SIE

NVIDIA

HPC + Artificial Intelligence to help overcome real life challenges:

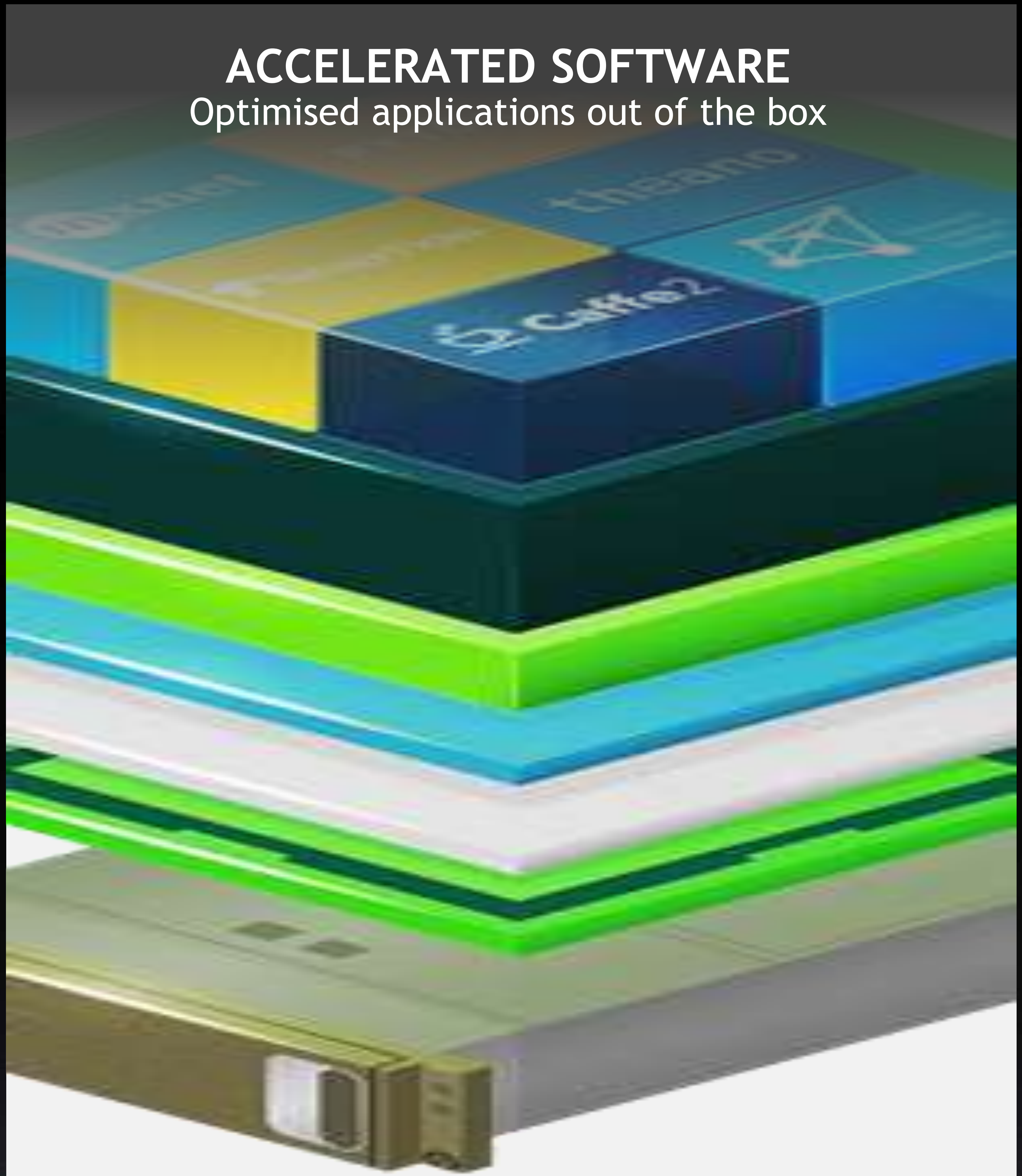
HOW NVIDIA ENGAGES IN DIFFERENT INDUSTRIES

HPC +AI INFRASTRUCTURE

Designing for customer use-cases

ACCELERATED SOFTWARE

Optimised applications out of the box



LIGHTHOUSE ENGAGEMENTS

Driving the field and gaining deep insight

HPC Across Industries

NVIDIA GPUs are optimizing over 700 applications across a broad range of industries and domains. See how GPU technology is tackling complex problems and transforming the global research community.

Supercomputing

Exploring supernova explosions. Mapping the Earth's interior. Predicting hurricanes. NVIDIA is powering the world's fastest supercomputers and HPC systems, giving researchers the power they need to simulate and predict our world.

Learn More ›

Healthcare & Life Sciences

Discovering drugs. Uncovering genetic mutations. Analyzing images. NVIDIA is equipping the world's leading healthcare institutions with advanced tools to accelerate precision medicine and build next-generation clinics.

Learn More ›

Energy

Producing energy. Refining and distributing oil. Reducing environmental impact. NVIDIA technologies are impacting world economies by fueling innovation in energy and enhancing individual ways of life.

Learn More ›

Public Sector

Cybersecurity. Disaster response. Humanitarian assistance. NVIDIA is building the technology for our world that will make communities safer and more connected everywhere.

Learn More ›

Run GPU-Accelerated Apps

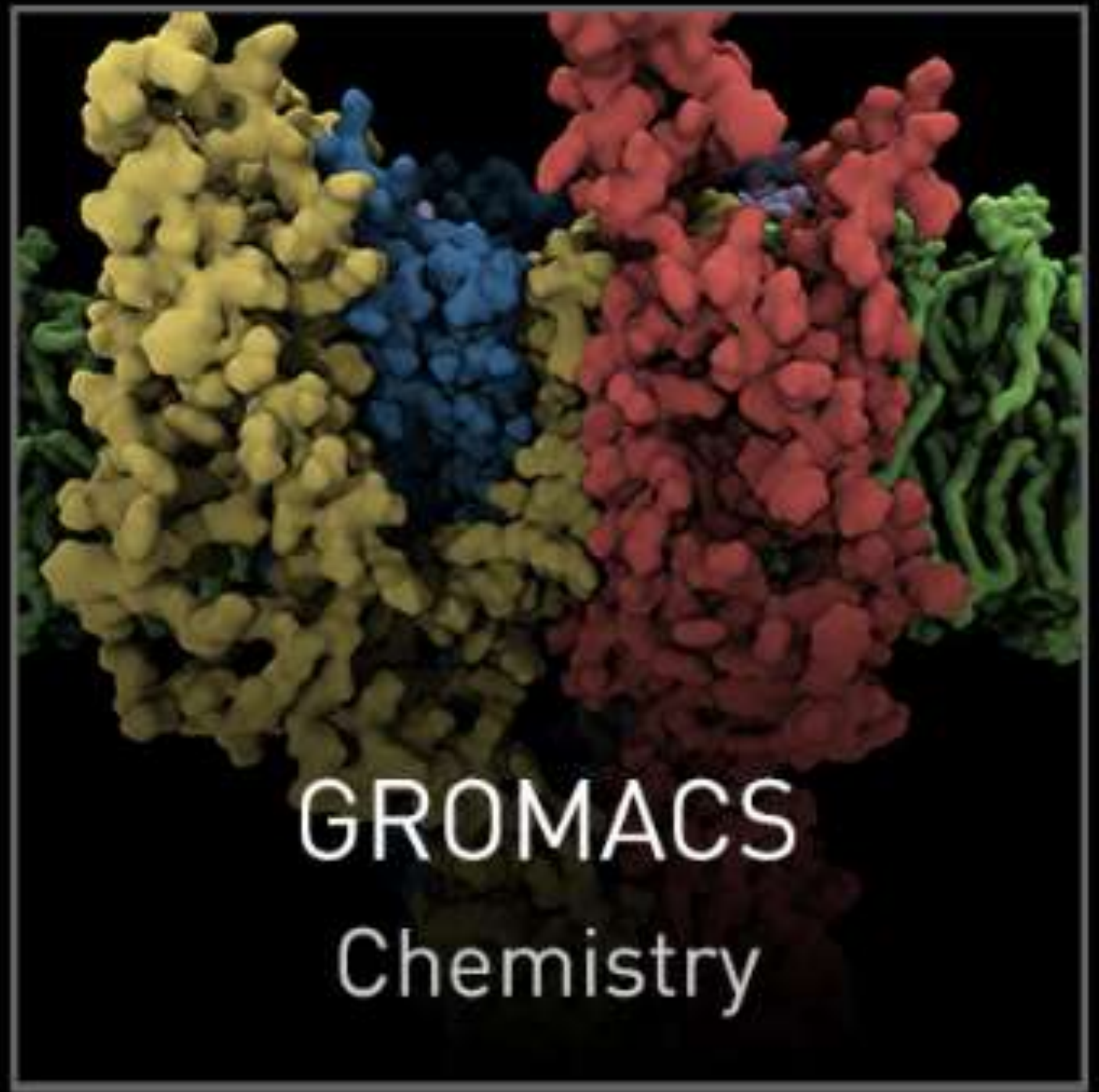
From weather prediction and materials science to wind tunnel simulation and genomics, NVIDIA GPU-accelerated computing is at the heart of HPC's most promising areas of discovery.

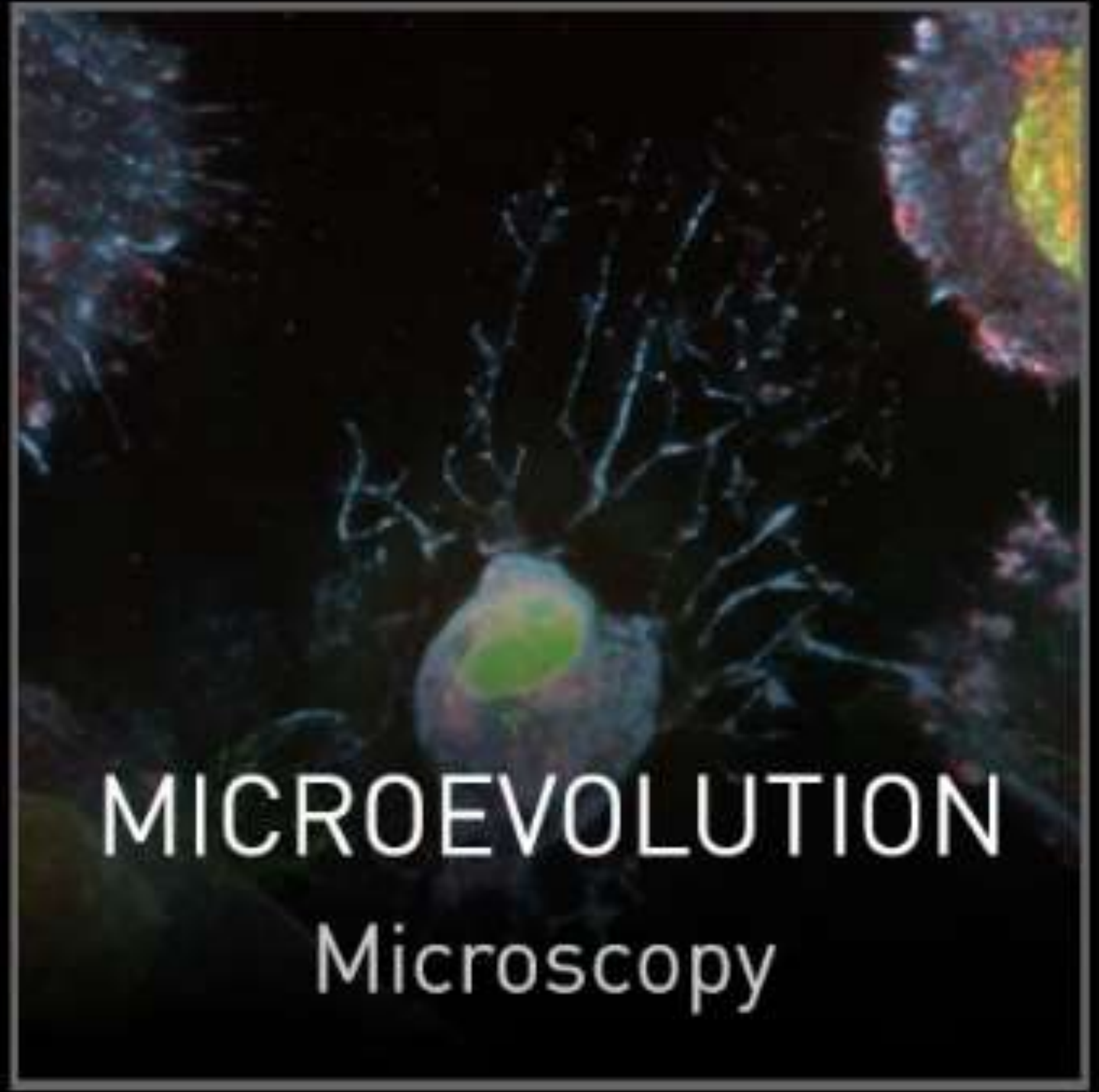
The NVIDIA CUDA® programming model is the platform of choice for high-performance application developers, with support for more than **700 GPU-accelerated applications**—including the top 15 HPC applications developers. Many of the top HPC applications are made available as pre-configured, containerized software on NGC.

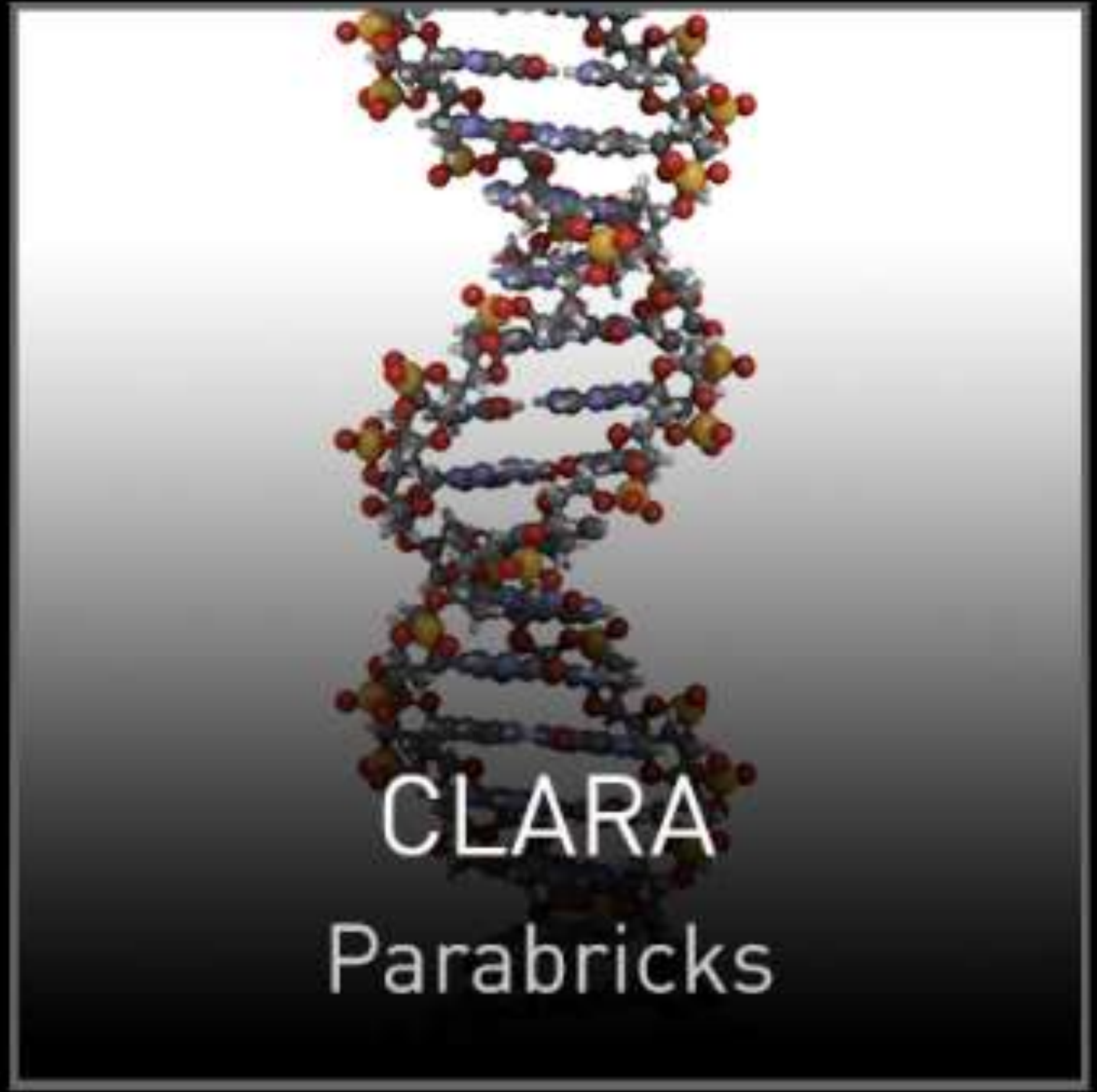
ⓘ **See HPC Application Performance**

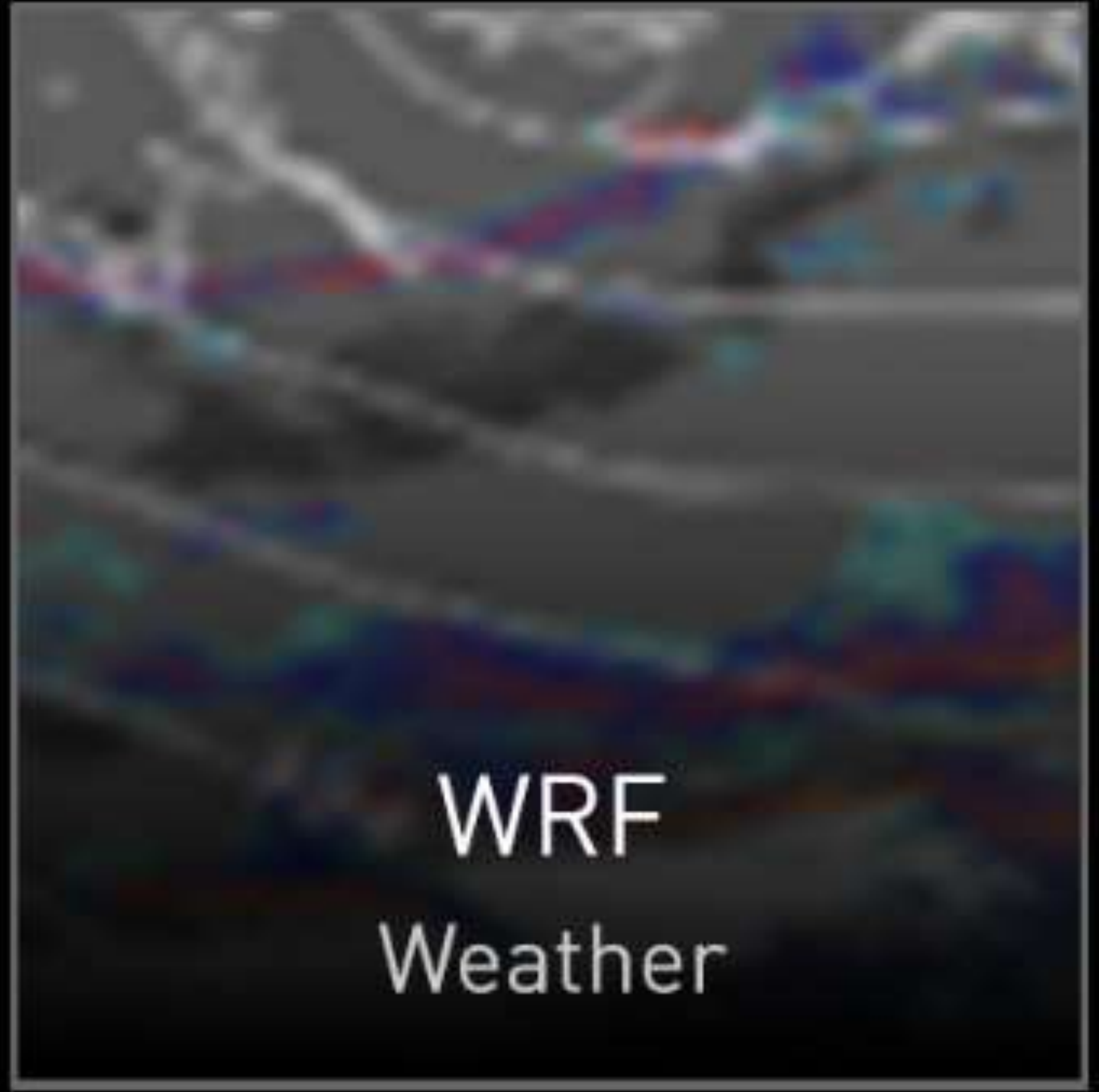
ⓘ **Explore HPC Software**

ⓘ **Explore Containers Available in NGC**

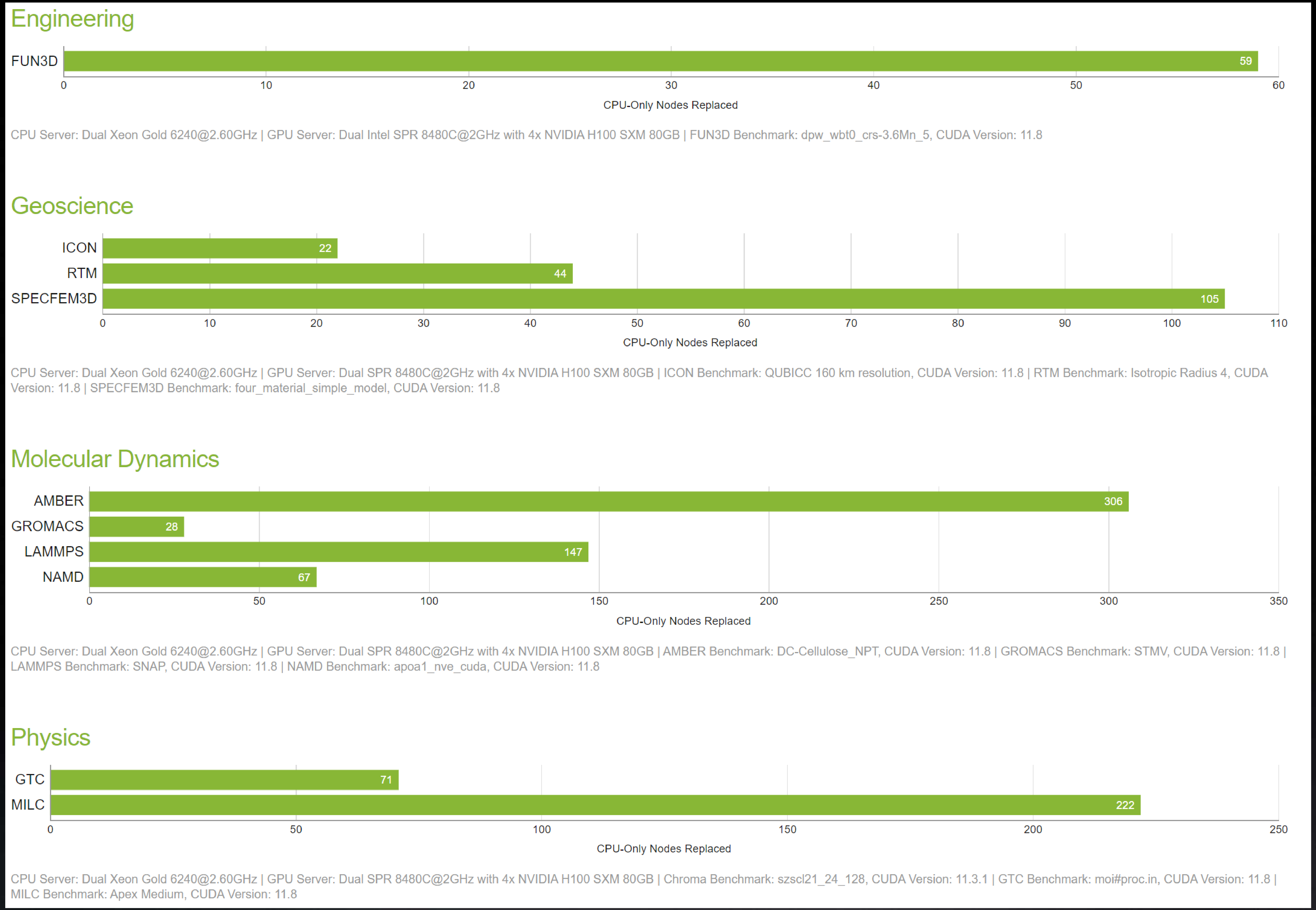








Classical HPC Modelling accelerated computing performance improvement,…



Engineering

CPU Server: Dual Xeon Gold 6240@2.60GHz | GPU Server: Dual Intel SPR 8480C@2GHz with 4x NVIDIA H100 SXM 80GB | FUN3D Benchmark: dpw_wbt0_crs-3.6Mn_5, CUDA Version: 11.8

Geoscience

CPU Server: Dual Xeon Gold 6240@2.60GHz | GPU Server: Dual SPR 8480C@2GHz with 4x NVIDIA H100 SXM 80GB | ICON Benchmark: QUBICC 160 km resolution, CUDA Version: 11.8 | RTM Benchmark: Isotropic Radius 4, CUDA Version: 11.8 | SPECFEM3D Benchmark: four_material_simple_model, CUDA Version: 11.8

Molecular Dynamics

CPU Server: Dual Xeon Gold 6240@2.60GHz | GPU Server: Dual SPR 8480C@2GHz with 4x NVIDIA H100 SXM 80GB | AMBER Benchmark: DC-Cellulose_NPT, CUDA Version: 11.8 | GROMACS Benchmark: STMV, CUDA Version: 11.8 | LAMMPS Benchmark: SNAP, CUDA Version: 11.8 | NAMD Benchmark: apoa1_nve_cuda, CUDA Version: 11.8

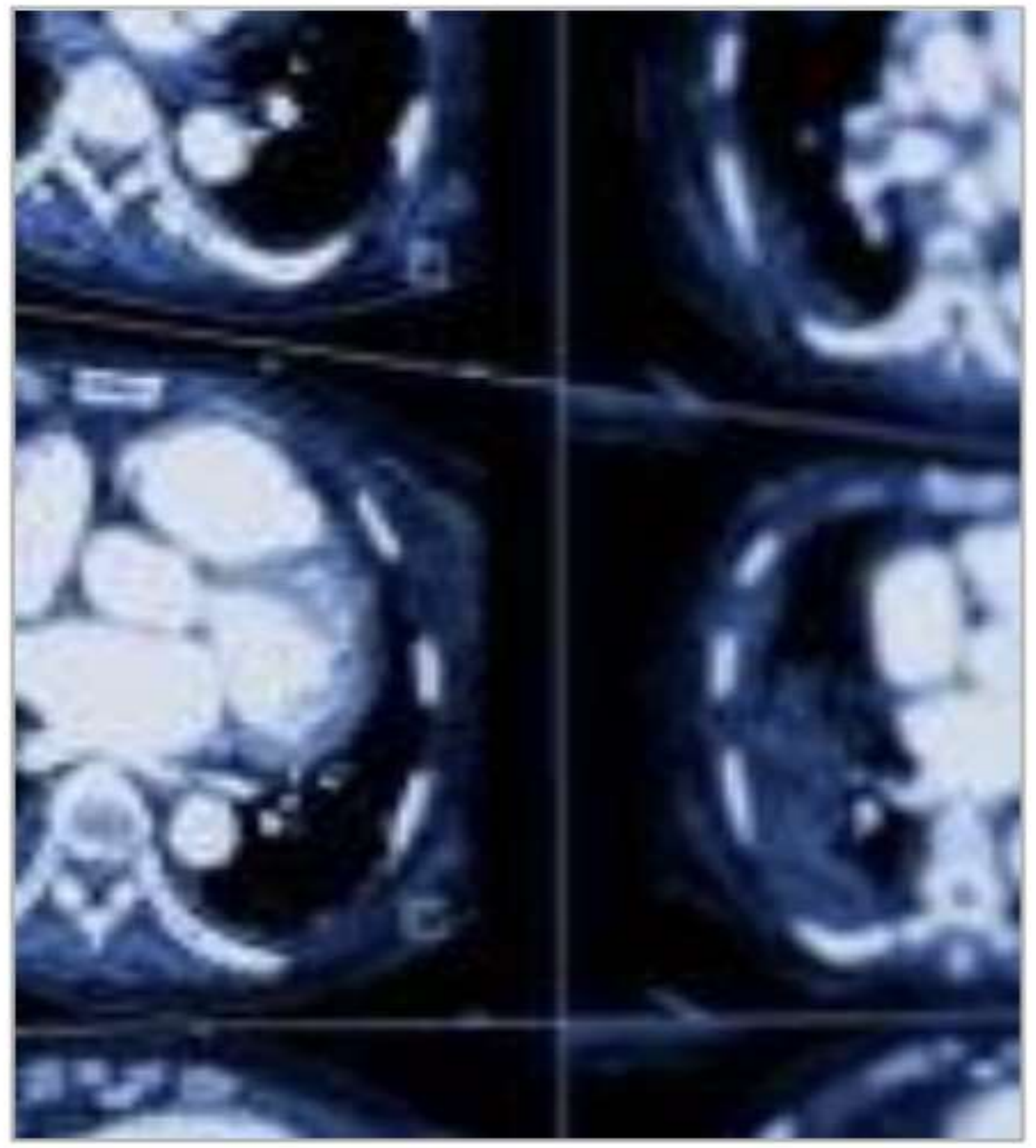
Physics

CPU Server: Dual Xeon Gold 6240@2.60GHz | GPU Server: Dual SPR 8480C@2GHz with 4x NVIDIA H100 SXM 80GB | Chroma Benchmark: szscl21_24_128, CUDA Version: 11.3.1 | GTC Benchmark: moi#proc.in, CUDA Version: 11.8 | MILC Benchmark: Apex Medium, CUDA Version: 11.8

Classical HPC Modelling accelerated computing is great but,…

DEEP LEARNING IS SWEEPING ACROSS INDUSTRIES

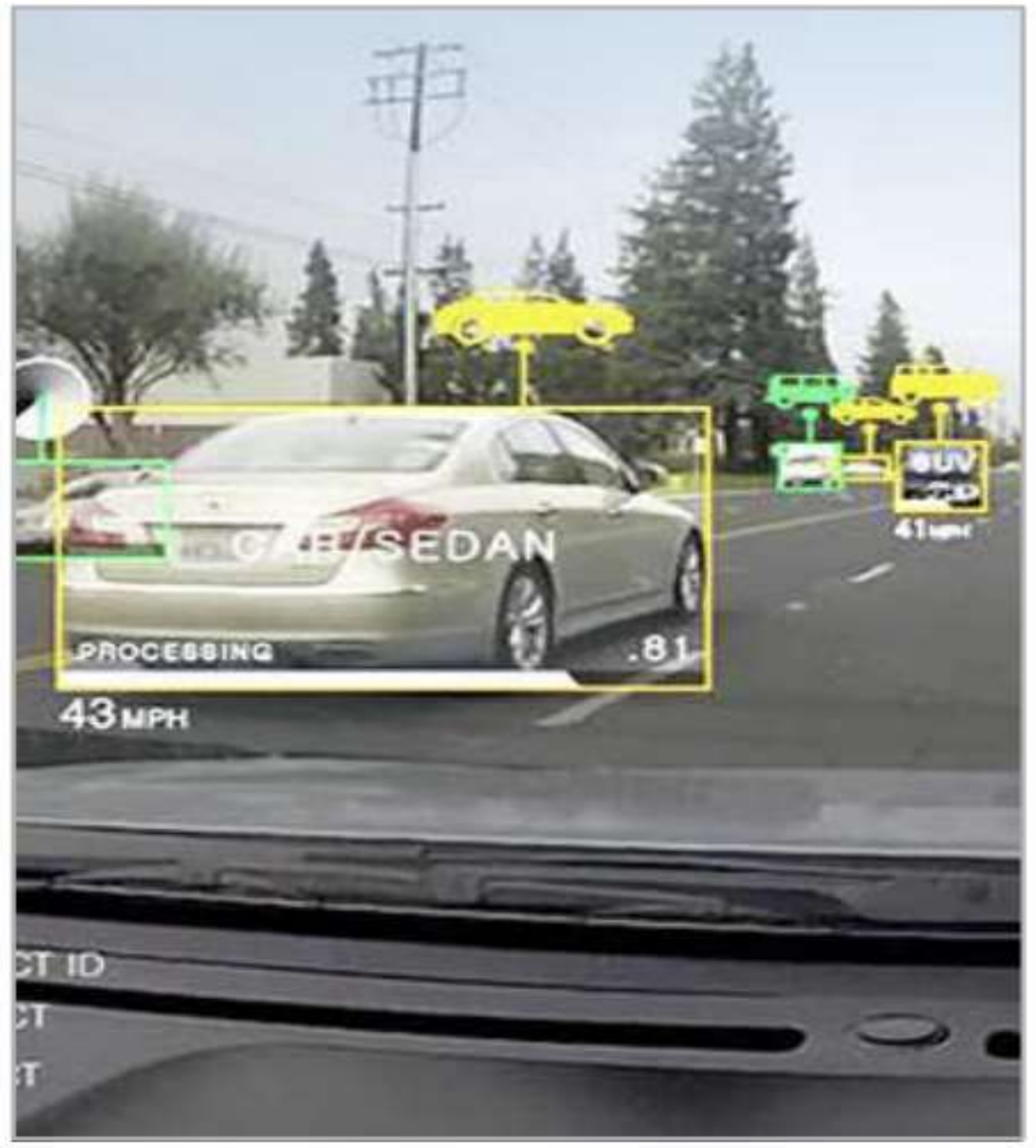
Internet Services
Image/Video Classification
Speech Recognition
Natural Language Processing



Medicine
Cancer Cell Detection,
Diabetic Grading,
Drug Discovery

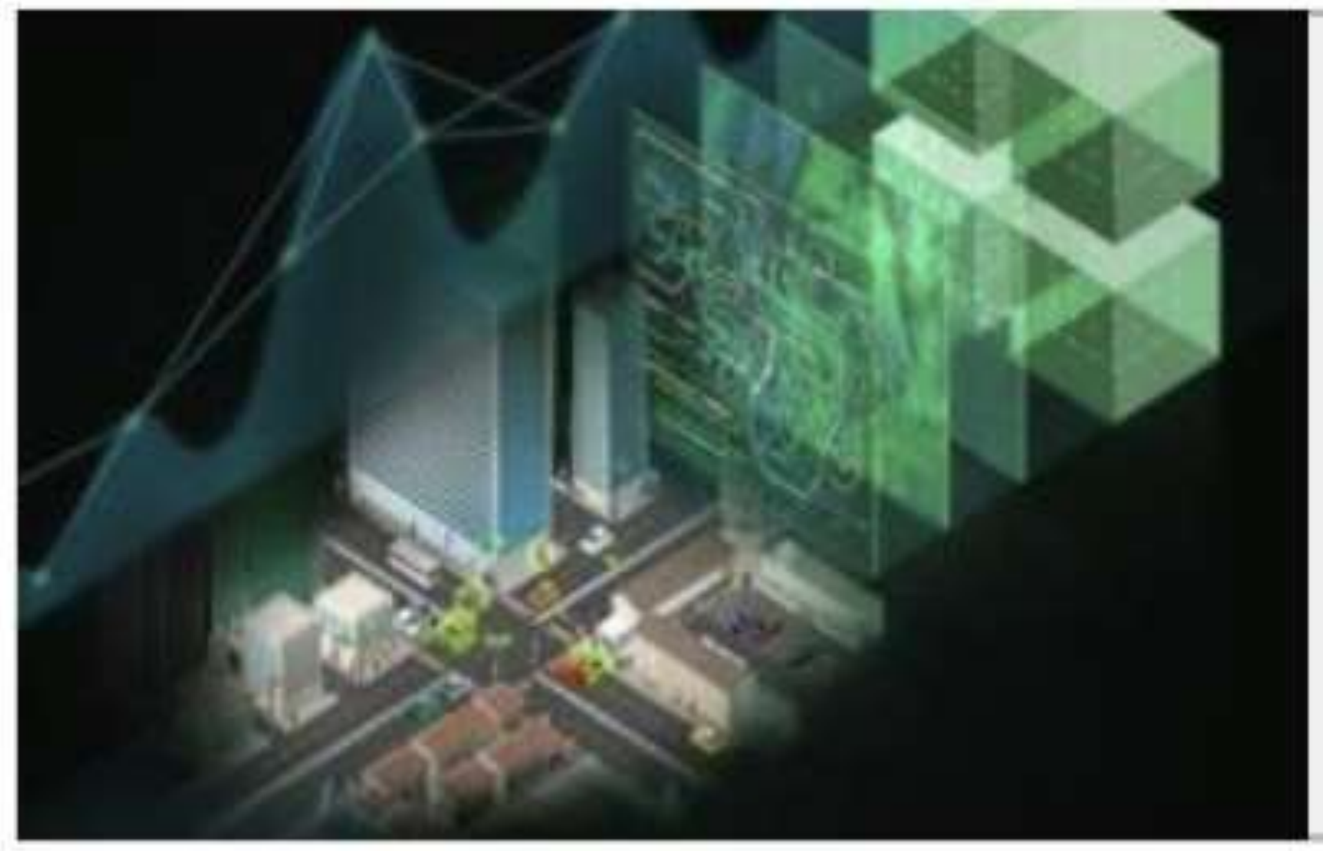
Media & Entertainment
Video Captioning
Content Based Search
Real Time Translation

Security & Defense
Face Recognition
Video Surveillance
Cyber Security



Autonomous Machines
Pedestrian Detection
Lane Tracking
Recognize Traffic Signs

Use Cases in Every Industry



CONSUMER INTERNET

Ad Personalization

Click Through Rate Optimization

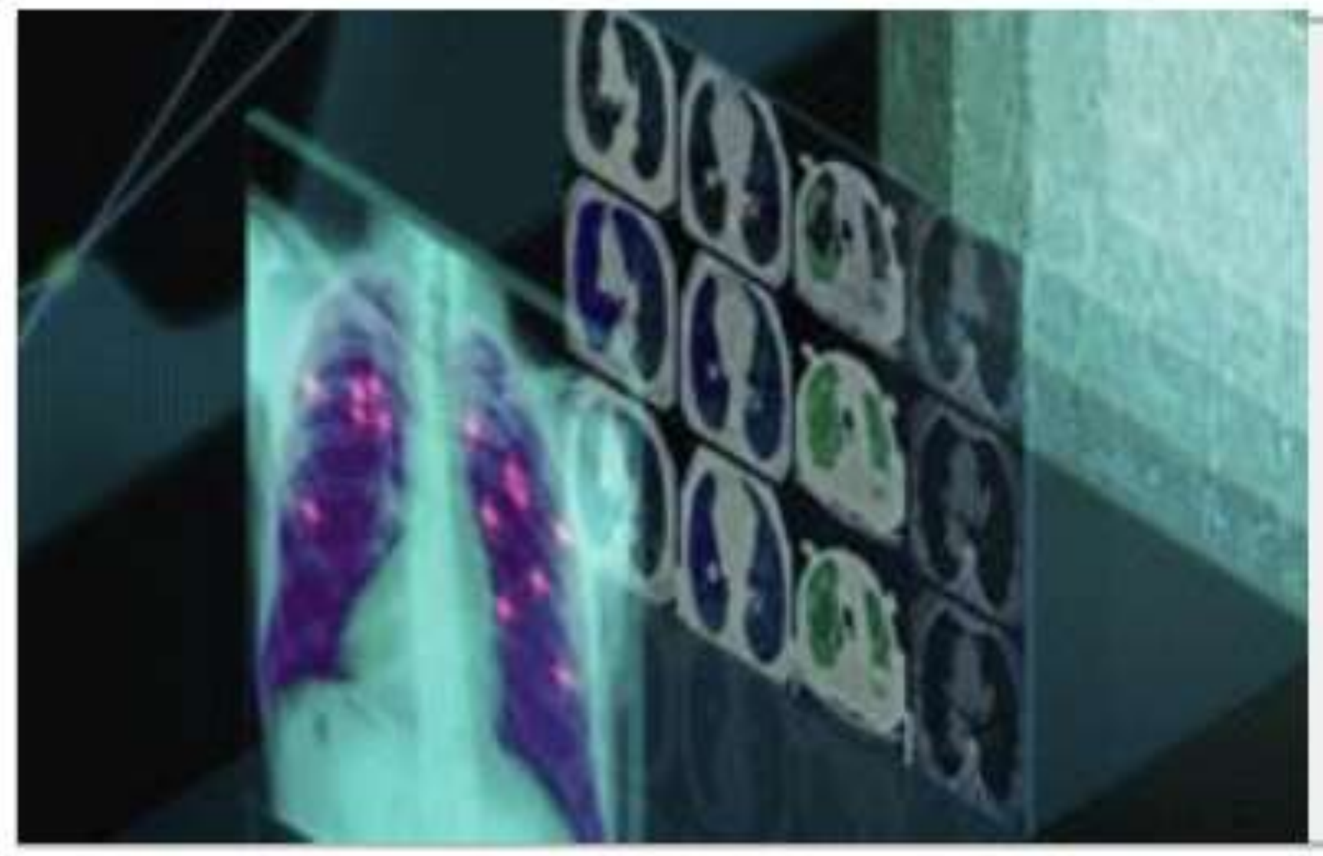
Churn Reduction

FINANCIAL SERVICES

Claim Fraud

Customer Service Chatbots/Routing

Risk Evaluation



HEALTHCARE

Improve Clinical Care

Drive Operational Efficiency

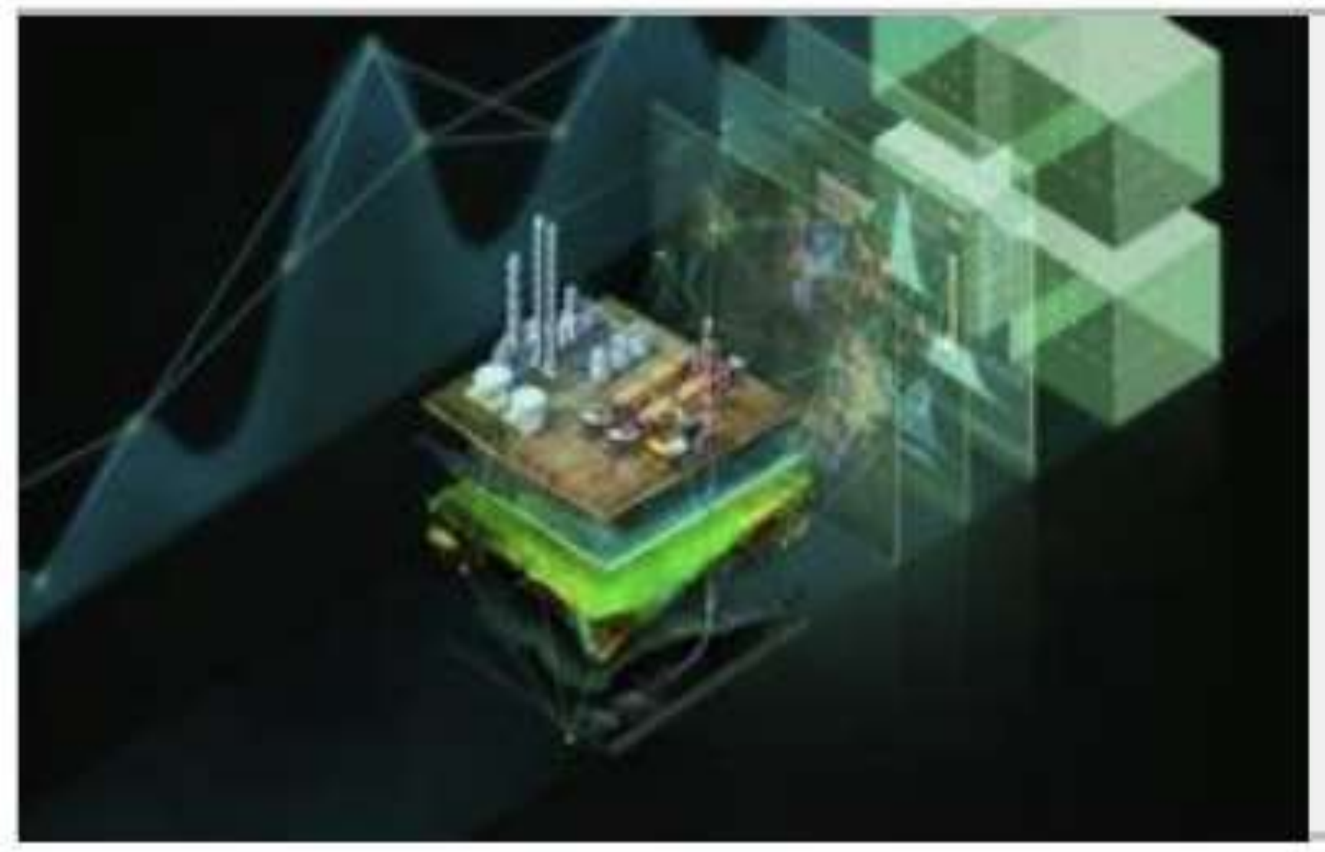
Speed Up Drug Discovery

RETAIL

Supply Chain & Inventory Management

Price Management / Markdown Optimization

Promotion Prioritization And Ad Targeting

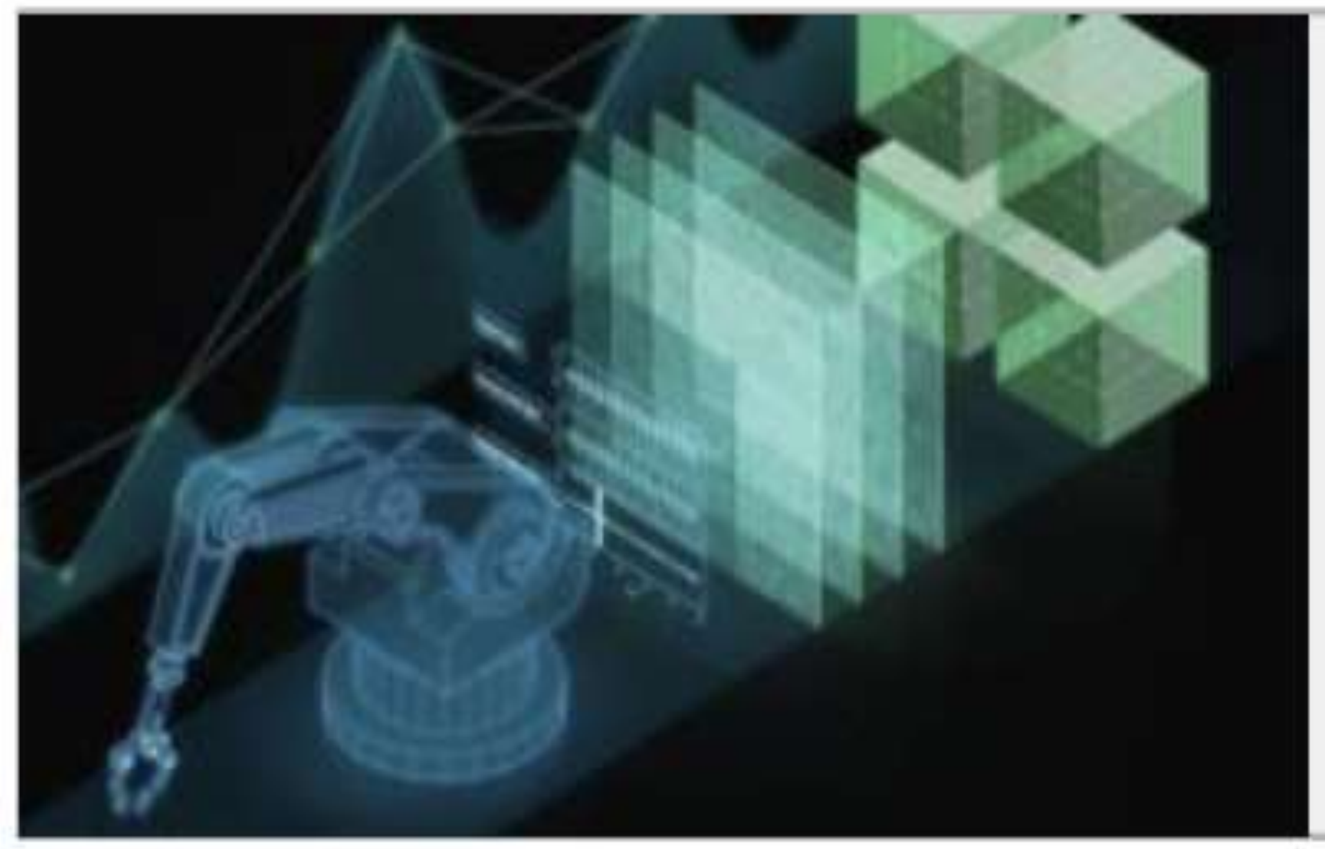


OIL & GAS

Sensor Data Tag Mapping

Anomaly Detection

Robust Fault Prediction



MANUFACTURING

Remaining Useful Life Estimation

Failure Prediction

Demand Forecasting

TELECOM

Detect Network/Security Anomalies

Forecasting Network Performance

Network Resource Optimization (SON)

AUTOMOTIVE

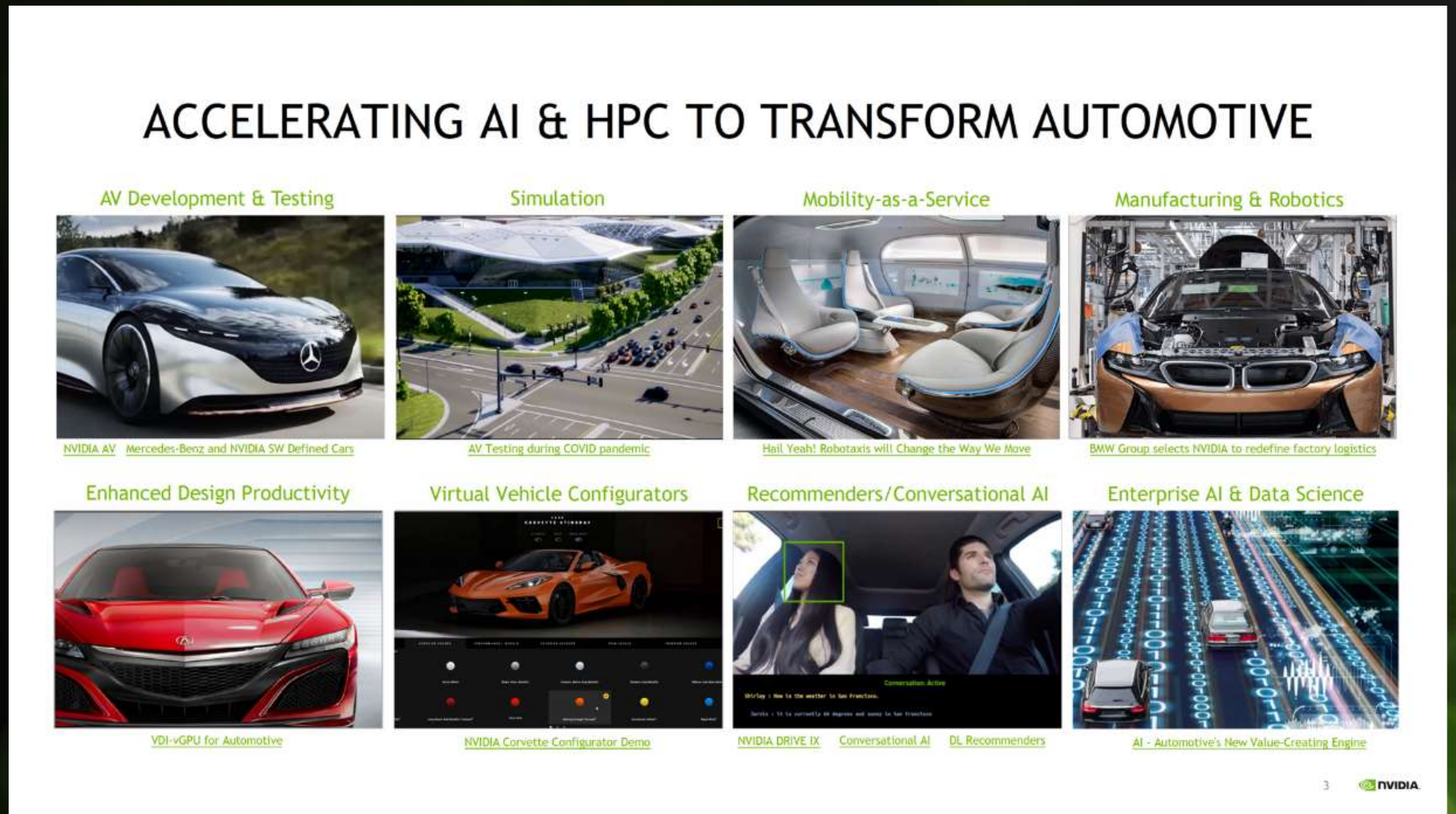
Personalization & Intelligent Customer Interactions

Connected Vehicle Predictive Maintenance

Forecasting, Demand, & Capacity Planning

Key NVIDIA AI use cases for:

- Automotive
- Financial Services (FSI)
- Energy: Oil & Gas / Utilities
- Healthcare
- Higher Education & Research (HER)
- Manufacturing
- Manufacturing Product Development
- Media and Entertainment
- Retail
- Telecommunications (Telco)
- Architecture, Engineering and Construction (AEC)
- HR and Education



ACCELERATING AI & HPC TO TRANSFORM AUTOMOTIVE

AV Development & Testing — NVIDIA AV · Mercedes-Benz and NVIDIA SW Defined Cars

Simulation — AV Testing during COVID pandemic

Mobility-as-a-Service — Hail Yeah! Robotaxis will Change the Way We Move

Manufacturing & Robotics — BMW Group selects NVIDIA to redefine factory logistics

Enhanced Design Productivity — VDI-vGPU for Automotive

Virtual Vehicle Configurators — NVIDIA Corvette Configurator Demo

Recommenders/Conversational AI — NVIDIA DRIVE IX · Conversational AI · DL Recommenders

Enterprise AI & Data Science — AI - Automotive's New Value-Creating Engine

ACCELERATING DIGITAL TRANSFORMATION IN FSI

AI/ML Optimizes Performance and Outcomes

Default Prediction

Fraud Detection

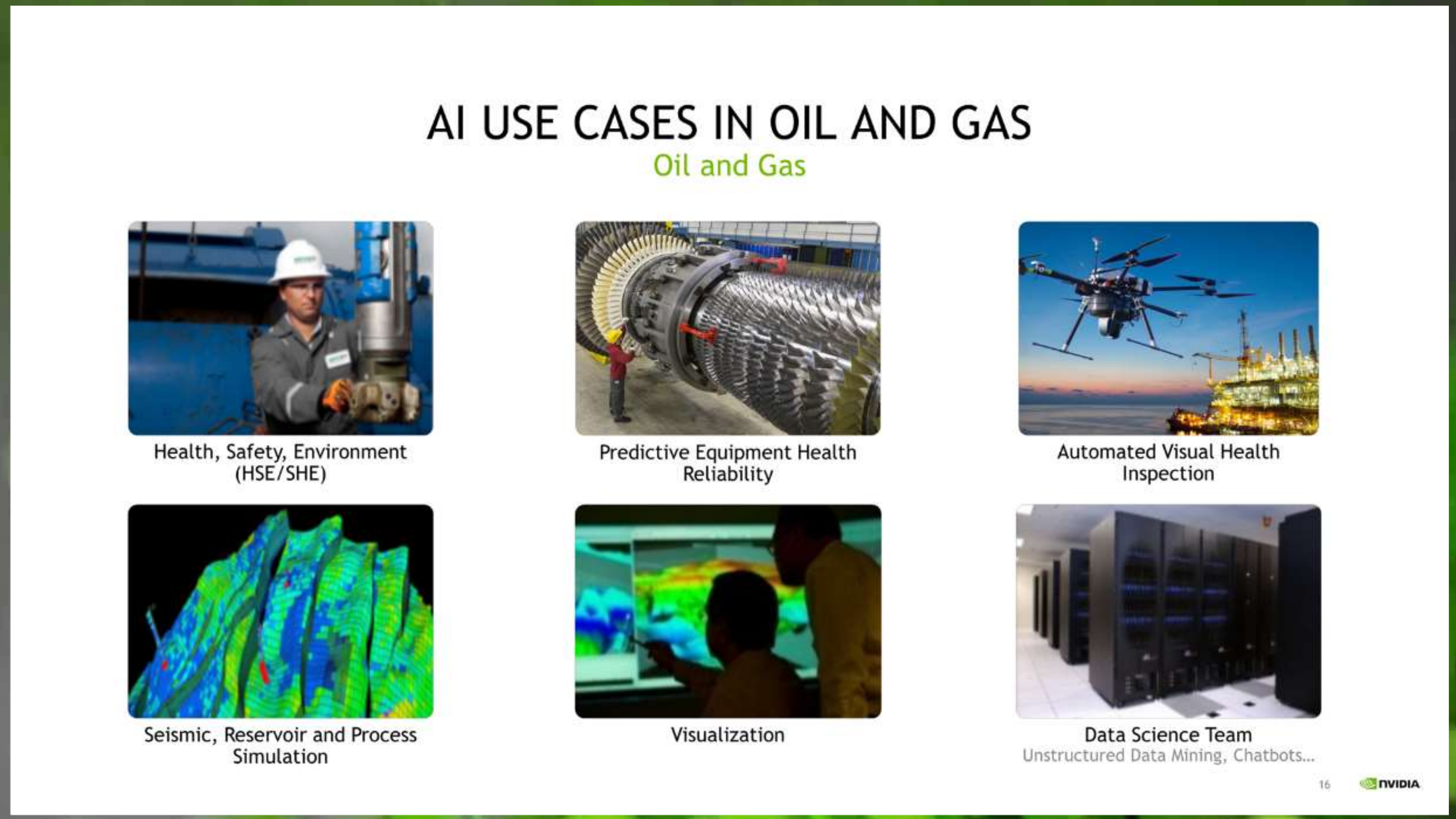
Virtualization (WFH)

Digital Payments

Recommendations

Customer Service

Algorithmic Trading



AI USE CASES IN OIL AND GAS

Oil and Gas

Health, Safety, Environment (HSE/SHE)

Predictive Equipment Health Reliability

Automated Visual Health Inspection

Seismic, Reservoir and Process Simulation

Visualization

Data Science Team
Unstructured Data Mining, Chatbots…

NVIDIA HAS SOLUTIONS FOR THE ENTIRE HEALTHCARE PATHWAY

AI Models & Frameworks | Accelerated Applications & Libraries

MEGAMOLBART: AI CHEMISTRY FRAMEWORK
Generate | Similarity Search | Property Prediction

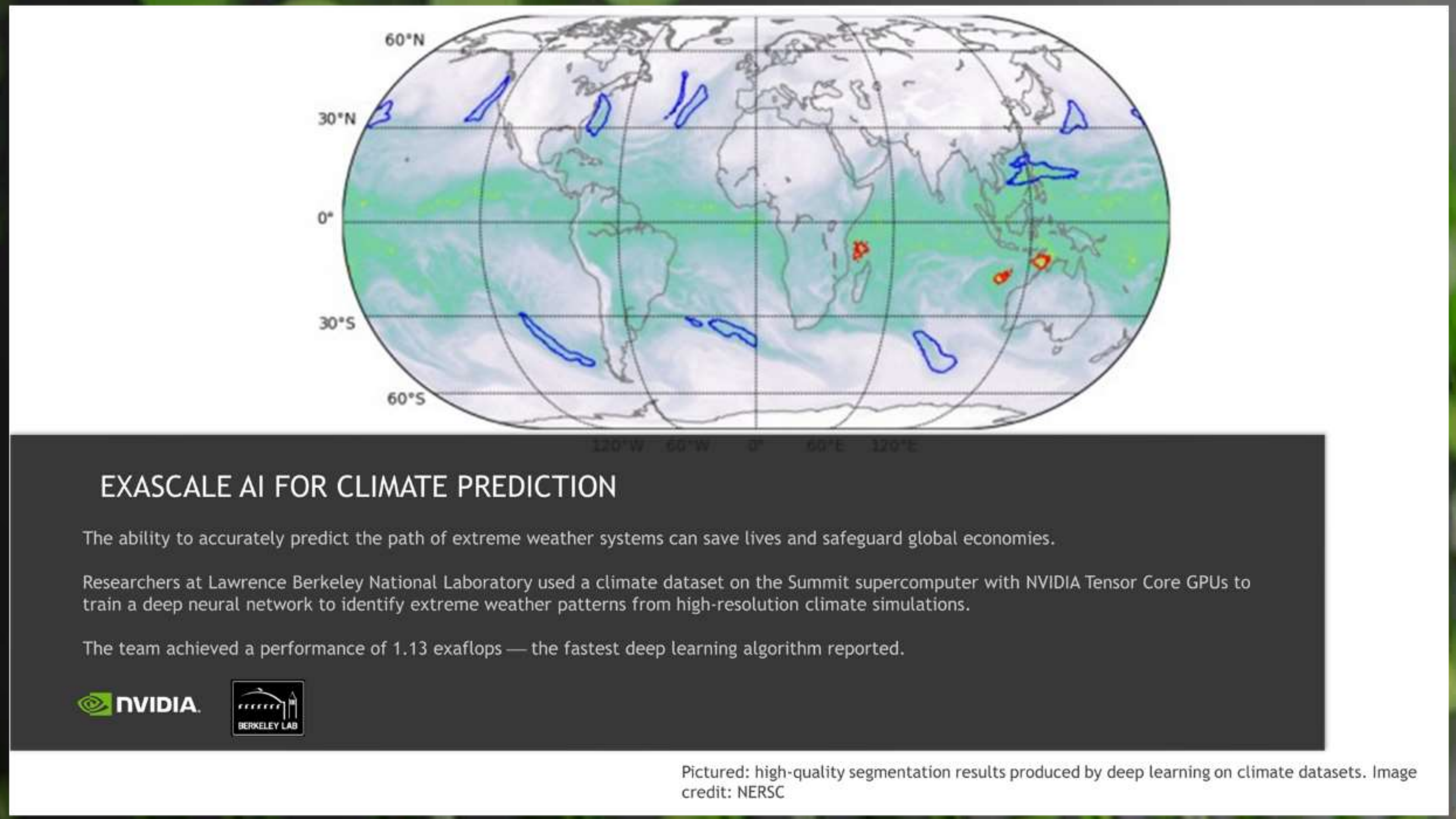
MONAI: AI IMAGING FRAMEWORK
Radiology | Pathology | Microscopy

NEMO: BIO & CLINICAL NLP FRAMEWORK
State-of-the-Art NLP Models

RAPIDS: GPU ACCELERATED DATA SCIENCE
ML | Graph Analytics | Data Prep

PARABRICKS: AI ACCELERATED GENOMICS
Accurate | Fast | Reduced Cost

OMNIVERSE: SMART HOSPITAL
Multiscale Simulation



EXASCALE AI FOR CLIMATE PREDICTION

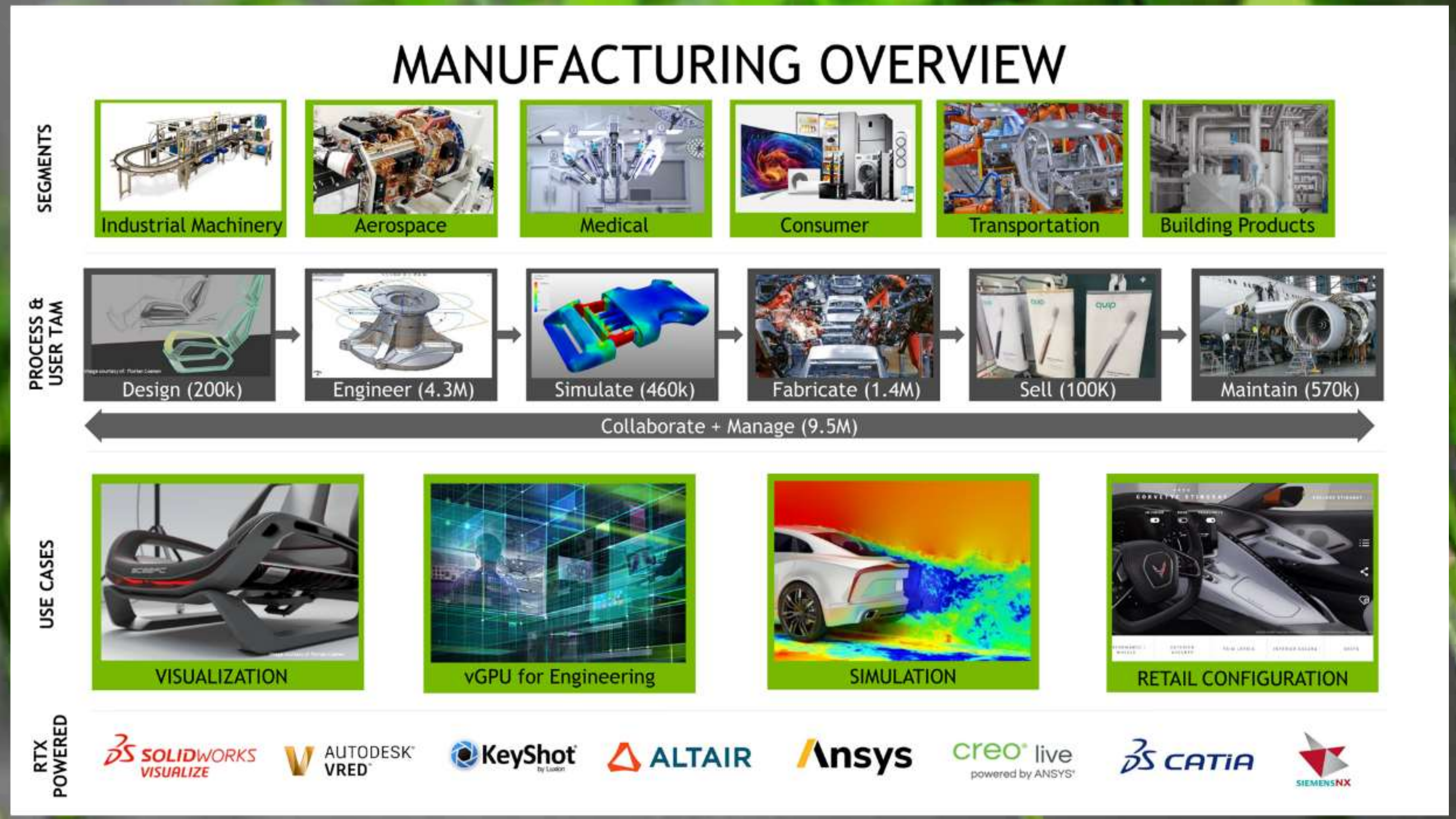
The ability to accurately predict the path of extreme weather systems can save lives and safeguard global economies.

Researchers at Lawrence Berkeley National Laboratory used a climate dataset on the Summit supercomputer with NVIDIA Tensor Core GPUs to train a deep neural network to identify extreme weather patterns from high-resolution climate simulations.

The team achieved a performance of 1.13 exaflops — the fastest deep learning algorithm reported.

NVIDIA. BERKELEY LAB

Pictured: high-quality segmentation results produced by deep learning on climate datasets. Image credit: NERSC



MANUFACTURING OVERVIEW

SEGMENTS: Industrial Machinery · Aerospace · Medical · Consumer · Transportation · Building Products

PROCESS & USER TAM: Design (200k) → Engineer (4.3M) → Simulate (460k) → Fabricate (1.4M) → Sell (100K) → Maintain (570k)

Collaborate + Manage (9.5M)

USE CASES: VISUALIZATION · vGPU for Engineering · SIMULATION · RETAIL CONFIGURATION

RTX POWERED: SOLIDWORKS VISUALIZE · AUTODESK VRED · KeyShot · ALTAIR · Ansys · creo live powered by ANSYS · CATIA · SIEMENS NX

NVIDIA AI ENTERPRISE IN MEDIA AND ENTERTAINMENT

EDITORS

REMOTE WORK / COLLABORATION

ARTIFICIAL INTELLIGENCE

POWERING RETAIL IN THESE CHALLENGING TIMES

Top 3 Segments of AI Use Cases in Retail

- Customer Engagement
- Operational Agility
- Seamless Omnichannel

$26T Global Retail Sales

2% Avg Net Profit Margin

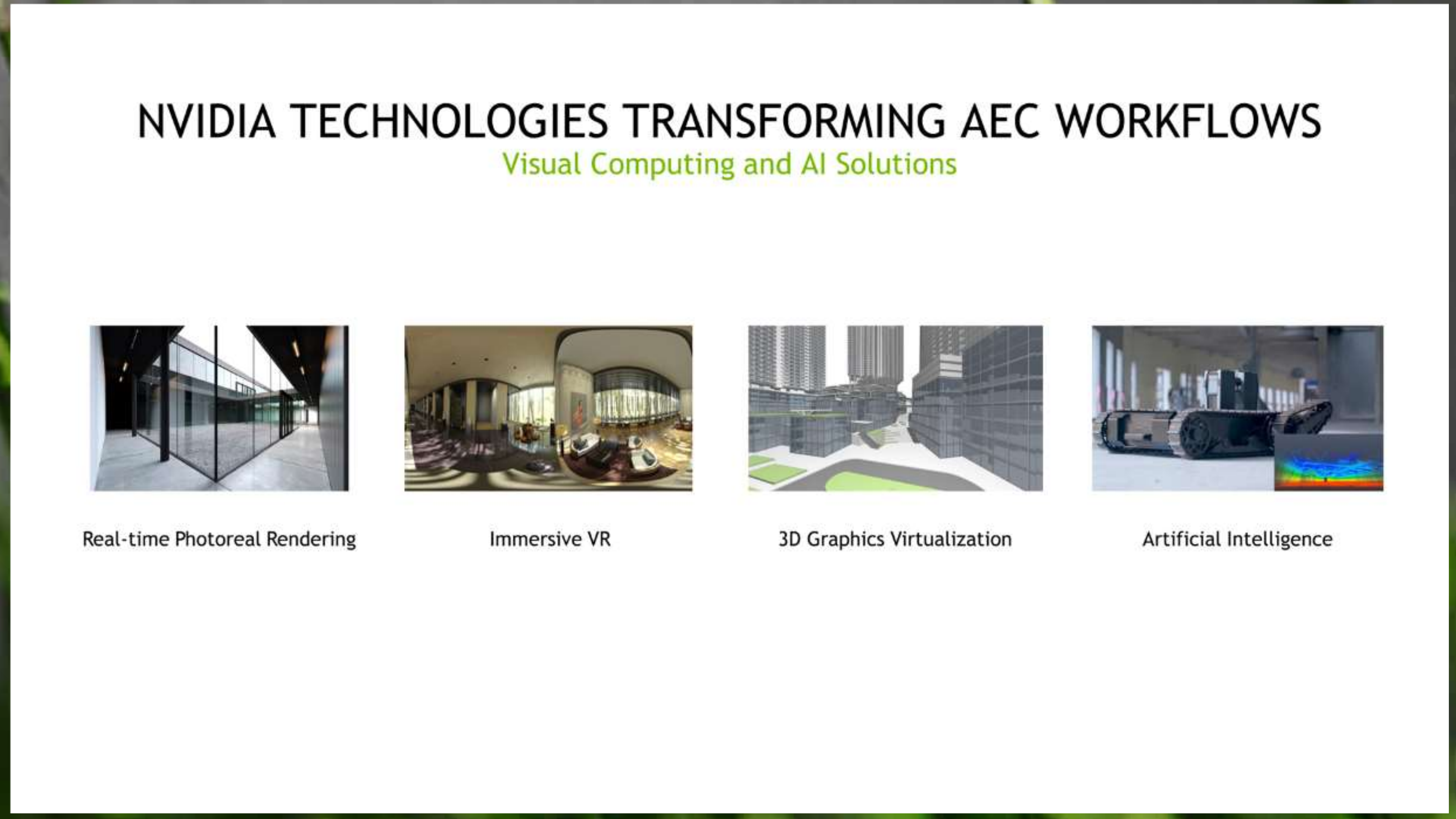
3X Increase in Profit with AI*

$1T Increase in Annual Profit

INTELLIGENT STORES/QSRS: Stock-Out & Curbside Pickup · Asset Protection & Frictionless Shopping

OMNI-CHANNEL MGMT: E-comm Recommenders & Conversational AI

INTELLIGENT SUPPLY CHAIN: Forecasting · Intra-Logistics and Last Mile Delivery



NVIDIA TECHNOLOGIES TRANSFORMING AEC WORKFLOWS

Visual Computing and AI Solutions

Real-time Photoreal Rendering

Immersive VR

3D Graphics Virtualization

Artificial Intelligence

TECHNOLOGY BEHIND IT ALL

AI APPLICATION FRAMEWORK

MODULUS

MONAI

RIVA

MAXINE

NEMO

MERLIN

CUOPT

MORPHEUS

TOKKIO

AVATAR

DRIVE

ISAAC

METROPOLIS

HOLOSCAN

PLATFORMS

NVIDIA HPC

NVIDIA AI

NVIDIA Omniverse

ACCELERATION LIBRARIES

cuNumeric

CV-CUDA

cuQuantum

Parabricks

Sionna

JetPack

RAPIDS

Spark

cuDNN

cuGraph

TensorRT

Triton

DeepStream

Flare

DOCA

Mag IO

Aerial

CLOUD-TO-EDGE DATACENTER-TO-ROBOTIC SYSTEMS

RTX

DGX

HGX

EGX

OVX

Super POD

AGX

3 CHIPS

GPU

CPU

DPU

It is all about platform: Our solutions catalog: NGC

Portal to AI services, freesoftware, support **NGC Catalog**

Cloud Services

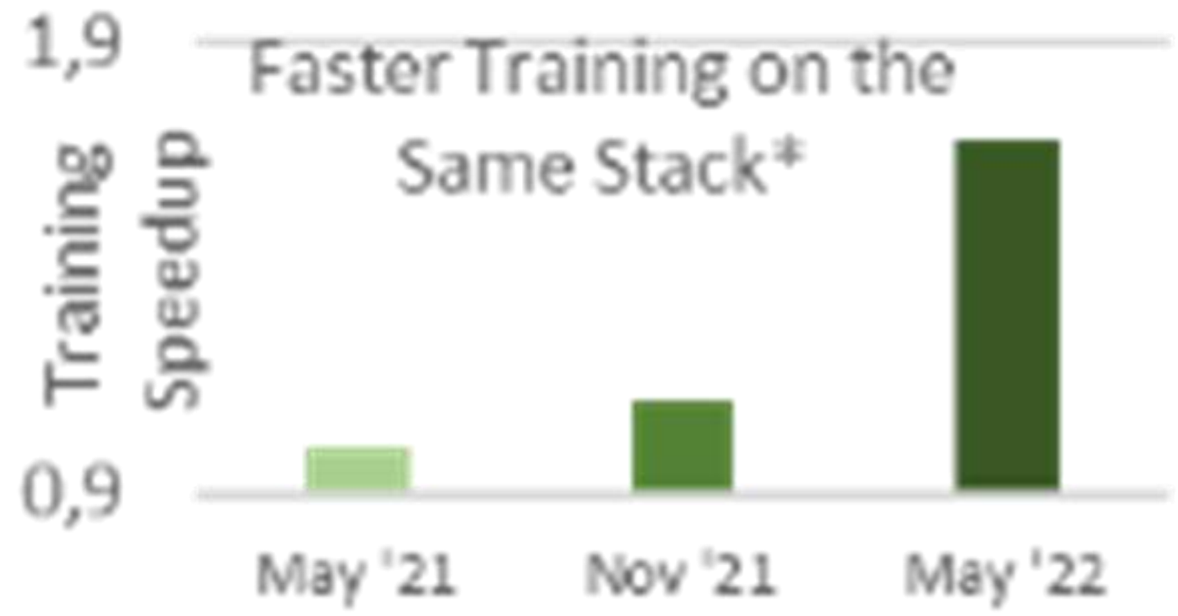
End-to-End AI development

AI Services for NLP, biology, speech

AI Workflow Management & Support

Performance Optimized

Tested across GPU-accelerated platforms

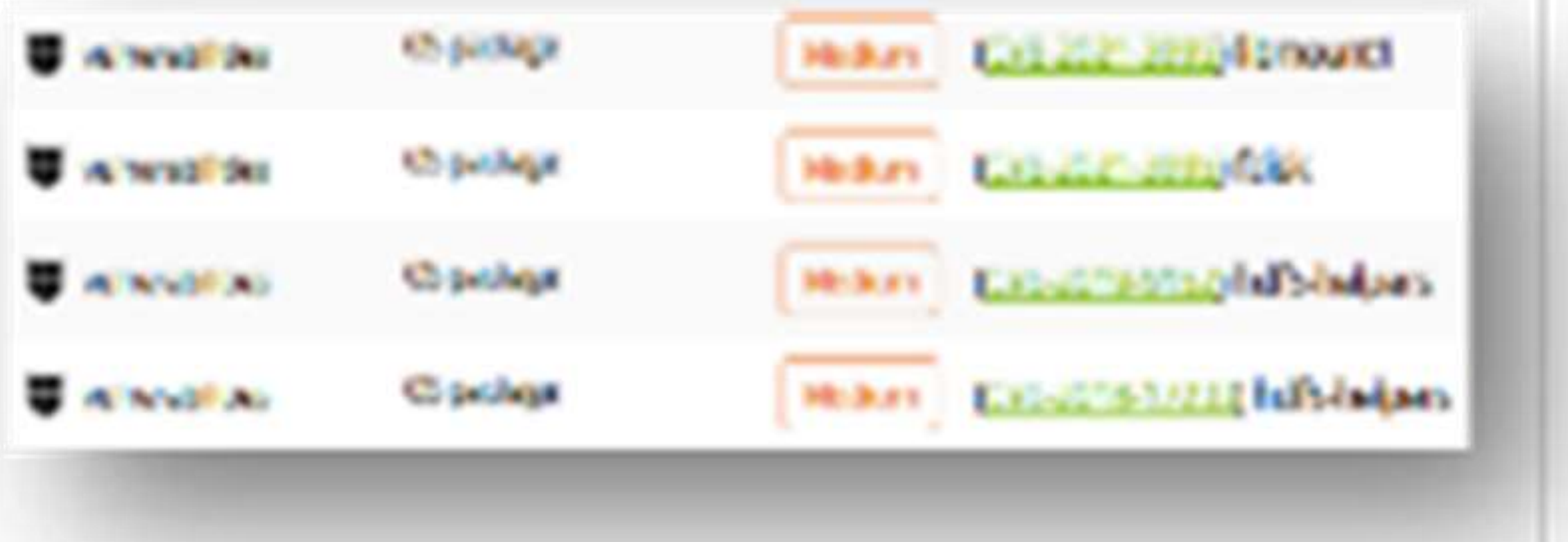


Monthly sw container updates

SOTA models

Fully Transparent

Quickly find and deploy the right sw



Detailed security scan reports

Model resumes

Accelerates Development

Focus on building, not setup

jupyter

One click deploy from NGC

aws | Google Cloud
Microsoft Azure | ORACLE CLOUD
Alibaba Cloud | OVHcloud

Develop once. Deploy anywhere w/ NVIDIA VMI

ngc.nvidia.com

Nvidia NGC Catalog

Getting Started

NVIDIA AI - End-to-End AI Development &...
Collection - Deep Learning
This is a collection of performance-optimized frameworks, SDKs, and models to build Computer Vision and Speech AI applications.
View Labels

HPC Collection
Collection - High Performance Computing
This collection provides access to the top HPC applications for Molecular Dynamics, Quantum Chemistry, and Scientific
View Labels

Deep Learning Frameworks
Collection - Deep Learning
This collection contains performance-optimized AI frameworks including PyTorch and TensorFlow
View Labels

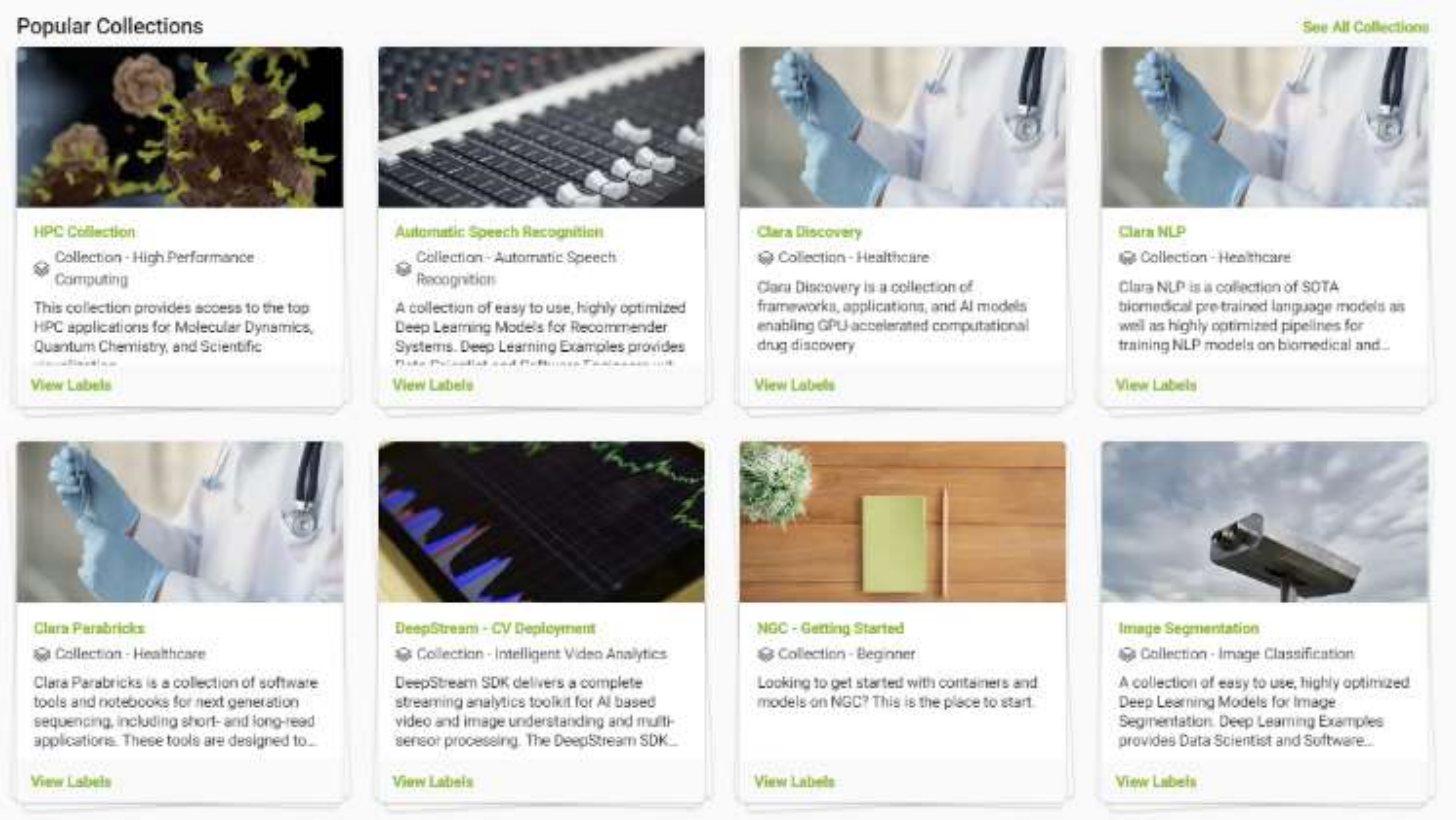
NGC - Getting Started
Collection - Beginner
Looking to get started with containers and models on NGC? This is the place to start.
View Labels

Documentation
We've got a whole host of documentation, covering the NGC UI and our powerful CLI. You can find out more here. Go to Documentation

Command Line Interface
Want to get more from NGC? Everything you see here can be used and managed via our powerful CLI tools. Download Now

NGC Private Registry
Private Registries from NGC allow you to secure, manage, and deploy your own assets to accelerate your journey to AI. Learn More

NGC Popular collections



Popular Collections

See All Collections

HPC Collection
Collection - High Performance Computing
This collection provides access to the top HPC applications for Molecular Dynamics, Quantum Chemistry, and Scientific
View Labels

Automatic Speech Recognition
Collection - Automatic Speech Recognition
A collection of easy to use, highly optimized Deep Learning Models for Recommender Systems. Deep Learning Examples provides
View Labels

Clara Discovery
Collection - Healthcare
Clara Discovery is a collection of frameworks, applications, and AI models enabling GPU-accelerated computational drug discovery
View Labels

Clara NLP
Collection - Healthcare
Clara NLP is a collection of SOTA biomedical pre-trained language models as well as highly optimized pipelines for training NLP models on biomedical and...
View Labels

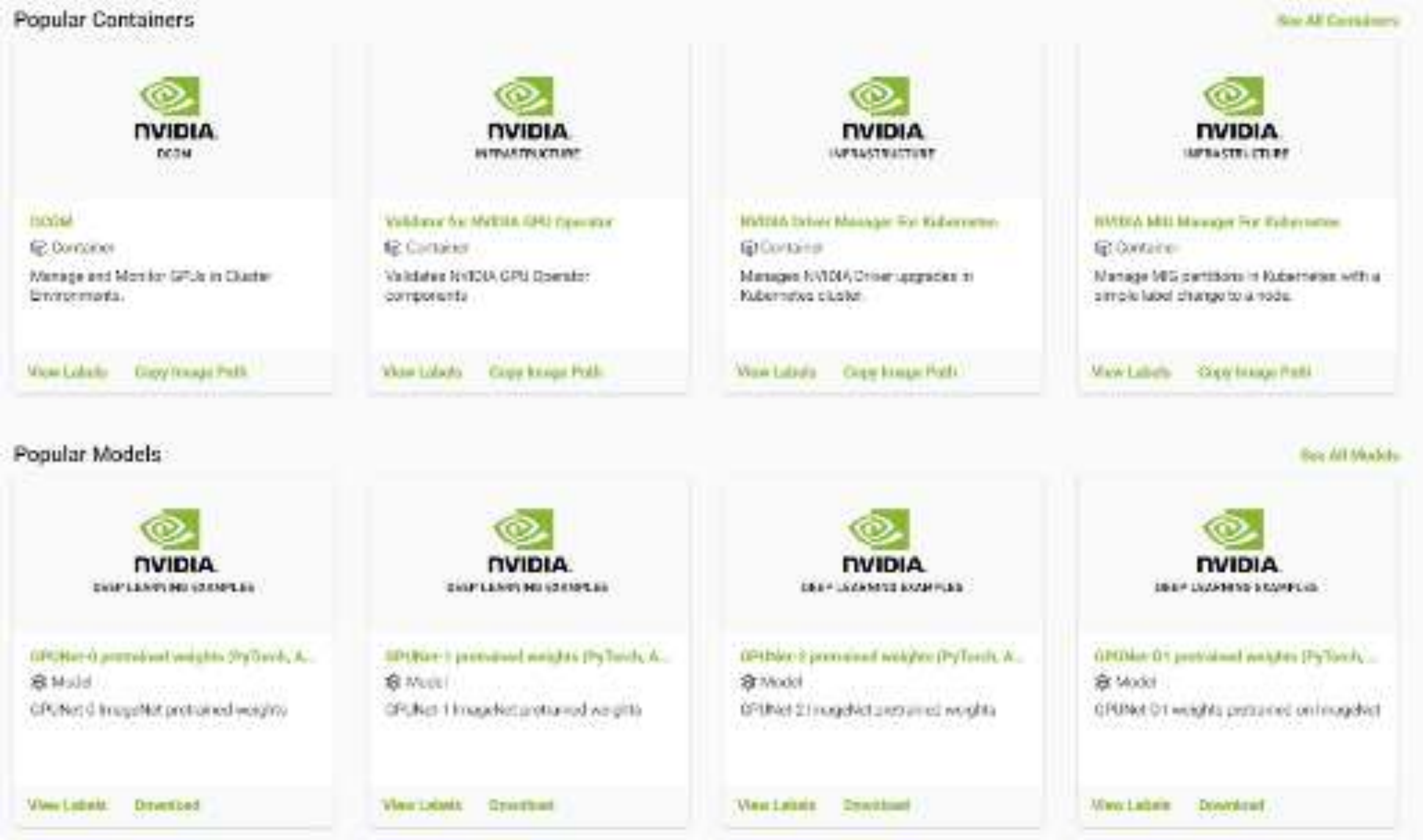
Clara Parabricks
Collection - Healthcare
Clara Parabricks is a collection of software tools and notebooks for next generation sequencing, including short- and long-read applications. These tools are designed to...
View Labels

DeepStream - CV Deployment
Collection - Intelligent Video Analytics
DeepStream SDK delivers a complete streaming analytics toolkit for AI based video and image understanding and multi-sensor processing. The DeepStream SDK...
View Labels

NGC - Getting Started
Collection - Beginner
Looking to get started with containers and models on NGC? This is the place to start.
View Labels

Image Segmentation
Collection - Image Classification
A collection of easy to use, highly optimized Deep Learning Models for Image Segmentation. Deep Learning Examples provides Data Scientist and Software...
View Labels

Nvidia free popular containers and resources



Popular Containers

See All Containers

DCGM
Container
Manage and Monitor GPUs in Cluster Environments.
View Labels Copy Image Path

Validator for NVIDIA GPU Operator
Container
Validates NVIDIA GPU Operator components
View Labels Copy Image Path

NVIDIA Driver Manager For Kubernetes
Container
Manages NVIDIA Driver upgrades in Kubernetes cluster.
View Labels Copy Image Path

NVIDIA MIG Manager For Kubernetes
Container
Manage MIG partitions in Kubernetes with a simple label change to a node.
View Labels Copy Image Path

Popular Models

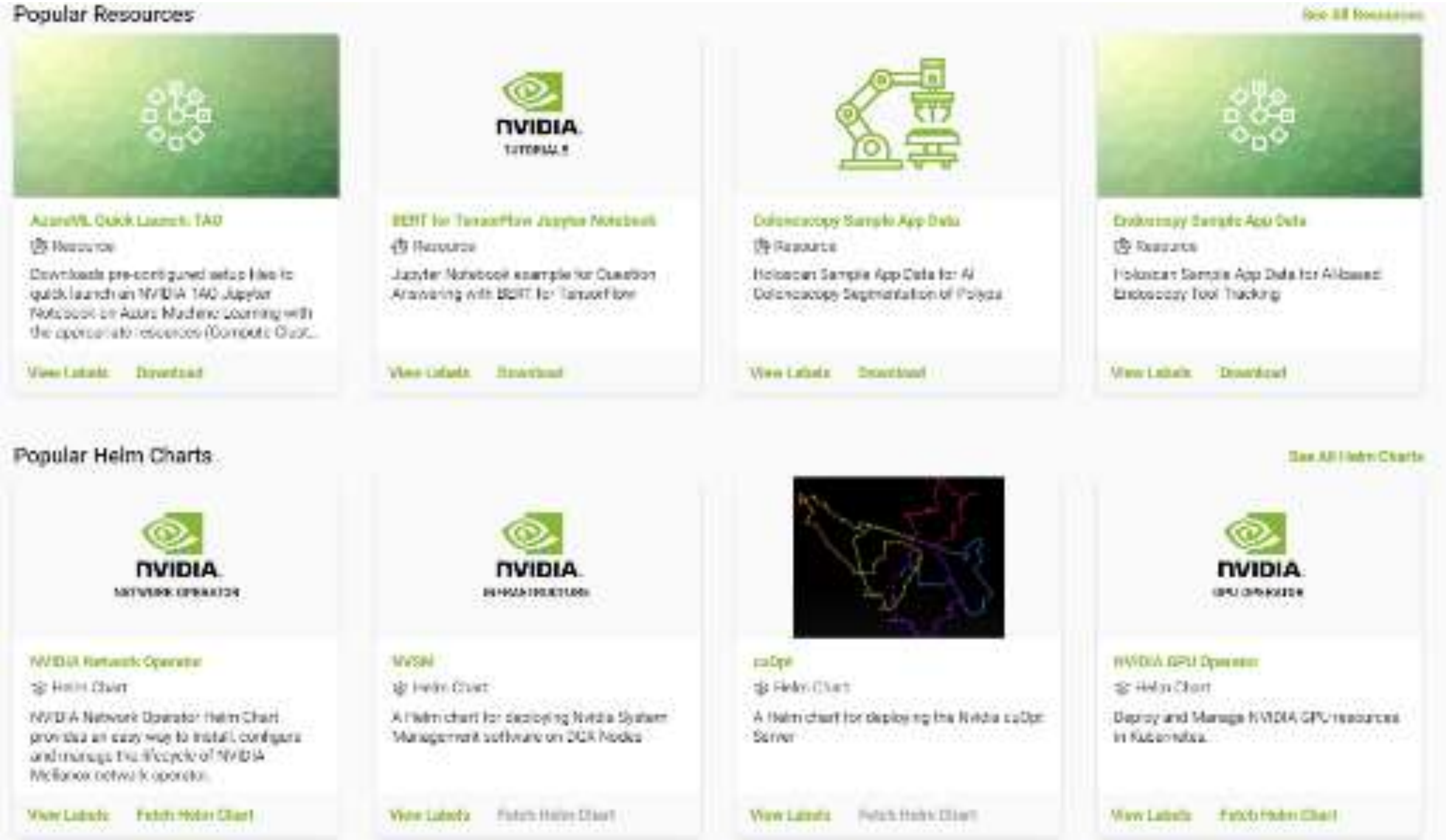
See All Models

GPUNet-0 pretrained weights (PyTorch, A...
Model
GPUNet 0 ImageNet pretrained weights
View Labels Download

GPUNet-1 pretrained weights (PyTorch, A...
Model
GPUNet 1 ImageNet pretrained weights
View Labels Download

GPUNet-2 pretrained weights (PyTorch, A...
Model
GPUNet 2 ImageNet pretrained weights
View Labels Download

GPUNet-D1 pretrained weights (PyTorch, ...
Model
GPUNet D1 weights pretrained on ImageNet
View Labels Download



Popular Resources

See All Resources

AzureML Quick Launch: TAO
Resource
Downloads pre-configured setup files to quick launch an NVIDIA TAO Jupyter Notebook on Azure Machine Learning with the appropriate resources (Compute Clust...
View Labels Download

BERT for TensorFlow Jupyter Notebook
Resource
Jupyter Notebook example for Question Answering with BERT for TensorFlow
View Labels Download

Colonoscopy Sample App Data
Resource
Holoscan Sample App Data for AI Colonoscopy Segmentation of Polyps
View Labels Download

Endoscopy Sample App Data
Resource
Holoscan Sample App Data for AI-based Endoscopy Tool Tracking
View Labels Download

Popular Helm Charts

See All Helm Charts

NVIDIA Network Operator
Helm Chart
NVIDIA Network Operator Helm Chart provides an easy way to install, configure and manage the lifecycle of NVIDIA Mellanox network operator.
View Labels Fetch Helm Chart

NVSM
Helm Chart
A Helm chart for deploying Nvidia System Management software on DGX Nodes
View Labels Fetch Helm Chart

cuOpt
Helm Chart
A Helm chart for deploying the Nvidia cuOpt Server
View Labels Fetch Helm Chart

NVIDIA GPU Operator
Helm Chart
Deploy and Manage NVIDIA GPU resources in Kubernetes.
View Labels Fetch Helm Chart

OK,… LETS TALK IRON !!!!

NVIDIA Data Center GPU Portfolio. Workload oriented

	GPU		DL Training & DA	DL Inference	HPC / AI	Omniverse / Render Farms	Virtual Workstation	Virtual Desktop (VDI)	Mainstream Acceleration	Far Edge Acceleration
Compute	H100		●	●	●				●	
	A100		◐	◐	◐				◐	
	A30			○	○				○	
Graphics / Compute	L40			●		●	●		●	
	A40					◐	◐		◐	
	A10			○		○	●	◐	○	◐
	A16						○	●		
Small Form Factor Compute/Graphics	A2			●			●	●	●	●
	T4			◐			◐	◐	◐	◐

○ ◐ ● **Price-performance** comparison in each product group (Compute, Graphics & Compute, SFF Compute & Graphics) and workload column

LET’S GET DEEPER

NVIDIA HOPPER

The Engine for the World’s AI Infrastructure

World’s Most Advanced Chip

Transformer Engine

2nd Gen MIG

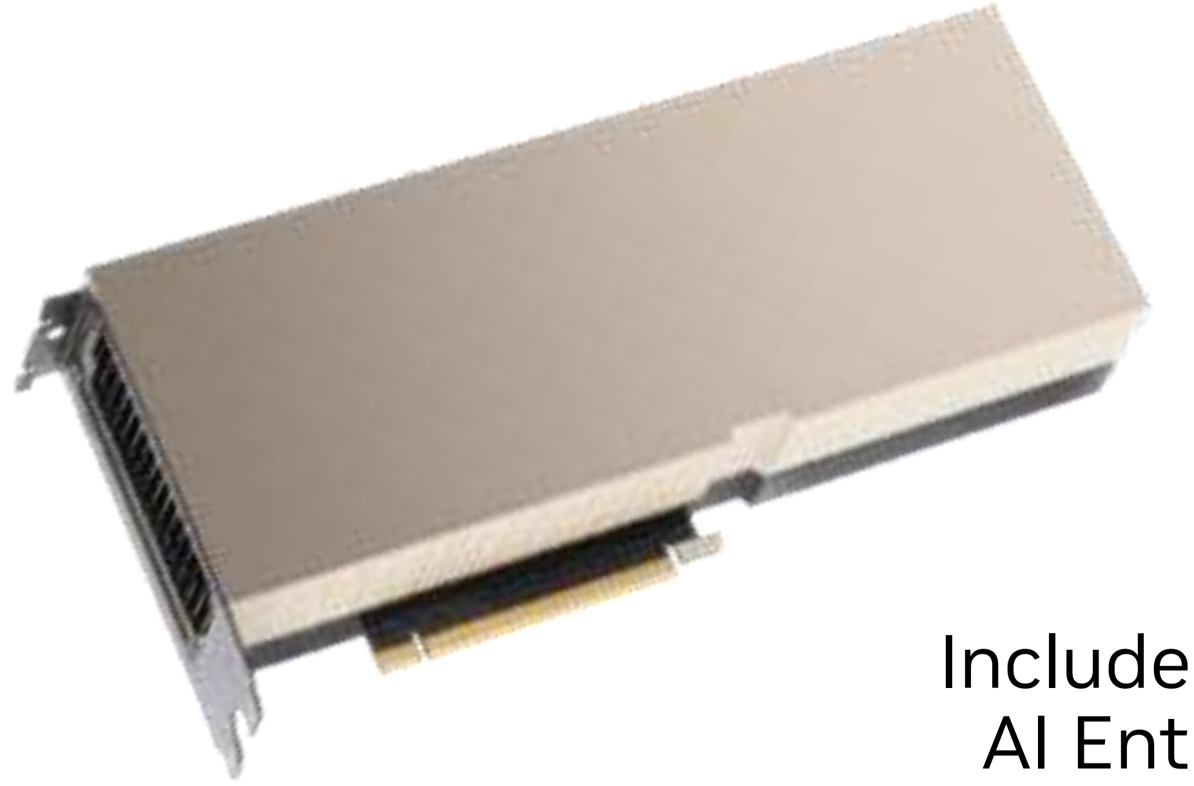
Confidential Computing

4th Gen NVLink

DPX Instructions

H100 SXM

H100 NVL



H100 PCIE

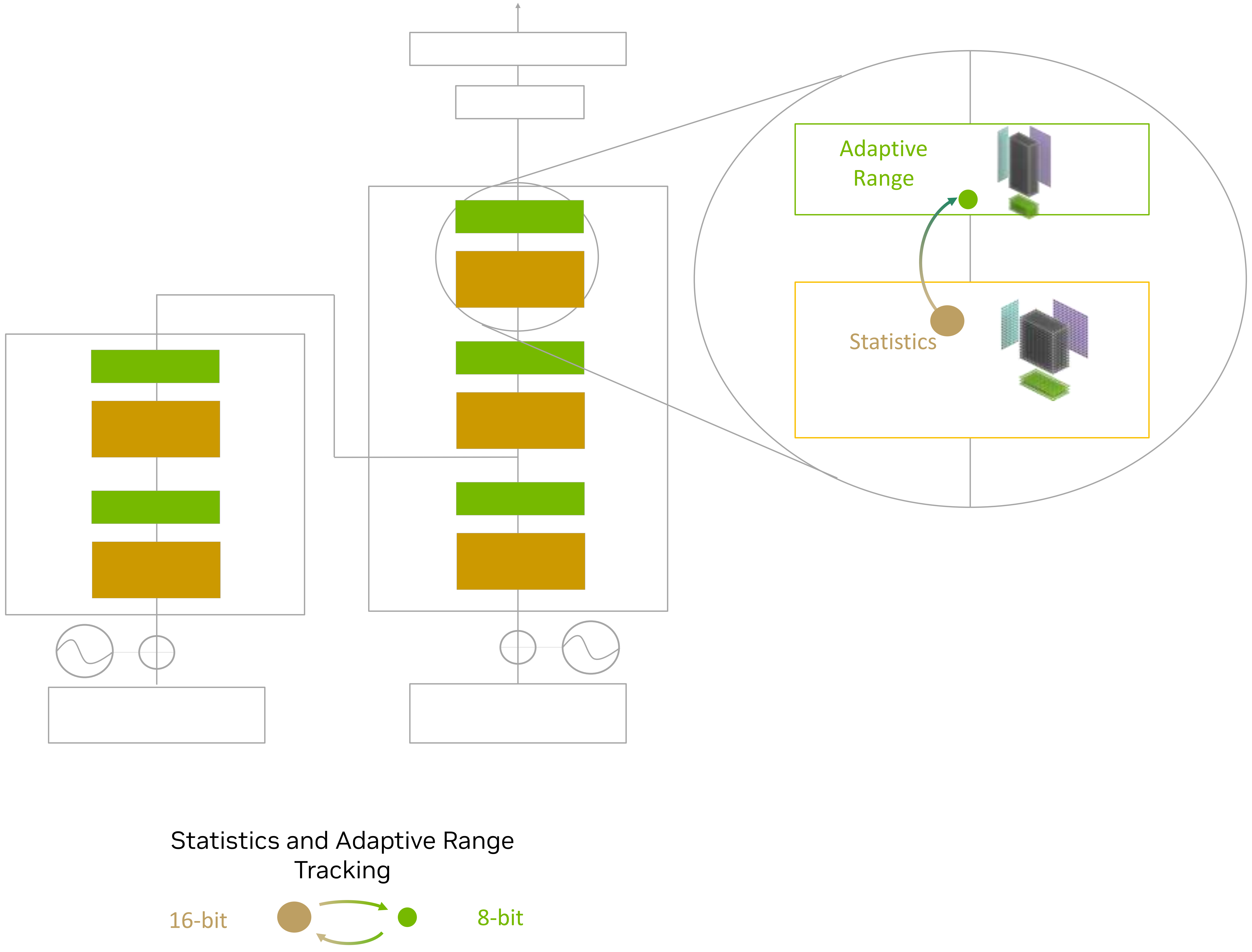
Includes NVIDIA AI Enterprise

NVIDIA AI Enterprise is a 5-Year Subscription

TRANSFORMER ENGINE

Tensor core optimized for transformer models

- 6X Faster Training and Inference of Transformer Models
- NVIDIA Tuned Adaptive Range Optimization Across 16-bit and 8-bit Math
- Configurable Macro Blocks Deliver Performance Without Accuracy Loss

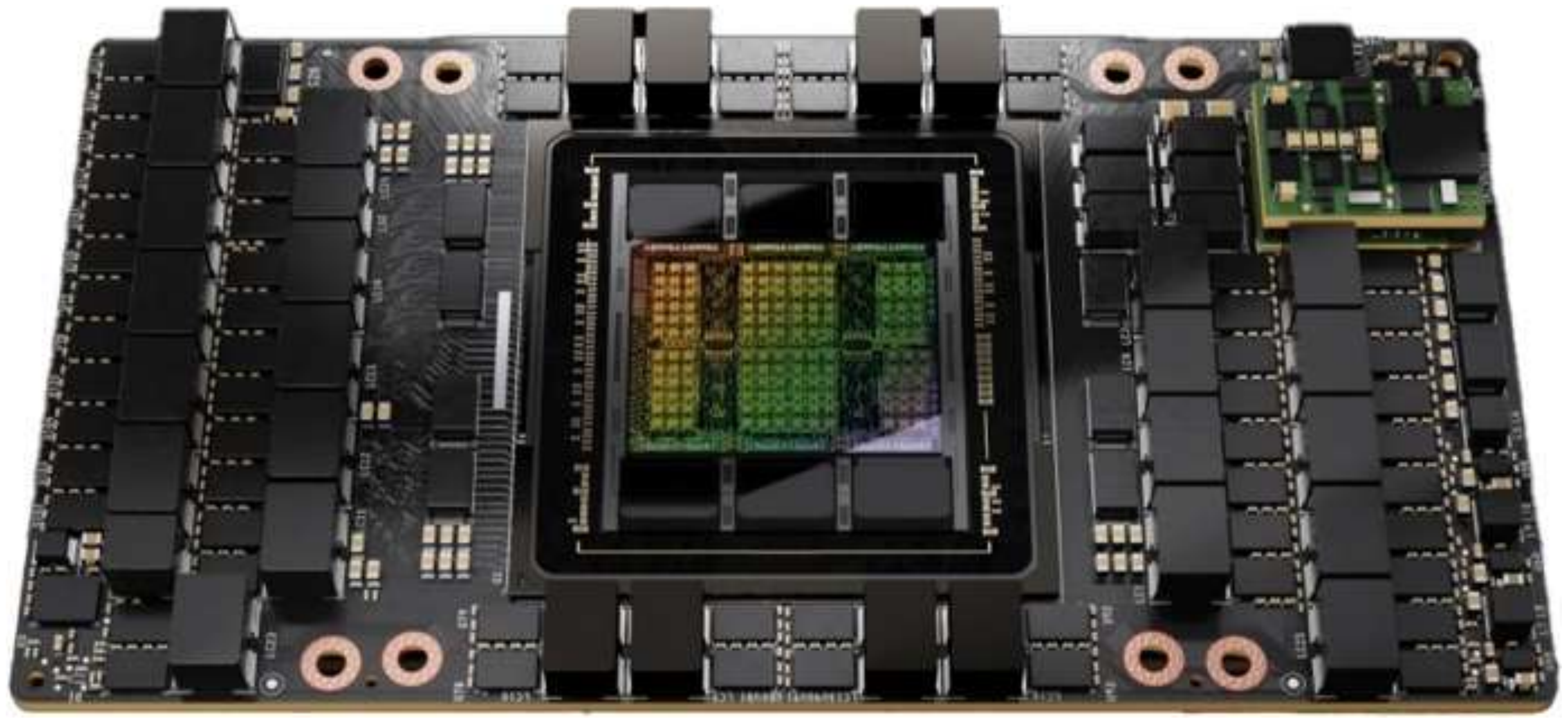


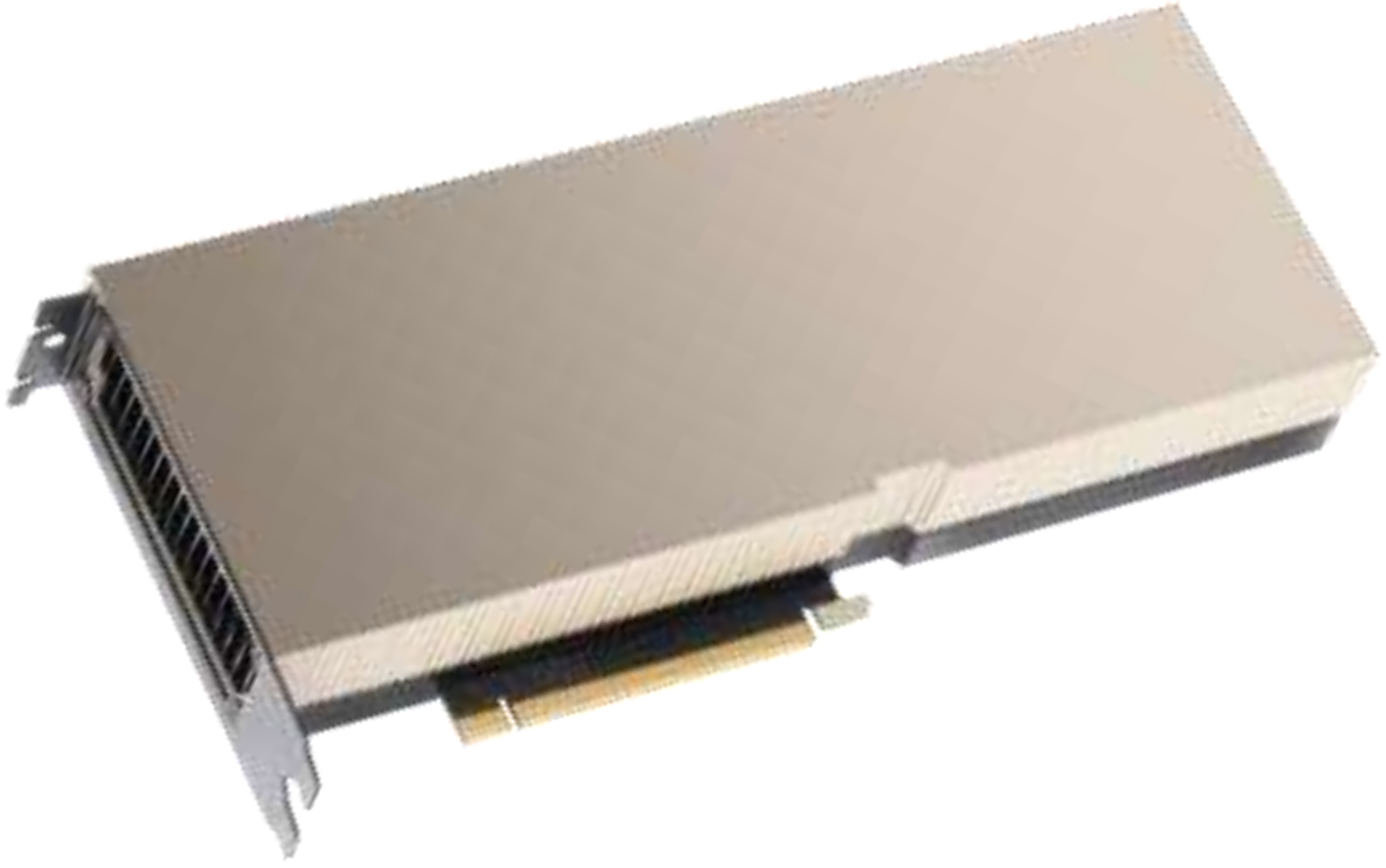
Statistics and Adaptive Range Tracking

NVIDIA H100 SXM5 AND PCIE

Unprecedented Performance, Scalability, and
Security for Every Data Center

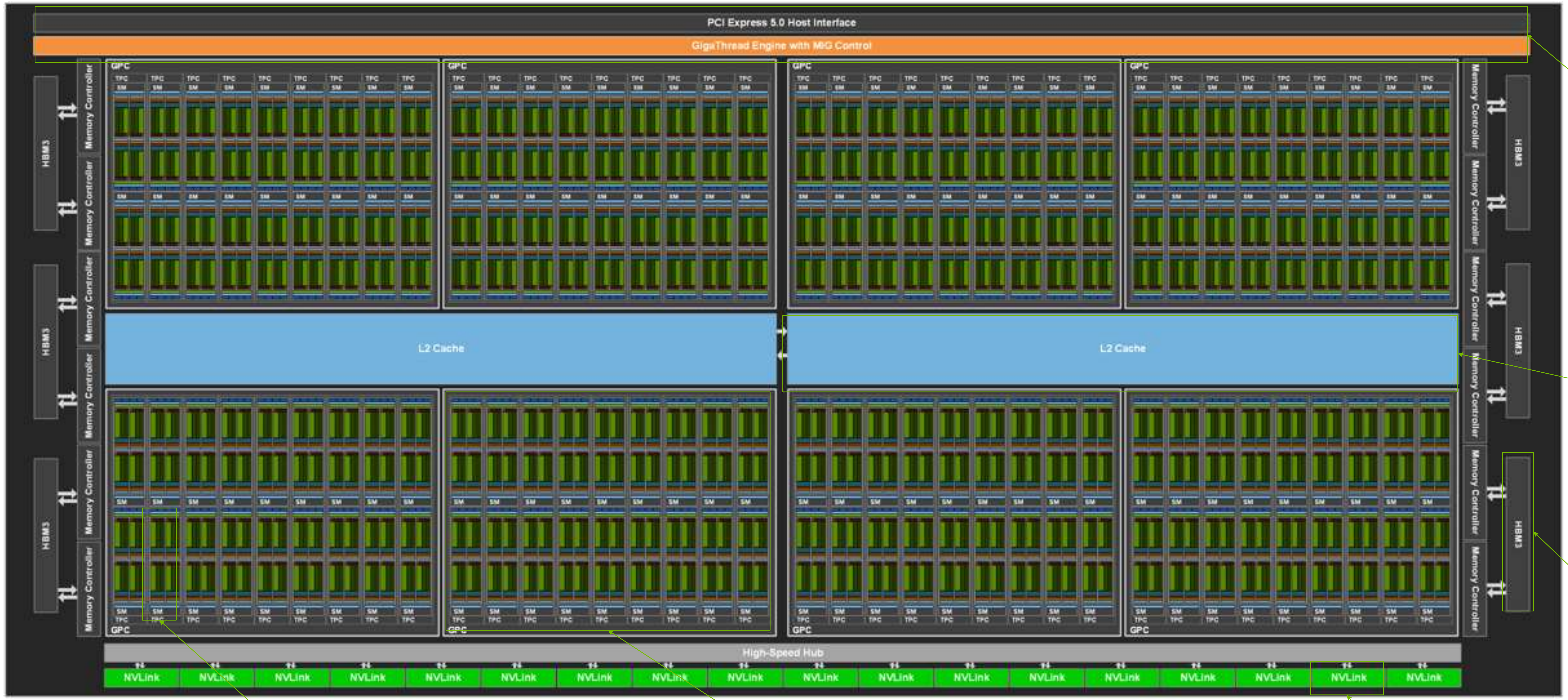
	H100 PCIe	H100-80 SXM5	H100-94 SXM5
New Features			
- Dynamic Programming Instructions	Supported	Supported	Supported
- Confidential Computing	Supported	Supported	Supported
- Transformer Engine with FP8	Supported	Supported	Supported
- Peak FP8 Tensor TFLOPS	1513/3026	1978/3957	1978/3957
- Peak FP16 Tensor TFLOPS	756/1513	989/1978	989/1978
- Peak TF32 Tensor TFLOPS	378/756	494/989	494/989
- Peak FP64 Tensor TFLOPS	51.2	67	67
- Peak INT8 Tensor TOPS	1513/3026	1978/3957	1978/3957
- Peak FP16 TFLOPS (non-Tensor)	102	134	134
- Peak BF16 TFLOPS (non-Tensor)	102	134	134
- Peak FP32 TFLOPS (non-Tensor)	51	67	67
- Peak FP64 TFLOPS (non-Tensor)	25	33	33
- Peak INT32 TOPS	25	33	33
Memory			
- Memory Interface	5120-bit HBM2e	5120-bit HBM3	6144-bit HBM2e
- Memory Size	80 GB	80 GB	94 GB
- Memory Bandwidth	2000 GB/sec	3300 GB/sec	2400 GB/sec
L2 Cache Size	50 MB	50 MB	50 MB
TDP	350 Watts	700 Watts	700 Watts





HOPPER ARCHITECTURE

H100 GPU Key features


PCI Express 5.0 Host Interface
GigaThread Engine with MIG Control
GPC
TPC
SM
L2 Cache
Memory Controller
HBM3
High-Speed Hub
NVLink

2nd Gen Multi-Instance GPU
Confidential Computing
PCIe Gen5

Larger 50 MB L2

80GB HBM3, 3 TB/s bandwidth

132 SMs
4th Gen Tensor Core

Thread Block Clusters

4th Gen NVLink
900 GB/s total bandwidth

Omniverse Enterprise

Build custom 3D metaverse applications , power large-scale simulations and operate photorealistic virtual worlds and complex digital twins

Rendering

Work with complex scenes and high-fidelity creative workflows with 3^{rd}-Gen RTX and 48GB of GPU memory

Virtualization*

Deliver high-performance workstation instances for high-end design, AI, and compute workloads

*vGPU support in Q1 2023

AI

Provision virtual AI/ML virtual workstations for model development, training, data exploration. Multi-GPU AI for larger workloads.

NVIDIA L40

Revolutionary capabilities for data center workloads

NVIDIA L40 GENERATIONAL COMPARISON

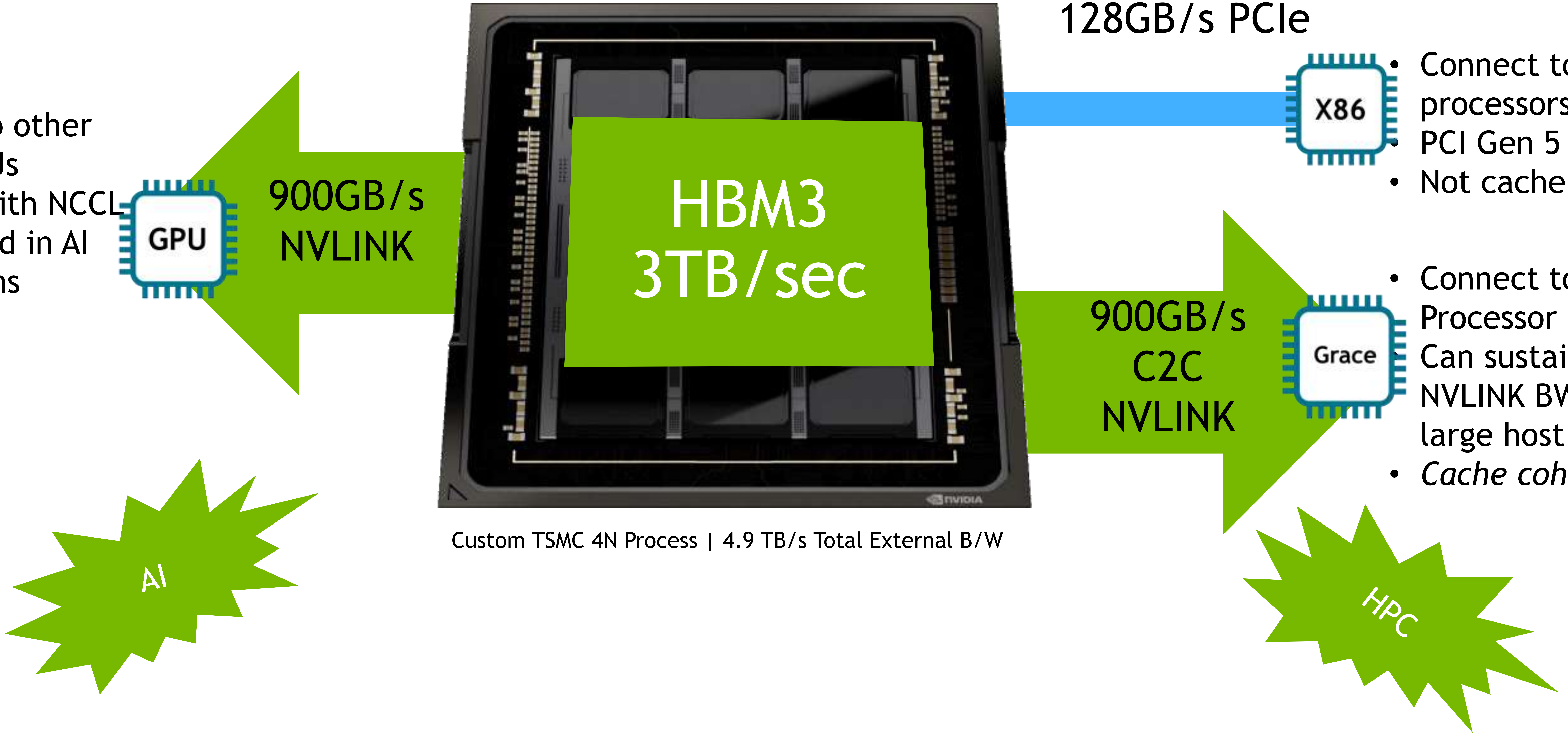
	NVIDIA L40	NVIDIA A40
GPU Architecture	NVIDIA Ada Lovelace Architecture	NVIDIA Ampere Architecture
FP32	90.5 TFLOPS	37.4 TFLOPS
RT Core	209 TFLOPS	73.1 TFLOPS
Tensor Float 32 (TF32)	90.5 | 181** TFLOPS	74.8 | 149.6* TFLOPS
BFLOAT16 Tensor Core	181 | 362** TFLOPS	149.7 | 299.4* TFLOPS
FP16 Tensor Core	181 | 362** TFLOPS	149.7 | 299.4* TFLOPS
FP8 Tensor Core	362 | 724** TFLOPS	NA
INT8 Tensor Core	362 | 724** TOPS	299.3 | 598.6* TOPS
INT4 Tensor Core	724 | 1448** TOPS	598.7 | 1197.4* TOPS
GPU Memory	48 GB GDDR6 w/ ECC	48 GB GDDR6 w/ ECC
GPU Memory Bandwidth	864 GB/s	696 GB/s
Max Thermal Design Power (TDP)	300 W	300 W
Form Factor	4.4” H x 10.5” L - Dual Slot	4.4” H x 10.5” L - Dual Slot
Interconnect	PCIe Gen4 x16: 64 GB/s	PCIe Gen4 x16: 64GB/s NVIDIA® NVLink® bridge for 2 GPUs:112.5 GB/s
Server Options	Partner and NVIDIA-Certified Systems™, NVIDIA® OVX™	Partner and NVIDIA-Certified Systems™, NVIDIA® OVX™

* Preliminary specifications, subject to change.
** Structural sparsity enabled

CAPTURING THE BOTTLENECK

HOPPER AND BANDWIDTH

For HPC and AI



- Connect to other Nvidia GPUs
- Program with NCCL
- *Mostly* used in AI applications

- Connect to x86 processors
- PCI Gen 5
- Not cache coherent

- Connect to GRACE Processor
- Can sustain full NVLINK BW into large host memory
- *Cache coherent*

Custom TSMC 4N Process | 4.9 TB/s Total External B/W

X86 GPU-GPU NVLINK

Architectures & Cost of *Connectivity*

1-Way

2-Way

H100 PCIe

H100 NVL

4-Way HGX

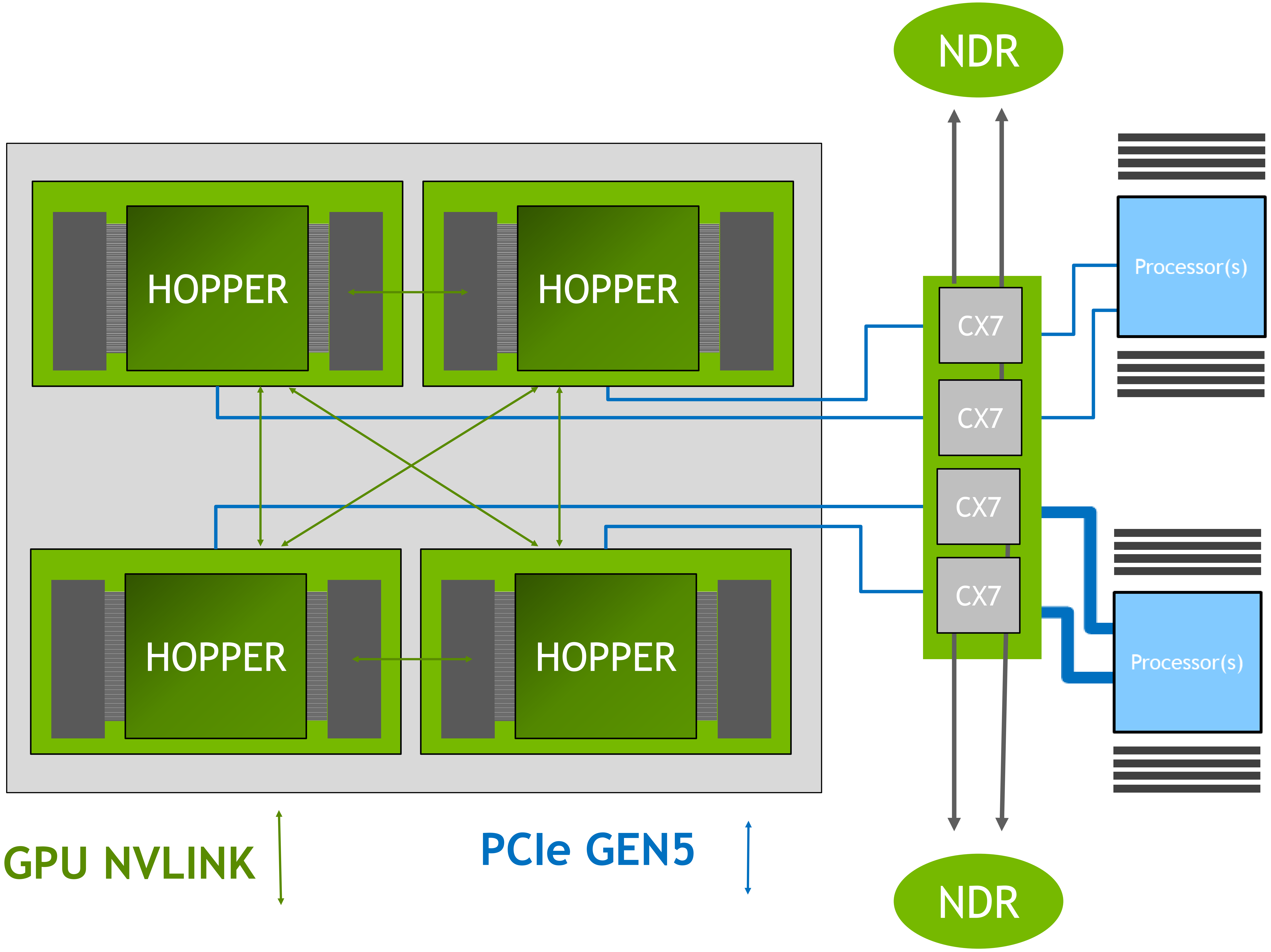
H100 80GB SXM5
H100 94GB SXM5

8-Way HGX

H100 80GB SXM5

H100 SXM5 NODE DESIGN

H100 80GB SXM5
H100 94GB SXM5



GPU NVLINK ↕

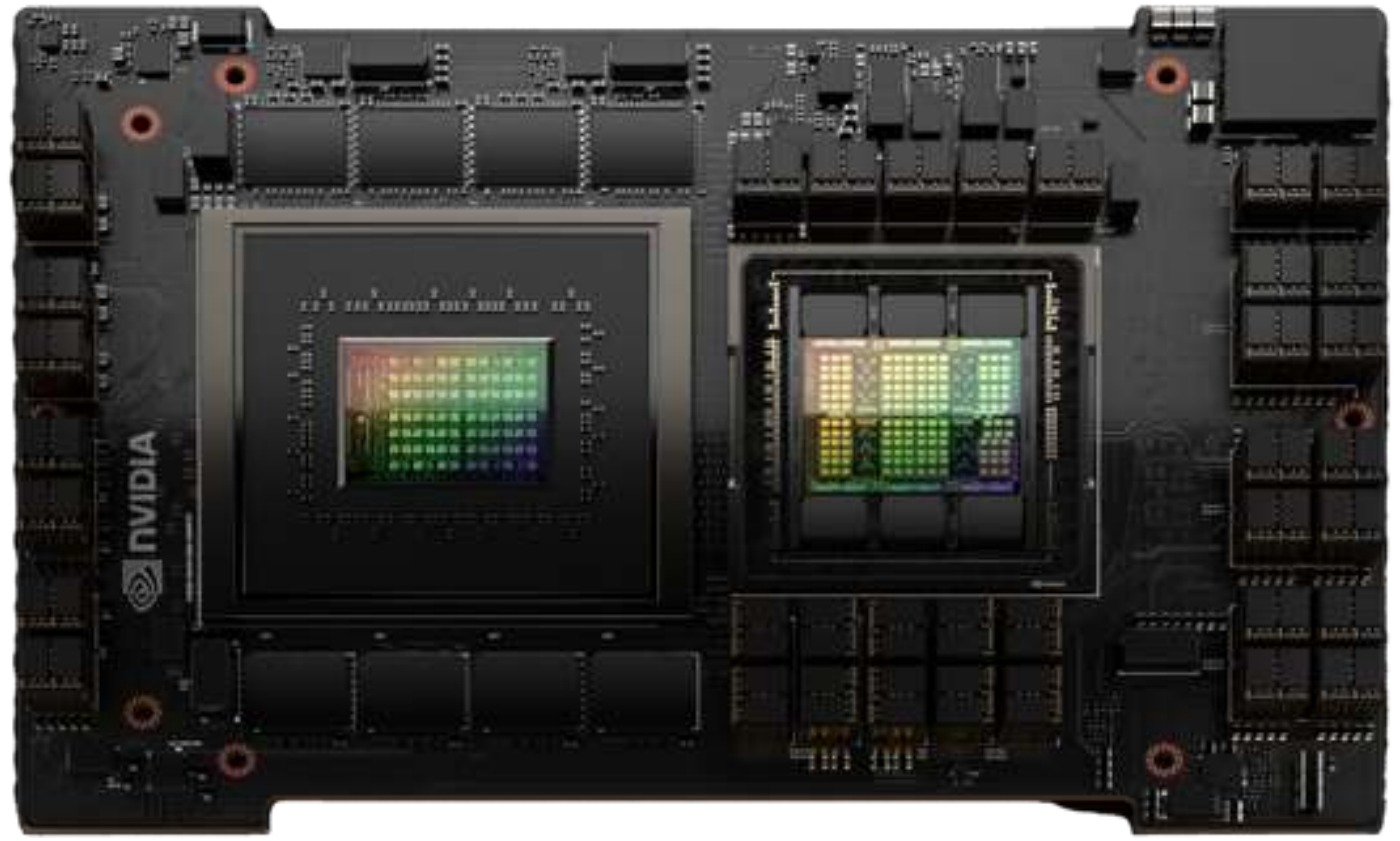
PCIe GEN5 ↕

WHAT IS NEXT?

NVIDIA GRACE PLATFORM

Grace Hopper Superchip

Giant Scale AI & HPC



Accelerated applications where CPU performance and system memory BW are critical since AI models continue to get bigger and our GPUs get even faster

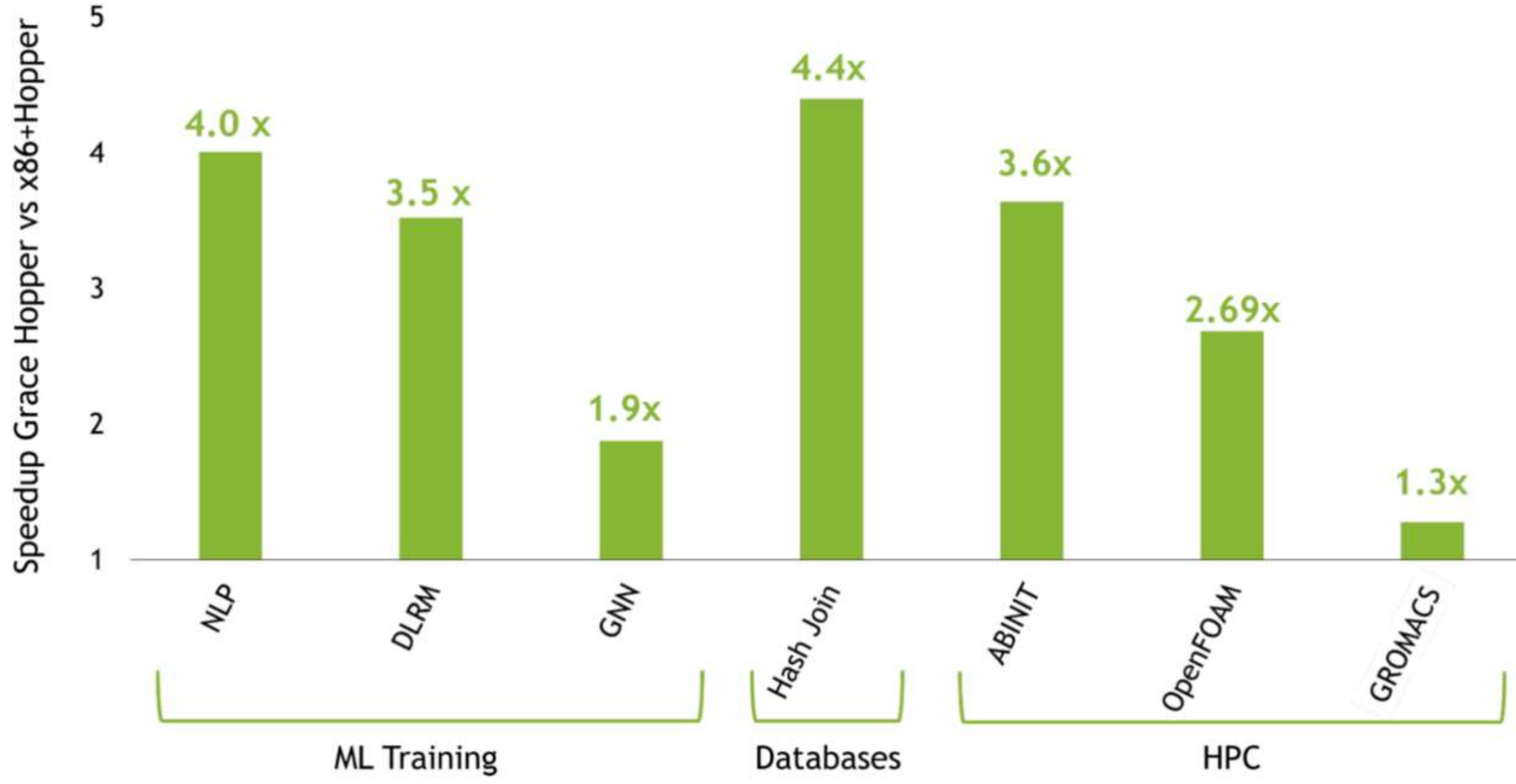
Grace CPU Superchip

CPU Computing

Applications that are not accelerated yet but where absolute performance, energy efficiency, and datacenter density matter, such as in scientific computing, data analytics, and hyperscale computing applications

NVIDIA Grace Hopper Superchip
Hardware Consistency
HIGH-SPEED I/O
4x 16x PCIe-5 512 GB/s
CPU LPDDR5X ≤ 512 GB
≤546 GB/s
GRACE CPU
CPU LPDDR5X ≤ 512 GB
NVLINK C2C 900 GB/s
GPU HBM3 ≤ 96 GB HBM3
≤3000 GB/s
HOPPER GPU
GPU HBM3 ≤ 96 GB HBM3
18x NVLINK 4 900 GB/s
NVLINK NETWORK ≤ 256 GPUs

NVIDIA NVLink-C2C is an NVIDIA memory coherent, high-bandwidth, and low-latency superchip interconnect. It is the heart of the Grace Hopper Superchip and delivers up to 900 GB/s total bandwidth. This is 7x higher bandwidth than x16 PCIe Gen5 lanes commonly used in accelerated systems.

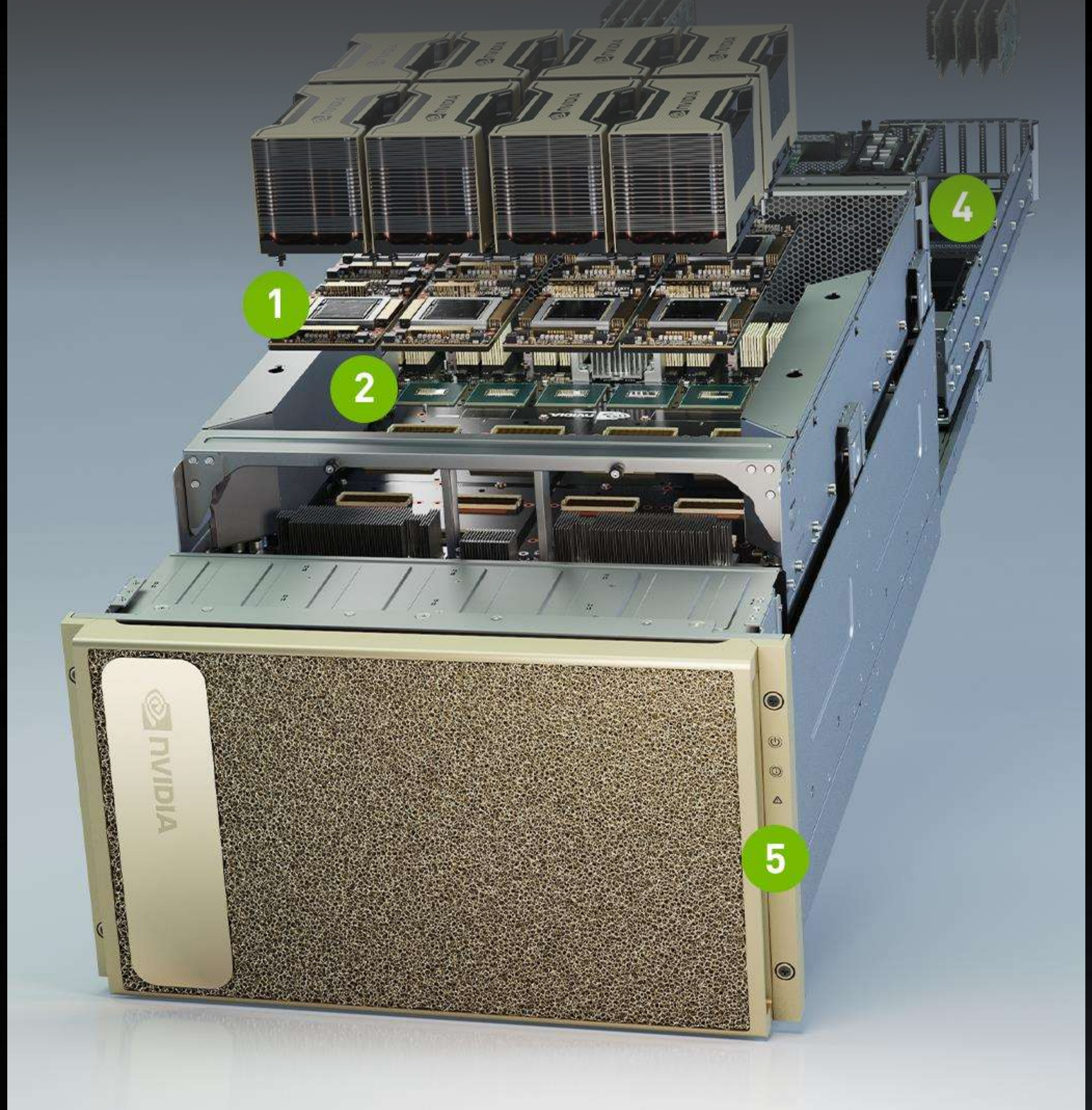


NVLink-C2C enables applications to oversubscribe the GPU's memory and directly utilize NVIDIA Grace CPU's memory at high bandwidth. With up to 512 GB of LPDDR5X CPU memory per Grace Hopper Superchip, the GPU has direct high-bandwidth access to 4x more memory than what is available with HBM. Combined with the NVIDIA NVLink Switch System, all GPU threads running on up to 256 NVLink-connected GPUs can now access up to 150 TB of memory at high bandwidth. Fourth-generation NVLink enables accessing peer memory using direct loads, stores, and atomic operations, enabling accelerated applications to solve larger problems more easily than ever.

BACK TO THE BEGINNING:
HOW NVIDIA HELPS IN DIFFERENT INDUSTRIES

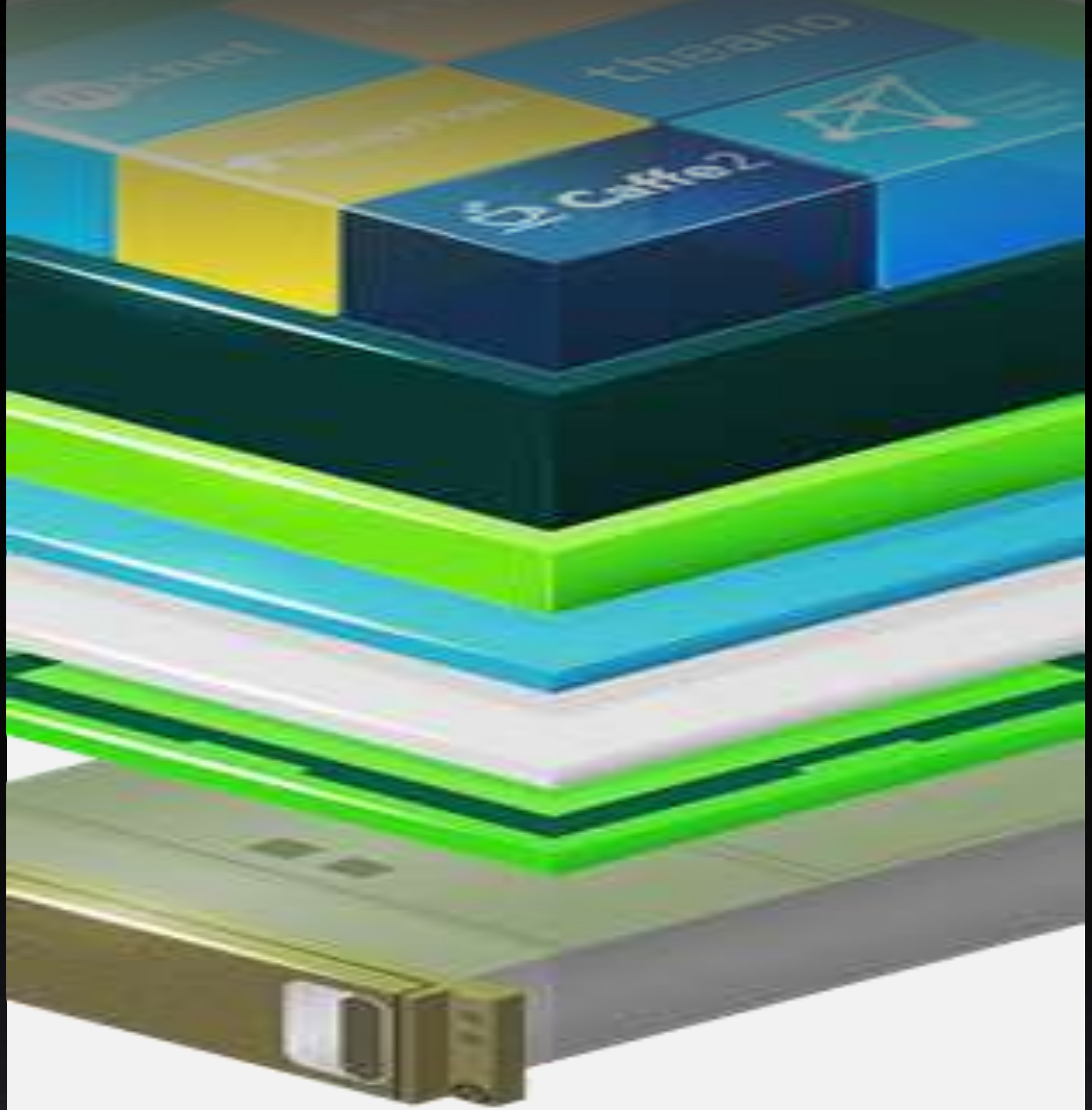
AI INFRASTRUCTURE

Designing for customer use-cases



ACCELERATED SOFTWARE

Optimised applications out of the box



LIGHTHOUSE ENGAGEMENTS

Driving the field and gaining deep insight

gsk
AstraZeneca
AI CENTRE for Value Based Healthcare
Oxford NANOPORE Technologies

A FEW VIDEOS

NVIDIA Omniverse

https://www.youtube.com/watch?v=Gn_IMIPrX9s

AMAZON Digital twin warehouse

https://www.youtube.com/watch?v=-VQLqs6s9y0

DriveSIm Mercedes

https://www.youtube.com/watch?v=UoPXzzK_g1Q

DRIVE Sim Scenario Reconstruction, Powered by Omniverse - YouTube

NVIDIA Healthcare

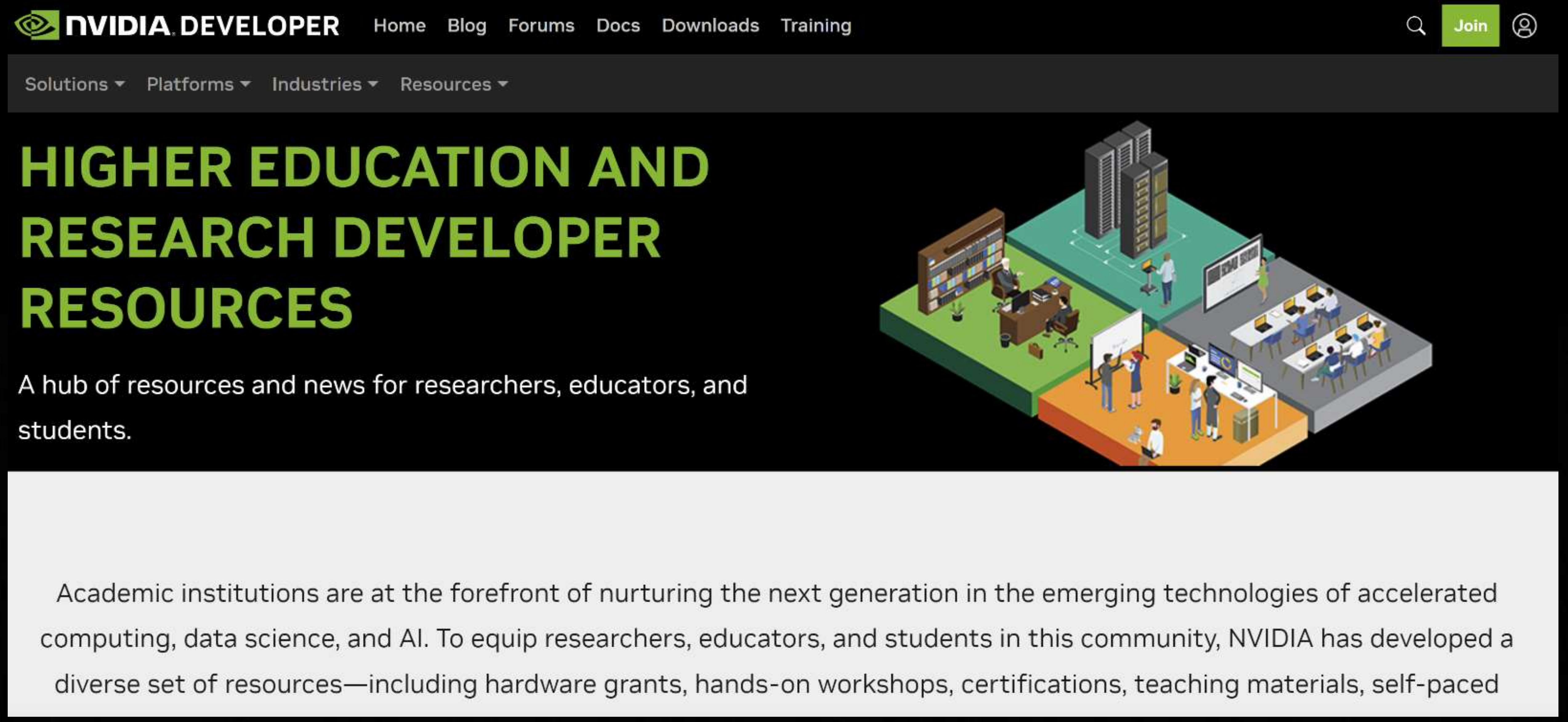
https://www.youtube.com/watch?v=qlNbC88SU7o

NVIDIA Clara/Holoscan

https://www.youtube.com/watch?v=RVFIDEuNtt0

https://www.youtube.com/watch?v=cGuh5XAdowg

INTERESTING NVIDIA LINKS FOR THE HER COMMUNITY

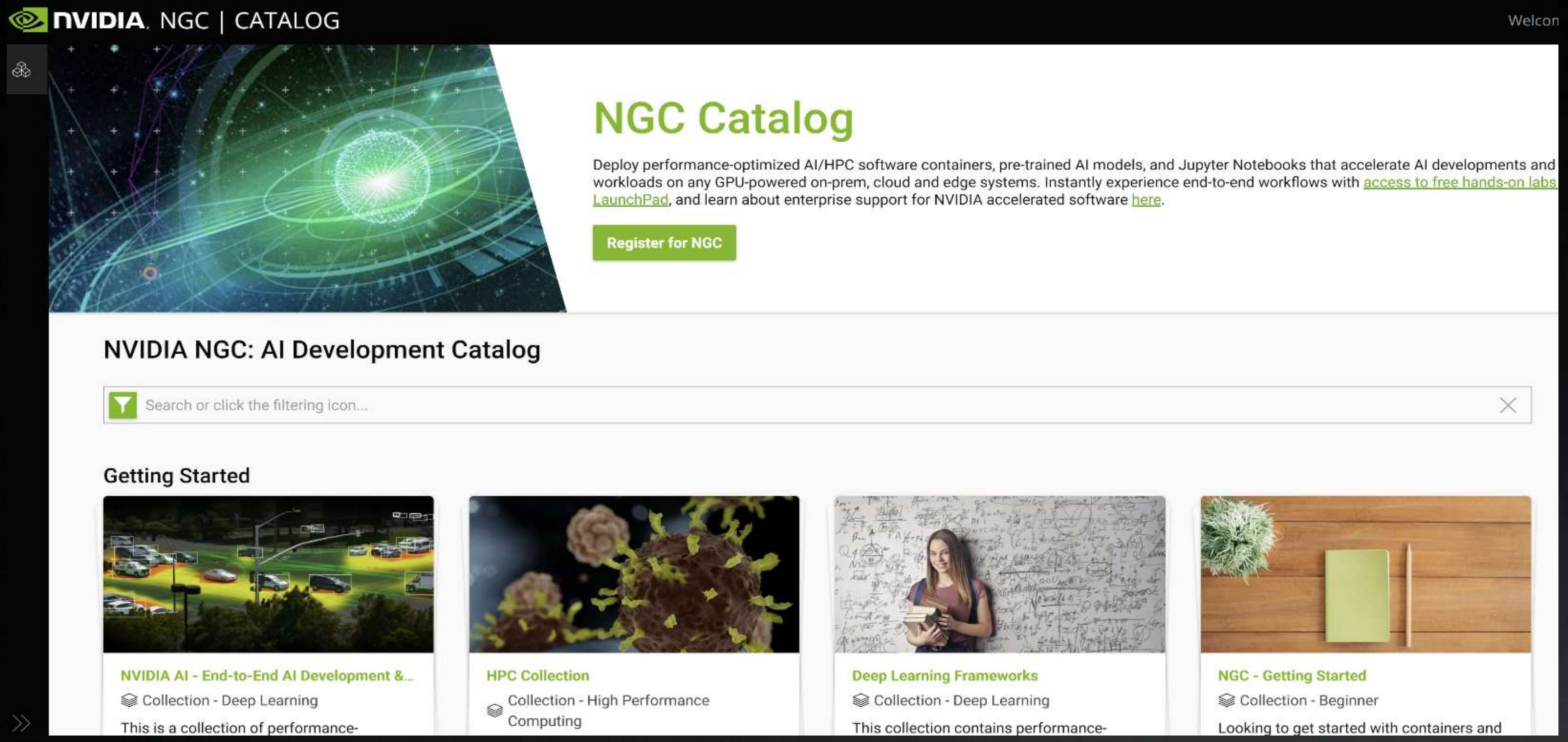


HIGHER EDUCATION AND RESEARCH DEVELOPER RESOURCES

A hub of resources and news for researchers, educators, and students.

Academic institutions are at the forefront of nurturing the next generation in the emerging technologies of accelerated computing, data science, and AI. To equip researchers, educators, and students in this community, NVIDIA has developed a diverse set of resources—including hardware grants, hands-on workshops, certifications, teaching materials, self-paced

https://developer.nvidia.com/higher-education-and-research



NGC Catalog

Deploy performance-optimized AI/HPC software containers, pre-trained AI models, and Jupyter Notebooks that accelerate AI developments and workloads on any GPU-powered on-prem, cloud and edge systems. Instantly experience end-to-end workflows with access to free hands-on labs LaunchPad, and learn about enterprise support for NVIDIA accelerated software here.

NVIDIA NGC: AI Development Catalog

Getting Started

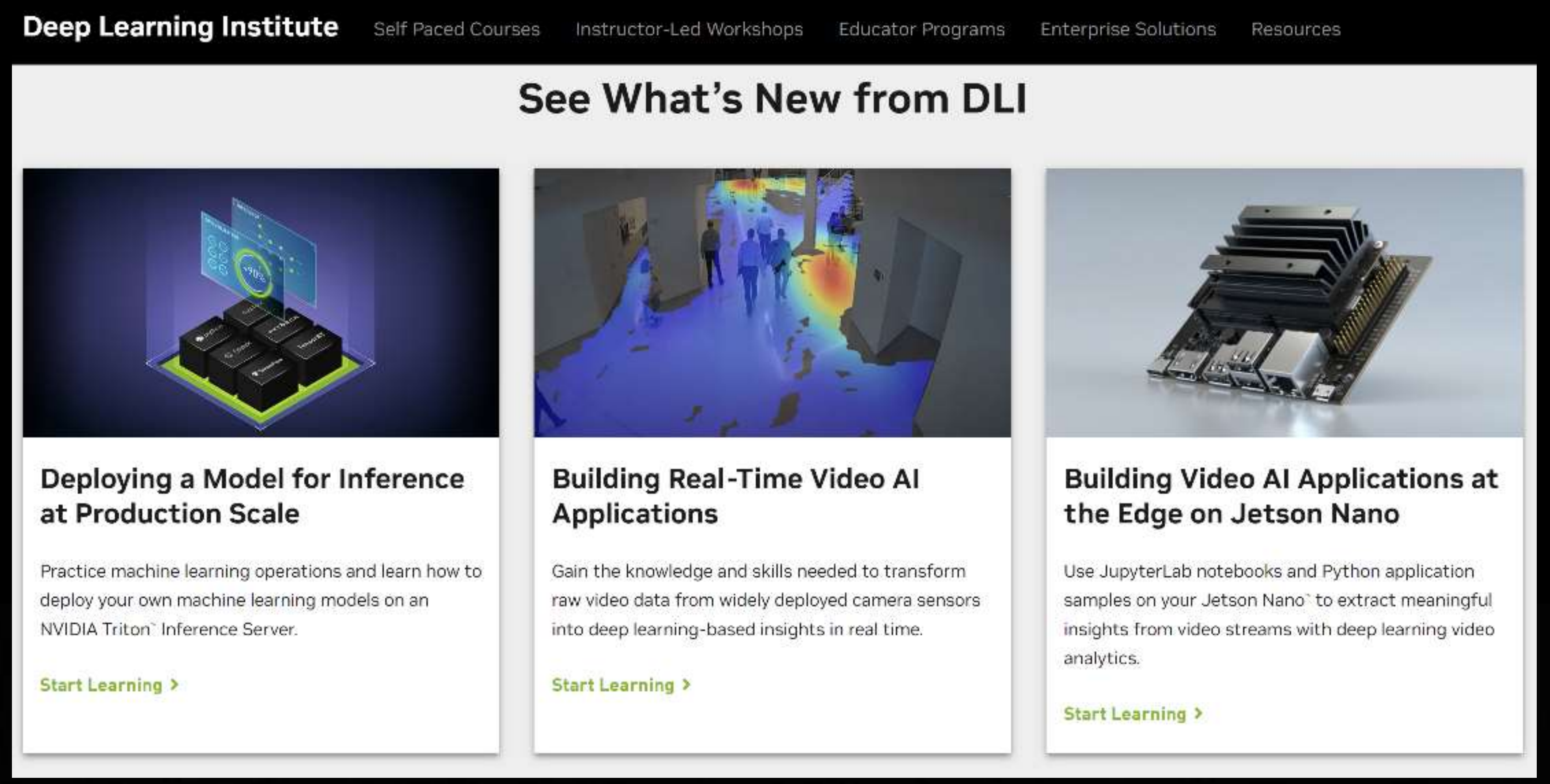
NVIDIA AI - End-to-End AI Development &...
Collection - Deep Learning
This is a collection of performance-

HPC Collection
Collection - High Performance Computing

Deep Learning Frameworks
Collection - Deep Learning
This collection contains performance-

NGC - Getting Started
Collection - Beginner
Looking to get started with containers and

https://catalog.ngc.nvidia.com/



See What's New from DLI

Deploying a Model for Inference at Production Scale

Practice machine learning operations and learn how to deploy your own machine learning models on an NVIDIA Triton` Inference Server.

Start Learning ›

Building Real-Time Video AI Applications

Gain the knowledge and skills needed to transform raw video data from widely deployed camera sensors into deep learning-based insights in real time.

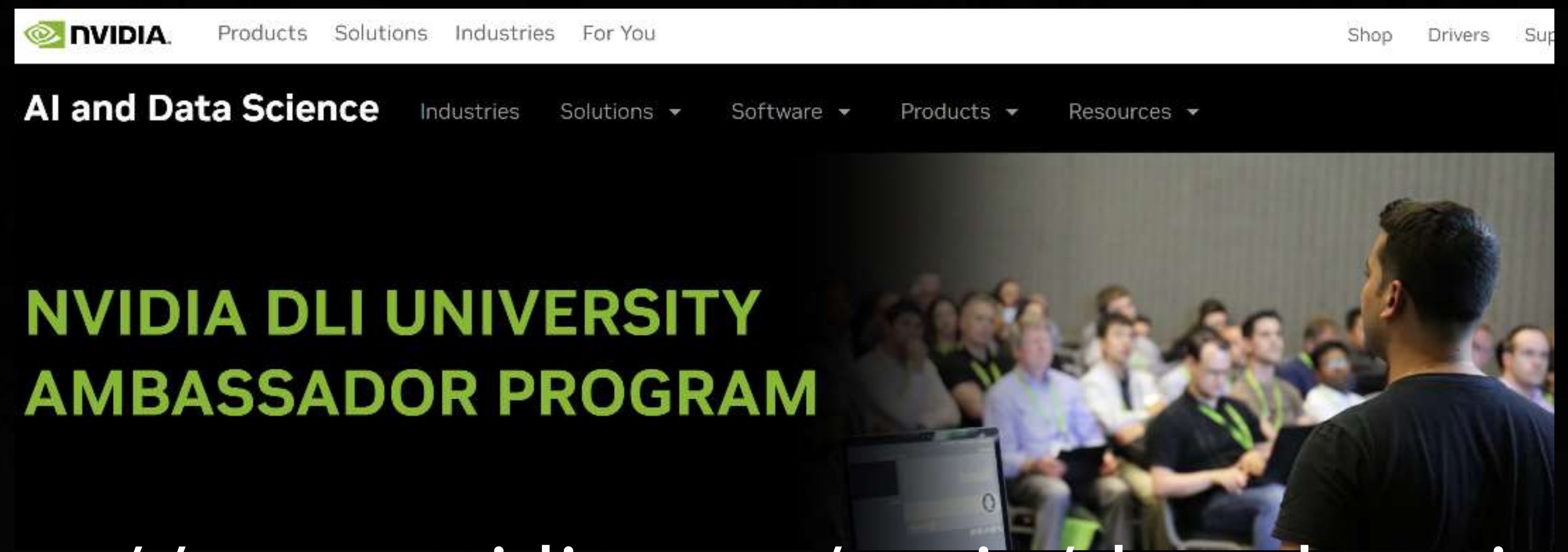
Start Learning ›

Building Video AI Applications at the Edge on Jetson Nano

Use JupyterLab notebooks and Python application samples on your Jetson Nano` to extract meaningful insights from video streams with deep learning video analytics.

Start Learning ›

https://www.nvidia.com/en-us/training/



NVIDIA DLI UNIVERSITY AMBASSADOR PROGRAM

https://www.nvidia.com/en-in/deep-learning-ai/education/ambassador-program/

Empowering the World's Cutting Edge Startups

Join NVIDIA Inception's global network of over 10,000 technology startups.

Apply Now

https://www.nvidia.com/en-us/startups/

Thank You!!!

JAVIERP@NVIDIA.COM

+34 635520529