compatibility, select “ESXi 5.5 and later” -> Next
  2e Select a guest OS, select “Windows” and “Microsoft Windows 7 (64-bit)” -> Next
- Move to “2f Customize hardware” to finish hardware settings.
Installing vDGA - 2f Customize Hardware

2f Customize hardware:
In “CPU” drop down option, define vCPUs,
E.g. give 4 CPU cores and 2 cores per socket
In “Memory” option, give enough RAM
In “Hard Disk 1” option, give enough storage
In “New CD/DVD Drive” drop down menu, select “Datastore ISO File”. “Select File” window pop up, click “datastore 1” and select ISO file -> OK
Check “Connect…” box -> Next
In “New device” drop down menu, select “PCI Device” then click “Add”
In “New PCI device” option, select one GPU then click “Reserve all memory”
“Reserve all guest memory (All locked)” box under “Memory” option should be checked automatically -> OK
Ready to complete -> Finish, run OS install and create VM
Installing vDGA - Finish VM

After OS finishes install...
- Install VMware tools, reboot
- Join domain if appropriate, reboot
- Install VMware View Agent, reboot
  - View Agent will enable RDP access during install if disabled

Note! RDP leverages the WDDM graphics driver, limited to DX, no OpenGL, so ultimately no access to NV Control Panel, etc. But useful for initial install steps.
Installing vDGA - Install NVIDIA Driver

- Log in as administrator
- Download the latest WHQL driver from www.nvidia.com
- Install driver, select “Custom (Advanced)” then check “Perform a clean installation”, finish installation then reboot

Tip: Before and after install NVIDIA driver
Installing vDGA - Horizon View

Using Horizon View to manage VDI:
- Add vDGA VM to pool
- Add entitlements
- Launch VMware View Client
- Right click on server and select connect
- Click “Continue” if security warning Log in as domain administrator
- Click “Settings” in upper right corner
- Select vDGA VM, set Connect Via: “PCoIP (Default)” and Display: “Fullscreen” -> OK
- Right click on VDI and select “Connect”

NOTE! Its critical to log in initially in Full Screen to enable NVIDIA driver!
Adding vDGA to existing VMs

- A VM must be updated to hardware version 9 or greater
  - Ensures maximum compatibility
  - Right click on VM in vSphere Web Client, choose “Upgrade Virtual Hardware”
  - ESXi 5.5 will upgrade the VM to version 10
For high end graphical and 3D workloads:
- Use high-performance thin client
  - Tera1 chip supports up to 30fps vs. Tera2 chip supports up to 60fps
- Or use a modern Core i3 or better Windows PC

Networking is critical, some PCoIP use cases spike to 70Mbps
- Optimize PCoIP to allow max 120fps, 30fps is the default
- Enable “Disable Build-To-Lossless to reduce overall PCoIP traffic
- Under certain extreme cases with CAD apps, change MaxAppFrameRate to “0”
  - HKLM\Software\VMware, Inc.\VMware SVGA DevTap\MaxAppFrameRate
  - Set to: dword:00000000 (default is 30)

Storage and data access
- Proximity to data, VLANs, IOPS, Solid State!
End-User Experience Tuning

- For VMs with more than 2GB of configured vRAM configure “pciHole.start”
  - Edit the *.vmx file for the VM, add “pciHole.start = “2048” to the bottom

- Enable Relative Mouse
  - On each connection user right clicks Horizon View Client Shade, selects Relative Mouse

- Improve video playback on VM’s with VMXNET3 Ethernet adapters
  - Locate or add HKLM\System\CurrentControlSet\Services\Afd\Parameters
  - Add/Edit:
    - Value Name: FastSendDatagramThreshold
    - Data Type: REG_DWORD
    - Value: 1500
  - Reboot!
POC Lessons Learned - GPU
Scoping Checklist — Success Factors

Establish well-defined scope, realistic goals, and success criteria

Validate delivery and configuration options

Plan for 2-3 weeks
  - First week — basic installation, initial testing and tuning, Google Earth test
  - Second week — end user testing
  - Third week — feedback session, troubleshooting, etc.

Limit to 1-2 use cases or user groups

Limit to 2-4 applications

Build basic Horizon View infrastructure, start simple!
  - Avoids unnecessary complexity and activities
  - Avoids potential infrastructure related performance issues

Vendor Support and Licensing
  - Not all vendors officially support their applications on a virtual environment
Implementation Phases
Implementation — Analysis and Planning Phase

Use Case Analysis: user groups, application sets, workflows, peripherals, network connections, and locations

User Segmentation

- Designers and Engineers — use vDGA
- Editors and Viewers — use vDGA, vSGA
- Knowledge Workers — use vSGA

Capacity Planning and Sizing

- Use POC Results
- Use Lakeside SysTrack
Implementation — Design Phase

Design Focus

- Network
  - type, bandwidth, latency, QoS
- Hypervisor configuration
  - HA options (manual vDGA)
  - Motion (vSGA)
- VM Configuration
  - vCPU & vRAM allocation
- User/VM Density per host
- Storage infrastructure
- K1 vs K2 GPU’s
Implementation — Rollout Phase

Monitor and Tune
- VM Density per Host, decrease or increase number of VMs
- VM Configuration (vCPU allocation, memory allocation, etc.)
- Increase GPU FB, never take it away!
Monitoring
Measuring GPU Utilization

- nvidia-smi.exe
  - Installed with all drivers
  - ESXi version as well (vSGA)

- Process memory usage == vGPU FB size

- Can be output to a file or looped with -l option

- Snapshot in time to see utilization

- Some settings for other cards
  - ECC - doesn’t apply
NVIDIA Control Panel & DirectX Diagnosis

dxdiag
Measuring GPU Utilization

CapsViewer

Free download

http://www.ozone3d.net/gpu_caps_viewer/

Has some demos to check OpenGL functionality

Measuring GPU Utilization
NVIDIA - Global Impact
NVIDIA GLOBAL IMPACT AWARD

- $150,000 annual award
- Categories include: disease research, automotive safety, weather prediction
- Submission deadline: Dec. 12, 2014
- Winner announced at GTC 2015

impact.nvidia.com

Recognizing groundbreaking work with GPUs in tackling key social and humanitarian problems
GTC Express Events
UPCOMING GTC EXPRESS WEBINARS

April 22: Challenges and Advances in Large-scale DFT Calculations on GPUs

May 13: An Overview of AMBER 14 - Creating the World's Fastest Molecular Dynamics Software Package

June 3: The Next Steps for Folding@home

Register at www.gputechconf.com/gtcexpress
GRID Resources

GRID Website
http://www.nvidia.com/vdi

Grid Forums
https://gridforums.nvidia.com

Sign up for the monthly GRID VDI Newsletter
http://tinyurl.com/gridinfo

GRID YouTube Channel
http://tinyurl.com/gridvideos

Questions? Email us
GRIDteam@nvidia.com

Follow us on Twitter
@NVIDIAGRID