Powering Intelligent Video Analytics (IVA) Using NVIDIA Virtual GPU and VMware vSphere

Charlie Huang - Sr. Product Marketing Manager

Shailesh Deshmukh - Sr. Solutions Architect
VGPU EVERYWHERE FOR EVERYONE, EVERY WORKLOAD

GPU ACCELERATED DATA CENTER & CLOUD
HOW VGPU WORKS
NVIDIA Virtual GPU Delivers a GPU Experience to Every Workload

CPU Only VDI

- Apps and VMs
- Hypervisor
- Server

With NVIDIA Virtual GPU

- Apps and VMs
- NVIDIA Virtual GPU
- NVIDIA virtualization software
- Hypervisor
- NVIDIA GPU
- Server

- NVIDIA Graphics Driver,
- NVIDIA Quadro Driver,
- NVIDIA Compute Driver
INTRODUCING NVIDIA vCOMPUTESERVER
Virtual GPU Software for Compute Workloads

Hypervisor virtualization accelerated by NVIDIA GPUs
Virtualize compute-intensive, server workloads
Advanced GPU features for AI, DL, Data Science, HPC
GPU performance with hypervisor management tools, flexibility, and security
Supported by all major hypervisor platforms VMware, Red Hat, Nutanix & Citrix¹
Software enabled with a vComputeServer Annual Subscription starting at $50 per GPU

¹. Nutanix and Citrix support coming soon, subject to hypervisor roadmaps.
EXPANDING THE NVIDIA VIRTUAL GPU PORTFOLIO

Use Case

Creative & Technical Professional

Knowledge Worker

Client Computing

Virtual GPU Software Edition

Virtual Data Center Workstation

AI, Deep Learning, & Data Science

vComputeServer

GRID Virtual PC

GRID Virtual Apps

Server Workloads

Compute Type

Client Computing

Creative & Technical Professional

Recommended GPU

NVIDIA V100 or T4

NVIDIA T4 or M10

NVIDIA T4 or Quadro RTX 6000, RTX 8000

NVIDIA T4 or M10

(NVIDIA P6 for blade form factor)

(NVIDIA P6 for blade form factor)
EASE DATA CENTER GPU ADOPTION

vComputeServer for Virtualization

**GPU Performance**
For Virtualization

**Maximize Utilization**
Ensure right-size allocation with GPU sharing & aggregation

**Management & Monitoring**
Use same hypervisor virtualization tools to manage GPU servers

**Live Migration**
Only vGPU solution to support live migration

**Security**
Hypervisor-based security extends to GPU workloads

**Multi-Tenant**
Isolate workloads to securely support multiple users

**Reliability**
ECC & Page Retirement
COMPUTE, ANY WAY YOU WANT
With vComputeServer & NGC Containers on vSphere
IVA USE CASES FOR EFFICIENCY AND SAFETY

Access Control
Managing operations
Parking Management
Traffic Engineering

Retail Analytics
Optical Inspection
Managing Logistics
Content Filtering
WHAT IS DEEPSTREAM?

Analyze Data From Cameras, Sensors and IoT Gateways in Real-Time

- NVIDIA’s DeepStream SDK delivers a complete streaming analytics toolkit for AI-based video and image understanding, as well as multi-sensor processing.

- DeepStream is an integral part of NVIDIA Metropolis, the platform for building end-to-end services and solutions for transforming pixels and sensor data to actionable insights.
NOTIONAL SOLUTION USING VCOMPUTE SERVER

Fractional vGPU for Multiple Streams, Jobs, and Flexibility via Virtual Machines or Delivered in Containers
Running DeepStream 3.0 in NCG Container and VM

- VMware ESXi 6.7 Update 3
- VMware vCenter 6.3
- VMware Horizon View 7.9
- NVIDIA vGPU 9.0 (vComputeServer)
- Guest OS: Ubuntu 16.04 / 18.04
- Container run: Docker / Nvidia-Docker
- Tensor RT 5.X
- DeepStream 3.0
Initiate the NCG container
NVIDIA Virtual Compute Server Website

NVIDIA Virtual GPU Website
www.nvidia.com/virtualgpu

Virtual GPU Test Drive
www.nvidia.com/tryvgpu

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
Thank You!!

Q & A