AS322463: BEST PRACTICES FOR DEPLOYING VIRTUAL WORKSTATIONS FOR ARCHITECTURAL DESIGN FIRMS

Jimmy Rotella – Sr. Solutions Architect, Virtual GPU Solutions
jrotella@nvidia.com
@jrotella

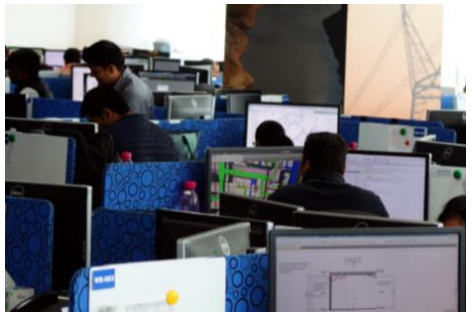
NVIDIA

AGENDA

- Why virtualize?
- Why vGPU?
- What is NVIDIA's vGPU solution?
- How do you size your environment?
- How do you plan and configure your environment?
- What are the first steps to get started?

TRENDS IN AEC

Unifying a Global Workforce



Collaboration

Global and field offices must collaborate on projects; version control is important when several designers work on the same project

IT Efficiency Translates to Revenue

Difficulties managing data and IT delivery can result in lost billable hours for architects and construction rework costs

Mobility

Architects & designers work more efficiently when freed from their desks; firms desire to hire talent globally, versus in the city the project resides

Bolstering Security

Firms hiring remote contractors to work on projects must find ways to keep project files accessible and secure

IMPROVING COLLABORATION & PRODUCTIVITY

WBCM wanted to deliver virtual desktops to users with no compromises in performance or user experience. vGPU enabled:

- *Performance monitoring and better troubleshooting*
- *Better agility and lower capital costs when setting up new offices and users*
- *Remote desktops with no decline in user experience*
- *Improved collaboration for geographically dispersed teams*

NVIDIA vGPU Technology: Tesla M60, Quadro Virtual Data Center Workstation

CONTINUOUS VALUE WITH QUADRO vDWS

Migrating from physical workstations to NVIDIA virtual GPU:

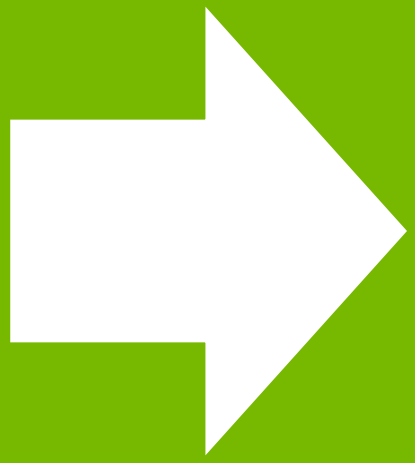
- *Enabled mobility and collaboration*
- *Users were more productive with a consistently great user experience*
- *CUDA enabled virtual workstations provide native Adobe experience*
- *Secured company IP in the data center*
- *Reduced help desk tickets with NVIDIA vGPU management and monitoring functionality*

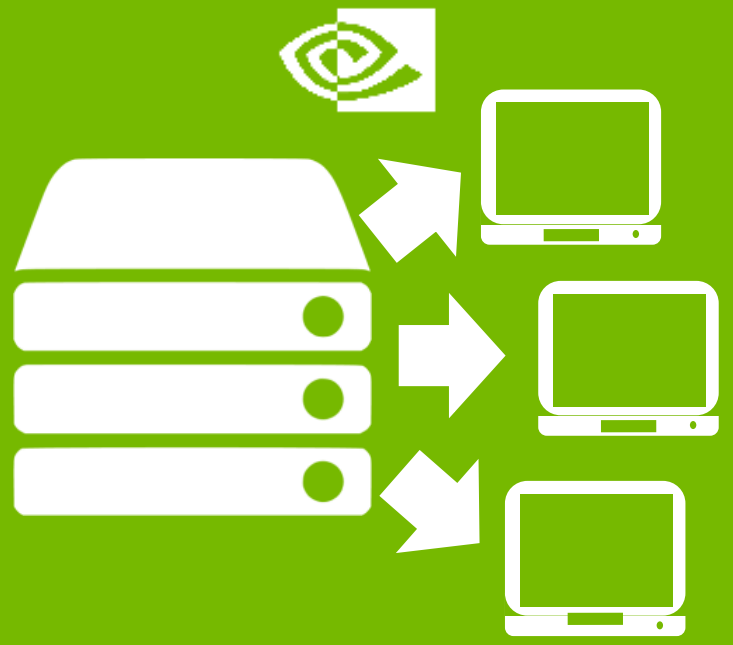
NVIDIA vGPU Technology: Tesla P4, Quadro Virtual Data Center Workstation

PERFORMANCE FROM THE DATA CENTER

NVIDIA virtual GPU technology delivers graphics accelerated virtual desktops and applications

Devices have graphics acceleration





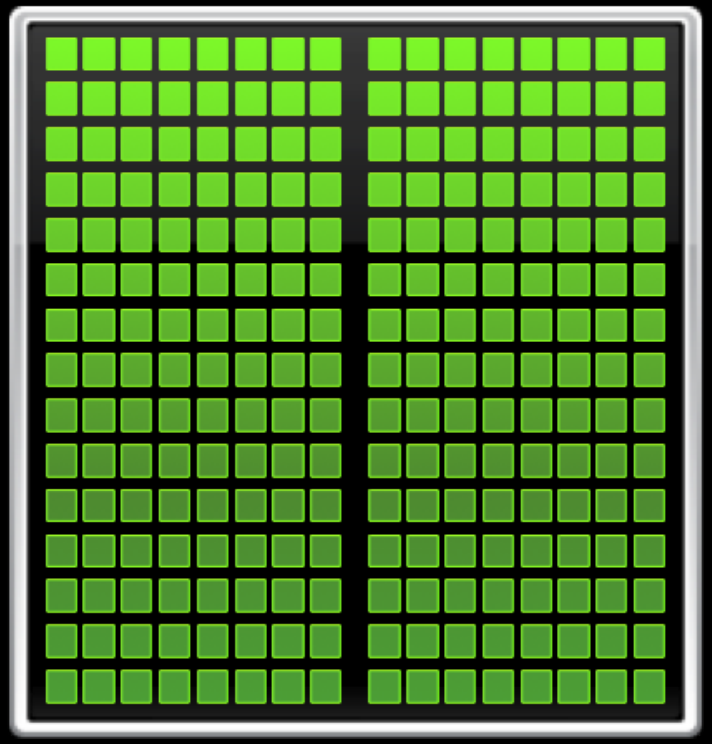
Virtual machines also need a GPU

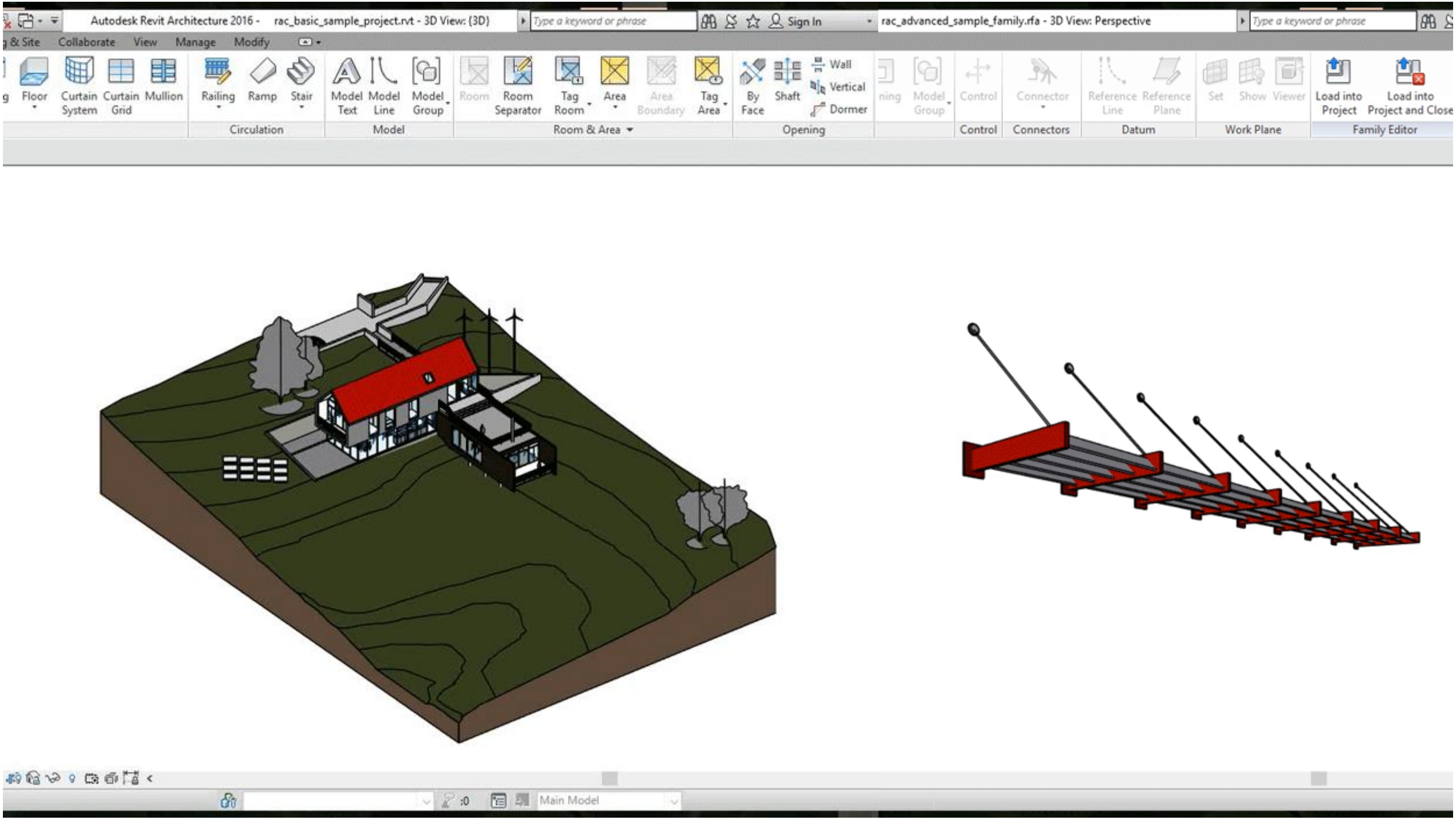
Why a GPU?

CPU
Optimized for
Serial Tasks

+

GPU
Optimized for Many
Parallel Tasks




Autodesk Revit Architecture 2016 - rac_basic_sample_project.rvt - 3D View: (3D)
Type a keyword or phrase
Sign In
rac_advanced_sample_family.rfa - 3D View: Perspective
Type a keyword or phrase
& Site
Collaborate
View
Manage
Modify
Floor
Curtain System
Curtain Grid
Mullion
Railing
Ramp
Stair
Model Text
Model Line
Model Group
Room
Room Separator
Tag Room
Area
Area Boundary
Tag Area
By Face
Shaft
Wall
Vertical
Dormer
ning
Model Group
Control
Connector
Reference Line
Reference Plane
Set
Show
Viewer
Load into Project
Load into Project and Close
Circulation
Model
Room & Area
Opening
Control
Connectors
Datum
Work Plane
Family Editor
Main Model

GPU VIRTUALIZATION FOR ANY WORKLOAD

NVIDIA delivers GPU virtualization for both graphics and compute workloads

ACCELERATE PRODUCTIVITY

For Every User, Any App

Knowledge Worker

Providing business users the highest level of experience for all their apps on any device

Office

Creative & Technical Professional

Uncompromised experience for professional graphics users allowing them to design on the go

AUTODESK AUTOCAD

SOLIDWORKS

Petrel

CATIA

AUTODESK REVIT

ArcGIS

PLM Software SIEMENS NX

AUTODESK MAYA

3DEXCITE

Scalability

Performance

NVIDIA GRID vPC and vApps
Excellent User Experience at Scale

Windows 10 Continues to Evolve

As new updates to Windows 10 aim to bring more and more 3D and 4D capabilities to base OSs, as well as everyday business applications such as Microsoft PowerPoint, the need for IT organizations to leverage virtual client computing with support for GPUs to deliver a superior virtual graphics experience to any device, anytime, anywhere will continue to grow exponentially.

20%
increase in GPU consumption in Windows 10 from 2017 to 2018

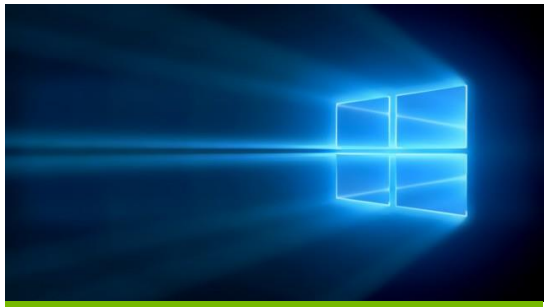
2X
increase in GPU consumption from Office 2010 to 2016 GPU

Source: Shannon Kalvar and Tim Grieser. IDC Worldwide Virtual Client Computing Software Forecast, 2018-2022. February 2018, IDC #US43530218

Lakeside SysTrack

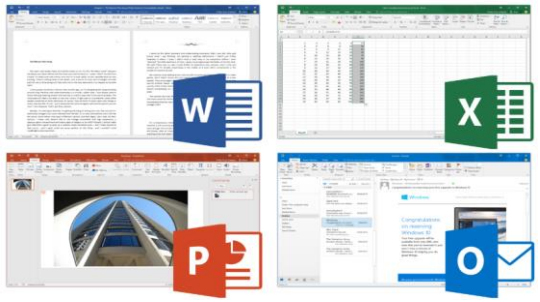
THE NEW DIGITAL WORKER

They Way we Work is Changing



Windows 10

32% increase in CPU requirement over Windows 7[1]

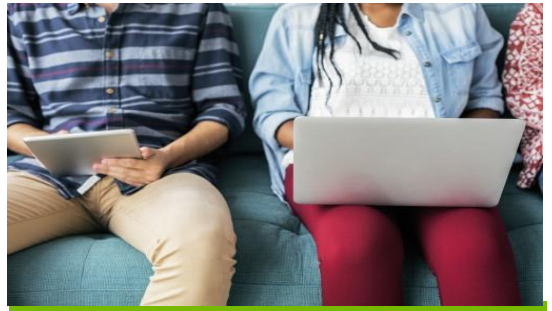


Office 365/Office 2016

50-85% increase in CPU requirements over Windows 7[1]

Web Browsers

Modern browsers are hardware accelerated by default



Latest Web Standards

Flash, HTML5, and WebGL are all very taxing to the CPU

PDF Viewers

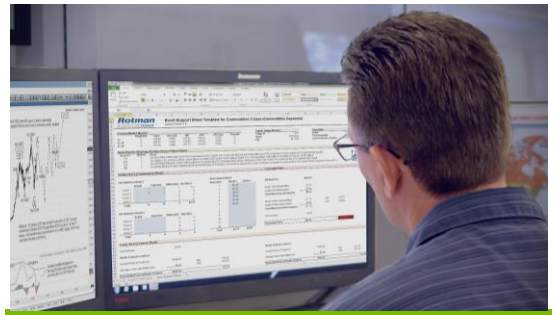
Adobe® Acrobat® and Microsoft Edge are hardware accelerated by default

Collaboration and Video

Skype and YouTube are now prevalent across the enterprise

Digital Imaging & Design

Some features in Adobe® Photoshop® won't work without a GPU[2]



Multi-, High Res Monitors

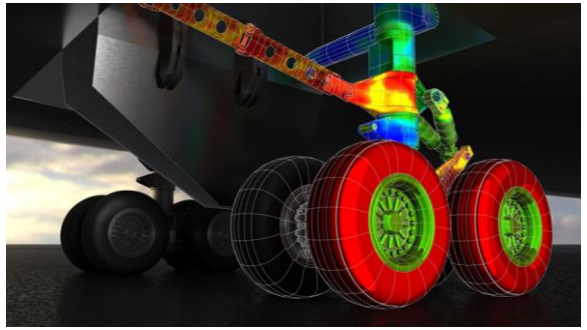
Multi-monitors is the new normal and 4K is becoming mainstream

NVIDIA.

NVIDIA QUADRO VIRTUAL DATA CENTER WORKSTATION POWERS CREATIVITY

THE MOST POWERFUL VIRTUAL WORKSTATION

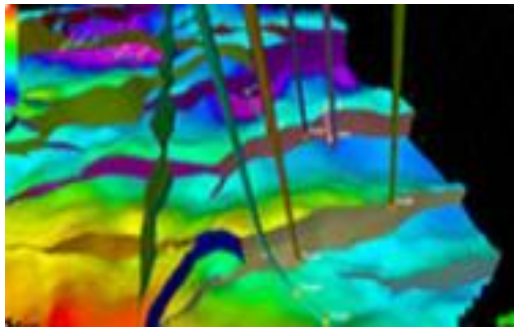
Quadro vDWS accelerates & streamlines professional workflows anywhere



Manufacturing & Automotive

Architecture

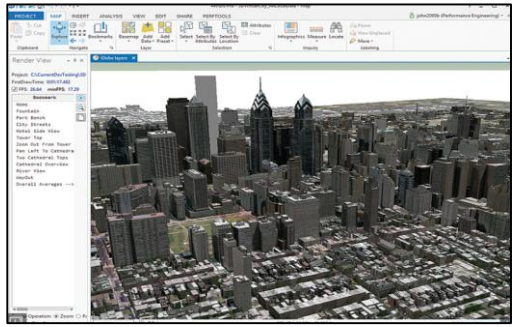
Medical Imaging



Energy

Video Editing

Photorealism

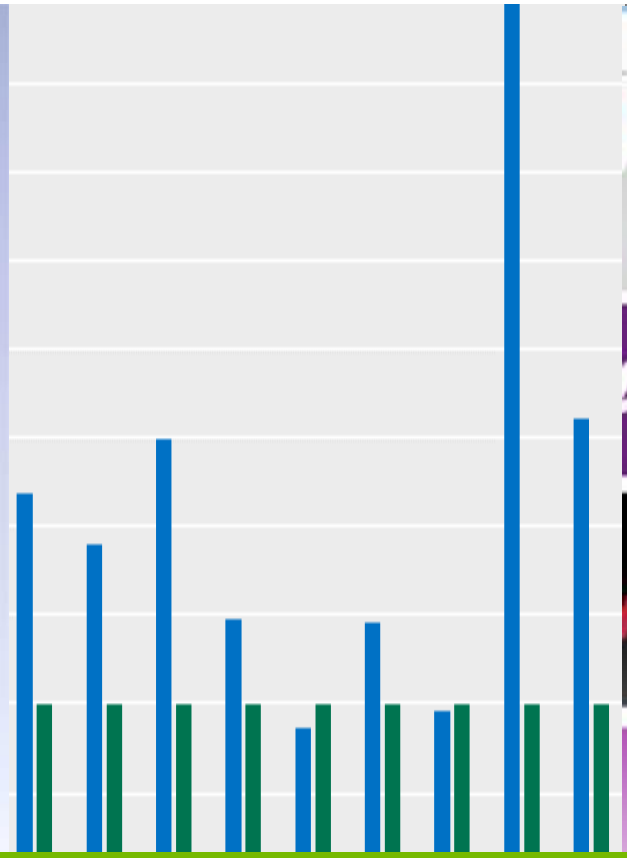


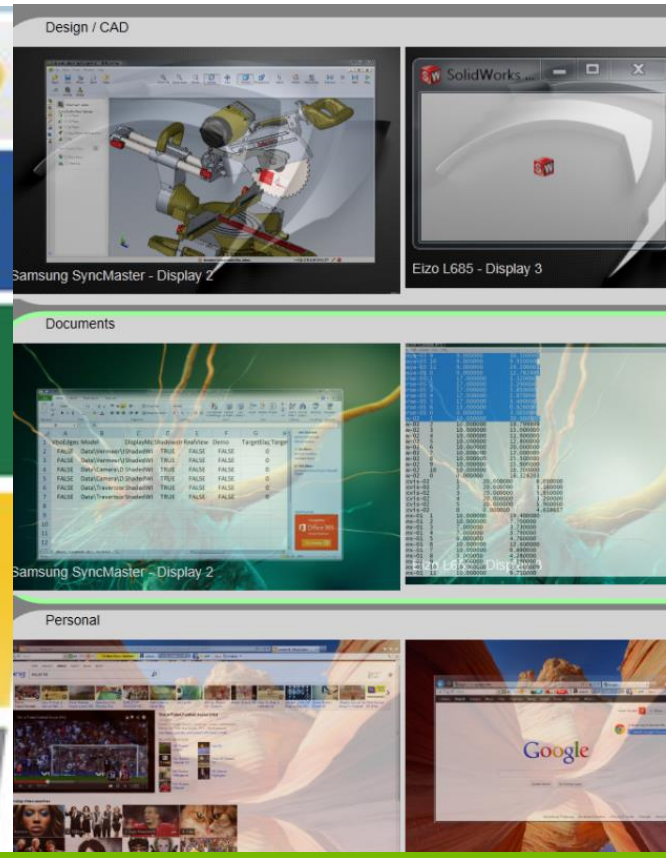
Mapping & Analytics

Simulated Training

QUADRO IN THE DATA CENTER

NVIDIA delivers Quadro features anywhere, on any device





Graphics & Compute | Performance Enhancements | Application Certification | Improved Productivity

QUADRO vDWS ACCELERATES ARCHITECTURE

"A balance of collaboration, performance, flexibility, and cost proved the true value for us. That is only made possible by a virtualized environment running on NVIDIA Quadro vDWS"
- Andrew Schilling, Chief Information Officer, CannonDesign

AUTOMOTIVE INNOVATION WITH QUADRO vDWS

"98-99% of our users can use the virtual environment just like a physical machine sitting in front of them. Users are actually reporting back that it performs exactly the same as a physical machine."

Wesley Struble, CAD system admin, DENSO International America

NVIDIA GPUs for Virtualization in AEC

	V100	RTX 8000	RTX 6000	P40	T4	M10	P6
GPUs / Board (Architecture)	1 (Volta)	1 (Turing)	1 (Turing)	1 (Pascal)	1 (Turing)	4 (Maxwell)	1 (Pascal)
CUDA Cores	5,120	4,608	4,608	3,840	2,560	2,560 (640 per GPU)	2,048
Tensor Cores	640	576	576	---	320	---	---
RT Cores	---	72	72	---	40	---	---
Memory Size	32 GB/16 GB HBM2	48 GB GDDR6	24 GB GDDR6	24 GB GDDR5	16 GB GDDR6	32 GB GDDR5 (8 GB per GPU)	16 GB GDDR5
vGPU Profiles	1 GB, 2 GB, 4 GB, 8 GB, 16 GB, 32 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 16 GB, 24 GB, 48 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 24 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 24 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB
Form Factor	PCIe 3.0 Dual Slot & SXM2 (rack servers)	PCIe 3.0 Dual Slot	PCIe 3.0 Dual Slot	PCIe 3.0 Dual Slot (rack servers)	PCIe 3.0 Single Slot (rack servers)	PCIe 3.0 Dual Slot (rack servers)	MXM (blade servers)
Power	250W/300W	295W	295W	250W	70W	225W	90W
Thermal	passive	active	active	passive	passive	passive	bare board
vGPU Software Editions	GRID vPC/vApps, Quadro vDWS, vComputeServer	Quadro vDWS	Quadro vDWS	GRID vPC/vApps, Quadro vDWS	GRID vPC/vApps, Quadro vDWS, vComputeServer	GRID vPC/vApps, Quadro vDWS	GRID vPC/vApps, Quadro vDWS

PERFORMANCE Optimized

DENSITY Optimized

BLADE Optimized

QUADRO RTX VIRTUAL WORKSTATIONS

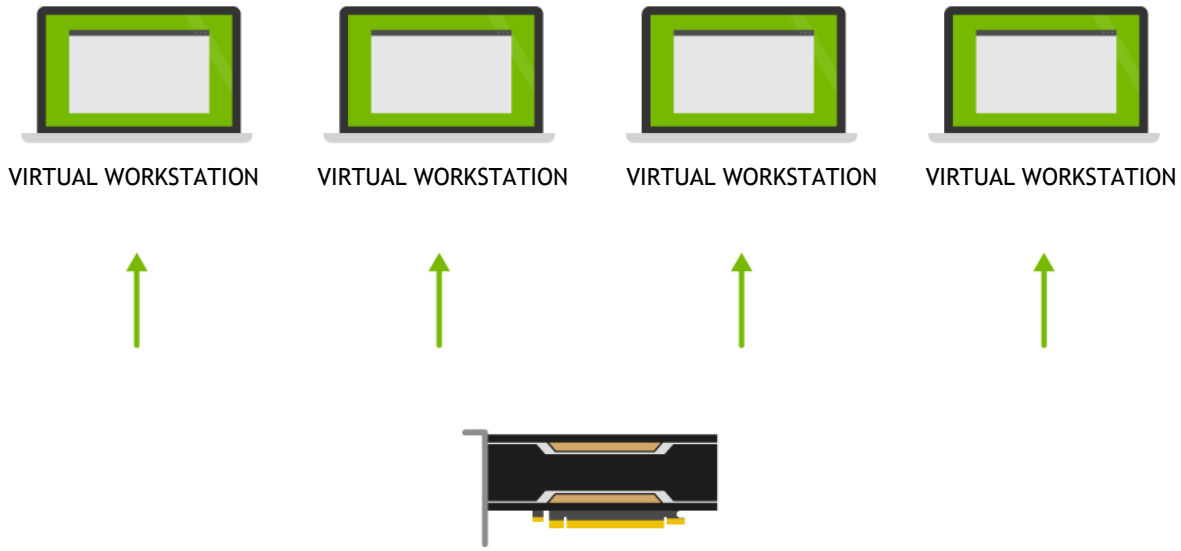
Positioning and Recommendations

Type of User	**Light Users** Small to medium models, scenes or assemblies with simple parts	**Medium Users** Large assemblies with simple parts or small assemblies with complex parts	**Heavy Users** Massive datasets, very large 3D models, complex designs, large assemblies
Recommended Solution	**NVIDIA T4 or P6 Quadro Virtual Data Center Workstation (Quadro vDWS)**	**NVIDIA T4 or P6 Quadro vDWS**	**NVIDIA Quadro RTX 8000, RTX 6000, P40 or V100 with Quadro vDWS**
GPU Memory	16 GB	16 GB	48 GB/32 GB/24 GB
Equivalent Performance	Multiple Quadro P1000	Up to Quadro P4000	Up to Quadro RTX 8000
Replaces	K2, M60, P4, M6	K2, M60, P4, M6	N/A

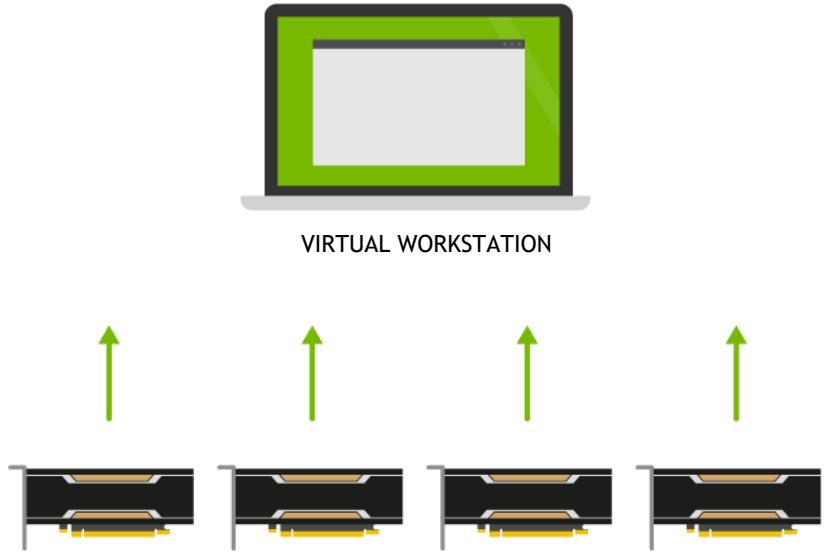
*Recommendations are based on Perf/$

WHAT IS MULTI-vGPU?

Multiple VM's Sharing a GPU
with NVIDIA Quadro vDWS



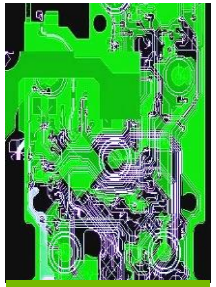
A Single VM Accessing Multiple Tesla GPUs
with NVIDIA Quadro vDWS (7.0)



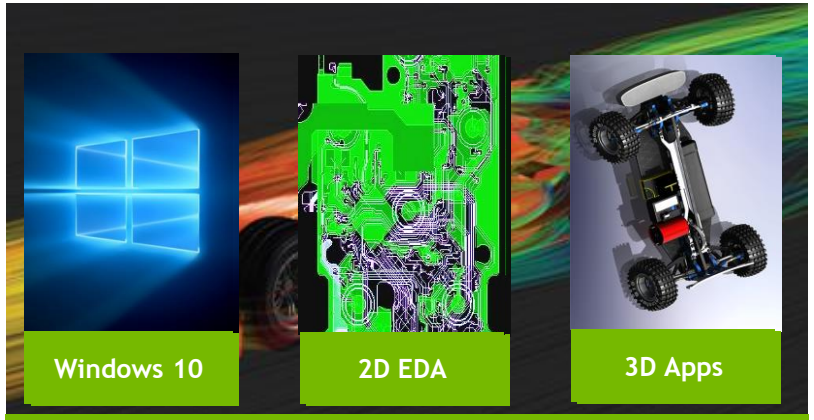
MIXED WORKLOADS WITH NVIDIA vGPU

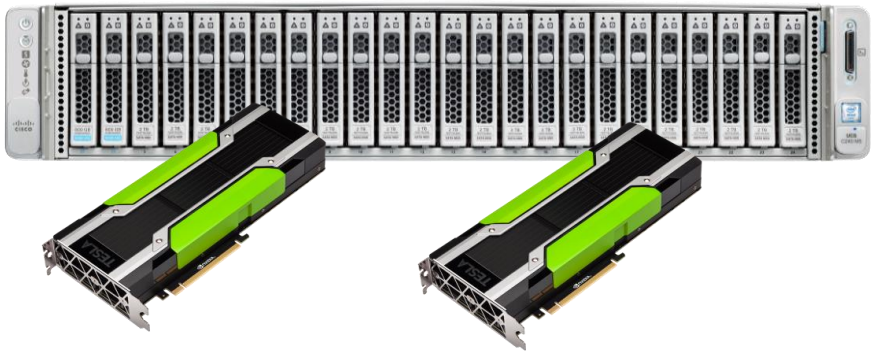
Increase productivity & utilization, decrease costs


Windows 10


2D EDA


3D Apps





vSphere vMotion with vGPU

End-to-end GPU insights with vROPS

NVIDIA.

RTX FOR QUADRO VIRTUAL WORKSTATIONS

Value of Virtualization Extended to RTX Server

Productivity:

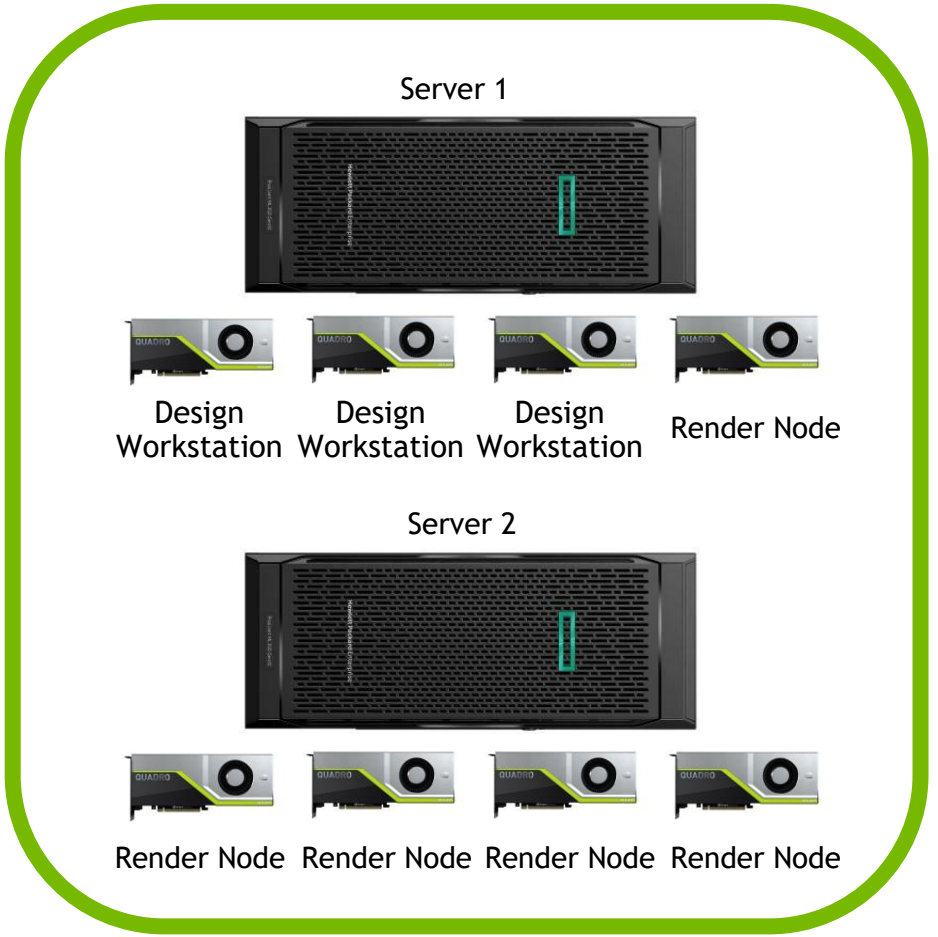
- Up to 5X faster rendering per 1U vs. CPU-only (up to 20X faster in 4U)

Utilization:

- Consolidate multiple individual physical workstations to one server with multiple virtual workstations
- Suspend design workstation nodes and spin up more render nodes as needed, suspend jobs and move to idle systems, utilize available virtual workstation resources when users log off

Cost:

- Buy fewer CPUs by offloading to GPU
- Buy one virtualized server vs. multiple CPU-only servers
- Easier management of VMs versus physical workstations




More than
4,000 VDI
systems take
advantage of
improved
performance and UX
Key Applications: CATIA, Cinebench
Graphics Acceleration: Quadro vDWS
GPU: NVIDIA Tesla M60
HONDA

AEC KEY USER GROUPS

Architects, Engineers, Designers

For remotely viewing and editing very large 3D project files and images

Recommend

Quadro vDWS on T4, P40, RTX6000, RTX8000, or P6
(supports up to four 4K displays)

Marketing, Creative, Design, Illustrators

For general purpose VDI, using virtualized design and creative apps such as Adobe Creative Cloud

Recommend

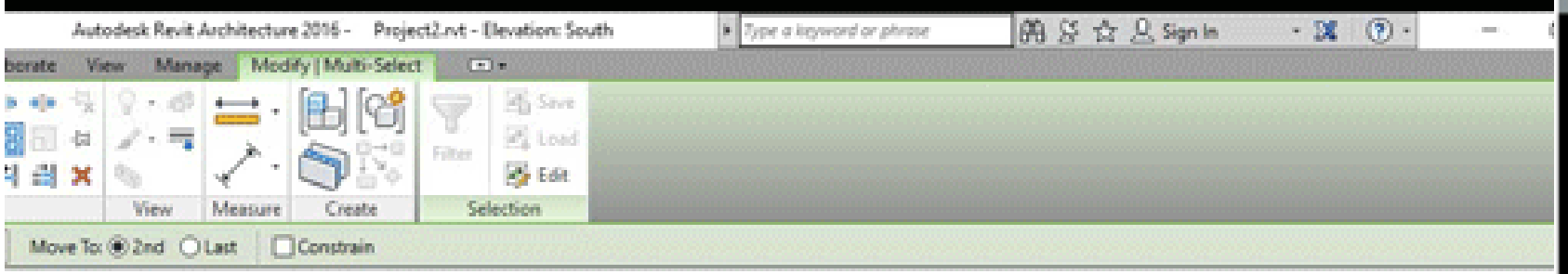
GRID vPC/vApps on T4, M10, or P6
(supports up to four HD or two 4K displays)

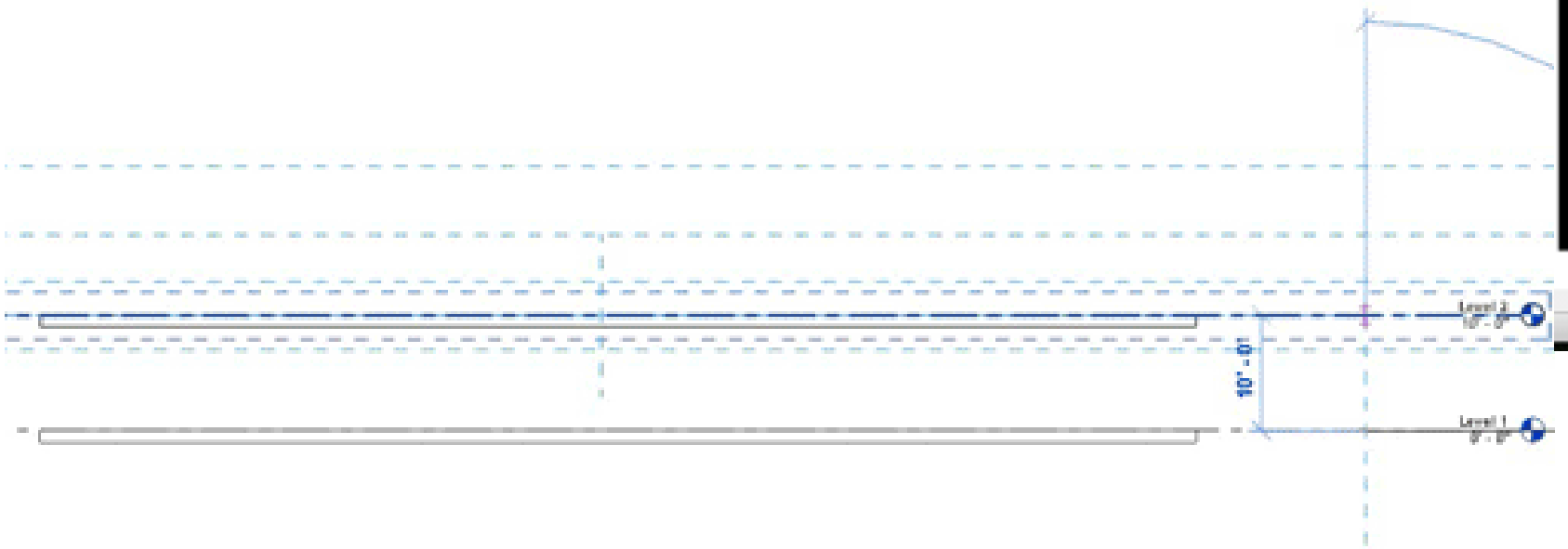
Accounting, Finance, Human Resources

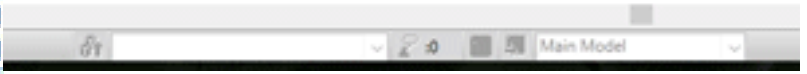
For general purpose VDI, using virtualized Windows 10 and common office productivity apps

Recommend

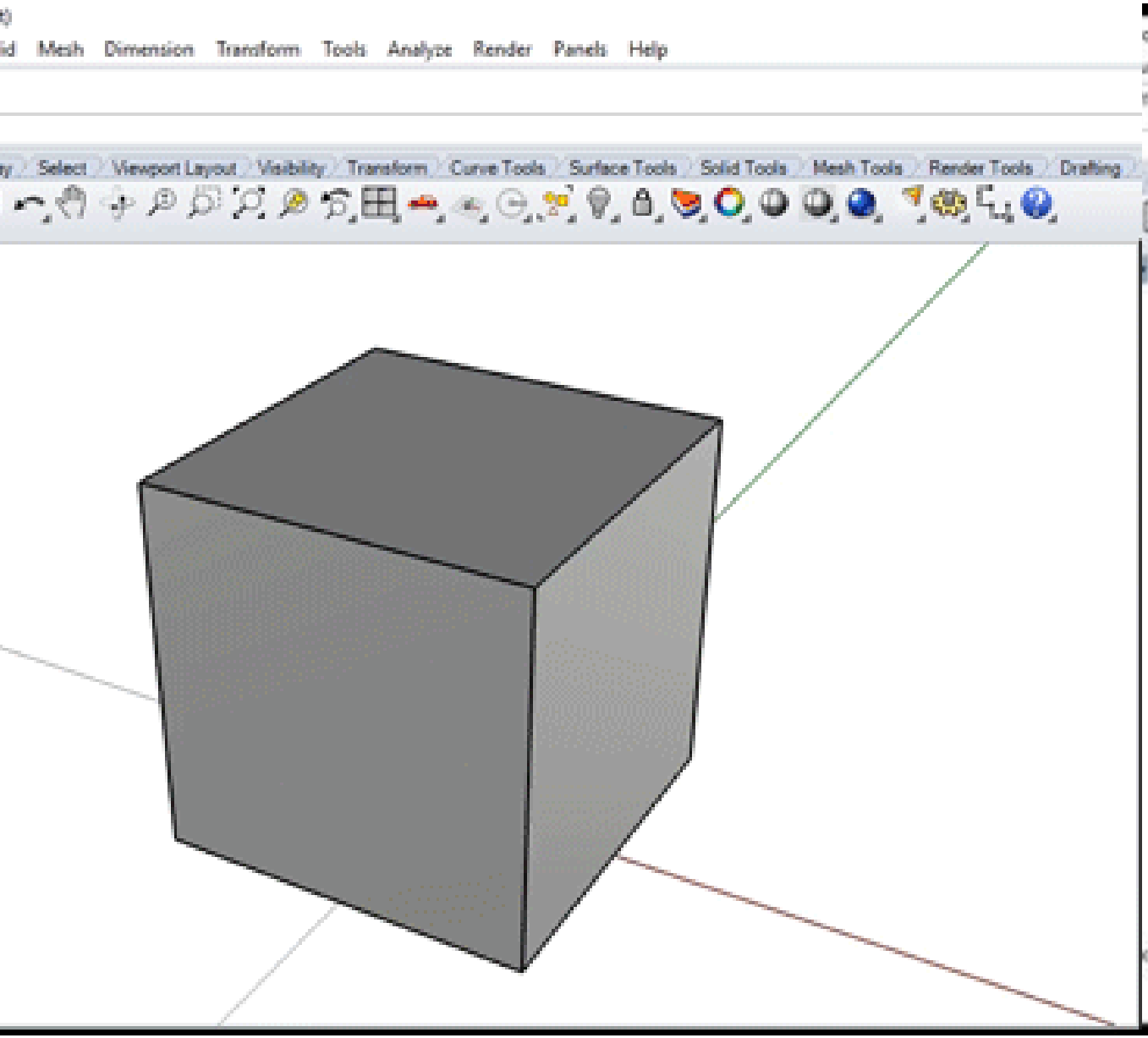
GRID vPC/vApps on T4, M10, or P6
(supports up to four HD or two 4K displays)

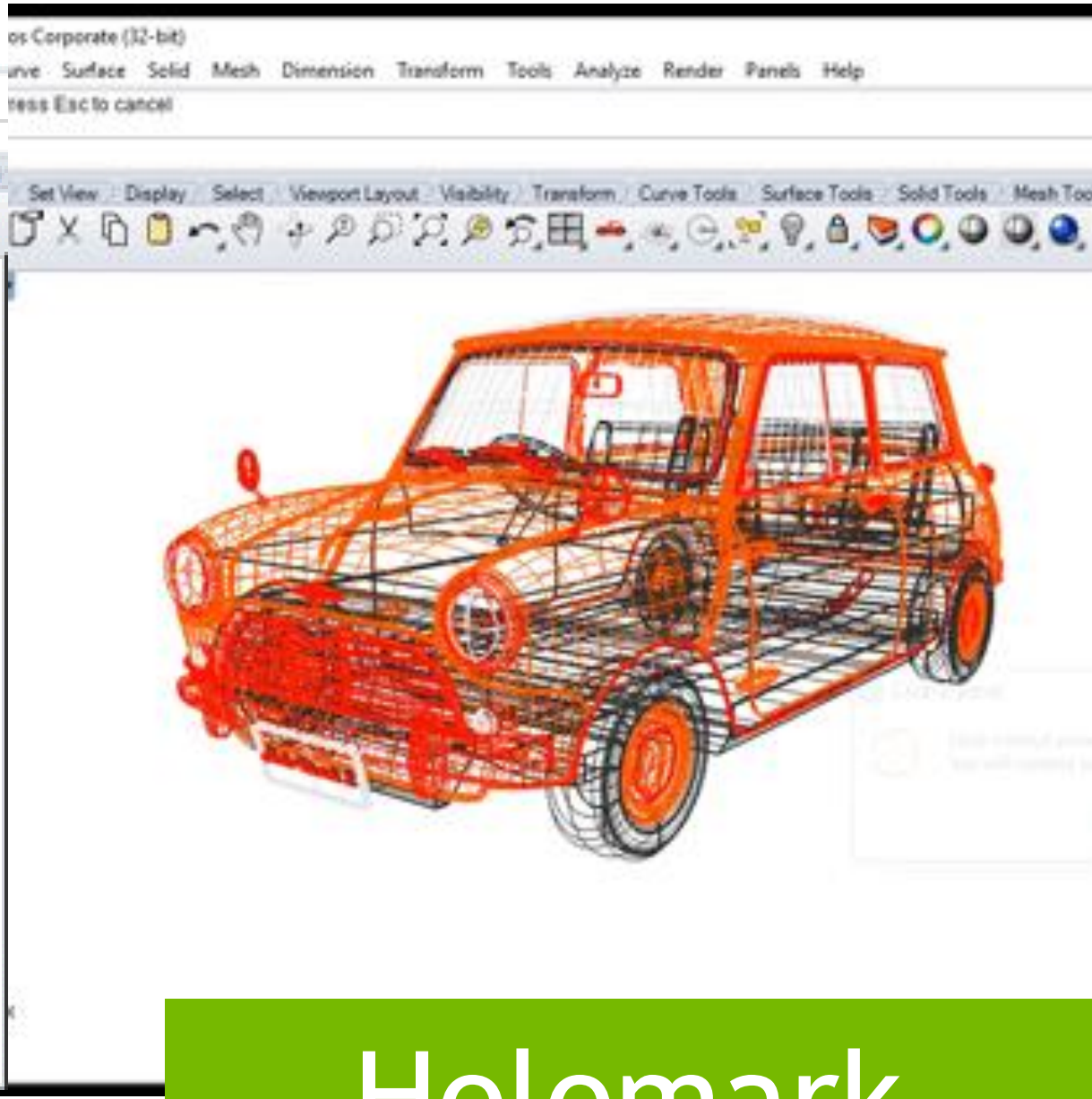






RFObenchmark


Mesh Dimension Transform Tools Analyze Render Panels Help
Select Viewport Layout Visibility Transform Curve Tools Surface Tools Solid Tools Mesh Tools Render Tools Drafting


Surface Solid Mesh Dimension Transform Tools Analyze Render Panels Help
Set View Display Select Viewport Layout Visibility Transform Curve Tools Surface Tools Solid Tools

Holomark

GPU PROFILER

https://github.com/JeremyMain/GPUProfiler/releases

- CPU Utilization
- GPU Utilization
- Frame Buffer Utilization
- RAM Utilization
- Video Encode
- Video Decode

Application Deployment Guide
Autodesk Revit 2015 with NVIDIA GRID vGPU on VMware Horizon
AUTODESK.
NVIDIA
vmware

These are general guidelines based on past and present customer use cases. Your mileage may vary. We strongly recommend running the **vGPU Profiler Tool** and a **POC** to determine final user requirements and server specifications.

Win10 Knowledge Workers (Office, WebGL, Teleconferencing, Training Videos, etc.)

- 2 vCPU per session
- 5:1 oversubscription of CPU cores
- 6-8GB of System RAM per session
- 1-2GB of frame buffer per guest (1B-2B profile)
- (1) vPC license per user
- GPU: M10 / T4 / P6

2D/3D Design (AutoCAD, Revit, BIM suites, etc.)

- 4 vCPU per session
- 3:1 oversubscription of CPU cores
- 16GB of System RAM per session
- 1-2GB of frame buffer per session (1Q-2Q profile)
- (1) vDWS license per concurrent session
- GPU: T4 / P6 / RTX6000

GIS (Esri ArcGIS)

- 4-8 vCPU per session
- 3:1 oversubscription of CPU cores
- 16GB or more
- 1-4 GB frame buffer per session (1Q-4Q profile)
- (1) vDWS license per concurrent session
- GPU: T4 / P6 / P40 / RTX6000 / RTX8000

Engineering (SolidWorks, Catia, Siemens NX, Creo, etc.)

- 4 vCPU user
- 3:1 oversubscription of CPU cores (max – less is better)
- Target 2.6GHz and above
- 16GB of RAM per session
- 1-4GB of frame buffer per session (1Q-4Q profile)
- (1) vDWS license per concurrent user
- GPU: T4 / P6 / P40 / RTX6000 / RTX8000

P6 for blade chassis only

NVIDIA GPUs for Virtualization in AEC

	V100	RTX 8000	RTX 6000	P40	T4	M10	P6
GPUs / Board (Architecture)	1 (Volta)	1 (Turing)	1 (Turing)	1 (Pascal)	1 (Turing)	4 (Maxwell)	1 (Pascal)
CUDA Cores	5,120	4,608	4,608	3,840	2,560	2,560 (640 per GPU)	2,048
Tensor Cores	640	576	576	---	320	---	---
RT Cores	---	72	72	---	40	---	---
Memory Size	32 GB/16 GB HBM2	48 GB GDDR6	24 GB GDDR6	24 GB GDDR5	16 GB GDDR6	32 GB GDDR5 (8 GB per GPU)	16 GB GDDR5
vGPU Profiles	1 GB, 2 GB, 4 GB, 8 GB, 16 GB, 32 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 16 GB, 24 GB, 48 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 24 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 24 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB
Form Factor	PCIe 3.0 Dual Slot & SXM2 (rack servers)	PCIe 3.0 Dual Slot	PCIe 3.0 Dual Slot	PCIe 3.0 Dual Slot (rack servers)	PCIe 3.0 Single Slot (rack servers)	PCIe 3.0 Dual Slot (rack servers)	MXM (blade servers)
Power	250W/300W	295W	295W	250W	70W	225W	90W
Thermal	passive	active	active	passive	passive	passive	bare board
vGPU Software Editions	GRID vPC/vApps, Quadro vDWS, vComputeServer	Quadro vDWS	Quadro vDWS	GRID vPC/vApps, Quadro vDWS	GRID vPC/vApps, Quadro vDWS, vComputeServer	GRID vPC/vApps, Quadro vDWS	GRID vPC/vApps, Quadro vDWS

PERFORMANCE Optimized

DENSITY Optimized

BLADE Optimized

NVIDIA T4 FOR VIRTUALIZATION

Powering 3D Professional Virtual Workstations

GPU Architecture	Turing
CUDA Cores	2,560
Tensor Cores	320
Memory Size	16 GB
vGPU Profiles	1 GB, 2 GB, 4 GB, 8 GB, 16 GB
Form Factor	PCIe 3.0 single slot
Power	70 W
Thermal	Passive

PLATFORMS ▸ DEVELOPERS ▸ INDUSTRIES ▸ SHOP DRIVERS ▸ SUPPORT ABOUT NVIDIA ▸

NVIDIA vGPU TECHNOLOGY RESOURCES TRY BUY

Subscribe

NVIDIA vGPU PARTNERS

FIND A VIRTUALIZATION PARTNER TO GET STARTED.

PREFERRED NVIDIA vGPU CERTIFIED SERVER VENDORS

Cisco | Dell EMC | HPE | Lenovo | Supermicro

NVIDIA VGPU CERTIFIED SERVERS

All Manufacturers | All Heights | All Form Factors | All Card Models | Search | Share URL...

Mixed GPU configurations within a server are not supported

* The maximum number of supported Tesla M10 and Tesla M60 cards per system when using NVIDIA GRID vGPU is two and four respectively. Please contact OEMs for 3x M10 configuration.

** With Expansion Chassis.

*** SXM form factor.

****NVLink with Quadro

For certified servers of EOL products, please visit Application Note

All NGC ready systems are part of vGPU certified server list. For complete list of NGC ready systems, please refer to NGC ready webpage

MANUFACTURER	MODEL	RACK UNITS	NODE PER CHASSIS	MAXIMUM CARDS PER NODE										MORE INFO
				Tesla M10	Tesla M60	Tesla P4	Tesla P6	Tesla P40	Tesla V100	Tesla V100 (32GB)	Tesla T4	RTX 6000	RTX 8000	

https://www.nvidia.com/en-us/data-center/resources/vgpu-certified-servers/

3 Components to NVIDIA Hardware and Software Setup

License Server

- Retrieve Eval Keys
- Register License Server
- Generate License File
- Install License Manager on Designated Server

Host Setup

- Bios Configurations (usually done by OEM)
- Download Zipped Bits
- Install VIB/Host Driver
- Nvidia-smi test
- Configure Host Graphics in Hypervisor

Guest Setup

- Add Shared PCI Device
- Assign vGPU profile
- Install Guest Drivers
- Point Guest to License Server

POC - Part 1

- Sign up for the NVIDIA Enterprise Portal and download the vGPU software and Vib. Please have login credentials available. https://nvid.nvidia.com/NvidiaUtilities/#/createNewuser
- Acquire NVIDIA license keys from OEM or NVIDIA partner. If keys have not been purchased or redeemed, we can utilize a trial key for initial setup.
- Rack and stack hosts with IPs and connectivity to guest images.
- Have someone available who has access to build the host, work within vCenter/XenCenter and who can build the base parent image.
- Clean 2012 R2 or 2016 guest 2 vCPU and 8GB of RAM.
- Download latest i586 offline installer version of Java.

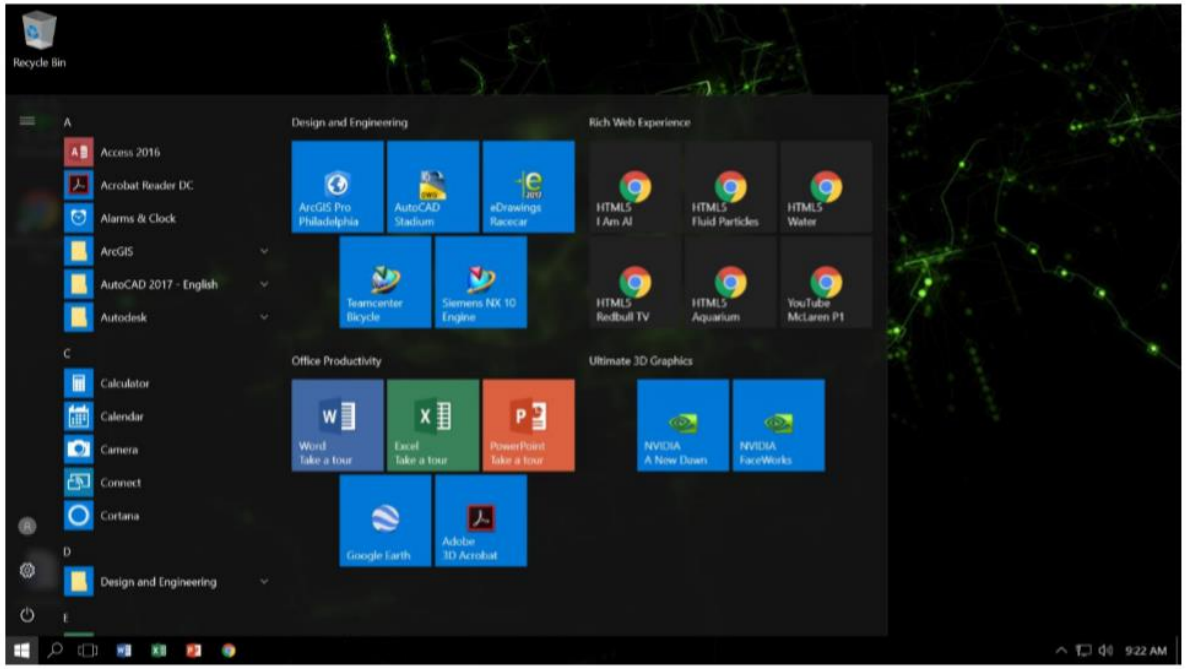
POC - Part 2

- Build ESXi/XenServer
- Load the VIBS/RPMs
- Review vCenter/XenCenter/Prism settings
- Review license server/NVIDIA portal. Rebuild/repair if necessary
- Build parent image process. Review vGPU profiles.
- Perform baseline tests.
- Answer questions you have about usage, profiles types, fudge factor vs. finite, etc.

TEST DRIVE NVIDIA VIRTUAL GPU TODAY

- Instant access on VMware vSphere and Horizon
- Pre-installed apps such as AutoCAD, SOLIDWORKS, ArcGIS, NX, etc.
- HTML demos and HD videos

http://www.nvidia.com/testdrive



NVIDIA VIRTUAL GPU RESOURCES

Virtual GPU Test Drive
https://www.nvidia.com/tryvgpu

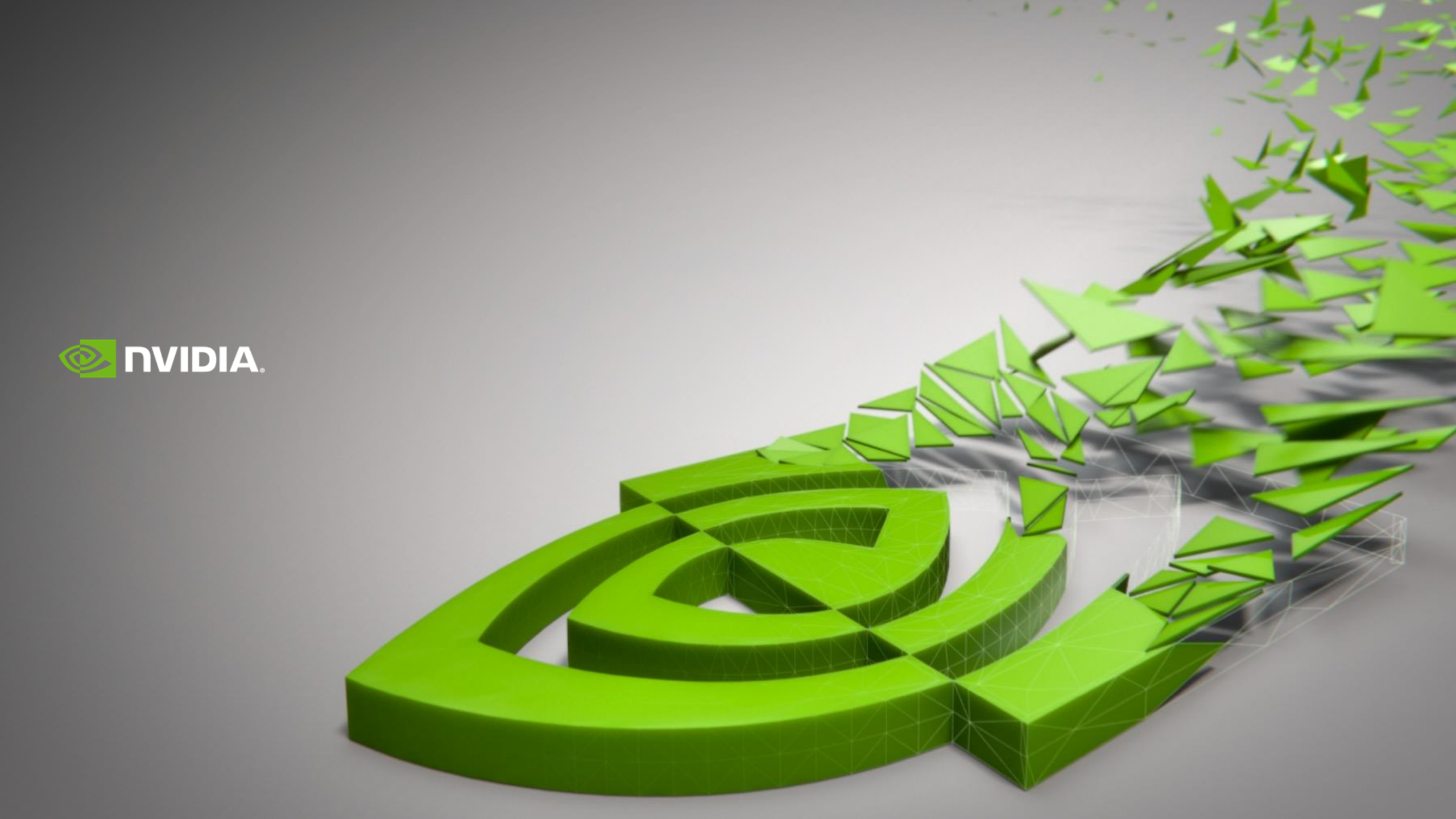
NVIDIA Virtual GPU Website
www.nvidia.com/virtualgpu

NVIDIA Virtual GPU YouTube Channel
http://tinyurl.com/gridvideos

Questions? Ask on our Forums
https://gridforums.nvidia.com

NVIDIA Virtual GPU on LinkedIn
http://linkd.in/QG4A6u

Follow us on Twitter
@NVIDIAVirt


NVIDIA