GPU in every workload

Mason Wu, Solution Architect
April. 2018

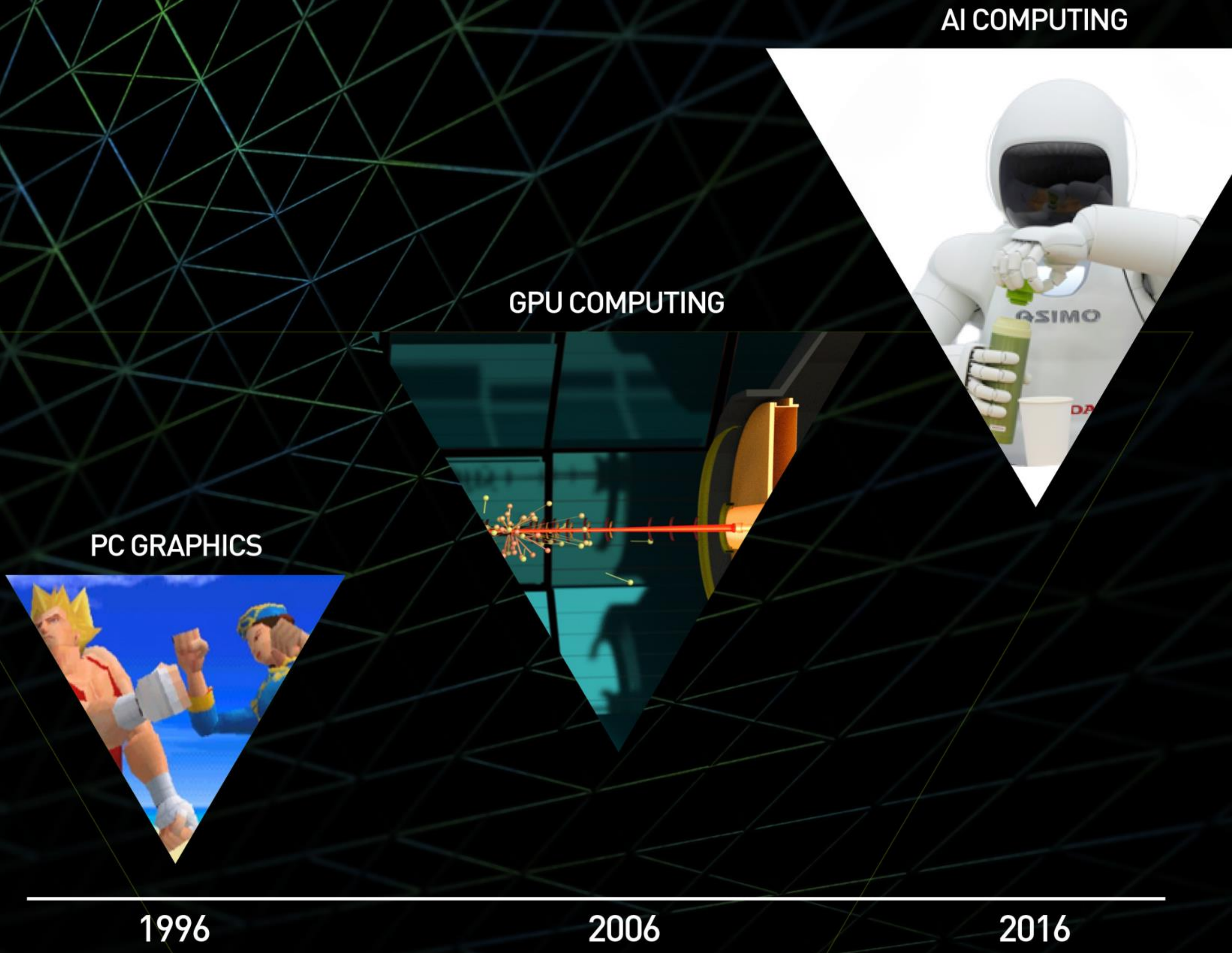
AGENDA

- NVIDIA Introduction
- Tesla for HPC and AI
- GRID vGPU for Virtualization environment
- Quadro for Professional visualization

NVIDIA
A LEARNING MACHINE

GPU sparked the growth of PC gaming, redefined modern computer graphics, and revolutionized parallel computing

GPU deep learning ignited modern AI — the next era of computing



PC GRAPHICS

GPU COMPUTING

AI COMPUTING

1996

2006

2016

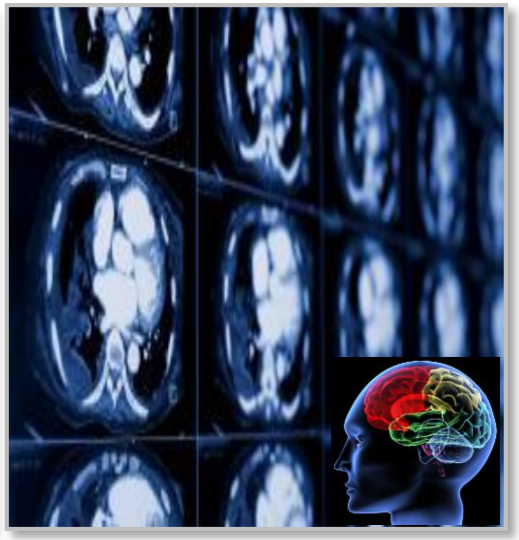
TESLA FOR HPC & AI

DEEP LEARNING IS SWEEPING ACROSS INDUSTRIES

Internet Services

- ➢ Image/Video classification
- ➢ Speech recognition
- ➢ Natural language processing

Medicine



- ➢ Cancer cell detection
- ➢ Diabetic grading
- ➢ Drug discovery

Media & Entertainment

- ➢ Video captioning
- ➢ Content based search
- ➢ Real time translation

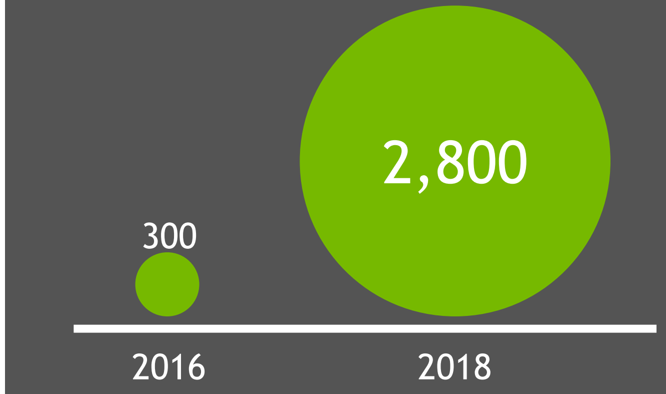
Security & Defense

- ➢ Face recognition
- ➢ Video surveillance
- ➢ Cyber security

Autonomous Machines

- ➢ Pedestrian detection
- ➢ Lane tracking
- ➢ Recognize traffic signs

MOST ADOPTED PLATFORM FOR ACCELERATING AI



9X STARTUPS ENGAGED
VIA INCEPTION PROGRAM

Caffe2

Chainer

KALDI

Microsoft
Cognitive
Toolkit

mxnet

PaddlePaddle

PYTORCH

TensorFlow

EVERY DEEP LEARNING
FRAMEWORK ACCELERATED

Alibaba Cloud
aliyun.com

aws

Google Cloud

IBM Cloud

Microsoft
Azure

Tencent Cloud

Cloud Services

DELL

Hewlett Packard
Enterprise

IBM

inspur

Lenovo

SUPERMICRO

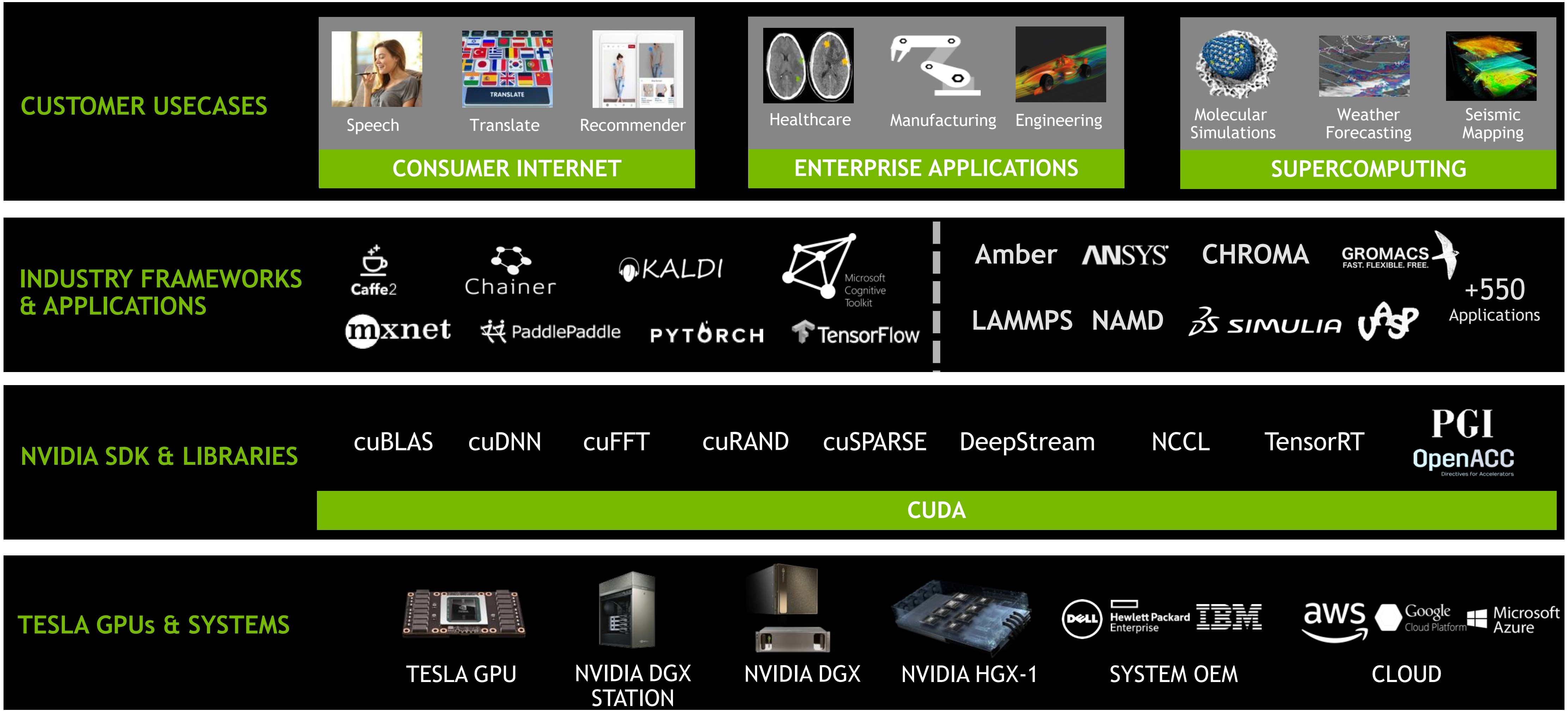
Systems

Desktops

AVAILABLE EVERYWHERE

TESLA STACK

World’s Leading Data Center Platform for Accelerating HPC and AI



CUSTOMER USECASES

Speech · Translate · Recommender — CONSUMER INTERNET

Healthcare · Manufacturing · Engineering — ENTERPRISE APPLICATIONS

Molecular Simulations · Weather Forecasting · Seismic Mapping — SUPERCOMPUTING

INDUSTRY FRAMEWORKS & APPLICATIONS

Caffe2 · Chainer · KALDI · Microsoft Cognitive Toolkit · mxnet · PaddlePaddle · PYTORCH · TensorFlow

Amber · ANSYS · CHROMA · GROMACS FAST. FLEXIBLE. FREE. · LAMMPS · NAMD · SIMULIA · VASP · +550 Applications

NVIDIA SDK & LIBRARIES

cuBLAS · cuDNN · cuFFT · cuRAND · cuSPARSE · DeepStream · NCCL · TensorRT · PGI OpenACC Directives for Accelerators

CUDA

TESLA GPUs & SYSTEMS

TESLA GPU · NVIDIA DGX STATION · NVIDIA DGX · NVIDIA HGX-1

SYSTEM OEM: DELL · Hewlett Packard Enterprise · IBM

CLOUD: aws · Google Cloud Platform · Microsoft Azure

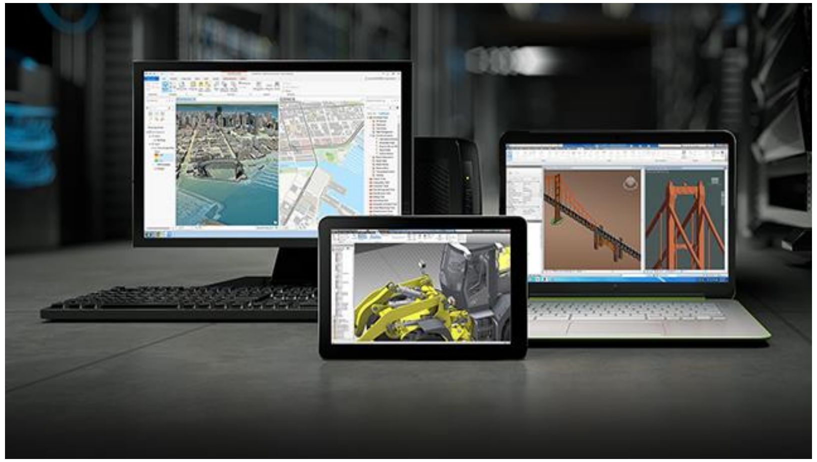
GRID VCPU FOR VIRTUALIZATION ENVIRONMENT

GPU VIRTUALIZATION FOR ANY WORKLOAD

NVIDIA delivers GPU virtualization for both graphics and compute workloads

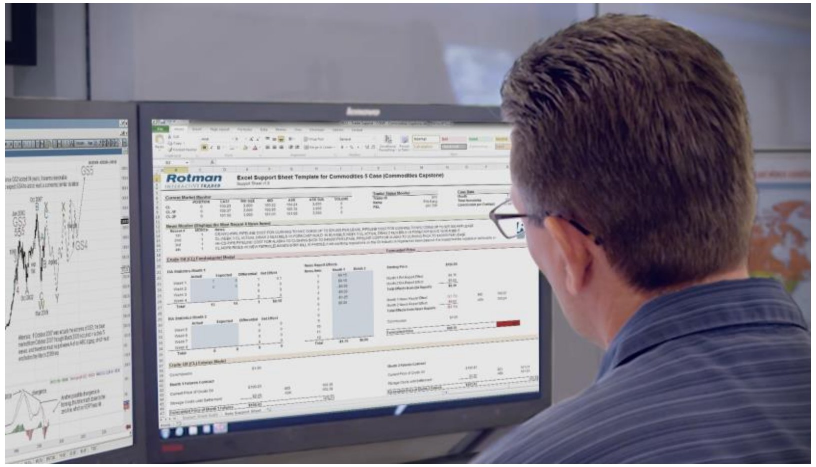
NVIDIA VIRTUAL GPU

Most Powerful Virtual Workstation



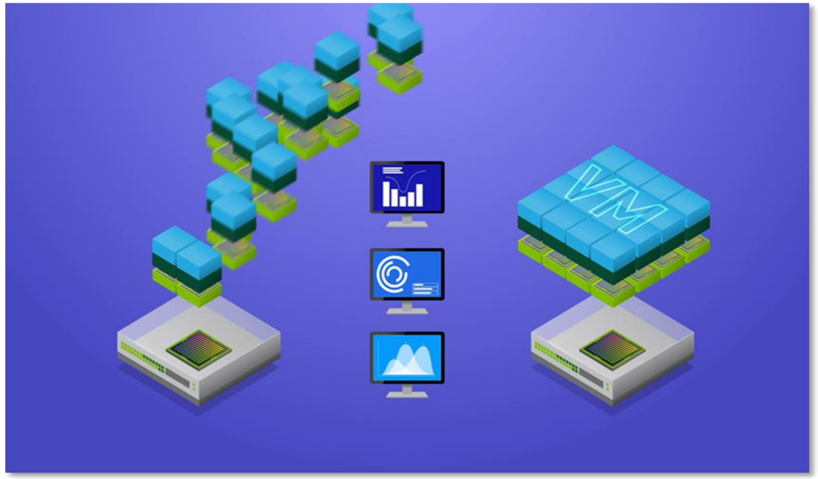
Quadro vDWS on Tesla V100, the most advanced data center GPU

Best Virtual PC for Knowledge Workers



Expanded GRID vPC delivers more value to more enterprise users

Agile GPU Accelerated Data Center



Live Migration for the Mission Critical Data Center

SOLUTION FOR EVERY USE CASE

Creative & Technical Professional

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

AUTODESK AUTOCAD · SOLIDWORKS · Petrel

CATIA · AUTODESK REVIT · ArcGIS ESRI

PLM Software SIEMENS NX · AUTODESK MAYA · 3DEXCITE

Knowledge Worker

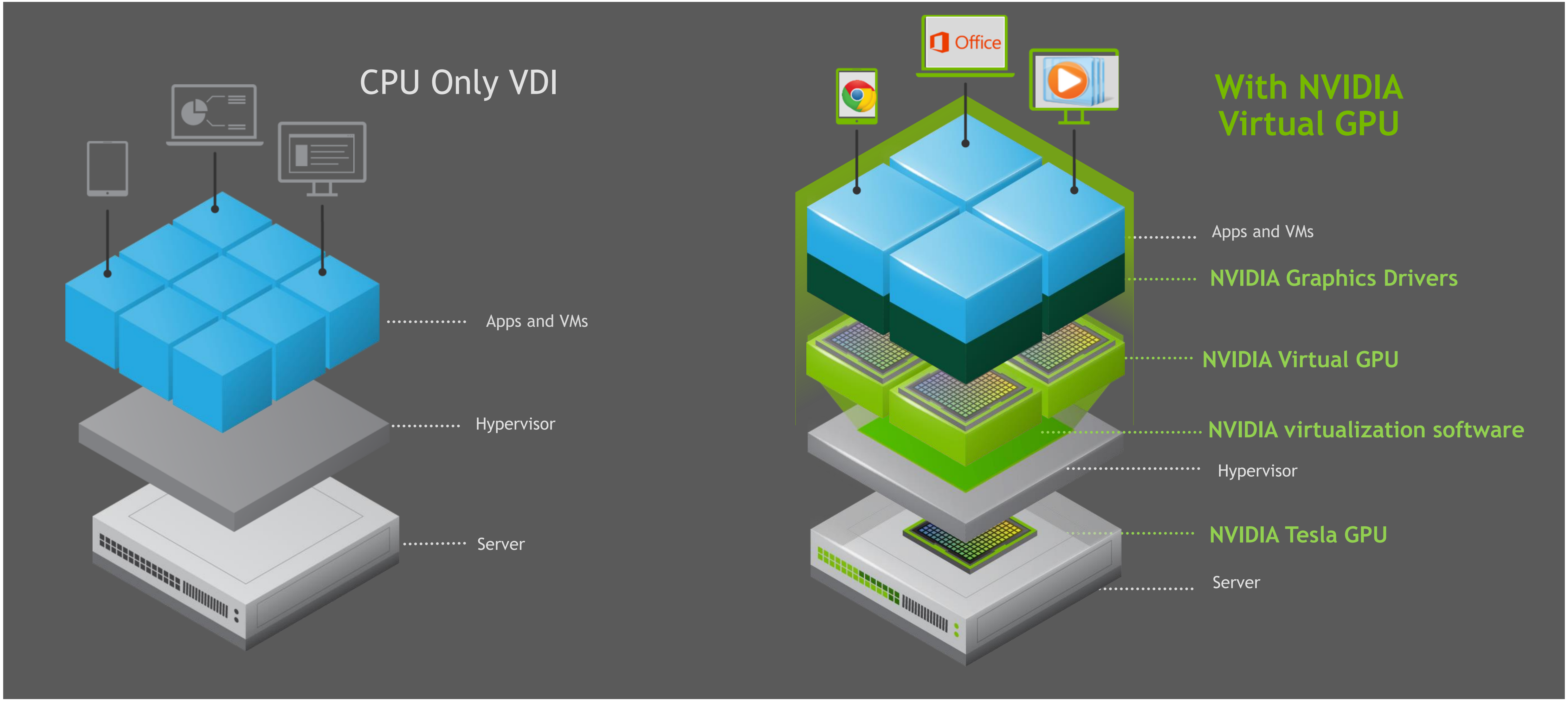
Providing business users the highest level of experience for all their apps on any device

Office · Ai · Ps · Id

Performance ↔ Scalability

HOW IT WORKS

NVIDIA virtual GPU products deliver a GPU Experience to every Virtual Desktop

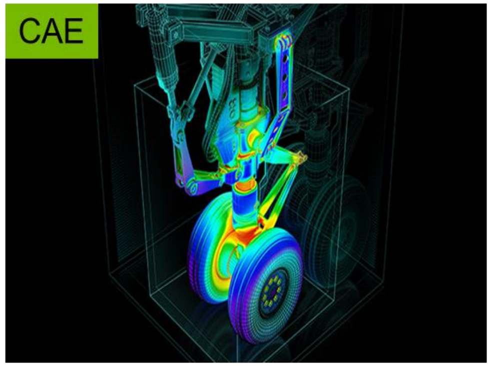


QUADRO FOR PROFESSIONAL VISUALIZATION

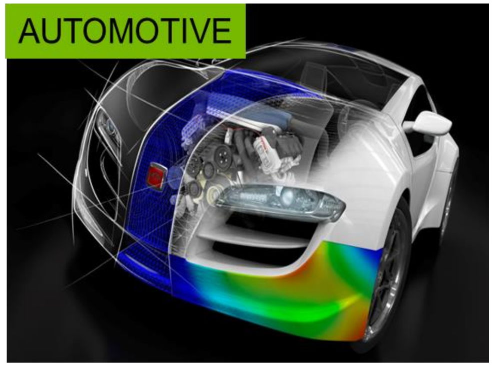
DEMAND FROM THE WORLD’S LARGEST INDUSTRIES

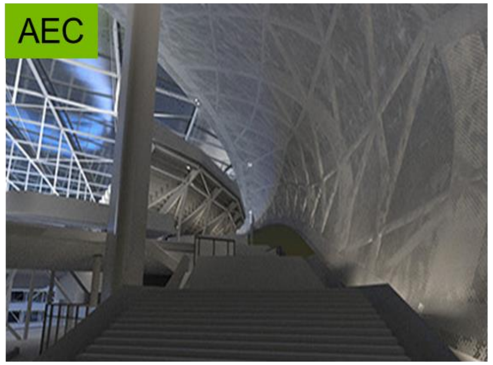
- Desire to work with full-fidelity models driving dramatic increase in data size
- Demand for photorealism throughout the design process for better insights
- Need for immersive environments for faster, higher-quality decision making
- Robust platform for rapid AI research, development & deployment

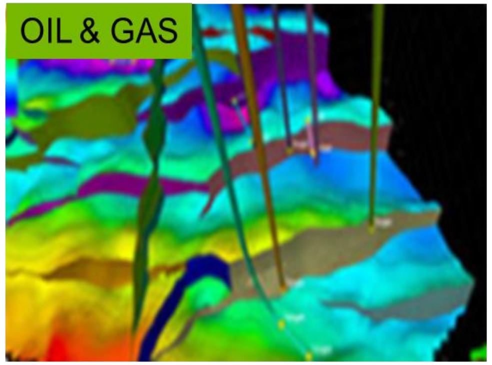
MANUFACTURING


CAE

MEDIA & ENTERTAINMENT


AUTOMOTIVE


AEC


OIL & GAS

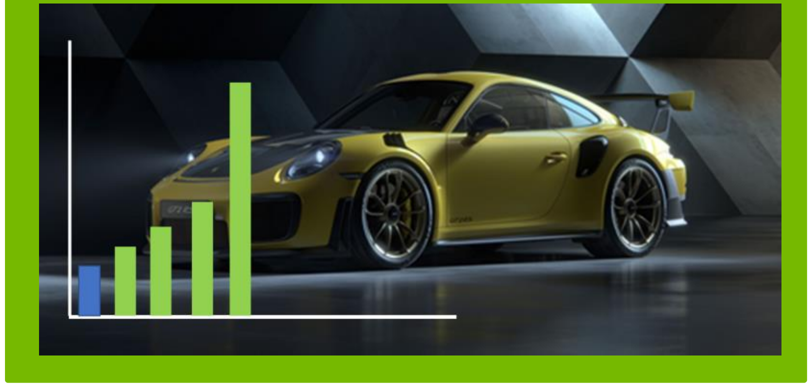
SCI-VIZ

HEALTHCARE

NVIDIA QUADRO GV100

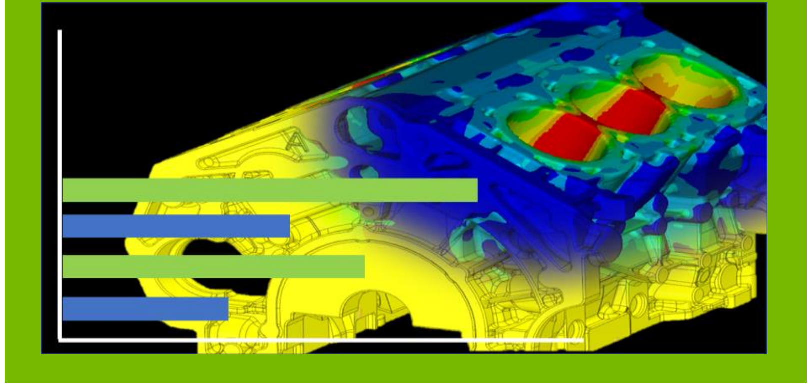
- Delivers world’s fastest real-time ray tracing on workstations for millions of artists & designers
- Opens new opportunities with accelerated AI for training and inference on workstations
- Accelerates CAE workloads for simulation analysts to bring optimal products to market faster

RAY TRACING



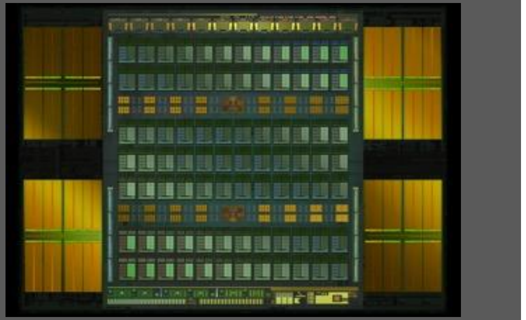
AI

SIMULATION



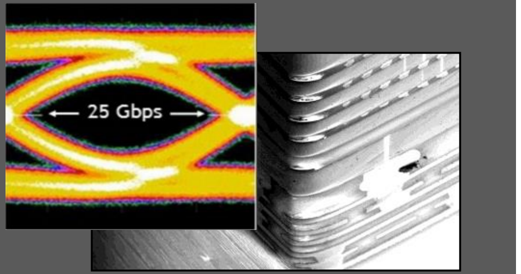
TECHNOLOGIES POWERING THE AI WORKSTATION

Volta GPU



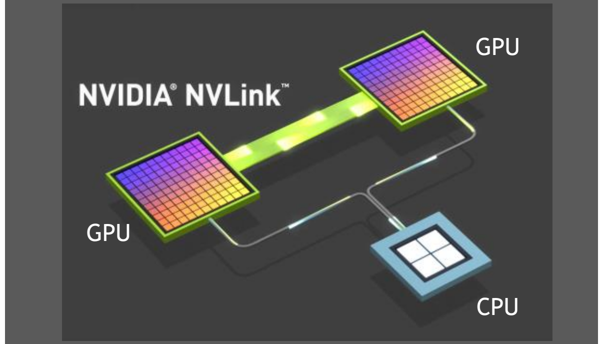
Efficient, Powerful

HBM2



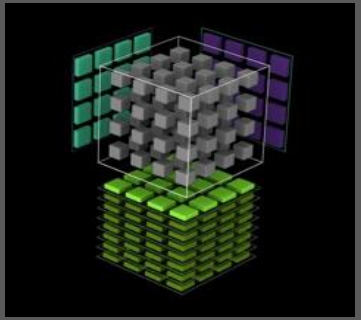
Largest GPU Memory

NVLink



Scalable Performance

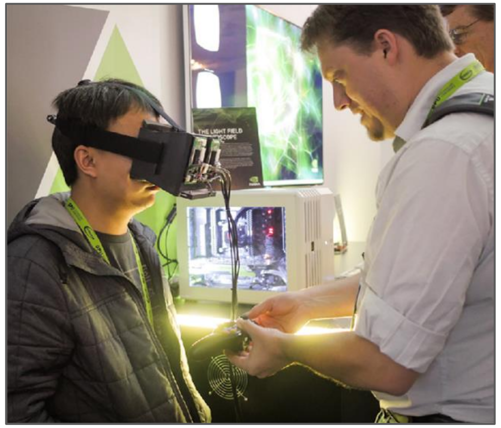
Tensor Cores

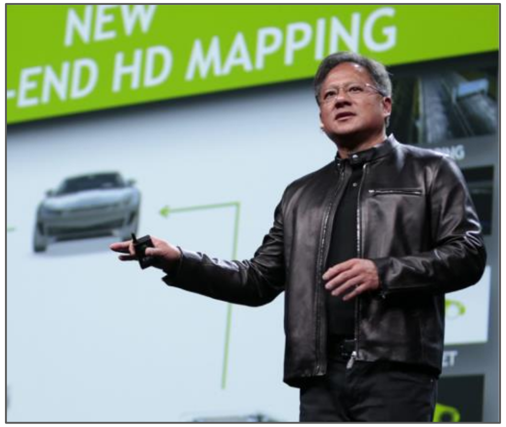


Accelerated AI

GPU TECHNOLOGY CONFERENCE

優惠代碼：GTCTWCHCH





2018年全球巡迴首站 – 台灣！

5月30、31日
探索各項人工智慧
與深度學習的頂級盛會

NVIDIA

NVIDIA