NVIDIA Quadro RTX
The Fusion of Graphics and AI
S9969 | Booth 1133
# NVIDIA Quadro RTX
The fusion of graphics and AI

<table>
<thead>
<tr>
<th>RT Cores deliver real-time ray tracing</th>
<th>Advanced shaders for graphics and VR</th>
<th>Tensor Cores power AI-augmented applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Model how light and materials behave in 3D</td>
<td>• Mesh shading</td>
<td>• AI denoising</td>
</tr>
<tr>
<td>• 25x faster than Pascal GPUs</td>
<td>• Variable rate shading</td>
<td>• Resolution scaling</td>
</tr>
<tr>
<td>• Renders film effects at 30x CPU speeds</td>
<td>• Texture space shading</td>
<td>• Video re-timing</td>
</tr>
</tbody>
</table>
NVIDIA RTX Technology
Next generation hybrid rendering

- Applications or Plug-ins
- MDL (Materials Definition Library) and USD (Universal Scene Description)
- OptiX | DXR | Vulkan
  - Rasterization (Graphics Pipeline)
  - Ray Tracing (RT Cores)
  - Compute (CUDA)
  - AI (Tensor Cores)

NVIDIA RTX
NVIDIA RTX RT Cores
Deliver hardware accelerated ray tracing

Functionality includes:

- Ray-triangle intersection checks
- Bounding Volume Hierarchy (BVH) manipulation
- Real-time performance in application viewports

American Gods image courtesy of Tendré
## NVIDIA RTX Advanced Shaders

More objects per scene, flexible control over detail and performance, better VR

<table>
<thead>
<tr>
<th>Mesh Shading</th>
<th>Texture Space Shading</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Mesh Shading Image" /></td>
<td><img src="image2" alt="Texture Space Shading Image" /></td>
</tr>
<tr>
<td>• Eliminates CPU call bottlenecks and draws triangles more efficiently</td>
<td>• Decouples shading from screen space, improving shading efficiency and reuse</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable Rate Shading</th>
<th>Multi-View Rendering</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Variable Rate Shading Image" /></td>
<td><img src="image4" alt="Multi-View Rendering Image" /></td>
</tr>
<tr>
<td>• Pixel shading rate control for effects like motion, blur, foveated rendering</td>
<td>• Extends Single Pass Stereo with unique view origin positions or directions</td>
</tr>
</tbody>
</table>
NVIDIA RTX Tensor Cores
Next generation hardware accelerated deep learning

Key benefits include:
- Hardware acceleration of deep learning enabled tools and applications
- Support for additional precision modes for improved performance
- Turing Tensor Cores tuned for fast training and inferencing performance
NVIDIA Quadro RTX VR
New capabilities for ultimate VR experiences

Variable Rate Shading
- Match lens optics
- Foveated rendering
- Places detail where most needed
- Context sensitive

Single Pass Stereo 2.0
- Even more efficient GPU use
- SMP for 4 independent views
- Wide FOV HMD’s
- Wrap-around multi displays

VirtualLink
-open industry standard
- Singe cable power, data, video
- Utilizes USB-C Alt-Mode
- Easy VR setup, reduced cabling

Performance
- Rasterization
- Ray tracing
- Physically-based audio
- Dual-input VR-5LI
NVIDIA Quadro RTX
Provides the memory required for professional VR

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometry</td>
<td>See entire urban scenes, facilities, airplanes and cars in VR with full fidelity</td>
</tr>
<tr>
<td>Textures</td>
<td>Use more and larger textures to create the most realistic VR environments possible</td>
</tr>
<tr>
<td>Panoramas</td>
<td>Move around without interrupting the VR experience by using multiple panoramas</td>
</tr>
<tr>
<td>Light Fields</td>
<td>Provides most realistic VR experience possible using photorealistic imagery</td>
</tr>
</tbody>
</table>

Gamers can cut down on fidelity and still play. Professional customers need full data fidelity to make the right decisions and create the best designs.
NVIDIA Quadro RTX Value Proposition
Spans all key markets and solutions

<table>
<thead>
<tr>
<th>M&amp;E</th>
<th>Manufacturing</th>
<th>AEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rendering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Creation</td>
<td>Product Design</td>
<td>Building Design</td>
</tr>
<tr>
<td>AI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-Painting Opres</td>
<td>Generative Design</td>
<td>Generative Design</td>
</tr>
<tr>
<td>VR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Creation</td>
<td>Design Review</td>
<td>Design Review</td>
</tr>
</tbody>
</table>

- Realtime rendering speeds up creative workflows
- AI-augmented tools accelerate the creative process
- VR content creation and design reviews
NVIDIA Quadro RTX | Create More, Wait Less

Note: CPU: Core i9-7900X, GPU: NVIDIA RTX, video playback at 2x speed
NVIDIA RTX Server
Highly configurable reference architecture for rendering pipelines

- Powered by Quadro RTX 8000 or 6000 and NVLink
- Ray traced global illumination of up to 96 GB scenes
- Remoting, batch and multi-GPU virtualization
- Rendering time reduced from hours to minutes

Suppliers are PNY and NVIDIA authorized, like the prior VCA Certified Rendering System program
NVIDIA RTX Server
Bare metal rendering with CUDA-X graphics

<table>
<thead>
<tr>
<th>RTX Server Validation</th>
<th>Arnold Core Test Suite + Project SOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Autodesk Arnold 5.3.0.0</td>
</tr>
<tr>
<td>Qualified System</td>
<td>8x Quadro RTX 8000 or RTX 6000 + NVLink</td>
</tr>
</tbody>
</table>

ARNOLD

BOXX | SUPERMICRO | EXXACT | TYAN
NVIDIA RTX Server
Virtual desktop, batch and remoting with CUDA-X graphics

<table>
<thead>
<tr>
<th>RTX Server Validation</th>
<th>Project Cirrus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Autodesk 3DS MAX and MAYA</td>
</tr>
<tr>
<td>Remoting Protocol</td>
<td>Teradici Cloud Access Plus</td>
</tr>
<tr>
<td>Virtualization</td>
<td>Quadro vDWS v8.0 + VMware vSphere 6.7</td>
</tr>
<tr>
<td>Qualified System</td>
<td>8x Quadro RTX 8000 or RTX 6000 + NVLink</td>
</tr>
</tbody>
</table>
VFX Film Pipeline
Complex, multi-phase workflow

Pre-Production
- Storyboards, Character Design
- Pre-Viz
- Shoot
- Dailies
- Offline Edit
- Scan Selects, Shot List

Asset Creation
- Motion Capture
- Model
- UVs
- Rig
- Fur, Cloth, Hair
- Shaders and Textures
- Matte Paintings

Clean Up
- Motion Cleanup
- Dust-Bust
- Pre-Grade
- Roto

Shots
- Camera Extract
- Layout
- Blocking
- Animation
- Simulation and FX
- Lighting
- Rendering
- Pre-Comp

Finishing
- Composite
- Online Edit Conform
- Grade
- Film or Digital Distribution
VFX Film Pipeline
Rendering used during many phases

Pre-Production: Storyboards, Character Design → Pre-Viz
Asset Creation: Motion Capture → Model → UVs → Rig → Fur, Cloth, Hair → Shaders and Textures → Matte Paintings
Clean Up: Motion Cleanup → Dust-Bust → Pre-Grade → Roto
Shots: Camera Extract → Layout → Blocking → Animation → Simulation and FX
Finishing: Composite → Online Edit Confirm → Grade → Final Cut Pro, Distribution
NVIDIA RTX Server

Explosion of content

<table>
<thead>
<tr>
<th>Year</th>
<th>VFX Shots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>63</td>
</tr>
<tr>
<td>2018</td>
<td>2680</td>
</tr>
</tbody>
</table>

Annual Spend on Original Content

- **Apple**
- **Amazon**
- **Netflix**

- **2017**: $2B
- **2018**: $2B
- **2019**: $4B
- **2020**: $8B
- **2021**: $12B
- **2022**: $12B
NVIDIA RTX Server
Rising quality bar requires SW and HW innovation

“After 6 months the Lightspeed and RenderMan team had a system that gives the effect of millions of lights and took the notational render time on the complex shots down from 1000 hours to 450 hours. The team continued and reduced this further to 125 hours and finally 75 hours a frame. With some additional work on the way the production team worked with the lighting in shots, the final per frame time at the end of production was just 50 hours per frame.*

From renderman.pixar.com

* Time it would take if the frame was rendered on a single core system. COCO logo, Disney, and Pixar are registered trademarks of Disney. COCO frame grab copyright Disney–Pixar.
NVIDIA RTX Server
GPU memory capacity meets industry needs

<table>
<thead>
<tr>
<th>Working Set Size</th>
<th>34 GB (Average)</th>
<th>75 GB (High)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 GB</td>
<td>48 GB</td>
<td>96 GB</td>
</tr>
<tr>
<td>Quadro RTX 6000</td>
<td>2x Quadro RTX 6000</td>
<td>2x Quadro RTX 8000</td>
</tr>
<tr>
<td>24 GB</td>
<td>Quadro RTX 8000</td>
<td></td>
</tr>
</tbody>
</table>
NVIDIA Quadro RTX Real-Time Rendering
 Resets datacenter technology and economic expectations

1/4 the cost | 1/10 the space | 1/11 the power

“The more you buy, the more you save!” – Jensen Huang, CEO NVIDIA
NVIDIA NGX
A fast, easy SDK for integrating AI features

NVIDIA Develops AI Model

NVIDIA Delivers AI Model and SDK for Application Integration and New Features Addition

Create Training Data
Train and Optimize AI Model

Installs Latest AI Model with NVIDIA Driver
NGX Aware Applications Present Features
NVIDIA NGX AI Upres
Exploring new ways to upscale content

Nearest Neighbor

Bicubic Filter

AI Super Res

AI Upres creates new pixels by interpreting the image and intelligently placing data, resulting in sharper enlargements by 2x, 4x or 8x.
NVIDIA NGX AI Slow-Mo
From 30fps to 120fps
NVIDIA NGX AI In-Painting

A magician, not a healer
Data Science Workstation
Powered by 2x NVIDIA Quadro RTX 8000 or 2x RTX 6000 with NVLink

260 TFLOPS Compute
48 or 96 GB GPU Memory

AMAX | BOXX | COLFAK | EXXACT | Microway | THINKMATE
NVIDIA CUDA-X AI and NVIDIA RAPIDS
Executes end-to-end data science and analytics pipelines entirely on RTX

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hassle-Free Integration</td>
<td>Accelerate your Python data science toolchain with minimal code changes and no new tools to learn</td>
</tr>
<tr>
<td>Improves Model Accuracy</td>
<td>Increase machine learning model accuracy by iterating on models faster and deploying them more frequently</td>
</tr>
<tr>
<td>Reduces Training Time</td>
<td>Drastically improve your productivity with near-interactive data science</td>
</tr>
<tr>
<td>Open Source, NVIDIA Optimizations</td>
<td>Customizable, extensible, interoperable open-source software is optimized and supported by NVIDIA and built on Apache Arrow</td>
</tr>
</tbody>
</table>

RTX

PNY

PNY PRO
NVIDIA RAPIDS with Anaconda
The new GPU data science pipeline

Apache Arrow
A columnar in-memory data structure that delivers efficient and fast data interchange with flexibility to support complex data models

cuDF Analytics
cuDF is a DataFrame manipulation library that accelerates loading, filtering, and manipulation of data for model training data preparation

cuML Machine Learning
cuML provides GPU-accelerated versions of all machine learning algorithms available in scikit-learn

cuGRAPH
A framework and collection of graph analytics libraries that seamlessly integrate into the RAPIDS data science platform

cuDNN
RAPIDS provides native array interface support, so data can be pushed to DL frameworks like PyTorch and Chainer

RTX Visualization
Native GPU in-memory data format provides high-performance, high-FPS data visualization, even with very large datasets
RAPIDS with Quadro RTX 8000
Unprecedented data science performance

End-to-End
XG Boost
Data Prep

CPU
RTX 8000
2x RTX 8000

*End-to-End time = ETL + conversion + training + validation. CPU: Xeon 6140 at 3.2 GHz, 3.7 GHz Turbo, 384 GB RAM, Ubuntu 16.04.4, NVIDIA driver 410.93*
Data Science Workstation
Opening up new vistas of discovery
NVIDIA Quadro RTX
The Fusion of Graphics and AI

S9969 | Booth 1133