ArcGIS Pro: Virtualizing in Citrix XenApp and XenDesktop

Emily Apsey – Performance Engineer
Presentation Overview

• What it takes to successfully virtualize ArcGIS Pro in Citrix XenApp and XenDesktop
  - Shareable GPU, hardware and software configuration and areas of optimization
• Esri Performance Testing initiatives
  - Performance and VM/GPU density
What and Why?

- **ArcGIS Pro**
  - New rendering engine
    - DirectX11 and OpenGL
  - Virtualizes differently than ArcMap
    - Based on GDI+
    - Heavily used in Citrix XenApp
  - Must virtualize well
    - Provide great UX
ArcGIS Pro Add-In

- Specifically designed for performance
- Scriptable and reproducible
- Self contained
- Packaging data and test package to business partners
- Great for Demos!
Testing Goals

• Show if a shareable GPU is needed (ArcMap + Pro)
  - ArcMap – single threaded, GDI+ rendering
  - Pro – multi threaded, DX or OpenGL rendering

• GPU needed for both 2D and 3D workflows?
  - How to monitor and communicate performance/ scalability

• Optimal VM configurations
  - vCPU, Memory, GPU Profile for optimal user experience
  - What and how to communicate information to users, IT administrators
  - How many VMs can a single GRID card accommodate?

• Administration
  - Know a little before you go on-site or create a demo of ArcGIS Desktop in virtualized environments
Configuration of Virtualization Hosts

- Adequate processors and shareable GPUs on the host
- Physical processors with enough cores to support VMs for modern multithreaded applications
  - Two Intel E5-2695 V2 processors. Each processor has 12 cores
- GPUs designed for virtualized environments such as those on the Nvidia GRID K2 cards

- Esri Performance Engineering Lab:
  - Dell R720 with 2 Intel Xeon E5-2695 V2@ 2.40GHz processors, 200Gb memory, 4TB storage, two Nvidia GRID K2 cards.

ArcGIS Pro uses both CPU and GPU

- GPU aids rendering performance by relieving CPU resources

![CPU Utilization graph](image1)

![GPU Utilization graph](image2)

*Scalability testing graphs using vGPU.

- If no GPU is present, all rendering commands are processed on CPU.

Virtualizing in Citrix XenApp & XenDesktop
Citrix Virtualization Solutions

• 7.x
  - Merged XenApp and XenDesktop technologies
  - Primarily uses MCS; can optionally use PVS or use pre-created VM’s
    - Machine Catalogs
      - Ability to create multiple VM’s based on Template
    - Delivery Groups
      - Defines Application or Desktop publication
      - Assigns users to VM’s

• Historically been separate products
  - Gives administrators the ability to admin both XA/XD via Citrix Studio
    - Licensing is still separate

  - [http://support.citrix.com/content/dam/supportWS/kA560000000TNDvCAO/XD_XA7.x_LicensingFAQ.pdf](http://support.citrix.com/content/dam/supportWS/kA560000000TNDvCAO/XD_XA7.x_LicensingFAQ.pdf)
VM Configuration - GPU

- VM’s can be configured to use GPU two ways:
  - Shareable (vGPU)
    - Many profiles to choose from
  - Passthrough (Dedicated)

- XenDesktop
  - Can be configured either Shareable or Passthrough

- XenApp
  - Passthrough (OS handles time slicing GPU across multiple terminal sessions)
  - Shareable is not officially supported by Citrix
Nvidia GRID K1 vs. K2 cards

• **K1**
  - GPU’s are equivalent to Quadro K600
    - Entry Level GPU’s
    - 4 K600 GPUs on board and 4GB memory per GPU

• **K2**
  - GPU’s are equivalent to Quadro K5000
    - High-end analyst who currently uses workstation w/ GPU
    - 2 K5000 GPUs on the board and 4GB memory per GPU

*K2 scales better than K1*

  Big difference is # of CUDA-cores on each GPU to process rendering task

  K1 (K600) has 192 per GPU
  K2 (K5000) has 1536 per GPU

Virtualizing in Citrix XenApp & XenDesktop
vGPU Technology

• Pass-through
  - Graphics commands of each virtual machine passed directly to the GPU
    - No translation by the hypervisor

• vGPU Manager (Hypervisor)
  - Assign the optimal amount of graphics memory to VM
  - Every virtual desktop has dedicated graphics memory
  - NVIDIA GRID K2 card (recommended): 2 GPUs, allowing 16 users to share a single card.
  - Controlled by vGPU Profile assigned to VM

• vGPU Profiles (used by VM’s)
  - Citrix XenDesktop
  - VMWare Horizon View

- See more at: http://www.nvidia.com/object/virtual-gpus.html#sthash.WN5id3FZ.dpuf
GRID vGPU Profiles

- Nvidia K2 Card: 2GPU’s (each GPU has 4GB of Memory)
Nvidia GRID vGPU Architecture

1. Install vGPU Manager

2. Configure VM to use vGPU

3. Install GRID vGPU drivers in VM
Step 1: vGPU Manager

- The NVIDIA GRID vGPU software package for Citrix XenServer

- Pre-requisites:
  - Citrix XenServer 6.2 SP1 with applicable hotfixes or later.
  - Citrix XenDesktop 7.1 or later

- NVIDIA Virtual GPU Manager runs in XenServer’s dom0.
  - Provided as an RPM
  - Copy to XenServer’s dom0 and then installed.

- Reboot Server

- Verify Installation
  - `lsmod | grep nvidia`
  - `nvidia-smi`
Step 2: VM Configuration (workflow for XD and XA)

1. Create VM
   - Assign Resources; CPU, Memory

2. Install OS
   - Install XenTools, Join Domain, Install VDA (Virtual Delivery Agent)

3. Install ArcGIS Pro

4. Create MCS Catalog
   - Hosting Connection / Resources must match currently installed vGPU/Passthrough GPU in the Template you are using.

5. Create delivery Group.

6. Shutdown VM, Assign GPU resources.
   - Virtual GPUs resident on a single physical GPU must be all of the same type

7. Power On VM, install Nvidia Driver
Step 2: VM Configuration (continued)

8. Verify GPU
   - Windows device manager:
     - Nvidia Control Panel
       - Right click on desktop.
   - Run Nvidia-smi on host
     - GPU Utilization is outputted.
     - Can also output log.
XenDesktop
Test Results
Virtualizing in Citrix XenApp & XenDesktop

3D Scalability Testing (240Q)
Host Avg CPU % Utilization

16VMS AVG CPU UTIL 75%
3D: Avg Draw Time and Min FPS

- **SumAvg**:
  - 1VM: 80 seconds
  - 4VM: 80 seconds
  - 8VM: 80 seconds
  - 12VM: 80 seconds
  - 16VM: 123 seconds

- **FPSMin**:
  - 1VM: 15 frames per second
  - 4VM: 15 frames per second
  - 8VM: 15 frames per second
  - 12VM: 15 frames per second
  - 16VM: 10 frames per second
Virtualizing in Citrix XenApp & XenDesktop

2D Scalability Testing (240Q)
Host Avg GPU % Utilization

- 4VMs:AvgGPU%Util
- 8VMs:AvgGPU%Util
- 12VMs:AvgGPU%Util
- 16VMs:AvgGPU%Util

16 VMS
AVG GPU
UTIL
28%
XenApp

Test Results
Nvidia GRID GPU Pass-Through Architecture

- Dedicates a GPU to a virtual machine
  - Full 4GB of GPU Memory (frame buffer)
- Recommended for Citrix XenApp
  - Currently not suitable for ArcGIS Pro
    - Heavily utilizes GPU Memory
    - OS does not split load evenly across multiple concurrent sessions
      - 1st one in takes as much as it needs
      - Not given back until session ends
ArcGIS Pro – XenApp 7.6

Requires GPU Pass-Through
- K2 GPU Memory = 4GB

OS does not do well splitting GPU resources across different terminal sessions

XenDesktop recommended
- Dedicated resources
- vGPU profile manages FB
- Dictates VM density

Esri collaborating with Nvidia and Citrix
VM Configuration Recommendations
Test Results: vCPU Recommendation

- Multiple threaded nature of ArcGIS Pro
  - Direct correlation between vCPU and amount of threads
  - Analysis of Total Task Time and Hung Time

- 6vCPU for most 3D and 2D data and workflows
- 4vCPU may be suitable for simple 2D data and workflows
Test Results: vGPU Recommendation

- **K280Q (4GB of FB, i.e. GPU Memory)**
  - Highest performing, lowest density
  - Lacked scalability, only 4VM’s on two K2 cards
  - Each gets an entire GPU

- **K260Q (2GB of FB)**
  - Great performance and UX experience, higher density
  - Scalability = 8VM’s; CPU/GPU Host Utilization had room for growth
  - Test results were within acceptable threshold

- **K240Q (1GB of FB)**
  - Good performance and UX experience.
  - Scalability = 16 VM’s

- **K220Q (512MB of FB)**
  - Least performing, highest density
  - 3D UX poor/performance slow; too little virtual RAM (VRAM).
  - May be suitable for simple 2D data and workflows

---

Recommended Profile

Virtualizing in Citrix XenApp & XenDesktop
Answers obtained from testing

- **ArcMap**
  - Rendering engine does not utilize GPU
    - GPU is not heavily utilized
      - Only when user explicitly executes hardware acceleration operations
Future testing

• Mixed workflows
  - Initial testing focused on rendering pipeline
  - Add analytical operations (CPU intensive)
    - How GPU Utilization and density is affected when users are simultaneously running a CPU intensive geoprocessing task?

• Mixed Virtualization Vendor Solutions
  - VMWare ESXi XenDesktop

• Storage – SAN
  - More realistic – large infrastructure don’t use SSDs

• Cloud based VDI (DaaS)
  - Increasing questions and interest
  - Distributors – Europe, NZ, US
Questions?

- Visit us at the Expo Hall
  - Desktop Island
  - ArcGIS Pro: virtualizing in VMWare Horizon View and Microsoft Hyper-V (Room 02)
    - Noon Tomorrow
    - eapsey@esri.com
Understanding our world.