



Graphics and Virtualization

Will Wade

Quadro Advanced Technologies

December, 2011



Why Virtualize with graphics?

Move desktop systems to the “cloud” to:

- **Improve Security – data and capital stay in the data center**
- **Better manageability – resources stay under IT control**
- **Increase Flexibility – work from anywhere, anytime, on any device**
- **Make better use of computing resources**

Move desktop systems to the “cloud” with graphics acceleration:

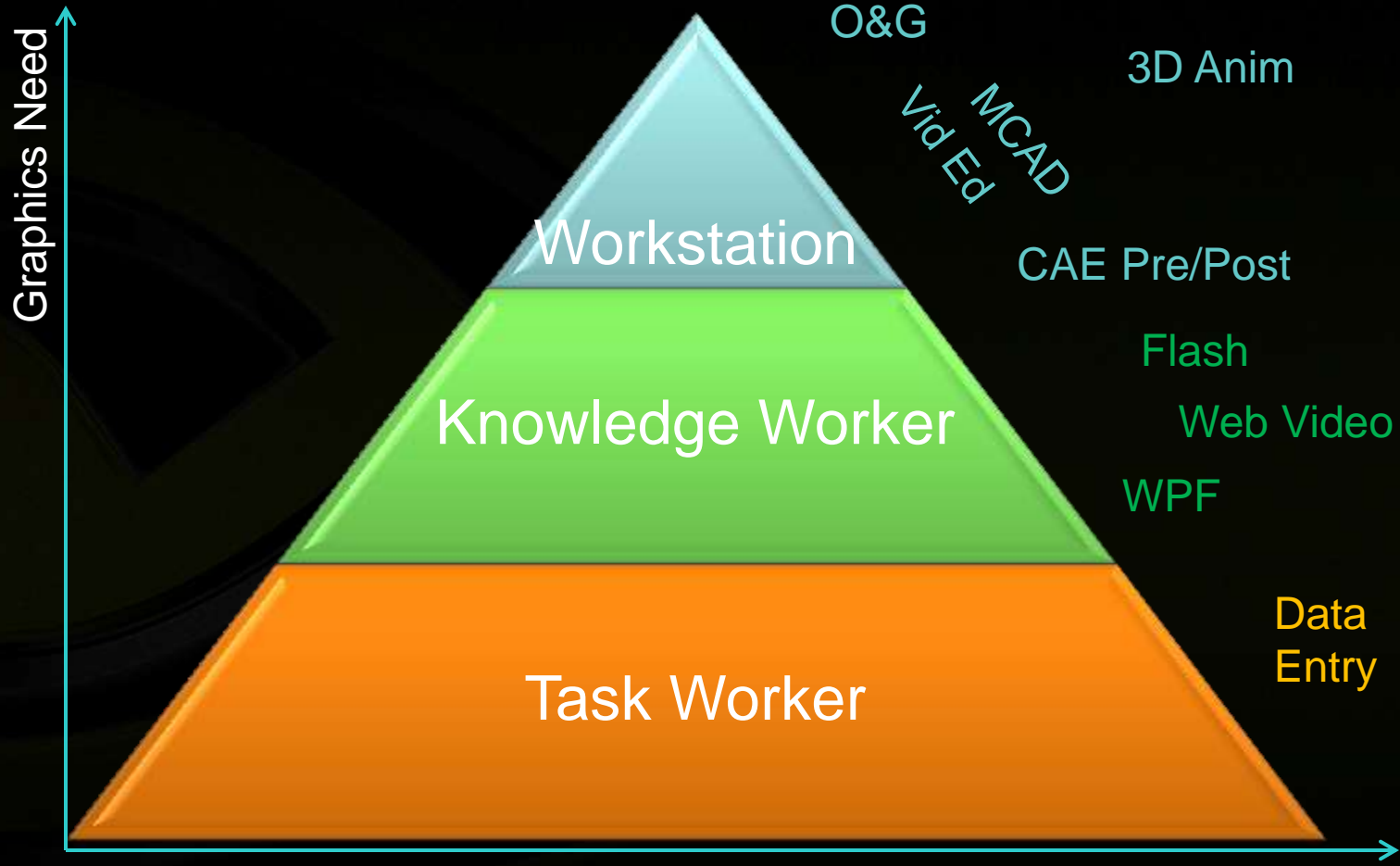
- **Best user experience**
- **Enhanced productivity**

Project “Monterey”



- **Graphics in the cloud toolset**
 - Software - SDK for virtualization, remoting, and data center management
 - Hardware - GPUs for data-center deployment
 - Clients - Tegra devices for client side optimization
- **Target Use Cases**
 - Remote Workstation
 - Graphics Accelerated VDI
 - Cloud Gaming

Desktop Virtualization Spectrum



Desktop Virtualization Spectrum



Graphics Need ↑

Solutions

Workstation

Knowledge Worker

VDI + graphics

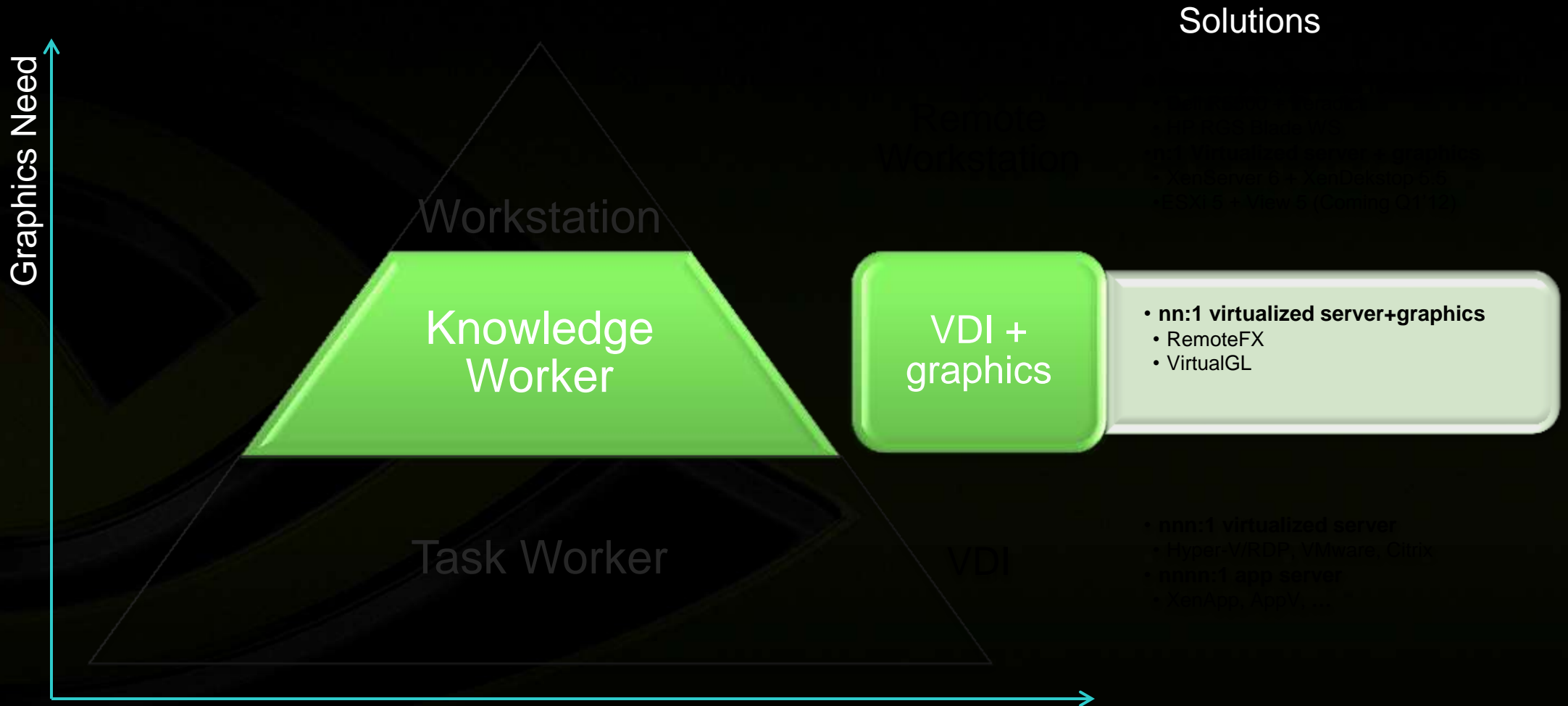
Task Worker

VDI

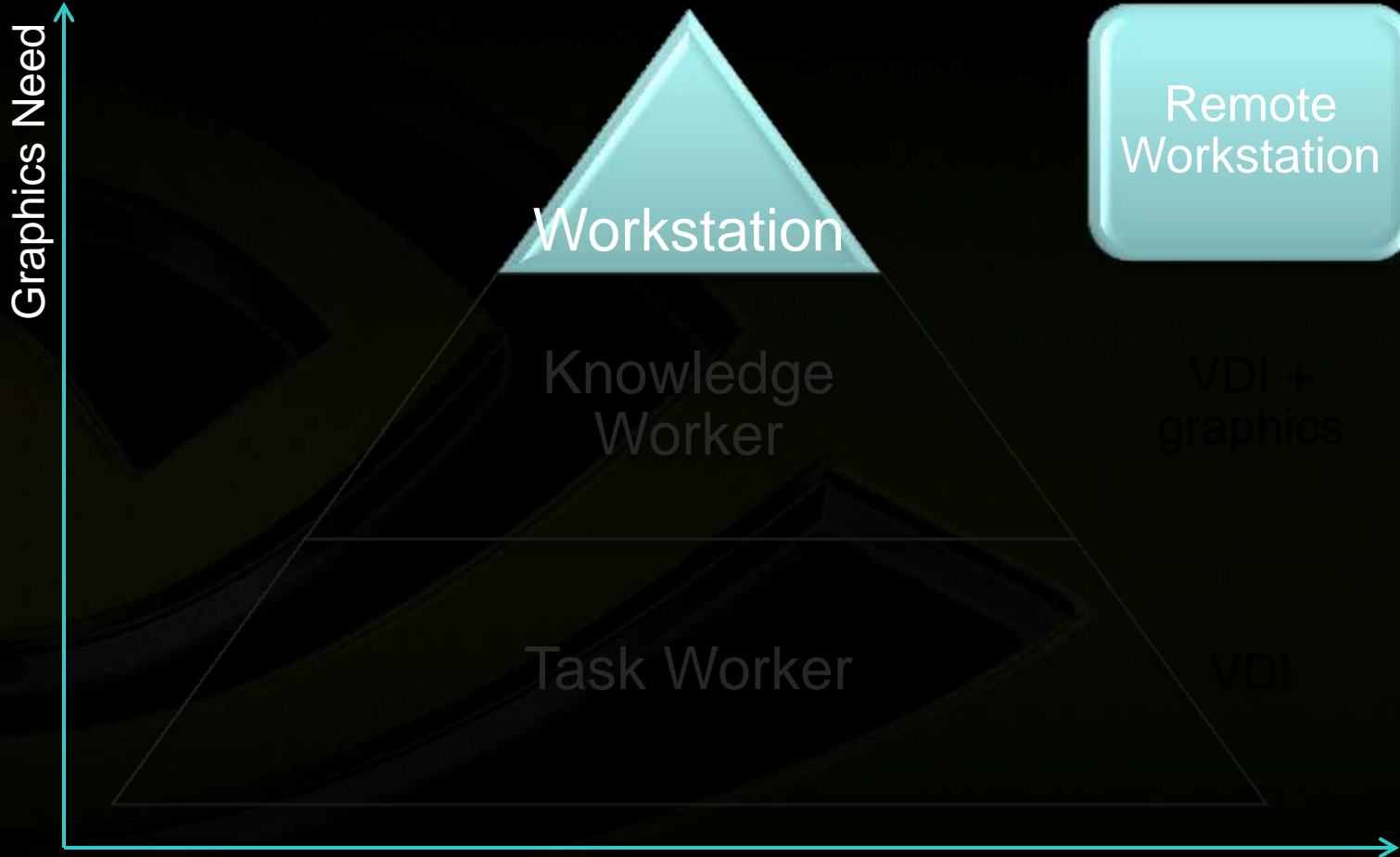
- n:n:1 virtualized server+graphics
- RemoteFX
- VirtualGL

- **n:n:1 virtualized server**
- Hyper-V/RDP, VMware, Citrix
- **n:n:n:1 app server**
- XenApp, AppV, ...

Desktop Virtualization Spectrum



Desktop Virtualization Spectrum



Solutions

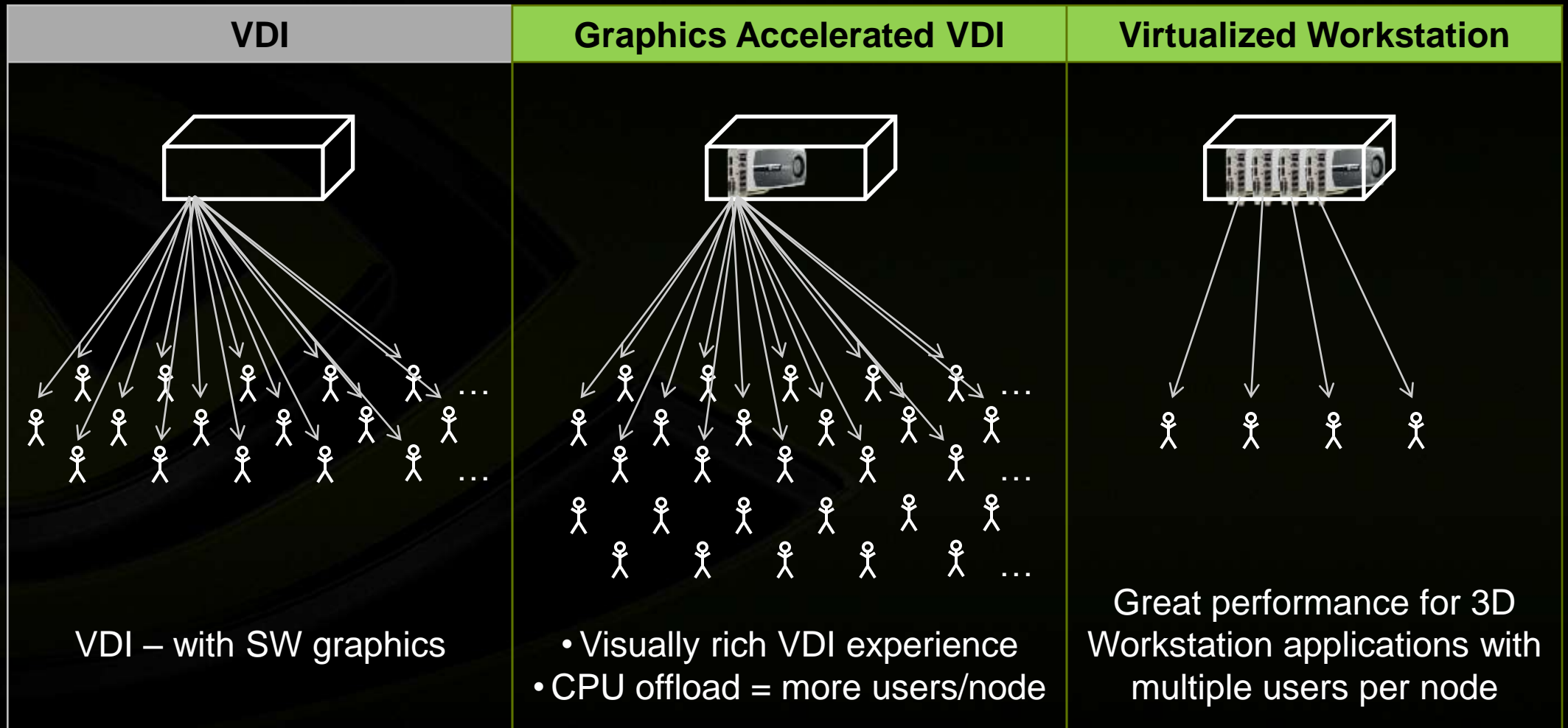
Remote Workstation

- **Remote dedicated workstation**
 - Dell R5500 + Teradici
 - HP Blade WS + RGS
- **n:1 Virtualized server + graphics**
 - XenServer 6 + XenDesktop 5.5
 - ESXi 5 + View 5 (Coming Q1'12)

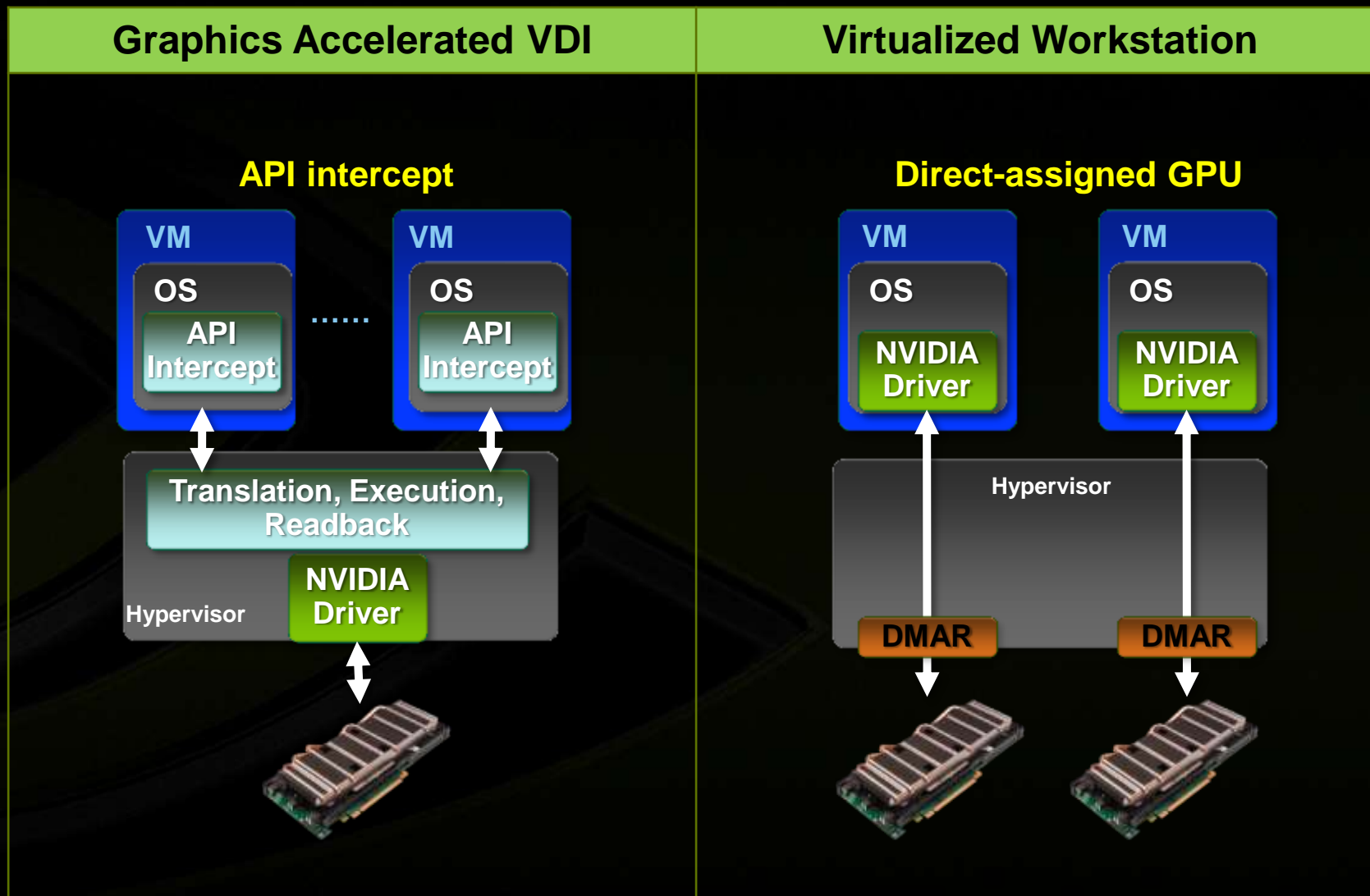
- **n:n:1 virtualized server + graphics**
 - RemoteFX
 - VirtualGL

- **n:n:n:1 virtualized server**
 - Hyper-V/VDI, VMware, Citrix
 - n:n:n:1 app server
 - Citrix AppCenter

NVIDIA Professional Cloud Visualization



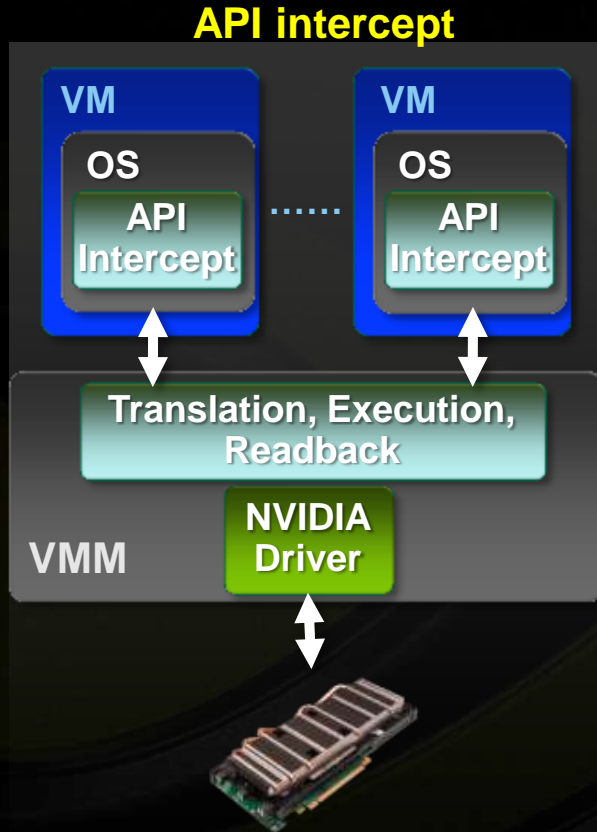
GPU virtualization technologies



Desktop Virtualization Solutions



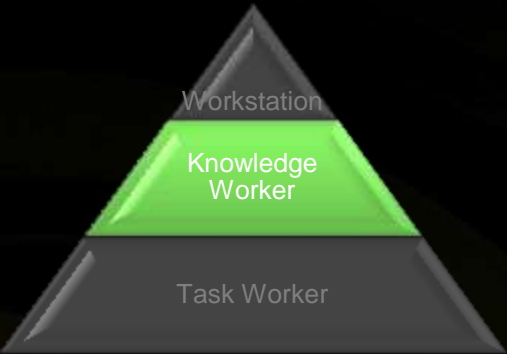
Desktop Virtualization



*e.g. NVIDIA Quadro with
Microsoft RemoteFX*

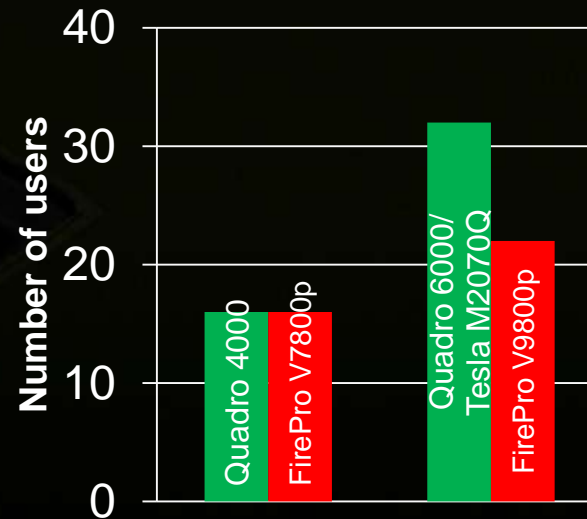
- Multiple VM/OS per GPU – scales up
- Highly portable – no IHV drivers in the VM
- Limited API support – DX9 only, limited OGL
- Moderate performance hit
- Shipping now in desktop-oriented products
 - Microsoft RemoteFX with Quadro and Tesla GPUs
 - Parallels Desktop for Mac, VMware Fusion / Workstation
 - VirtualGL, VirtualBox

NVIDIA and Microsoft RemoteFX



- 1:n GPU Virtualization via API intercept model
- Remoting via RDP 7.1 / RFX
- Limited API support – DX9, OGL 2.1
- Good office perf, not WS class

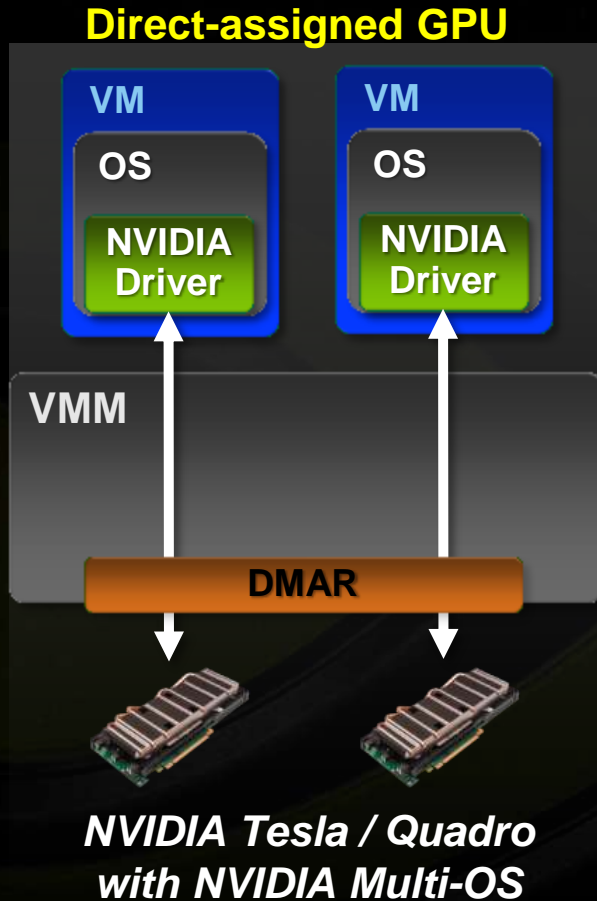
	# Users per card	Street Price	\$ / user
FirePro V7800p	16	\$1,000	\$63
Quadro 4000	16	\$723	\$45
FirePro V9800p	22	\$2,100	\$95
Quadro 6000 / Tesla M2070Q	32	\$3,000	\$94



Workstation Virtualization Solutions

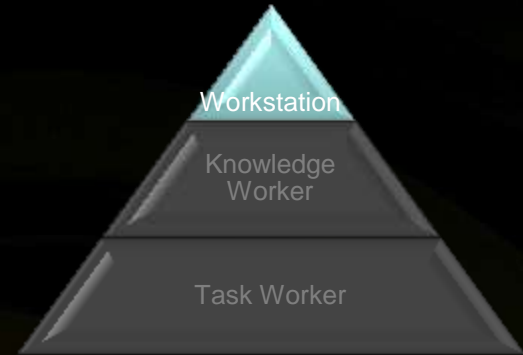


Workstation Virtualization



- One OS per GPU – baremetal performance
- All APIs supported - CUDA, DX, OGL, ...
- Shipping now with Quadro GPUs and Tesla M2070Q
 - Parallels Workstation 6 Extreme
 - Citrix XenServer 6
- Additional solutions in 2012
 - VMware ESXi in Q1'12
 - Xen open-source hypervisor “how to” whitepaper

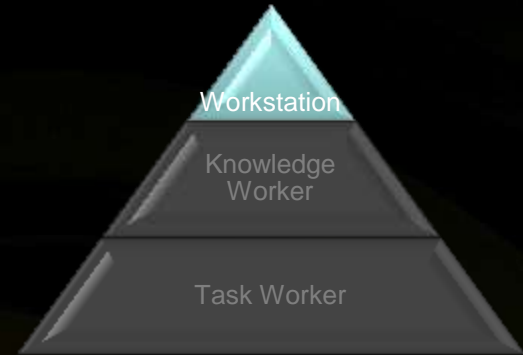
NVIDIA and Citrix



- **1:1 GPU to user**
- **Virtualization**
 - **N-MOS GPU pass through in XenServer 6**
- **Remoting**
 - **Available today in XenDesktop 4.5 HDX 3D Pro**
 - **GPU acceleration for graphics rendering and compression**



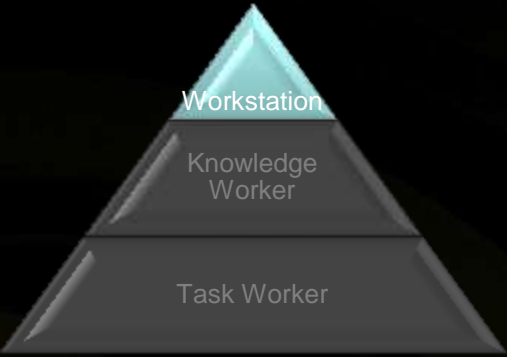
NVIDIA and VMware



- **1:1 GPU to user**
- **Announced support for Project Monterey at VMworld Copenhagen**
 - **N-MOS GPU pass through in ESXi**
 - **Project Monterey support in View**
- **Private beta in Dec '11**
- **Targeting public launch Q1'12**



NVIDIA and Parallels



- **1:1 GPU to user – desktop WS focused**
- **Parallels WS Extreme 6**
 - N-MOS GPU pass through
- **No remoting solution**



Roadmap – GPU enabled Virtualization



NVIDIA Project Monterey for Remoting



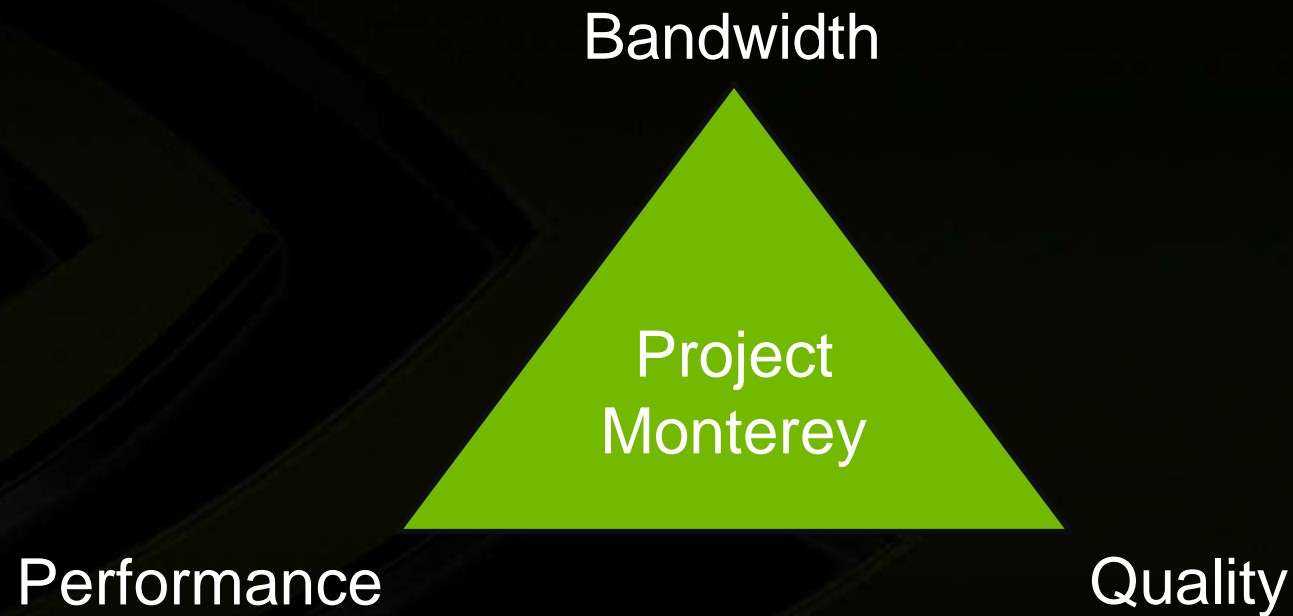
Problem Definition



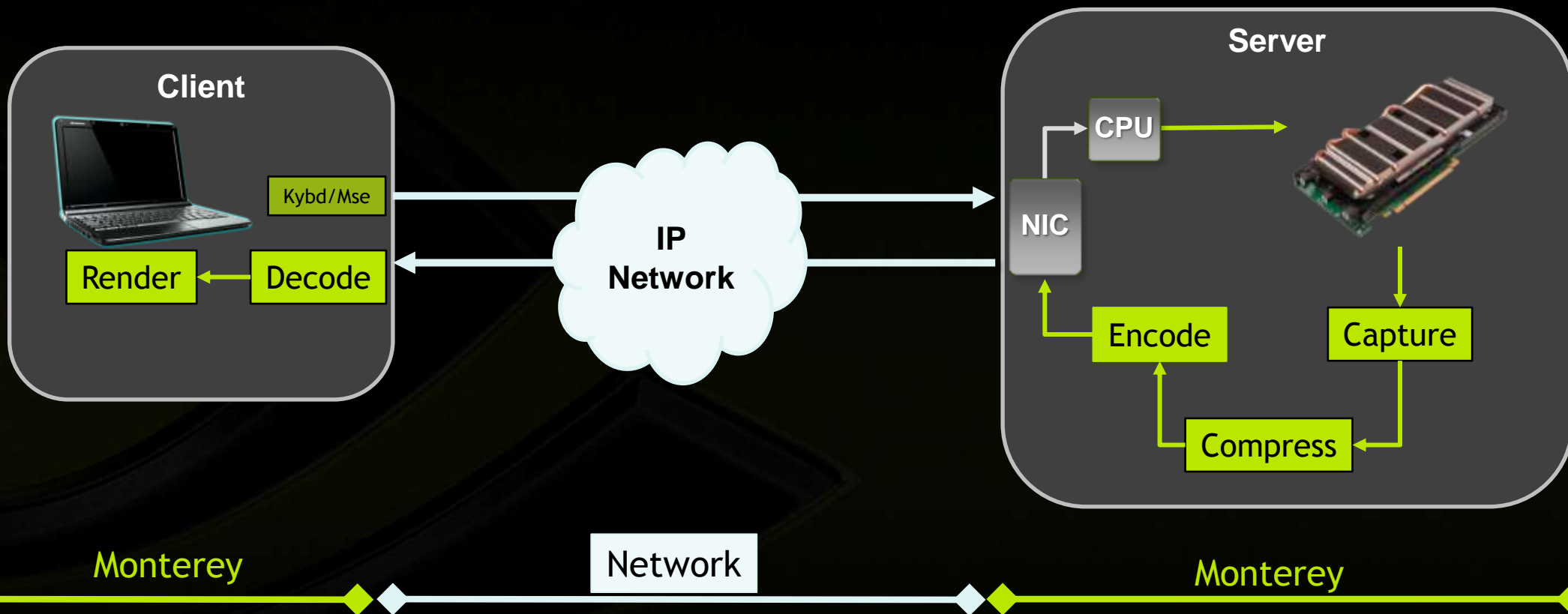
IT organizations are quickly moving to a VDI infrastructure for PC users, but can't for workstation users due to the inability to get "just like local" responsiveness for WS applications.



Remote Graphics Metrics



Remoting Architecture



GPU Technology Conference 2012

May 14-17 | San Jose, CA

The one event you can't afford to miss

- Learn about leading-edge advances in GPU computing
- Explore the research as well as the commercial applications
- Discover advances in computational visualization
- Take a deep dive into parallel programming

Ways to participate

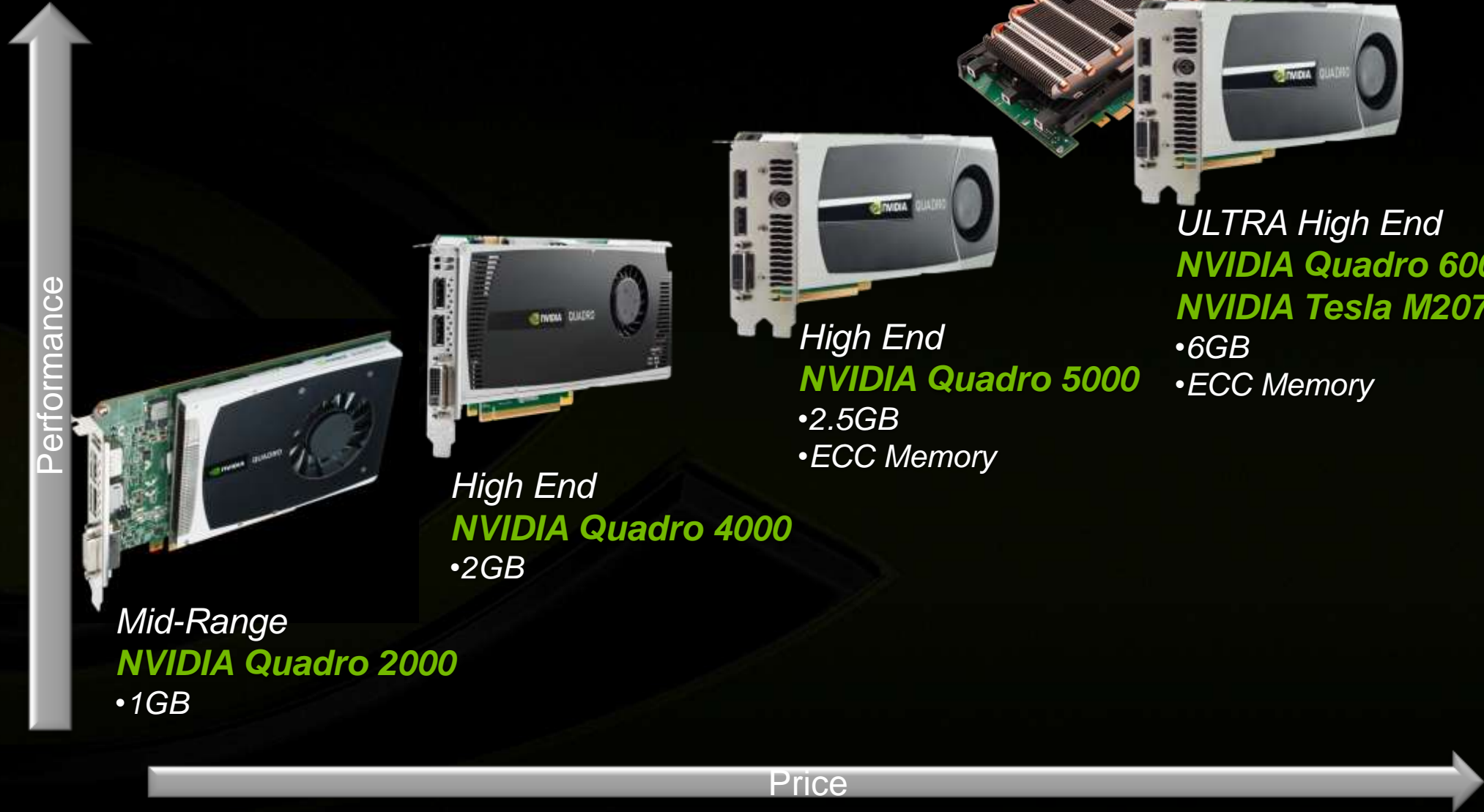
- Speak - share your work and gain exposure as a thought leader
- Register - learn from the experts and network with your peers
- Exhibit/Sponsor - promote your company as a key player in the GPU ecosystem



Server Graphics Solutions



NVIDIA Quadro for Virtualization



Tesla M2070-Q

Quadro for the Data Center



- Quadro 6000 performance in a passively cooled server module
- For integration in OEM servers
- 6 GB frame buffer memory
 - Enables very large data-set visualization or
 - Enables many concurrent users for remoting
- No physical display connectors – remote display only



Quadro 4000



- Single slot width
 - For space constrained servers
 - Higher GPU density in single-wide server layouts
- 142W for efficient cooling
- Sweet spot for
 - Mid-range dedicated graphics in virtualized servers
 - High-density single-wide configs for RemoteFX



Thank You