Delivering Real World 3D Applications with VMware Horizon, Blast Extreme and NVIDIA Grid

Sebastian Brand – Lead Systems Engineer EUC at VMware
Luke Wignall – Sr. Manager, Performance Engineering at NVIDIA
Disclaimer

• This presentation may contain product features that are currently under development.
• This overview of new technology represents no commitment from VMware to deliver these features in any generally available product.
• Features are subject to change, and must not be included in contracts, purchase orders, or sales agreements of any kind.
• Technical feasibility and market demand will affect final delivery.
• Pricing and packaging for any new technologies or features discussed or presented have not been determined.
## Agenda

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Understanding Graphics Use Cases</td>
</tr>
<tr>
<td>2</td>
<td>What’s new with VMware Horizon and NVIDIA GRID.</td>
</tr>
<tr>
<td>3</td>
<td>Sizing Your Virtual Desktop for High Performance Graphics</td>
</tr>
<tr>
<td>4</td>
<td>Resources</td>
</tr>
</tbody>
</table>
Understanding Graphics Use Cases
Virtualization Across the Enterprise
NVIDIA and VMware drive additional benefits to both user and IT

User benefits
- Increased performance
- Mobility and flexibility
- Improved productivity

IT benefits
- Centralized management
- Simplified support
- Data security
Challenges for Virtual Graphics

Professional graphics workloads require great user experience

**UX**
- Require “snappy” experience

**Chip**
- Rely on heavy encoding and decoding

**People**
- User density is limited by CPU bottleneck
Graphics Accelerated Digital 3D Workspace
VMware & NVIDIA bring the power of GPUs to virtual workloads

Physical devices have graphics power that drive user experience

Graphics acceleration allows you to deliver virtual workflows with native performance
Evolution of Graphics in Windows OS

Today's operating systems and applications are optimized for GPU. Win10 has 32% more graphics consumption compared to Win7.
## Graphics Consumption of Windows Applications

Applications that are accelerated by graphics has doubled since 2011

<table>
<thead>
<tr>
<th>Application</th>
<th>Windows 7</th>
<th>Windows 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrome</td>
<td>36%</td>
<td>59%</td>
</tr>
<tr>
<td>Excel</td>
<td>53%</td>
<td>64%</td>
</tr>
<tr>
<td>Firefox</td>
<td>59%</td>
<td>85%</td>
</tr>
<tr>
<td>PowerPoint</td>
<td>64%</td>
<td>409%</td>
</tr>
<tr>
<td>Outlook</td>
<td>85%</td>
<td></td>
</tr>
</tbody>
</table>

*Percent of time consuming GPU comparing Windows 7 to Windows 10 (DirectX or OpenGL)*
QUADRO POWERING THE MOST ADVANCED WORKFLOWS

Photorealistic. Immersive. Collaborative.

Media & Entertainment  Manufacturing & Design  AEC  Oil & Gas  Visual Simulation
Enterprise transformation has begun

**VISUAL WORKSPACE**

**MOBILITY**

**COLLABORATION**

**LARGE DATA INTERACTIVE HPC**

**VR**

**PHOTOREALISM**

**AI**

**VISUAL COMPUTING SPECTRUM**

Information Workers/Students

Designers/Scientists
What’s new with VMware Horizon and NVIDIA Grid
Growth in NVIDIA and VMware Partnership

- **2014**: vGPU integration for vSphere and Horizon
- **2015**: NVIDIA announces NVIDIA GRID software
- **2016**: VMware launches Blast Extreme with NVENC
- **2016**: Highest density vGPU solution with up to 128 users
- **2017**: Instant clones supported for vGPU
- **2017**: High availability for vGPU supported
Growth in NVIDIA and VMware Partnership

- **2014**: vGPU support for vSphere and Horizon
- **2015**: NVIDIA announces NVIDIA GRID software
- **2016**: VMware launches Blast Extreme with NVENC
  - Highest density vGPU solution with up to 128 users
- **2017**: Instant clones supported for vGPU
  - High availability for vGPU supported

Drumroll please…
What’s New with NVIDIA GRID and VMware Horizon
Advancing innovation for high performance graphics accelerated virtual desktops and applications

GRID
August 2017 Release

Tech Preview: Session Collaboration

Tech Preview: Suspend & Resume, vMotion for vGPU VM’s

Tech Preview: Suspend & Resume for vGPU VM’s

Blast Extreme Enhancements

vGPU Insights Integrated in vROPs
WORLDS MOST POWERFUL VIRTUAL WORKSTATION

NVIDIA Quadro vDWS Software
For Tesla GPU Servers

Tesla Pascal support
CUDA for compute accelerated workstation applications
Hardware encode (NVENC) accelerated Linux workstations
Certified on 120+ servers from 30+ server vendors
NVIDIA GRID GPU VIRTUALIZATION PLATFORM

Industry Leading Virtualization Platform

GRID
- vApps
- vPC

Quadro vDWS
- Design
- Rendering
- HPC/AI

vROPs: vGPU MONITORING, INSIGHT, & MANAGEMENT
DATA CENTER and/or CLOUD ACCESSIBLE

VMware vSphere + Horizon

NVIDIA VIRTUALIZATION SOFTWARE
- GPU Sharing
- GPU QoS

NVIDIA TESLA GPUS
- MAXWELL: M60, M6, M10 (graphics sharing only)
- PASCAL: P40, P6, P4, P100 (graphics & compute sharing)
NVIDIA GRID August 2017 Release

• Feature/Capability Overview

Industry leading user experience
• Entire Pascal GPU Lineup
  • Up to 2x Graphics Perf (pending)
  • Larger profiles (3, 6, 12, 24 GB)
• Guaranteed QoS – Fair share scheduler
• Compute (CUDA) Accelerated Graphics Apps

Cost effective large scale deployment
• Up to 24 vGPU per GPU
  • 50% increase over today
• New P6 – 16GB FB
• 24:1 Graphics/CUDA VMs
• 24:1 Linux NVENC support

Enterprise-wide GPU Value
• VDI During the Day
• Run HPC, AI at Night

End-to-end management and insight
• End-user and help desk insight
• Application level monitoring
• vROPs Integration
• In-branch driver interoperability
• Highly available license server
Improved Performance for High Performance Applications

Tesla P40 delivers up to 2x the performance of Tesla M60

GPU Throughput of a P40 Compared to a Single M60 GPU. The maximum throughput per GPU compares the overall performance of a GPU that can be shared across multiple virtual machines. The score differs to a single SPEC ViewPerf 12.1 score because the GPU is only consistently and fully utilized with multiple virtual machines.
Choosing the Right GPU for Your Workload

Density and performance requirements will determine hardware requirements.
Blast Extreme Enhancements

Experience
Blast Extreme Adaptive Transport Makes the Workspace Better

- 6X faster file transfers across WAN
- 50% lower bandwidth utilization
- 13x higher framerate under extreme network conditions
Deliver Secure, Immersive 3D Workflows from the Cloud

- Advanced 3D experience with NVIDIA GRID
- Virtual workflows delivered from the data center
- Scale on demand for great cost flexibility
NVIDIA Blast Extreme Acceleration

Increased scalability and performance for graphics workloads

Up to 51 ms latency reduction

Up to 89% bandwidth reduction

Up to 18% increase in density

Source: NVIDIA GRID Performance Engineering Lab
Maintain a Great User Experience Across

- WAN/Cloud
- Corporate LAN
- Public Wi-Fi
- Mobile networks

Dynamically Adjust to:

- Varying speeds
- Latency
- Severe packet loss

Cater to different workloads

- Virtual Apps
- Virtual Desktops
- Virtual Workstation

Perform well across all client (any device) and server Platforms

Attend #ADV1609BU for details
Accelerating Linux 3D Applications with NVIDIA GRID

Linux 3D application acceleration

- NVENC support for 3D Graphics applications
- Leverages Blast Extreme Adaptive Transport
- Supported with Tesla M60/M6/M10 cards
vGPU Insights integrated with vROps
NVIDIA GRID Support with vRealize Operations for Horizon

Entire Stack Monitoring
Insights into Users, Apps and Infrastructure

Single Pane of Glass
Monitor both Horizon and XenApp Stacks

Right-Sized Resources
Utilization Metrics and Management

End-User Viewpoints
Optimize Performance and Meet SLAs

GPU Support
Monitor GPU status
v4H & PA - User Experience Dashboard: Nvidia GRID vGPU – 3D & Memory Utilization

- User Experience Dashboard “weather map” proactively offers insight into key user experience metrics:
  - vCPU, vDisk, vRAM and sessions
  - Nvidia vGPU 3D and memory utilization (new!)
  - Easy to understand object relationship topology view
  - Charts and spark lines for trending insight
SIZING FOR BEST USER EXPERIENCE
Example: Tesla M10 FB, GPU, and Encode Engine for Knowledge Worker

- 1 GB GRID vGPU Framebuffer Utilization
- Tesla M10 GPU Utilization for 8 VMs
- Tesla M10 Encoder Utilization for 8 VMs

Cirrus Knowledge Worker Workload (Excel, Word, PowerPoint, Chrome, Media Player, PDF)
Tech Previews
Tech Preview – Blast Extreme Session Collaboration.

Session Collaboration

- Tech preview only
- Simplifies development workflows
- Requires Blast Extreme protocol
- Multicast any desktop to many simultaneous users
- Collaboration UI on VDI desktop to initiate collaboration sessions
Considered Milestones for VMware vSphere with NVIDIA GRID

- GRID vPC (VDI)
- Quadro vDWS (VDI)
- High Performance Computing
- Machine Learning

Suspend & Resume
Tech Preview
See @booths
VMW Cloud Platform - New Workloads
VMW EUC 3D Experience
NVIDIA GRID

Snapshots
Roadmap
See @booths
VMW Cloud Platform - New Workloads
VMW EUC 3D Experience
NVIDIA GRID

vSphere vMotion
Roadmap

vSphere DRS
Roadmap

The information in this presentation is intended to outline our general product direction and should not be relied on in making a purchasing decision. It is for informational purposes only and may not be incorporated into any contract.
Tech Preview: Suspend and Resume
Tech Preview: vSphere vMotion with vGPU.
Resources

• NVIDIA GRID & VMware Horizon Deployment Guide

• VMware Horizon with View on Virtual SAN Reference Architecture

• NVIDIA GRID Certified Servers

• ESRI ArcGIS Pro App Guide with VMware Horizon and NVIDIA GRID vGPU
  – http://www.nvidia.com/esriappguide

• Autodesk Revit App Guide with VMware Horizon and NVIDIA GRID vGPU
Thank You