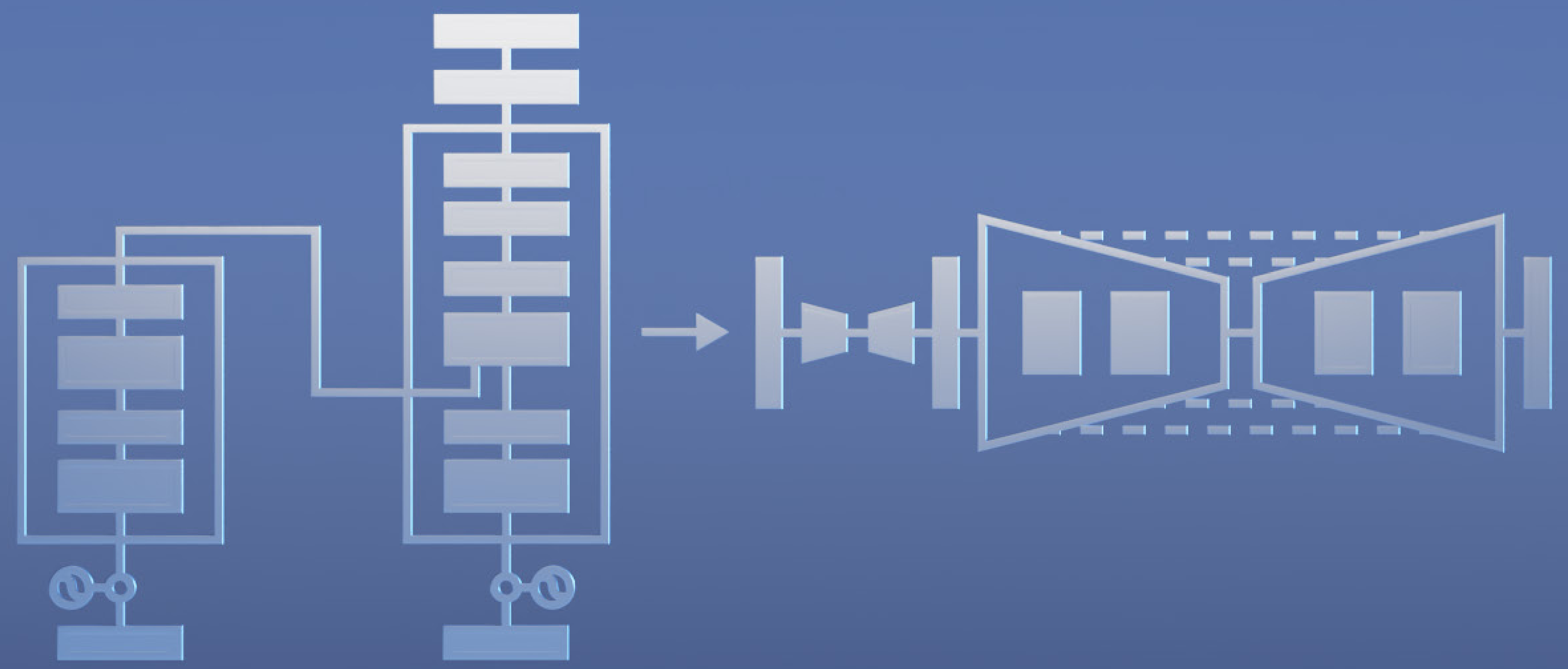
NVIDIA

SUSTAINABILITY

GENERATIVE AI



INDUSTRIAL DIGITALIZATION

cuQuant
Spark
RAPID
RAFT
cuOpt
Holoscan
cuLitho
ensorRT
Triton
TMS
V-CUDA
rabricks
NVIDIA
AI
NVIDIA
Omniverse

DIAMOND SPONSORS

accenture
Alibaba Cloud
aws
飞桨
火山引擎
DELL Technologies

Deloitte.
DOMINO
GIGABYTE™
Google Cloud
Hewlett Packard Enterprise
IBM

快手 · StreamLake
Lenovo
Microsoft
ORACLE CLOUD Infrastructure
QCT
run:ai

SUPERMICRO®
Tencent Cloud
vmware®
xFUSION
Z hp

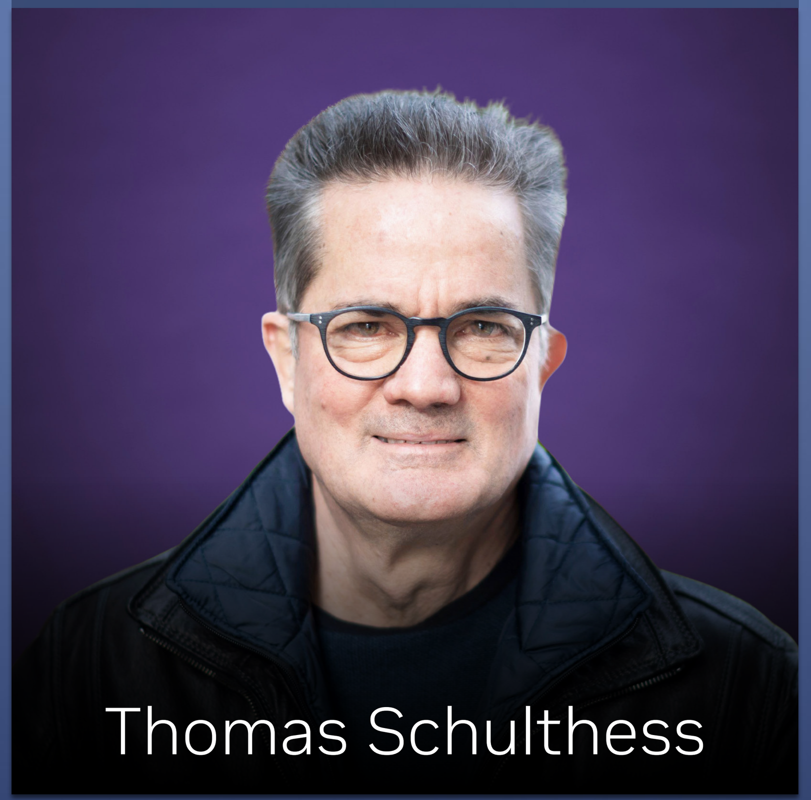
DeepMind

Argonne National Laboratory

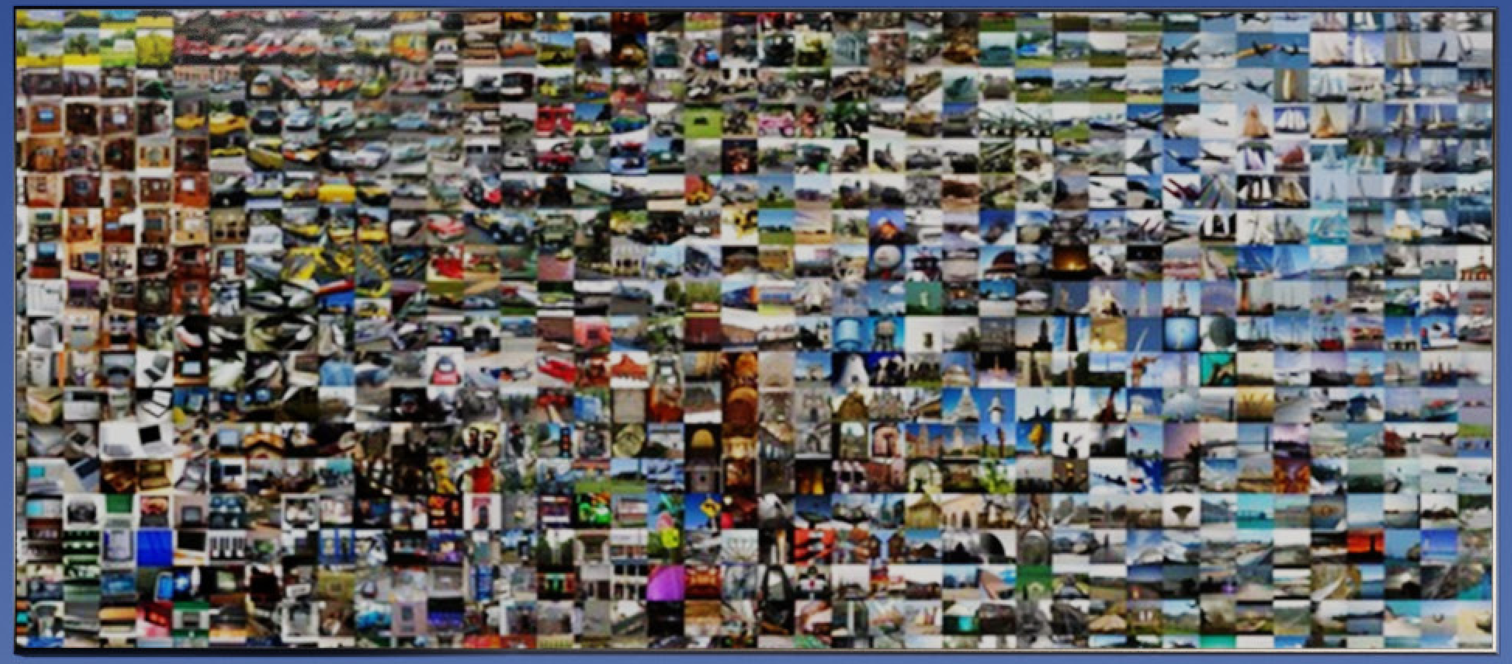
Adobe

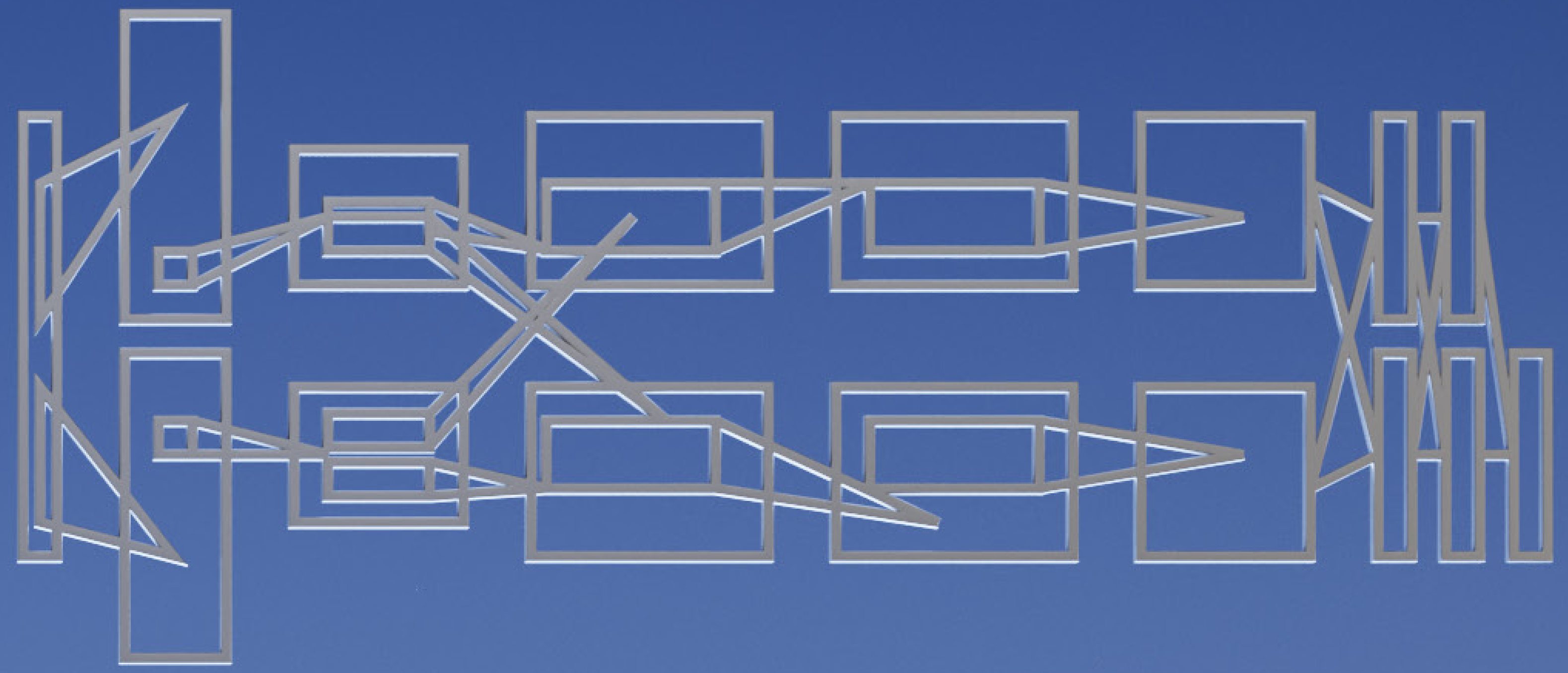
Netflix

ETH Zurich



OpenAI

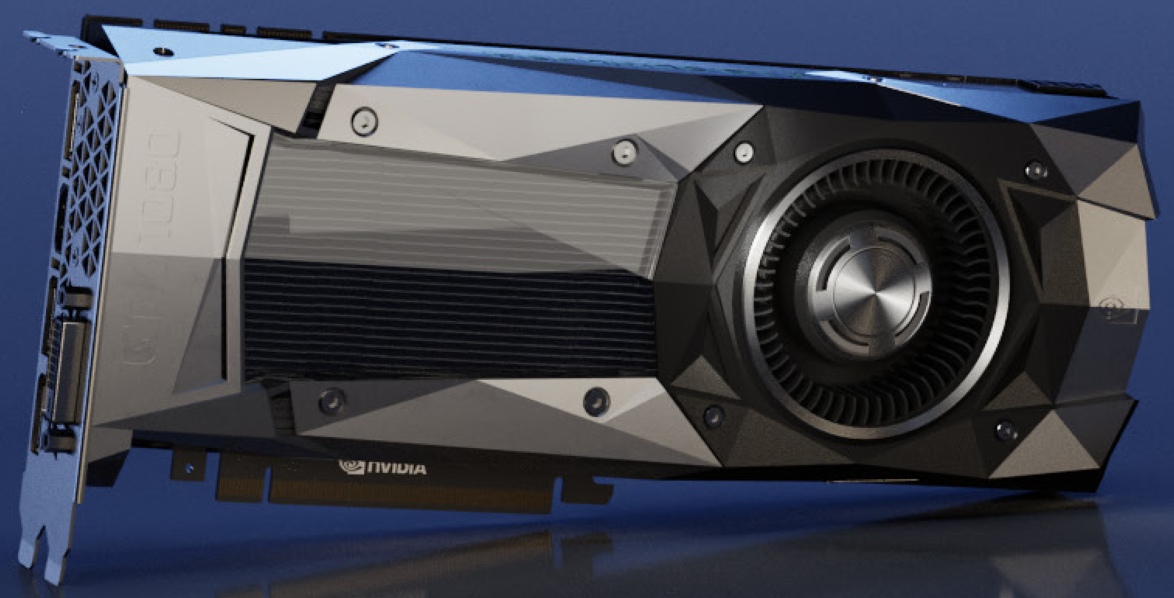


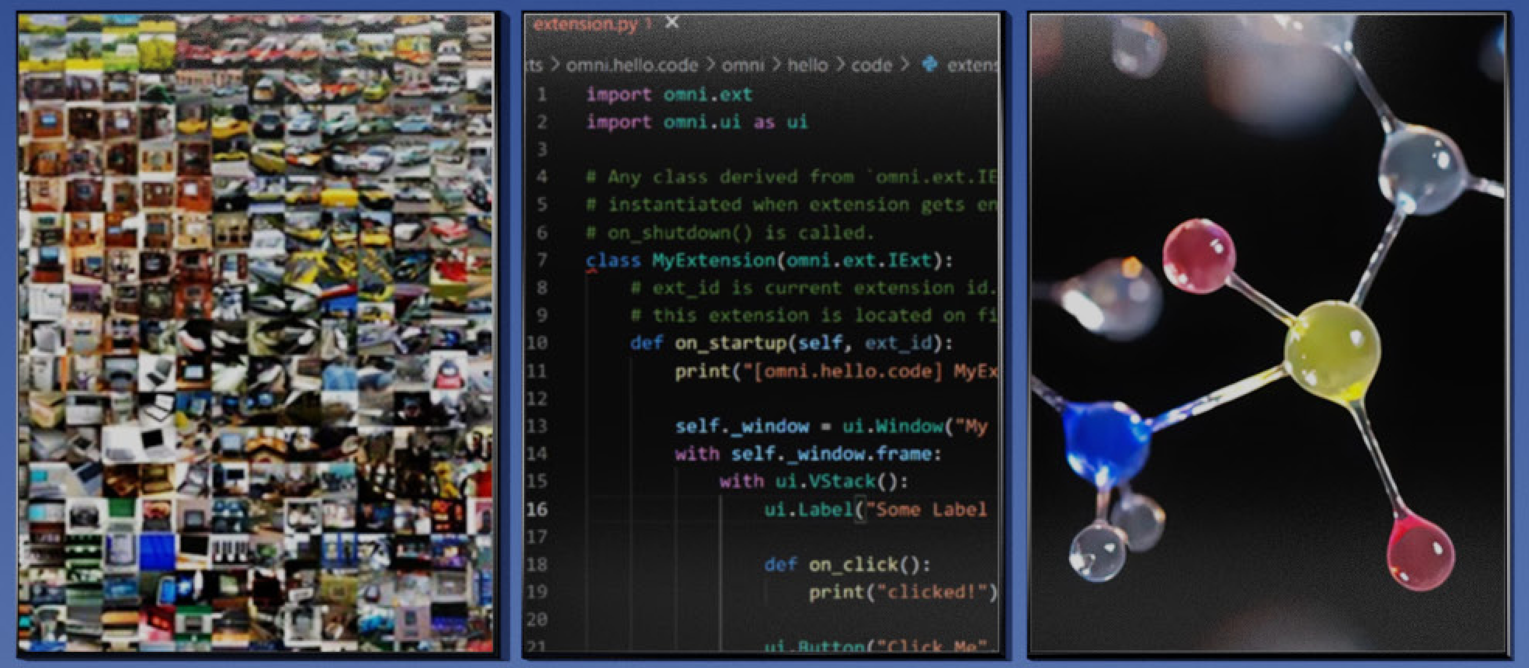


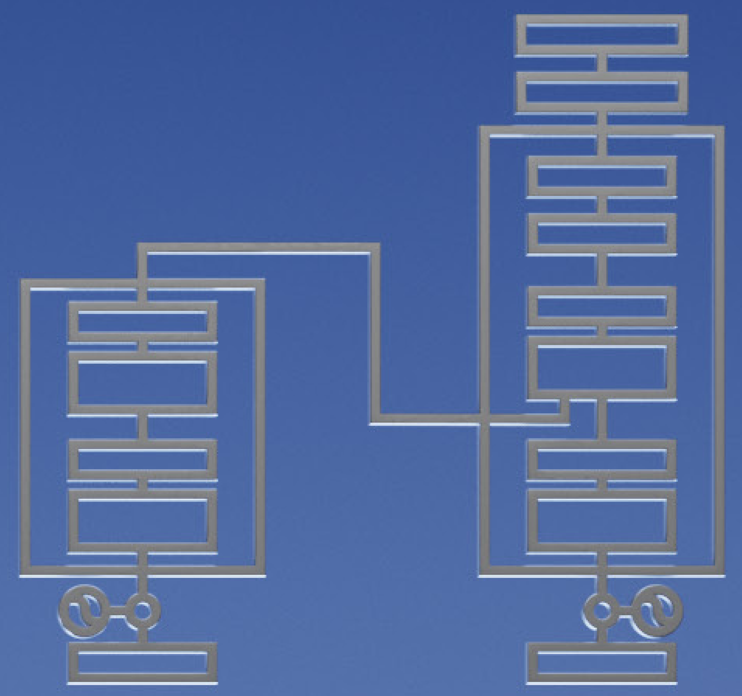
AlexNet

61M Parameters

262 PetaFLOPS





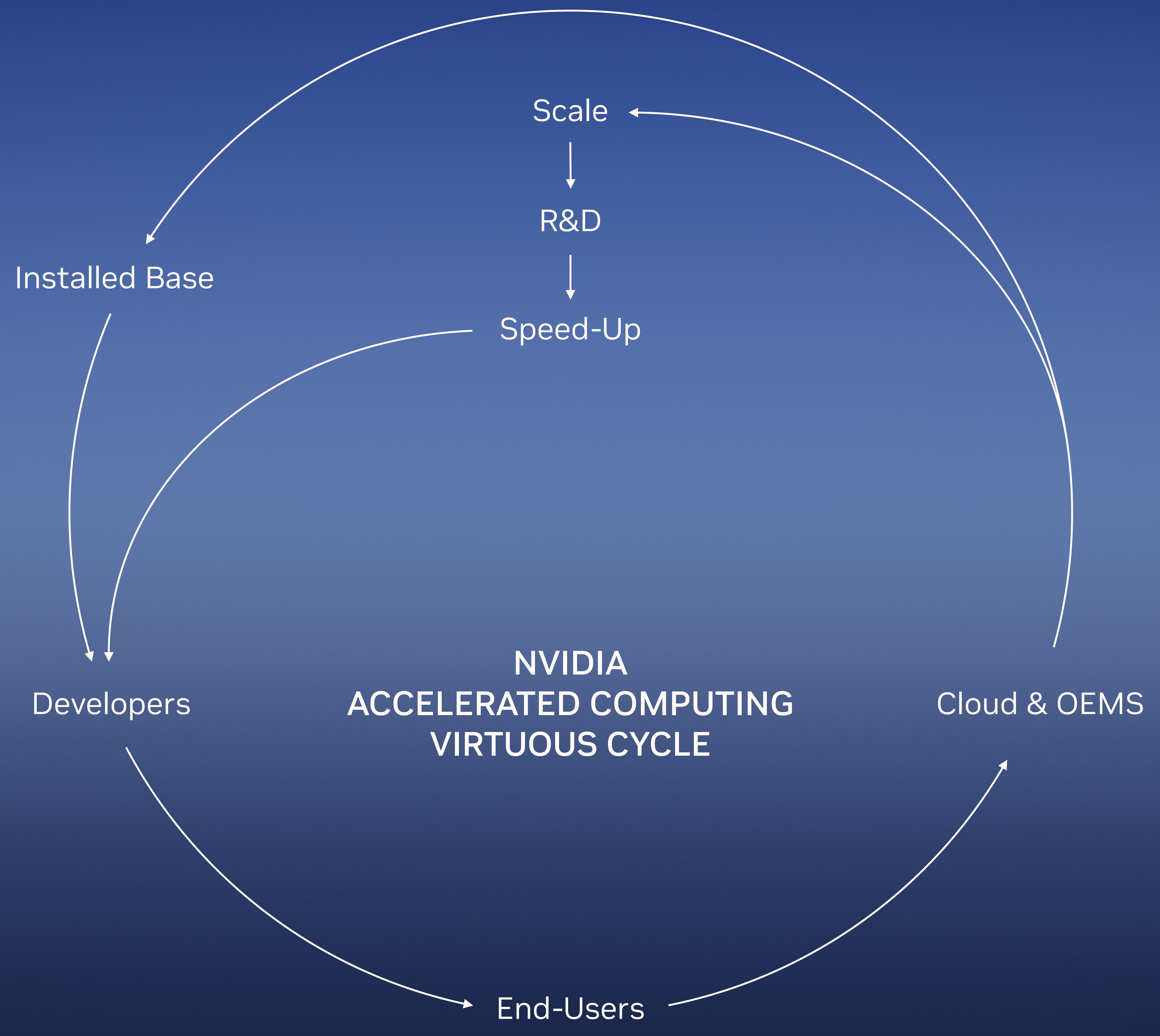


GPT-3

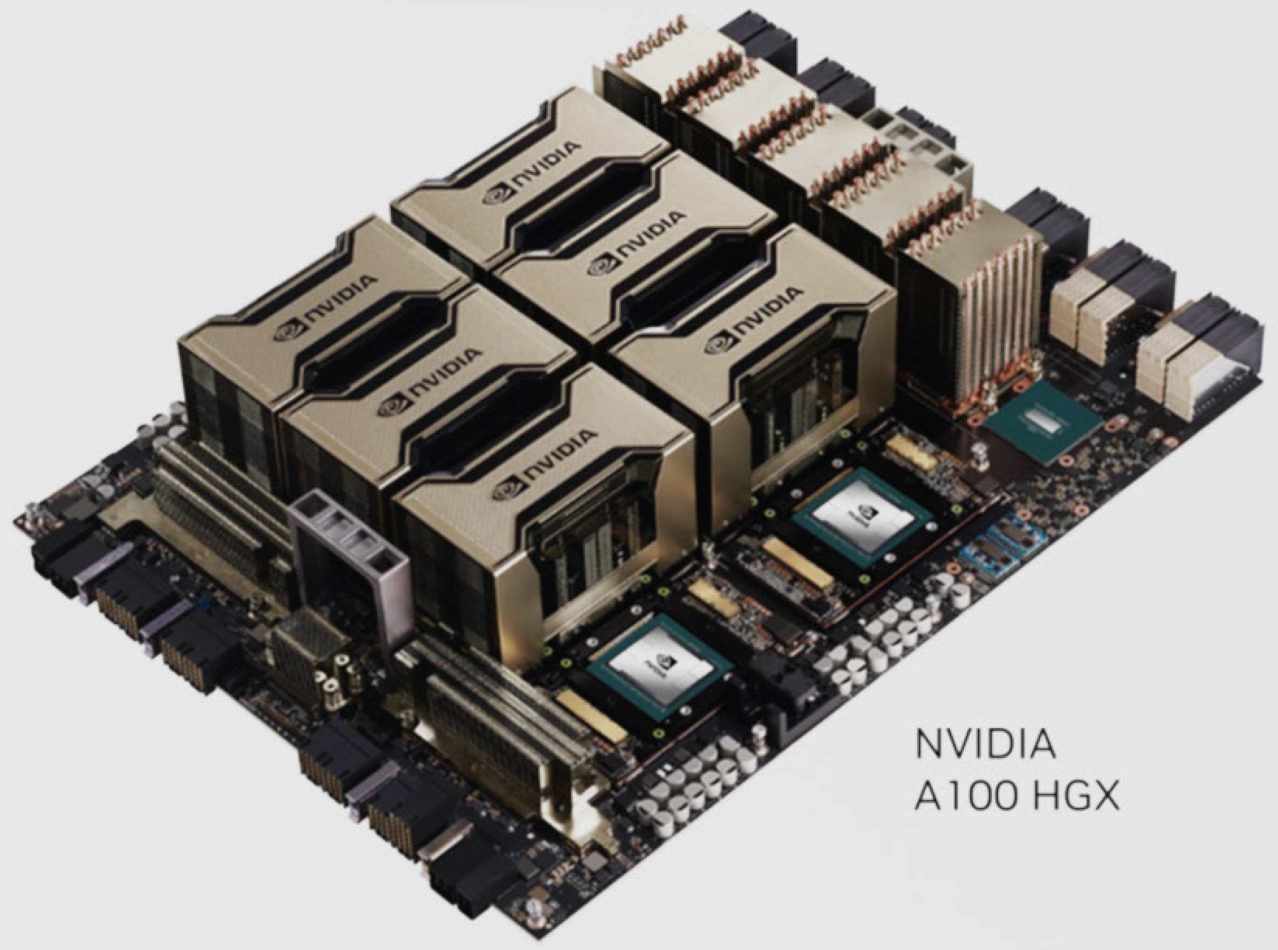
175B Parameters

323 ZettaFLOPS

cuQuantum
Acceleration Libraries
NVIDIA AI
NVIDIA Omniverse


Scale
R&D
Speed-Up
Installed Base
Developers
NVIDIA
ACCELERATED COMPUTING
VIRTUOUS CYCLE
Cloud & OEMS
End-Users

GPU Acceleration: A Tipping Point for High-fidelity CFD



NVIDIA
A100 HGX

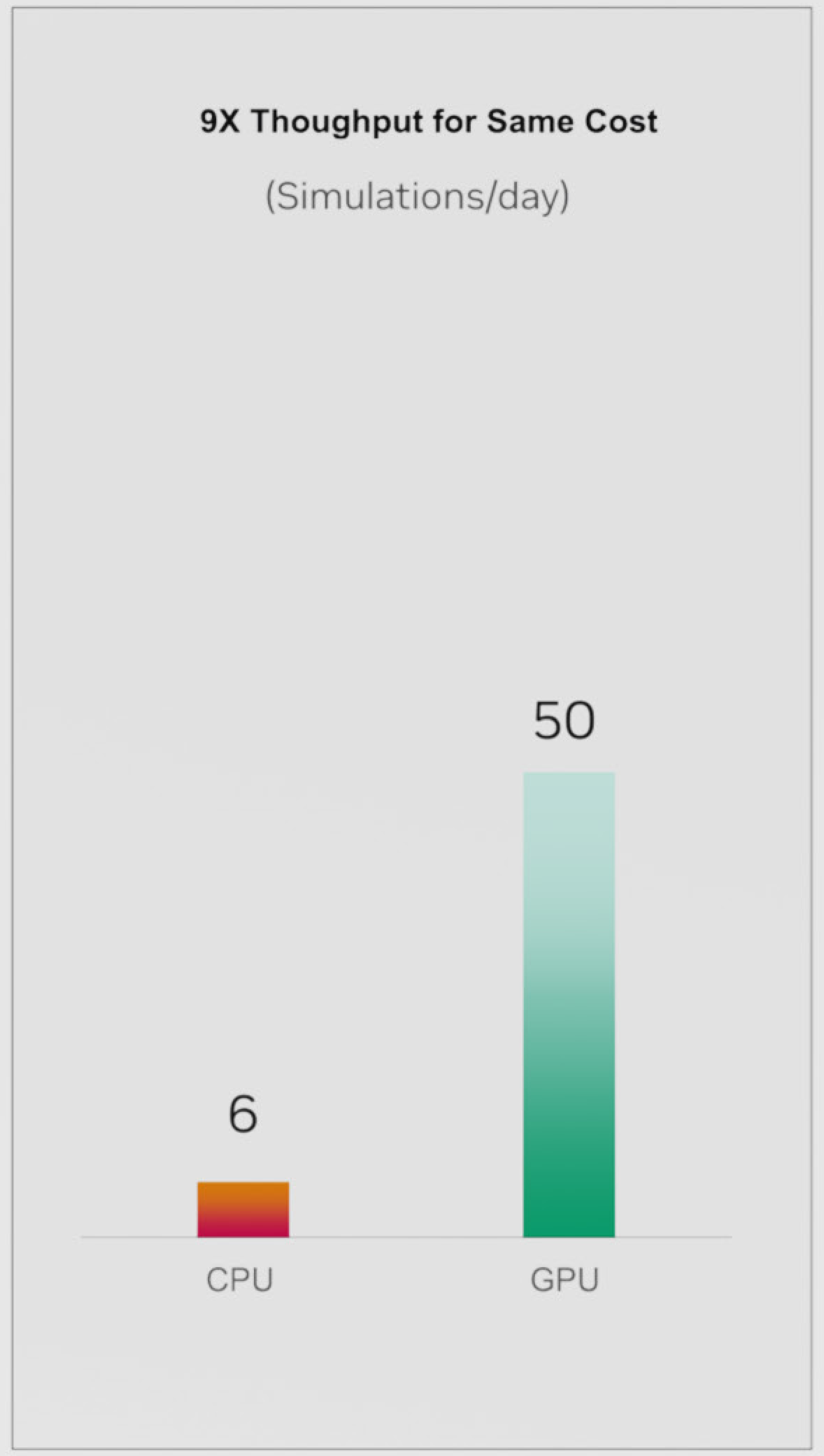
GPU acceleration is transforming high-fidelity CFD

9x throughput for the **SAME cost** of CPU*

9x lower cost and **17x less energy** for the **SAME throughput** of CPU*

NVIDIA's mature SDKs have been critical in supporting this transformation

Impacting designs in aerospace, automotive, and many other industries



*Basis: CFD reacting compressible charLES solver on 50M cell grid running for 50ms per case.
Hardware: CPU: c6a.32xlarge AMD Epyc Gen2 , GPU p4d.24xlarge 8 x NVIDIA A100 GPU

cādence


cuQuantum
NVIDIA AI
NVIDIA Omniverse
agnostiq
Cirq
QUANTUM LEAF
aws
QuEra
QM ware
Qiskit
PENNYLANE


cuQuantum
NVIDIA AI
NVIDIA Omniverse
ANYON
IONQ
ORCA COMPUTING
atom computing
IQM
OQC
QUANTINUUM
PASQAL
QUANTUM BRILLIANCE
rigetti
QuEra
XANADU
NVIDIA


cuQuantum
NVIDIA AI
NVIDIA Omniverse
QUANTUM
MACHINES


Spark
RAPIDS
NVIDIA AI
NVIDIA Omniverse
CLOUDERA
Amazon EMR
databricks
Dataproc

RAFT
NVIDIA AI
NVIDIA Omniverse
FAISS
milvus
redis


cuOpt
NVIDIA AI
NVIDIA Omniverse

AT&T

NVIDIA


cuOpt
NVIDIA AI
NVIDIA Omniverse
accenture
Capgemini
Deloitte.
softserve
quantiphi
NVIDIA


TensorRT
Triton
TMS
CV-CUDA
NVIDIA AI
NVIDIA Omniverse
amazon
Microsoft
AMERICAN EXPRESS
UNITED STATES POSTAL SERVICE


Triton
CV-CUDA
NVIDIA AI
NVIDIA Omniverse
Uber
ROBLOX

CV-CUDA
VPF
Parabricks
NVIDIA AI
NVIDIA Omniverse
Microsoft Bing
runway
Tencent 腾讯


Parabricks
NVIDIA AI
NVIDIA Omniverse
FORM BIO
DNAnexus
Terra


Parabricks
NVIDIA AI
NVIDIA Omniverse
bionano
PacBio
nanoString
Oxford NANOPORE Technologies
SINGULAR GENOMICS
ULTIMA GENOMICS


cuQuantum
Holoscan
NVIDIA AI
NVIDIA Omniverse

Medtronic

NVIDIA

ASML EUV LITHOGRAPHY

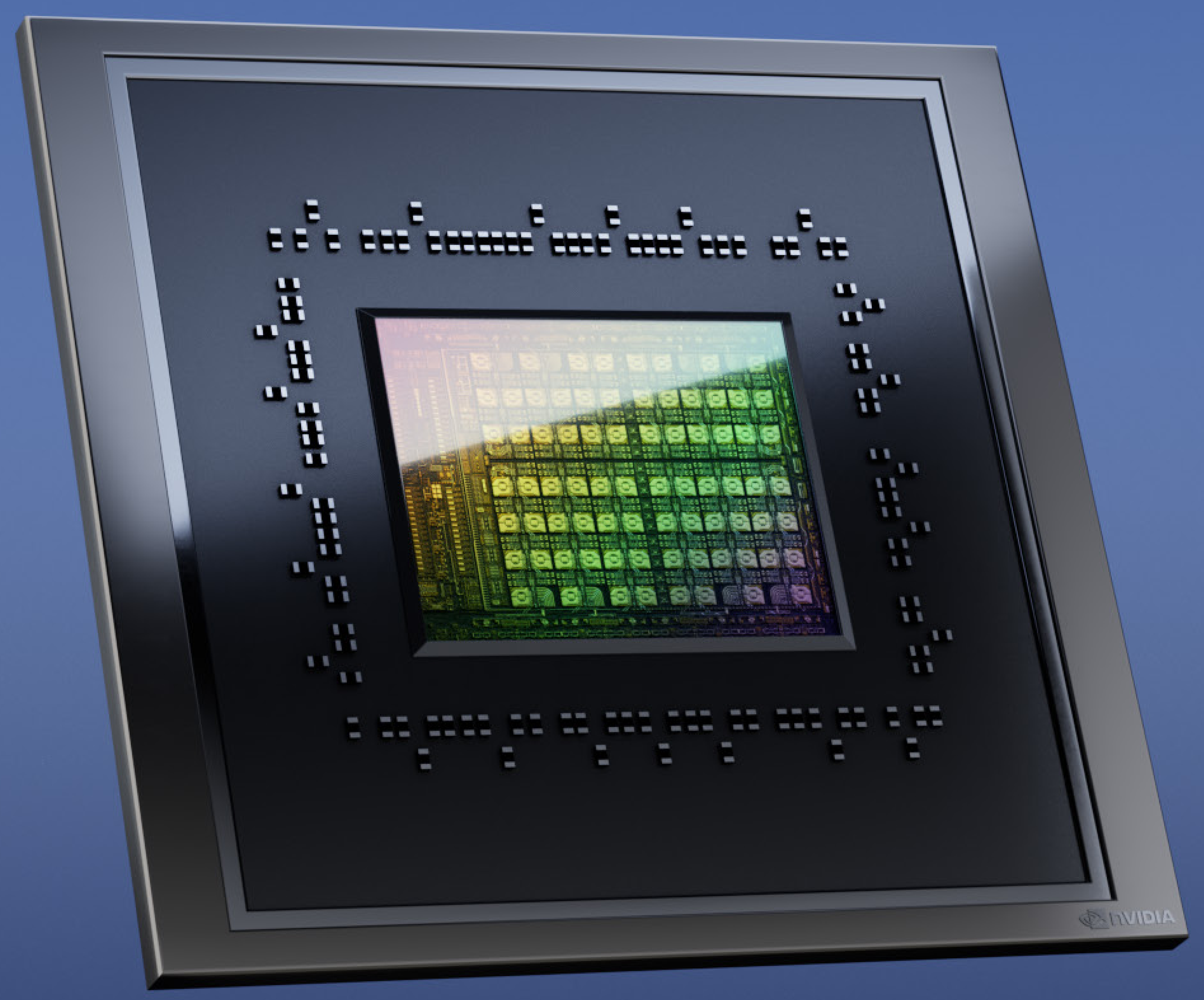
Operating at the Limits of Physics

ASML

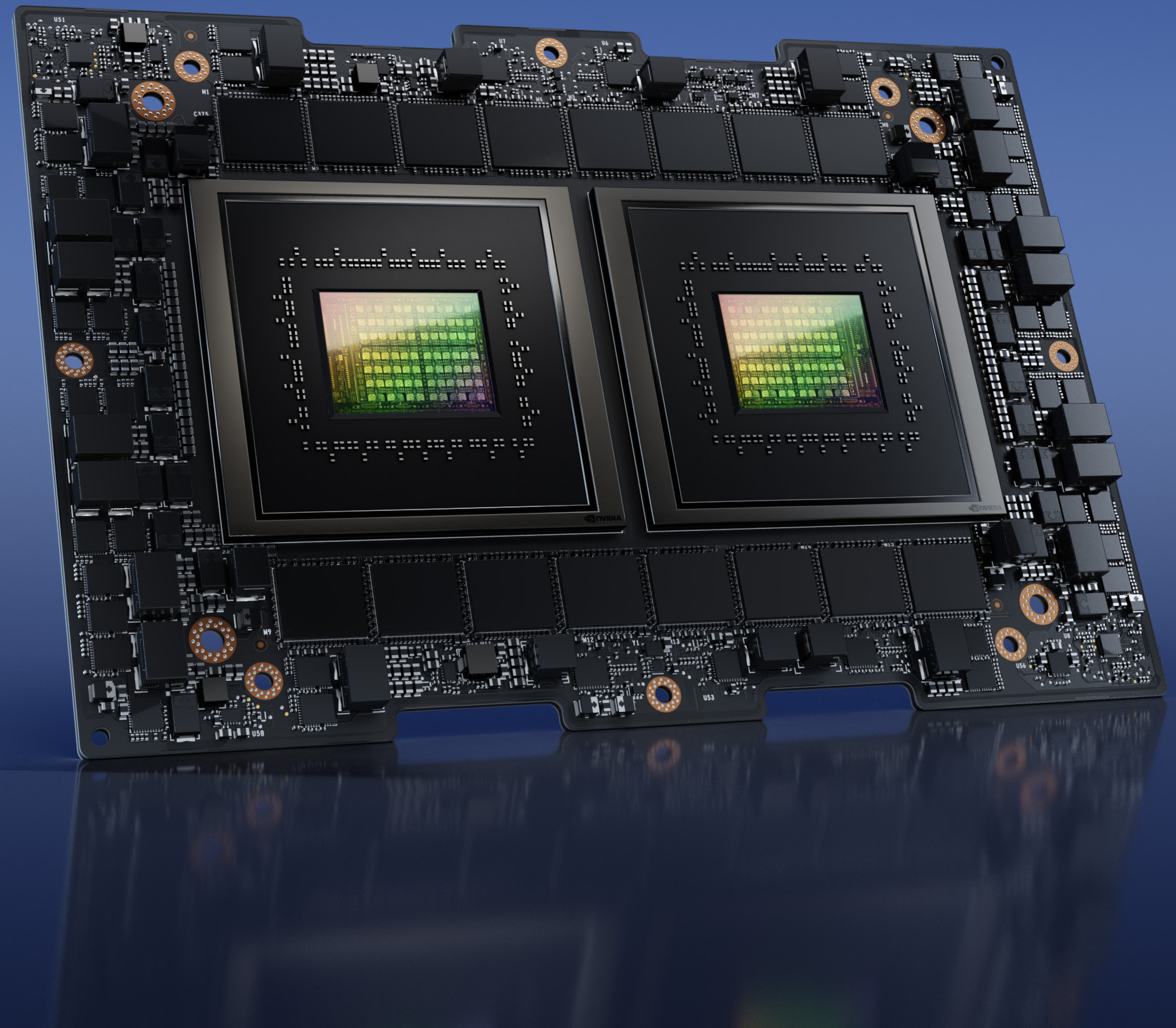
tsmc®

SYNOPSYS®

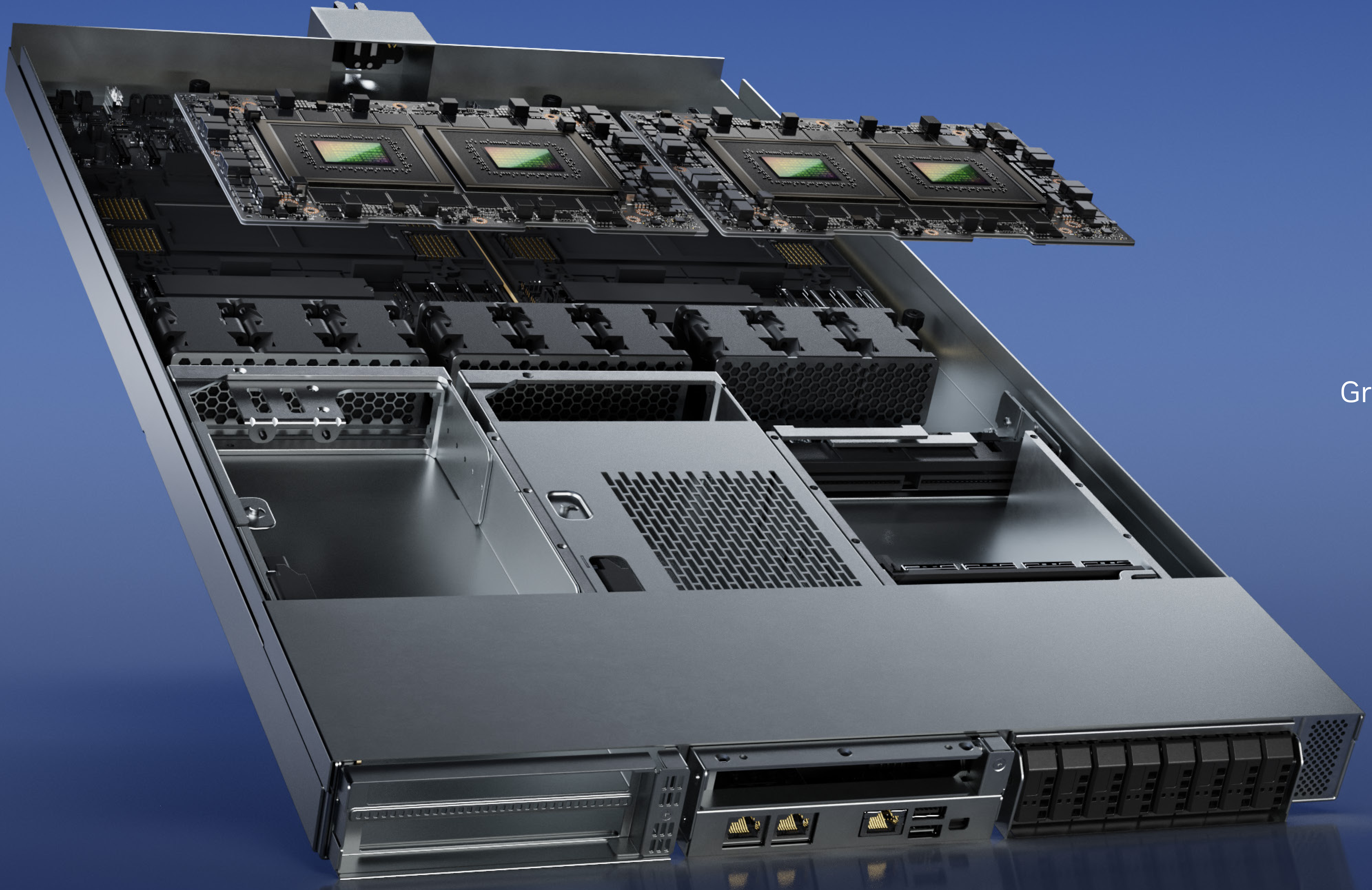
NVIDIA



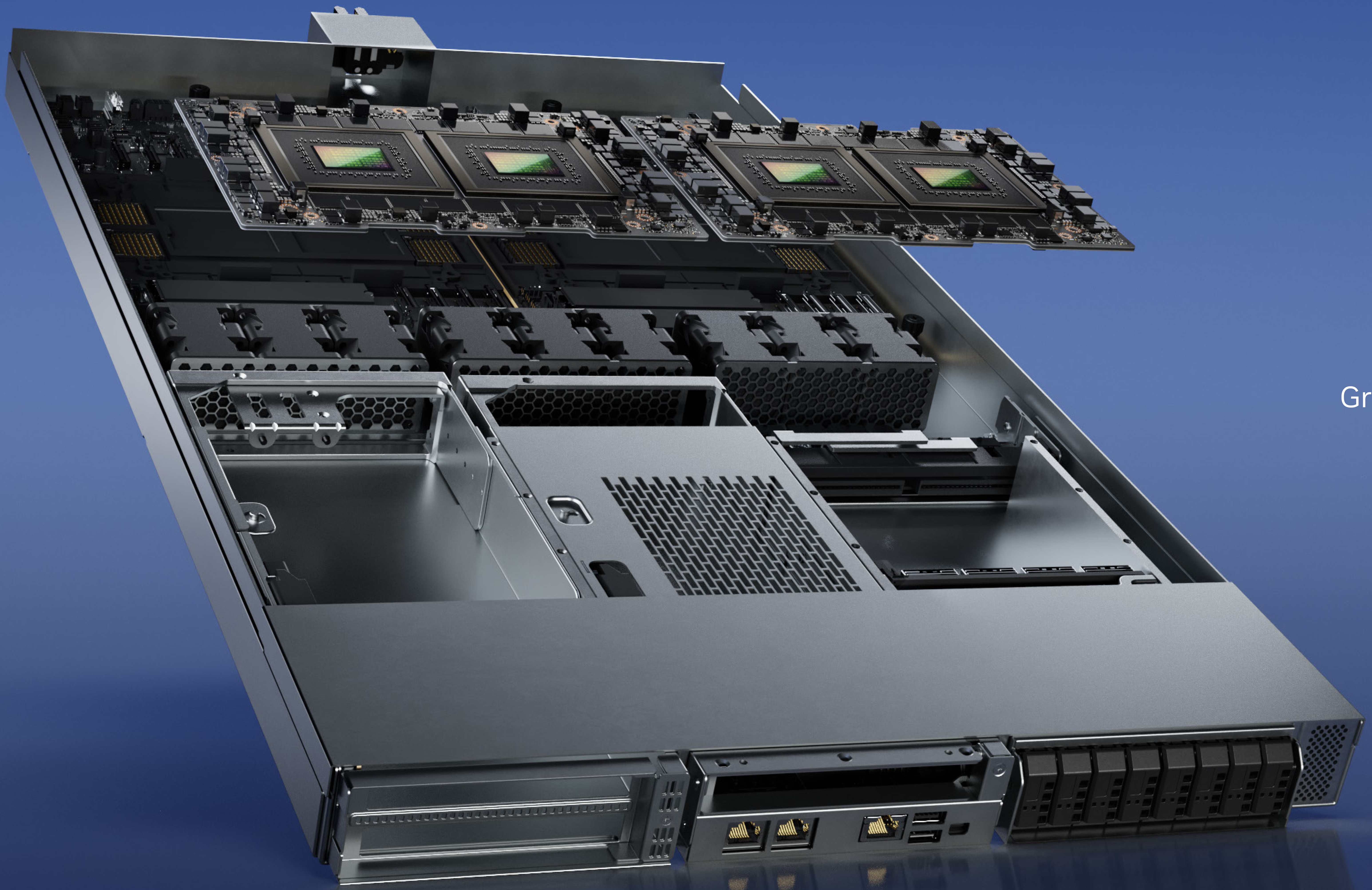
Grace CPU



Grace CPU Superchip



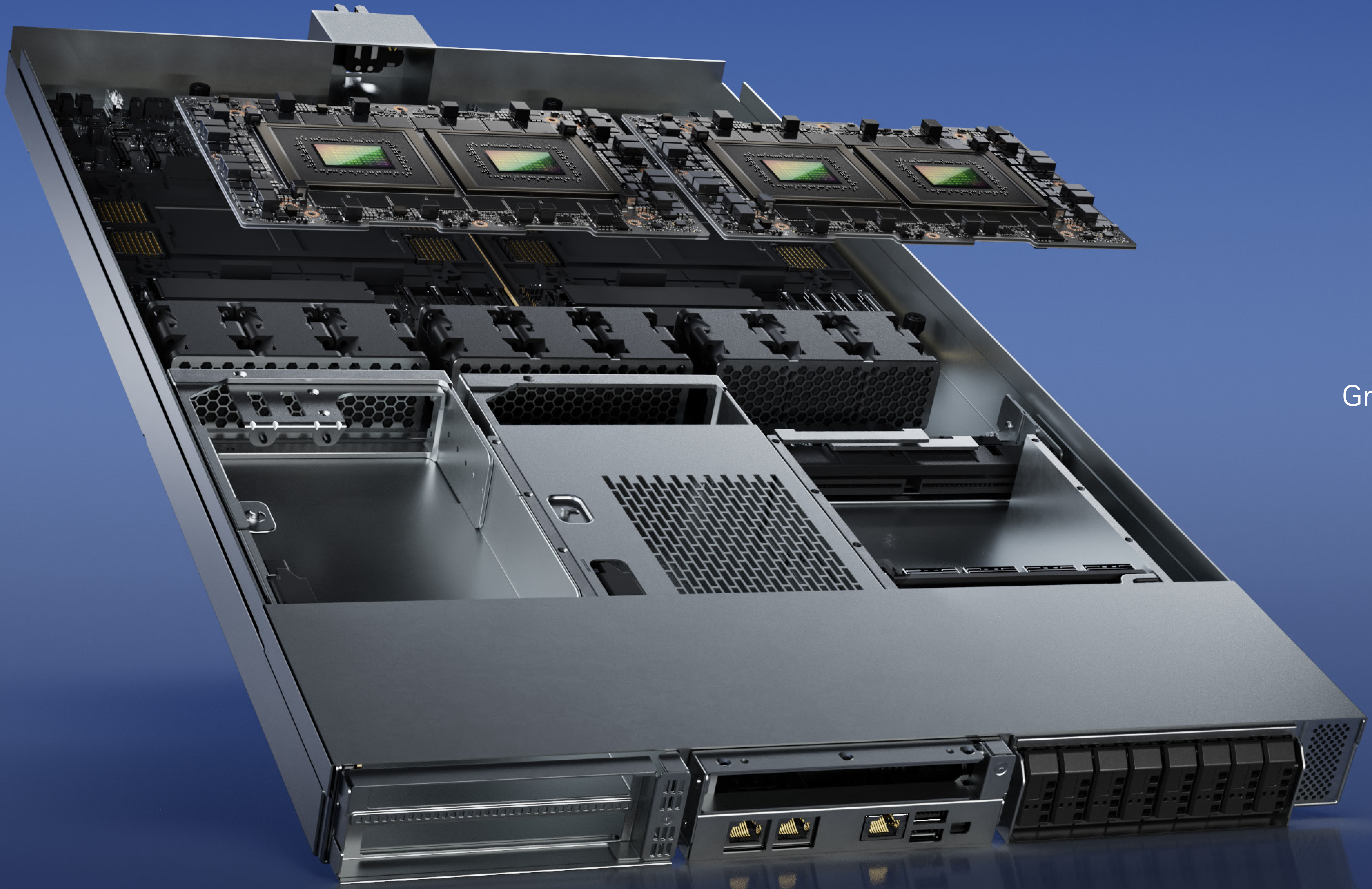
Grace CPU Superchip 1U Server



Grace CPU Superchip 1U Server

NVIDIA GRACE
VS
NEXT-GEN X86

<table>
<tr><th></th><th>Performance</th><th>Energy Efficiency</th><th>Data Center Throughput</th></tr>
<tr><td>Microservices</td><td>1.3X</td><td rowspan="2">1.7X</td><td rowspan="2">2X</td></tr>
<tr><td>Big Data</td><td>1.2X</td></tr>
</table>



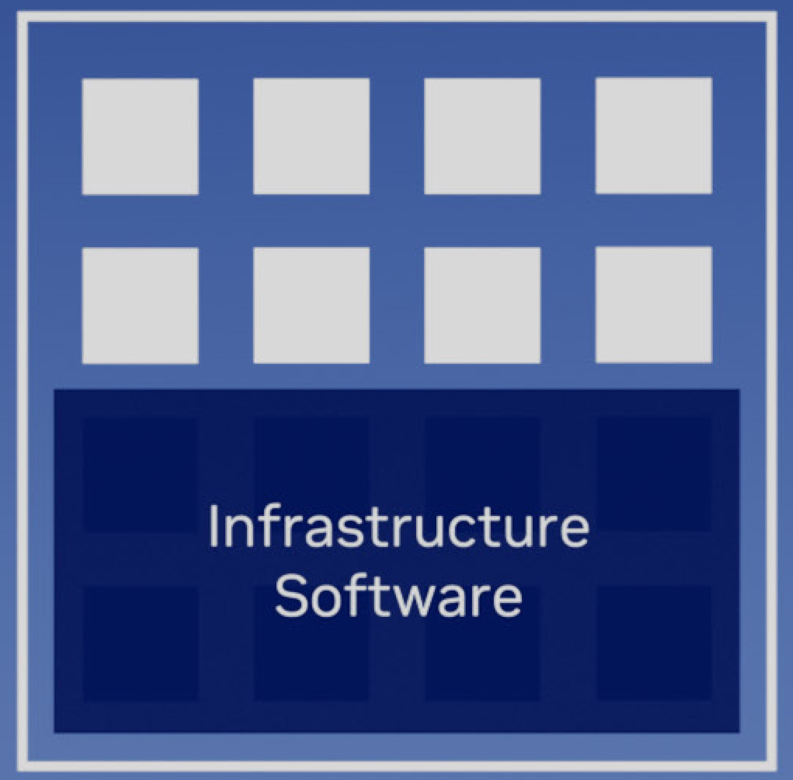
Grace CPU Superchip 1U Server

NVIDIA GRACE
VS
NEXT-GEN X86

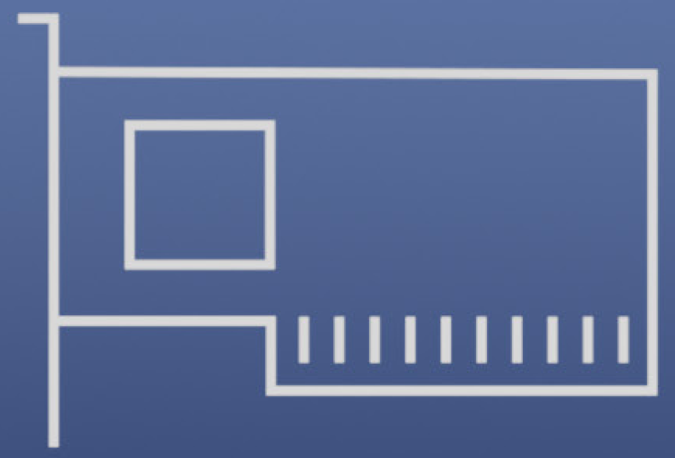
	Performance	Energy Efficiency	Data Center Throughput
Microservices	1.3X	1.7X	2X
Big Data	1.2X		

ASUS Atos GIGABYTE Hewlett Packard Enterprise

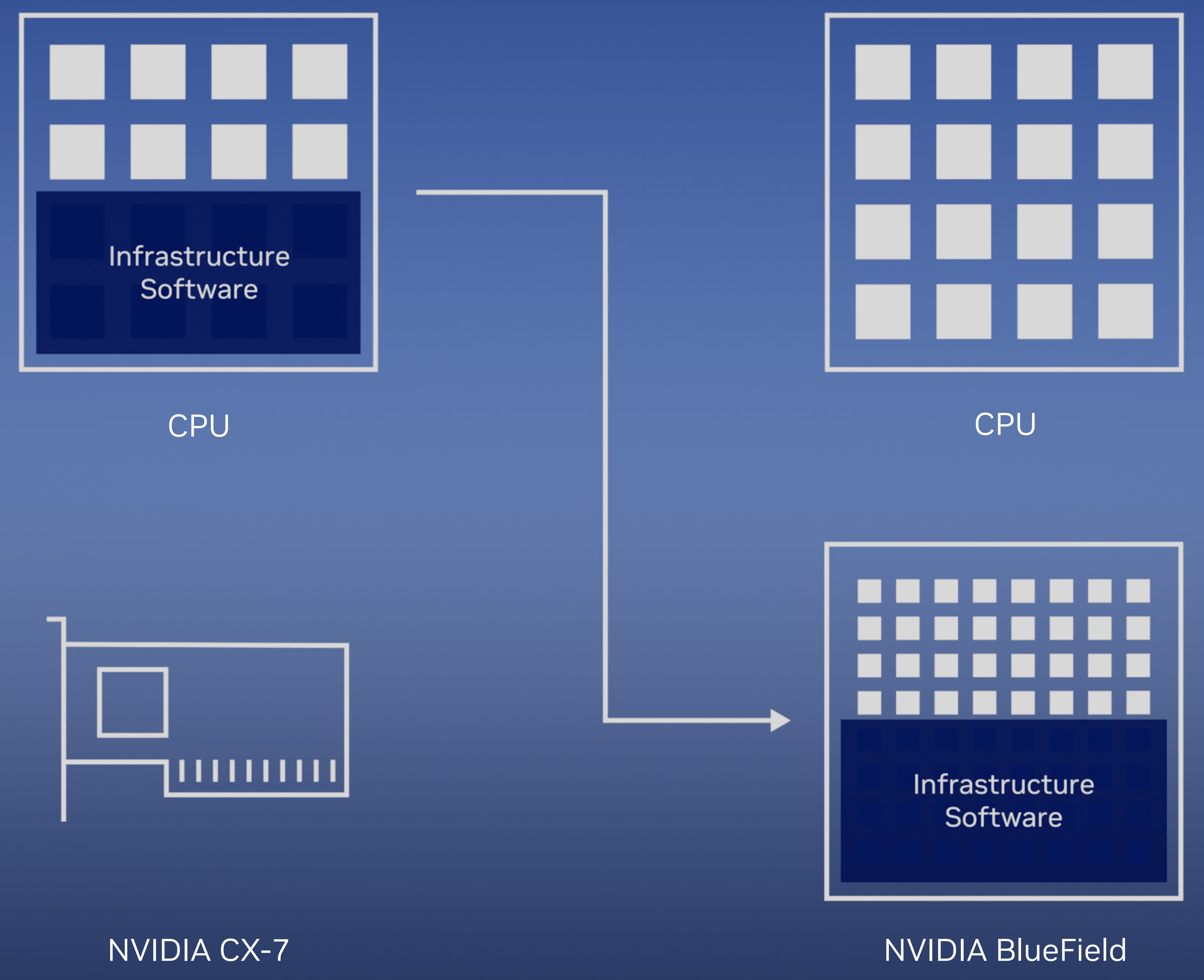
QCT SUPERMICRO wistron ZT Systems



CPU



NVIDIA CX-7


Infrastructure Software
CPU
CPU
Infrastructure Software
NVIDIA CX-7
NVIDIA BlueField

CHECK POINT™

cisco

ddn

DELL EMC

JUNIPer® NETWORKS

paloalto® NETWORKS

Red Hat

vmware®

NVIDIA BLUEFIELD-3

BAIDU AI CLOUD

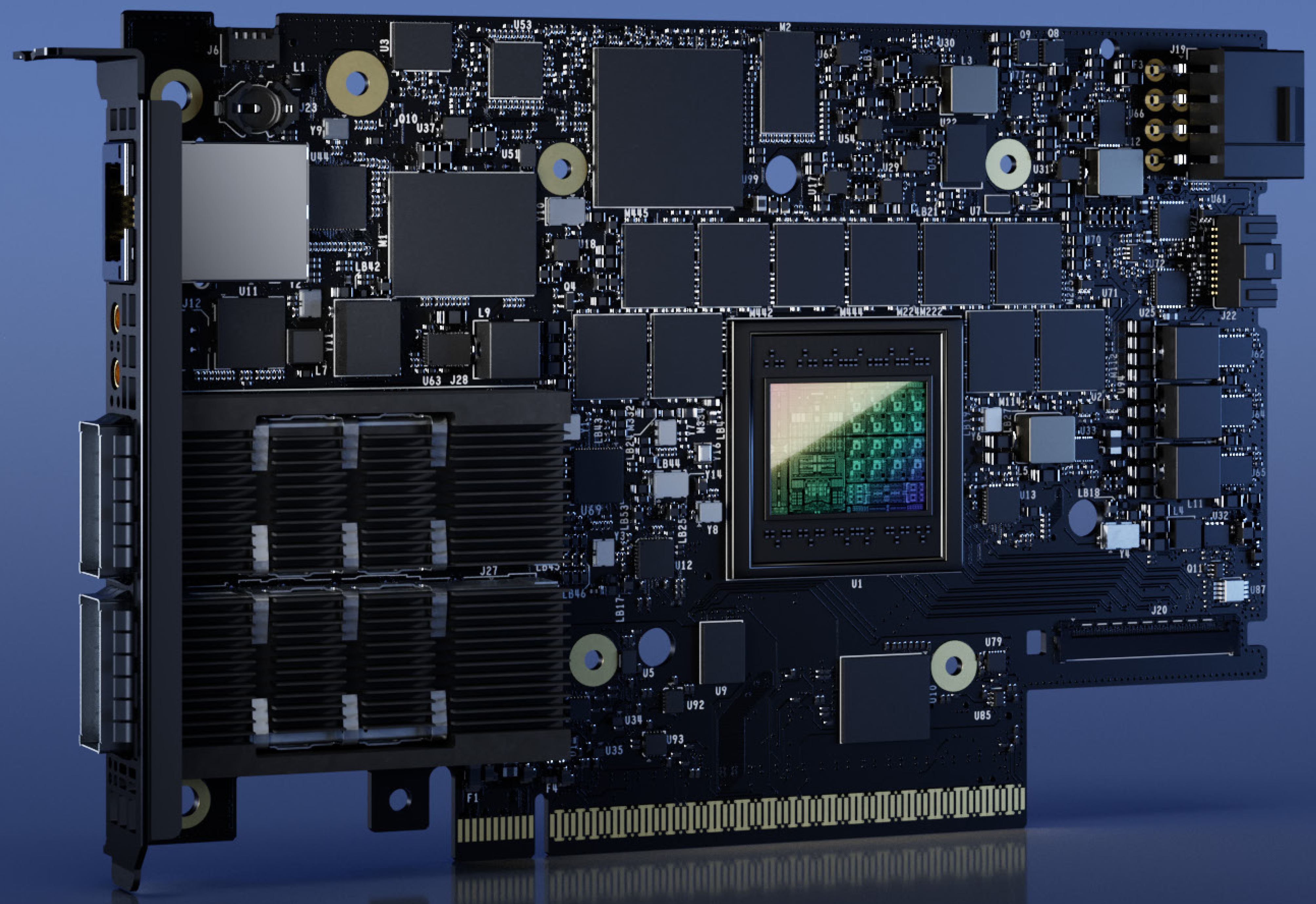
CoreWeave

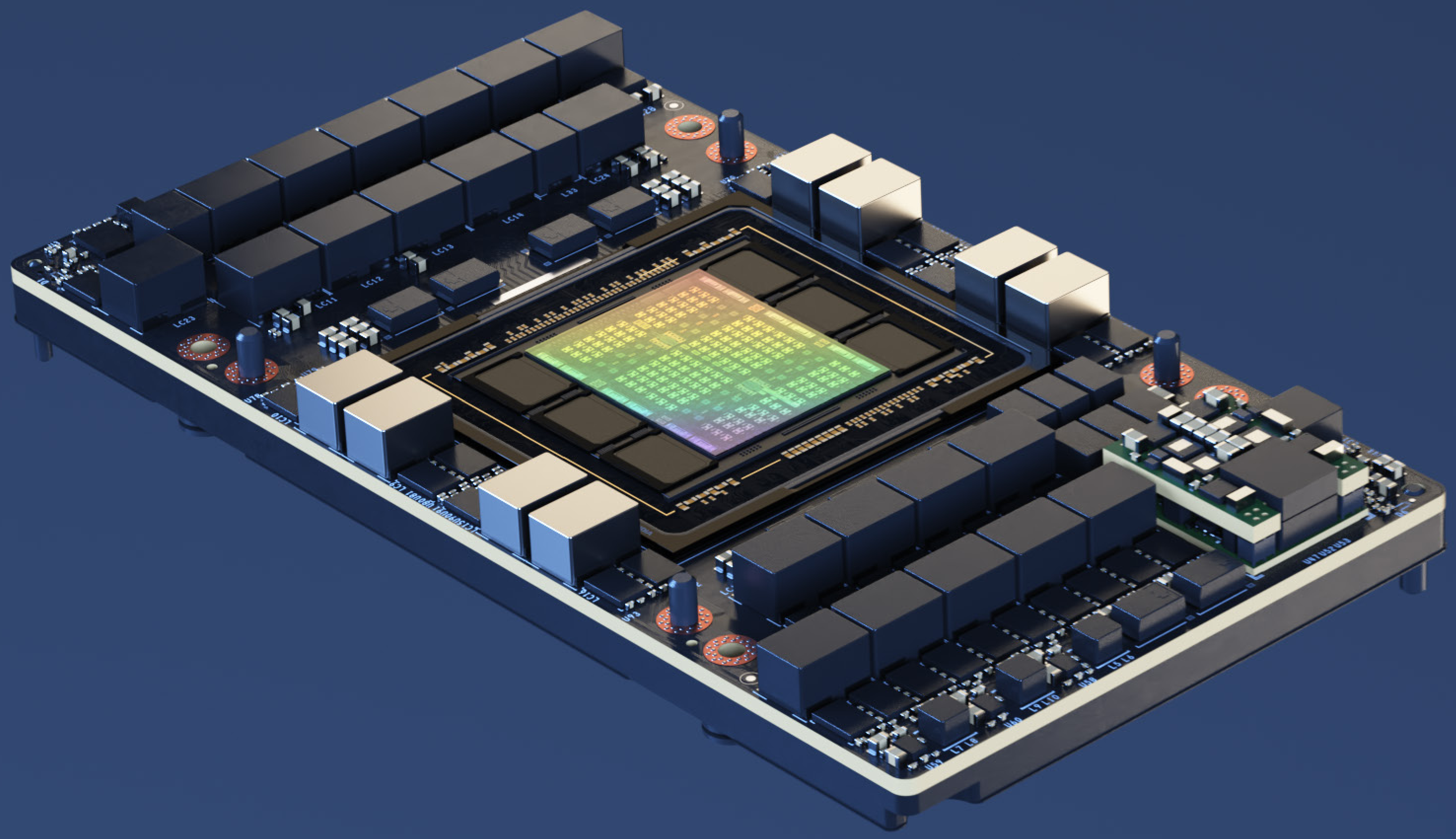
JD.COM

Microsoft Azure

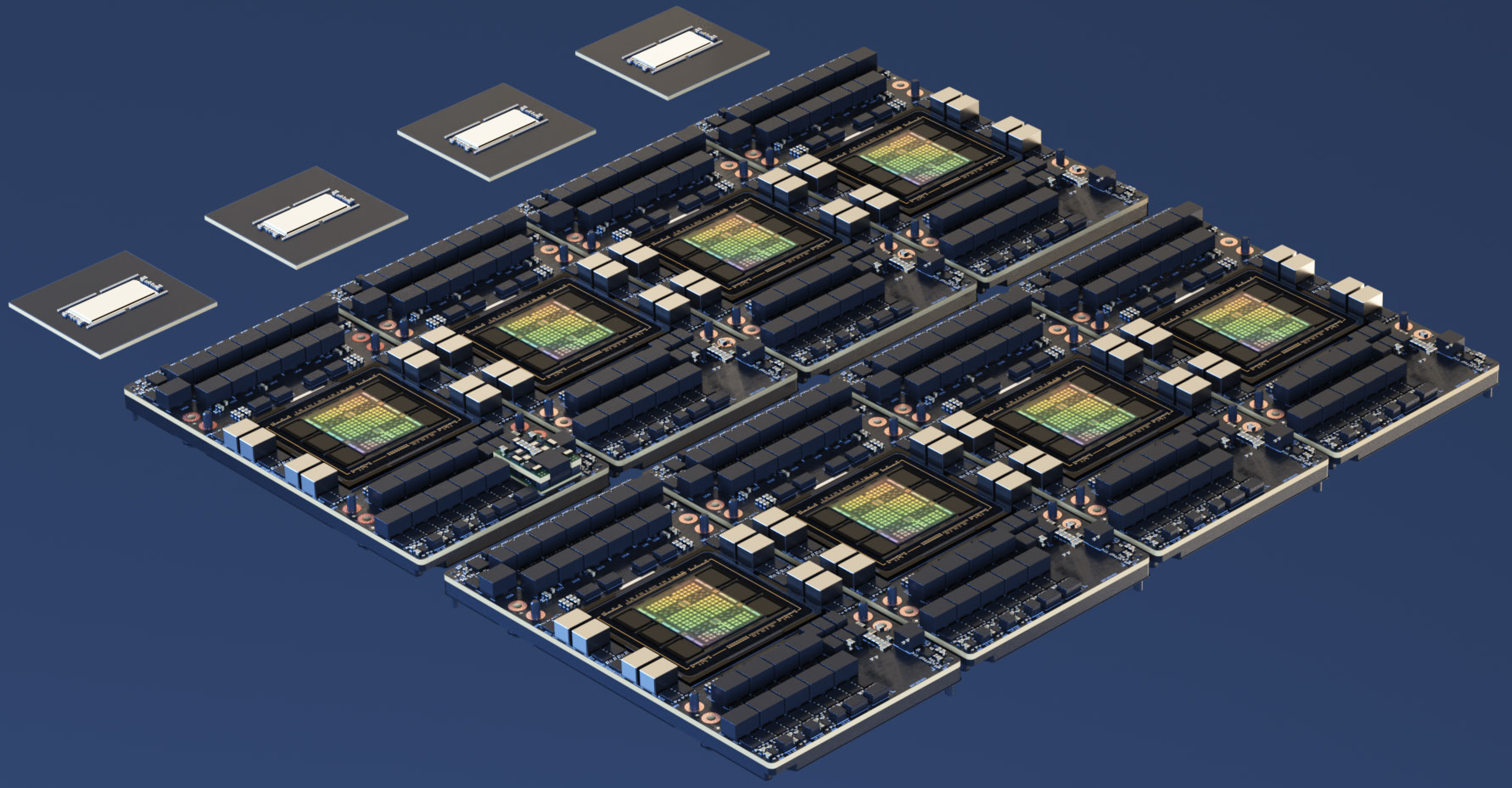
ORACLE CLOUD Infrastructure

Tencent 腾讯

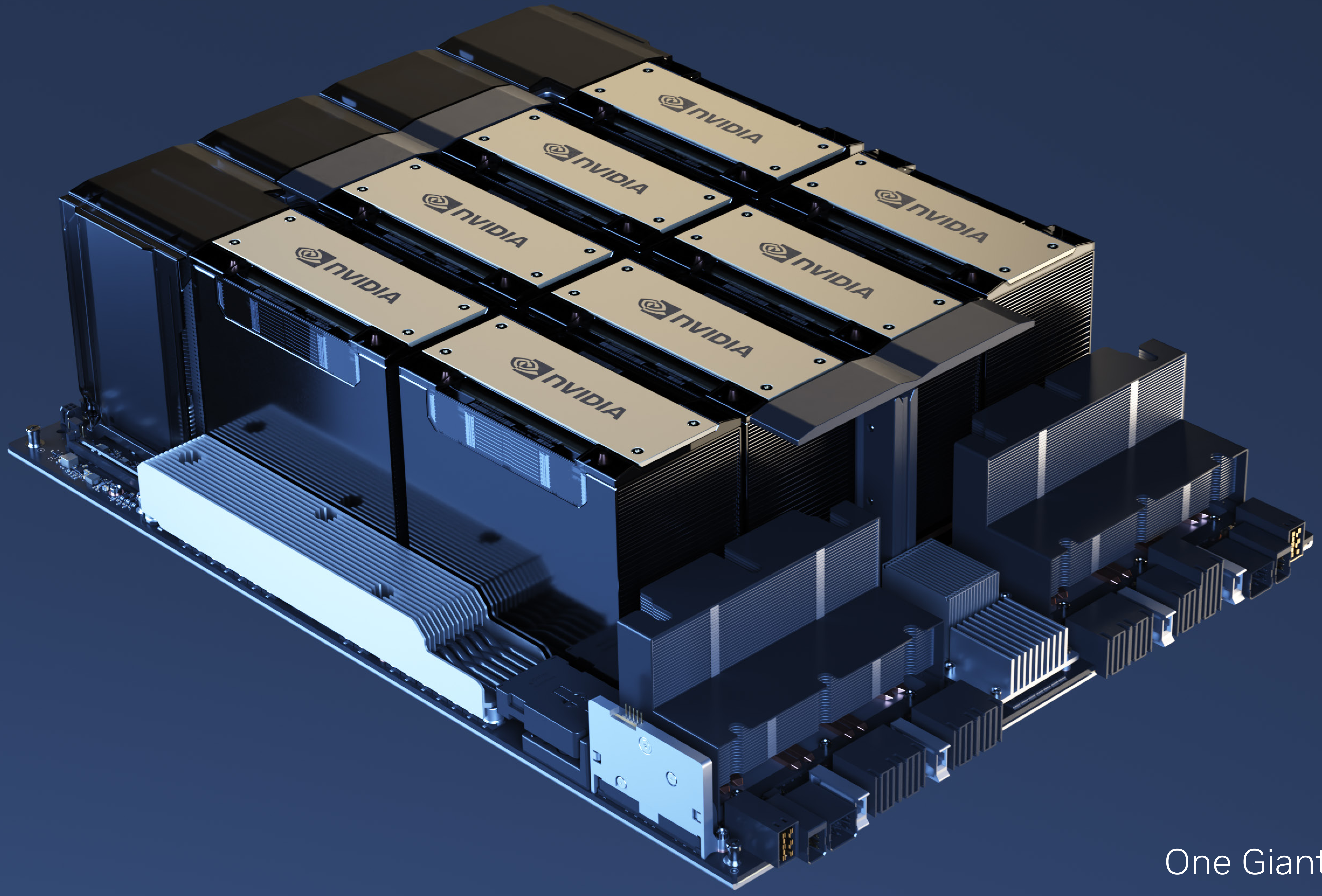




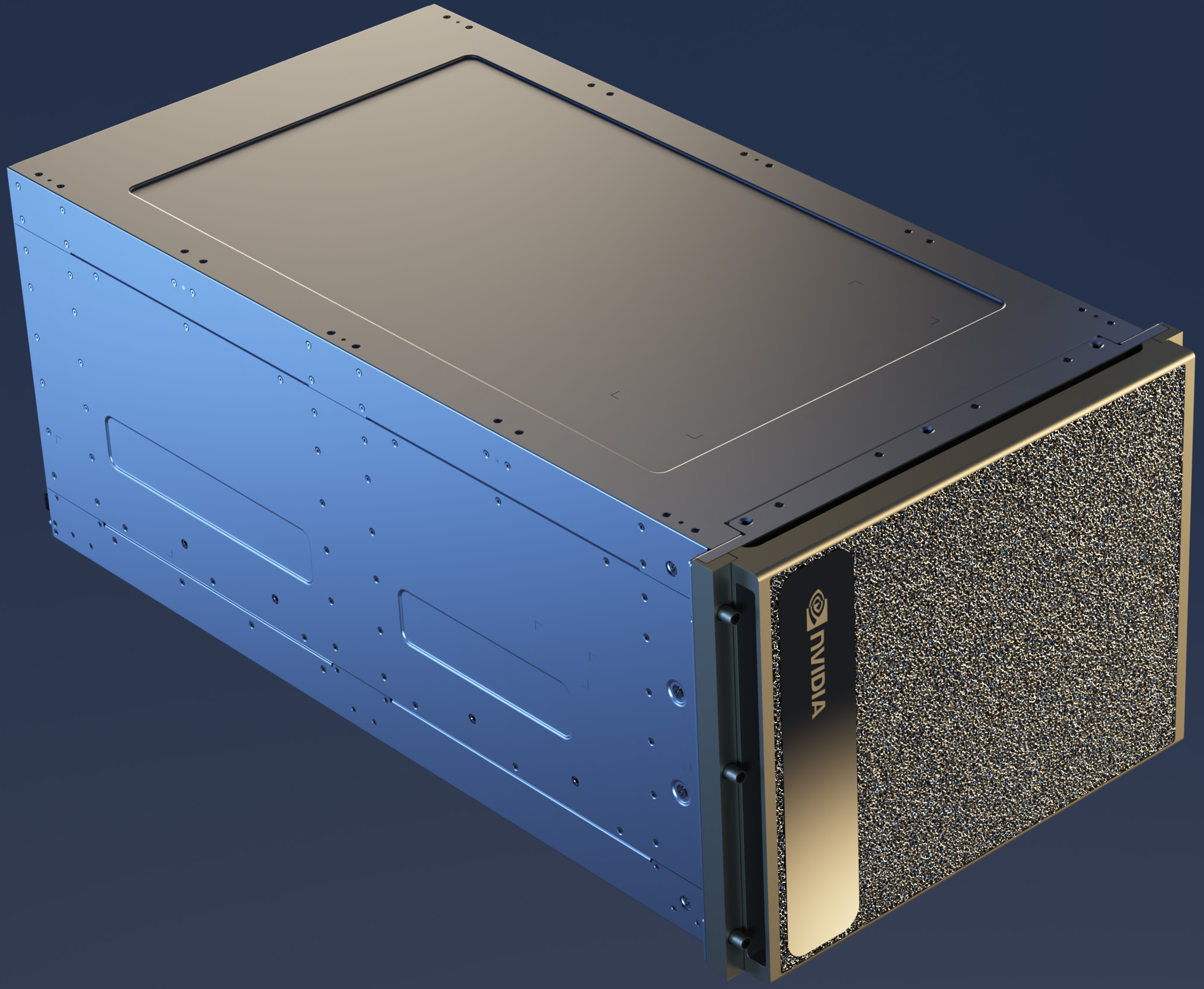
NVIDIA H100 GPU



Fully Connected with
NVLink Switches

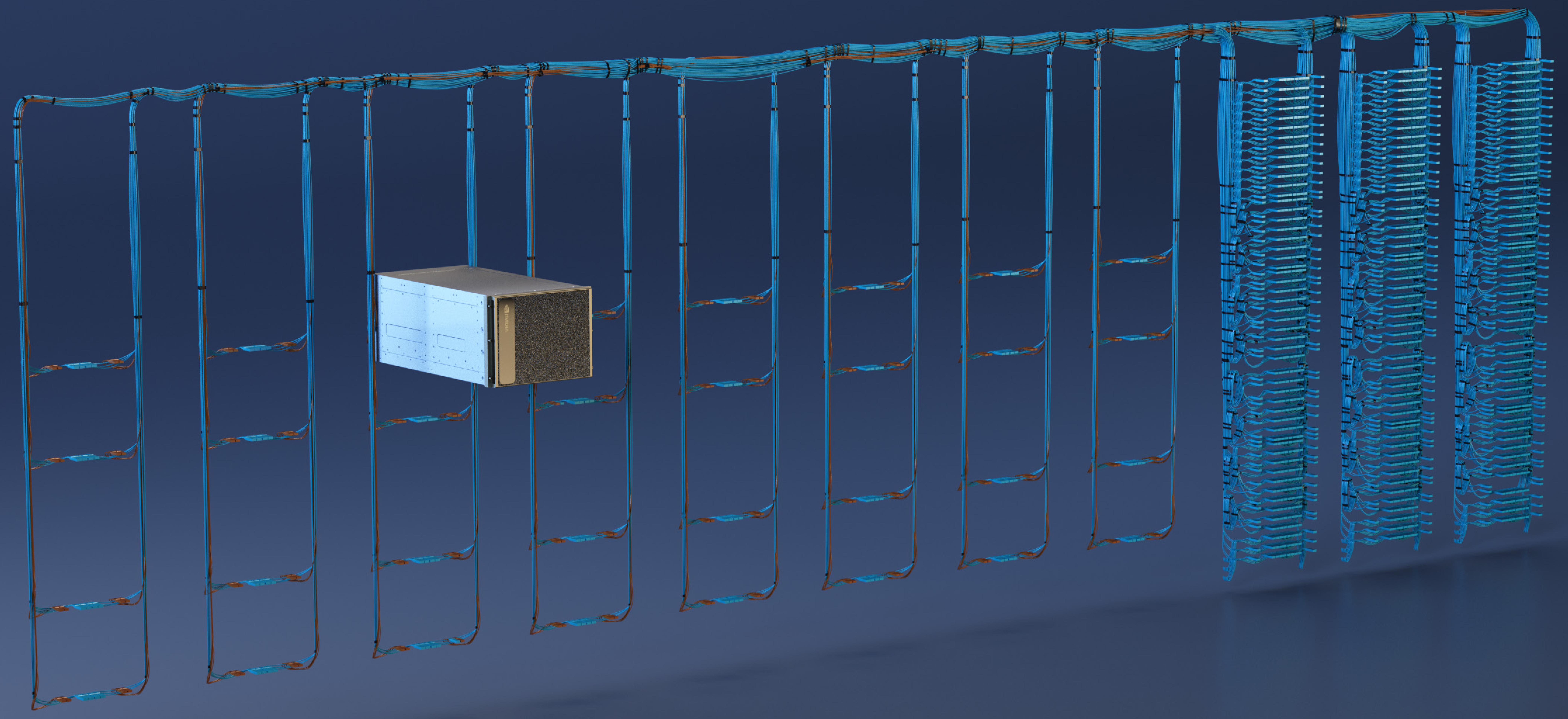


One Giant GPU

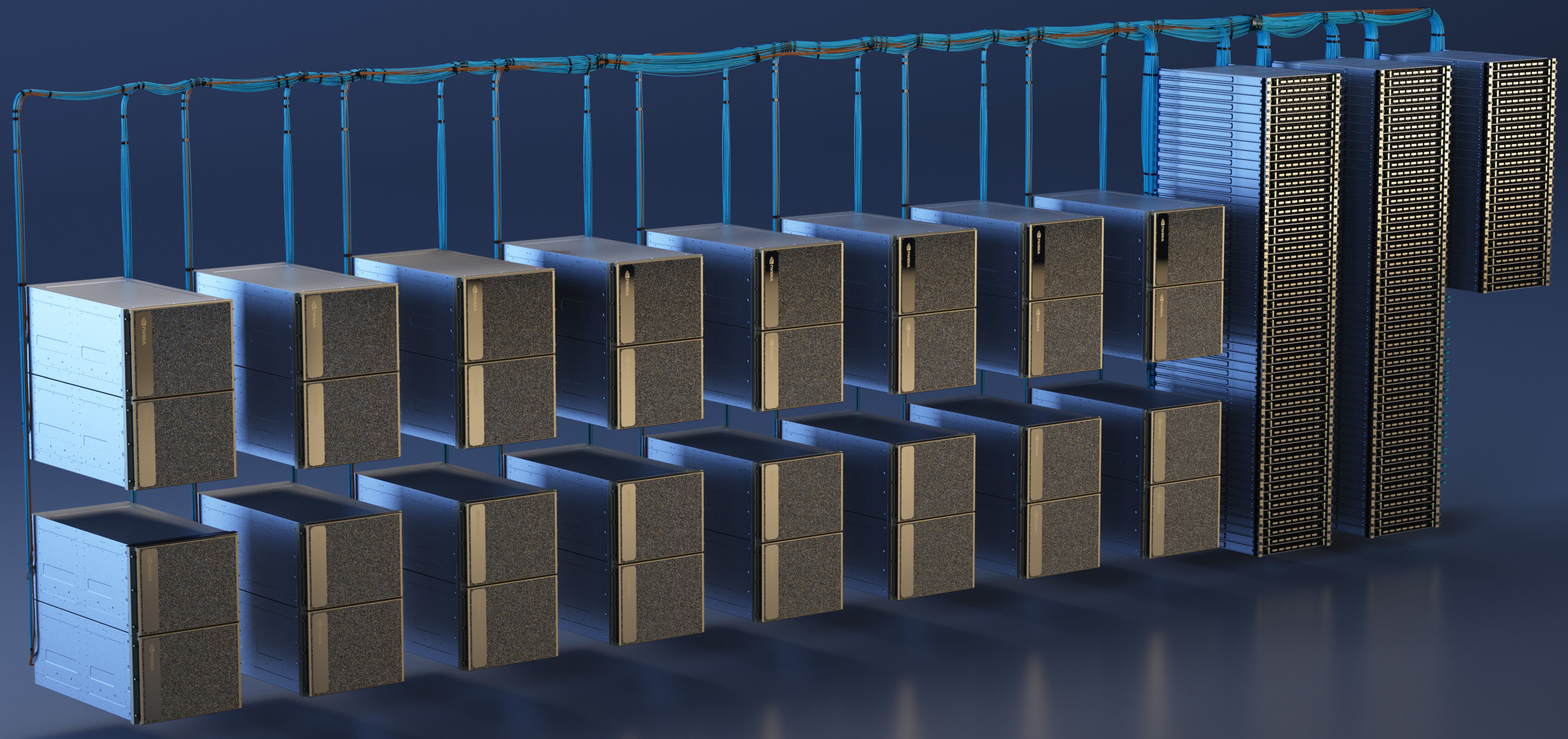


NVIDIA DGX H100


NVIDIA

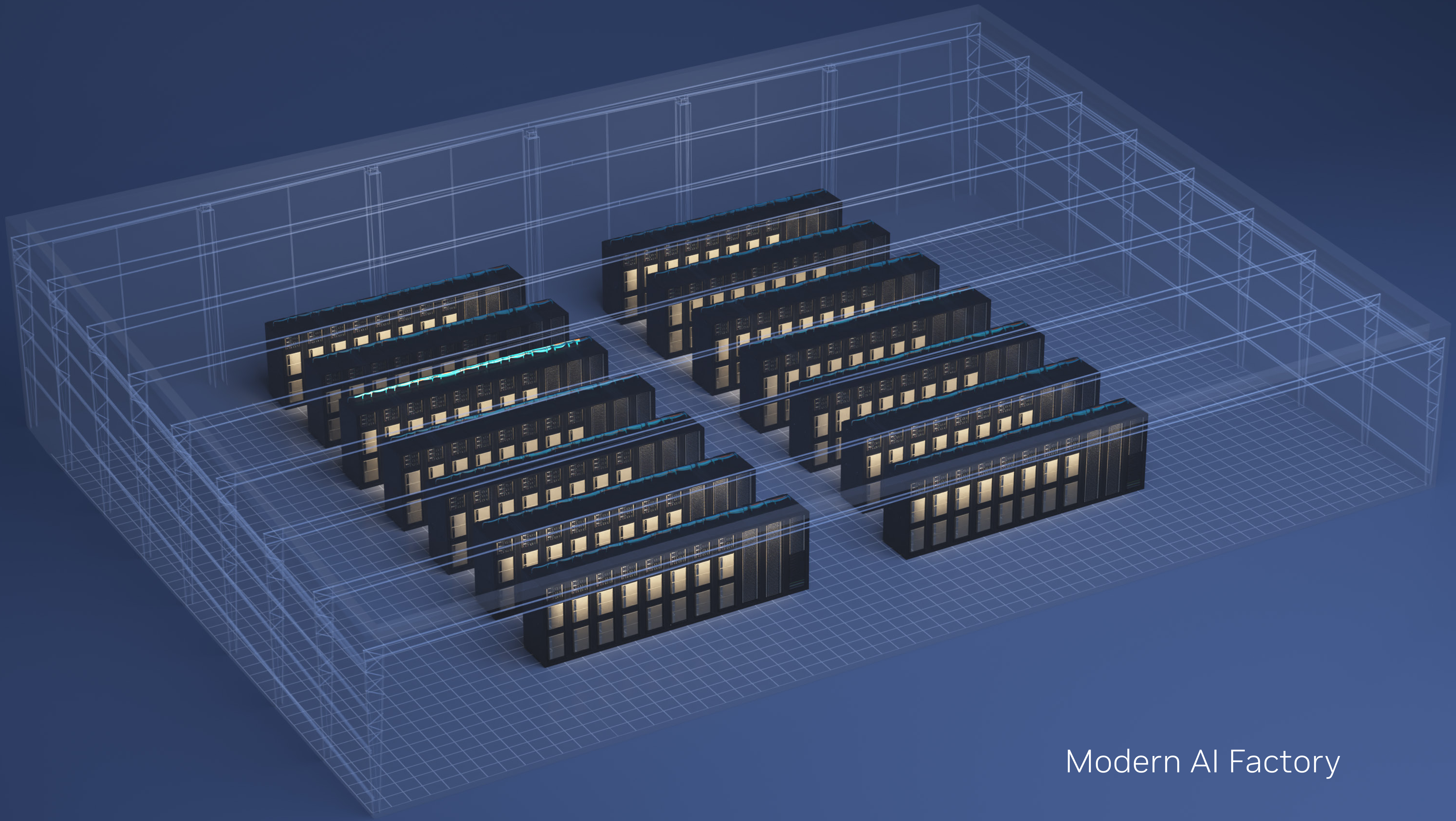


NVIDIA Quantum-2
InfiniBand Networking



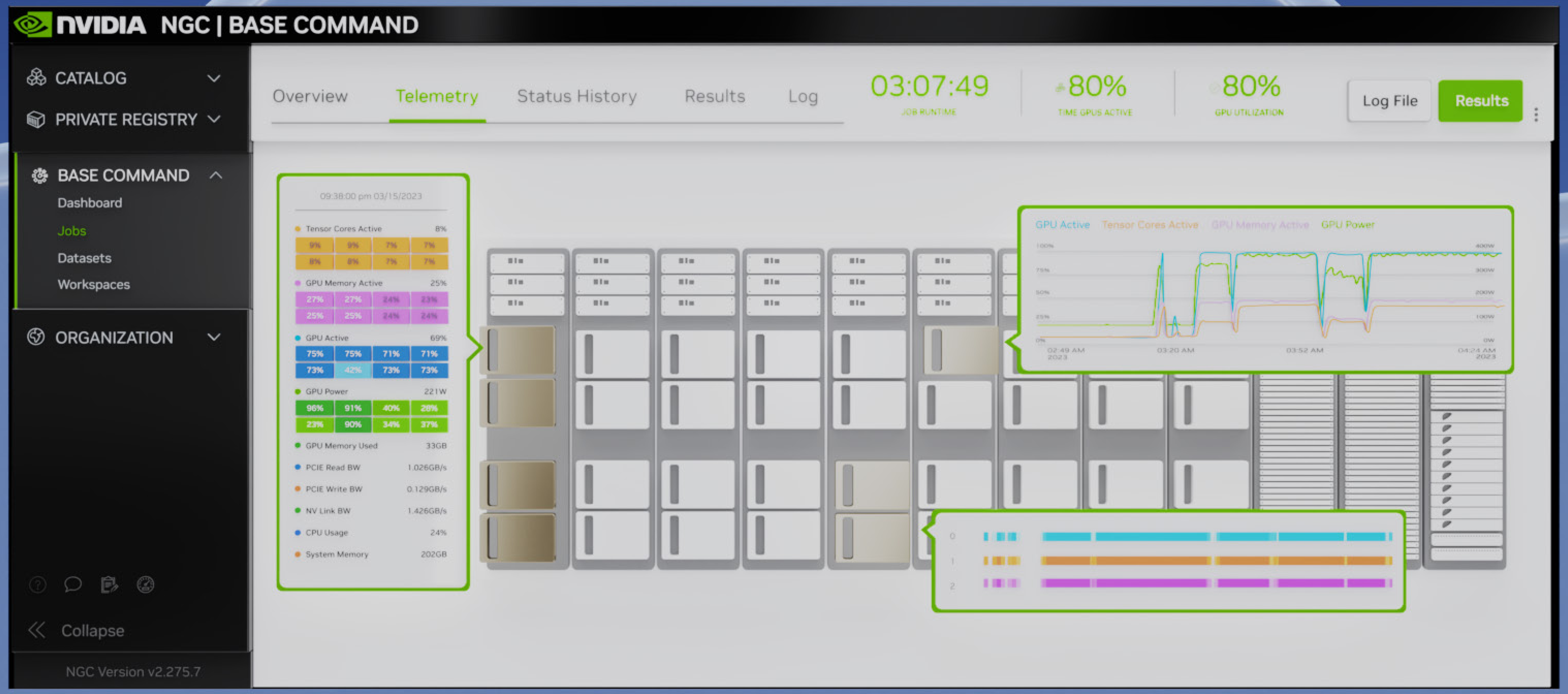
Scaling to Hundreds and Thousands of DGX Nodes


DGX H100
AI Supercomputer



Modern AI Factory

Atos
aws
Cirrascale CLOUD SERVICES
CoreWeave
DELL Technologies
GIGABYTE™
Google Cloud
Hewlett Packard Enterprise
Lambda
Lenovo
ORACLE CLOUD Infrastructure
QCT™
SUPERMICRO®



NVIDIA DGX CLOUD

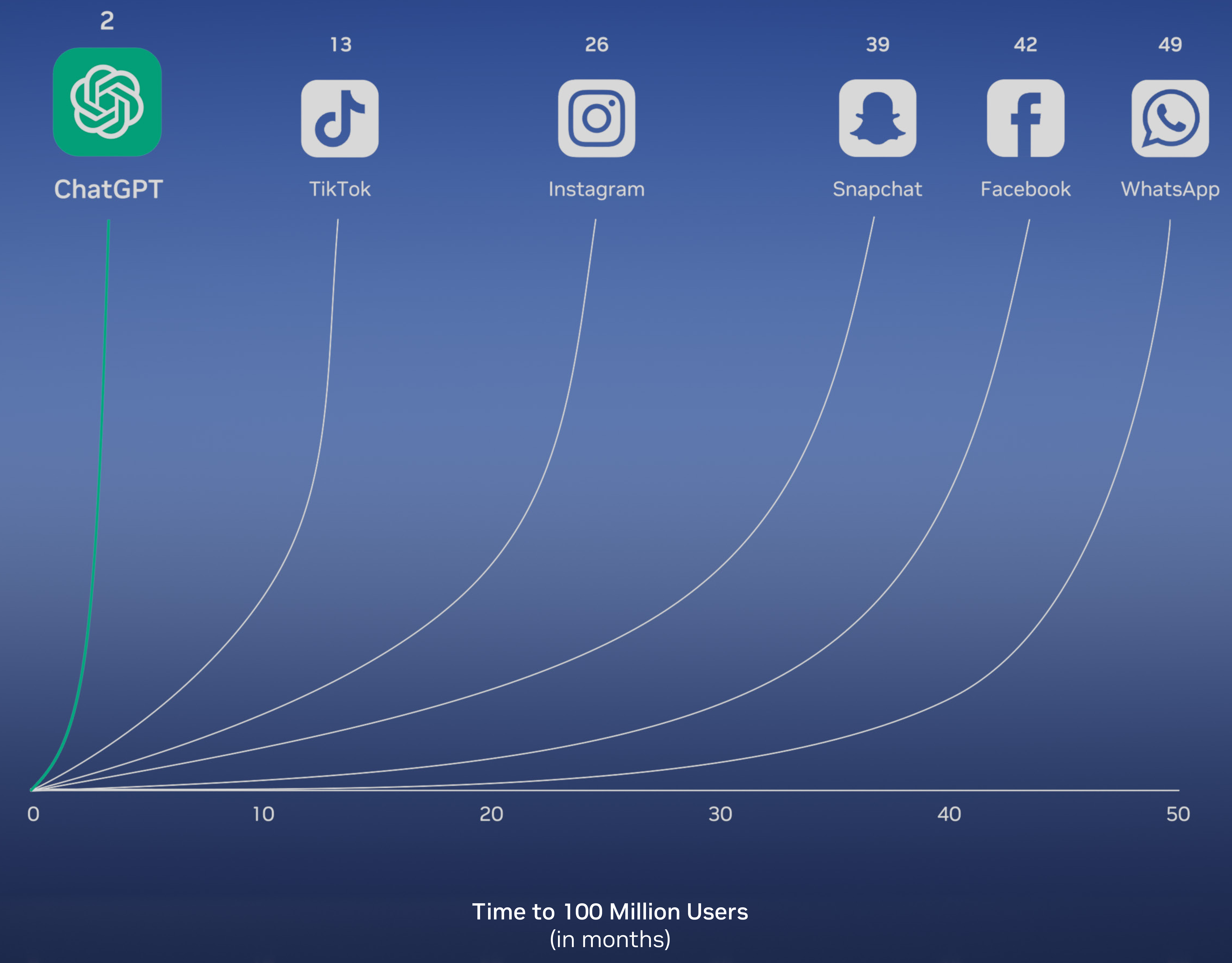
Google Cloud

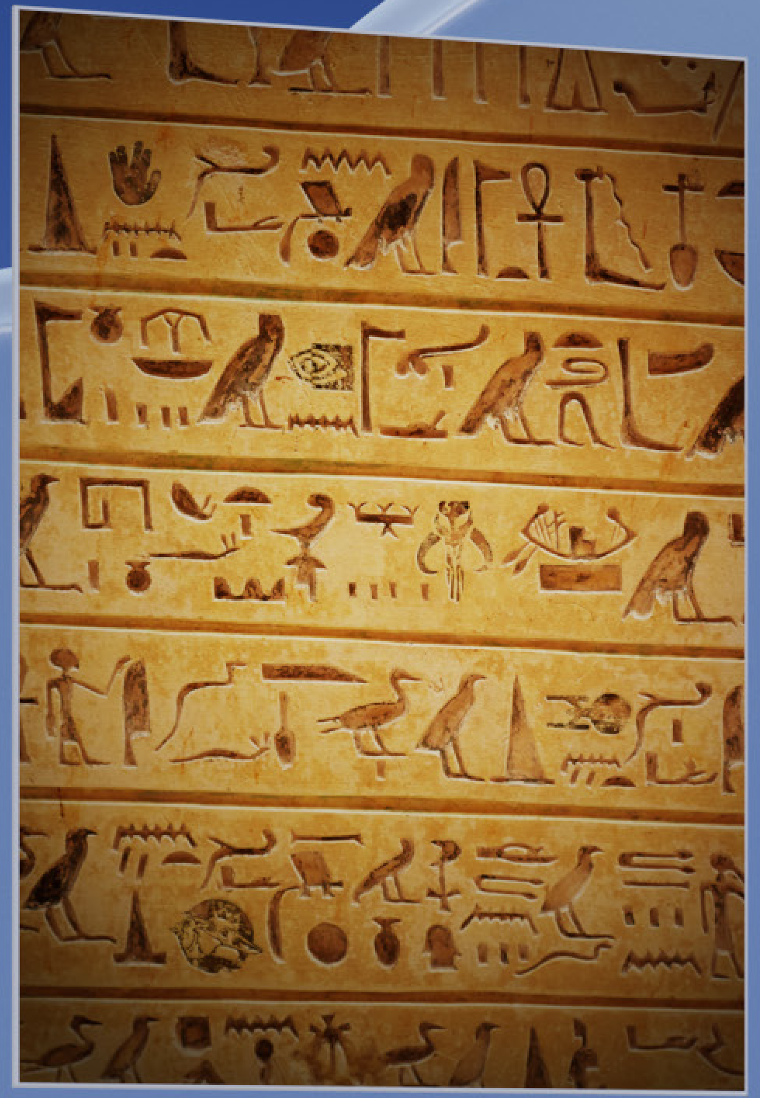
Microsoft Azure

ORACLE CLOUD Infrastructure

ORACLE
CLOUD
Infrastructure

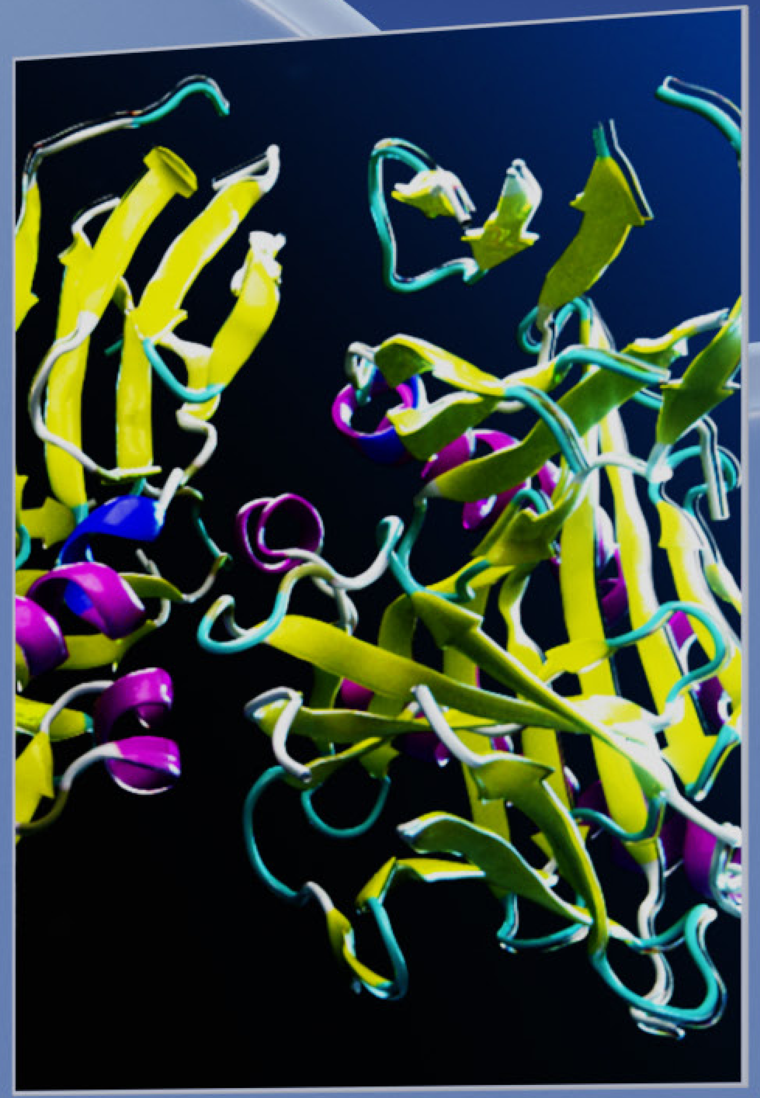
NVIDIA


2
ChatGPT
13
TikTok
26
Instagram
39
Snapchat
42
Facebook
49
WhatsApp
0
10
20
30
40
50
Time to 100 Million Users
(in months)



NEMO

PICASSO



BIONEMO

NVIDIA AI FOUNDATIONS

Google Cloud

Microsoft Azure

ORACLE CLOUD Infrastructure

NEMO

PICASSO

BIONEMO

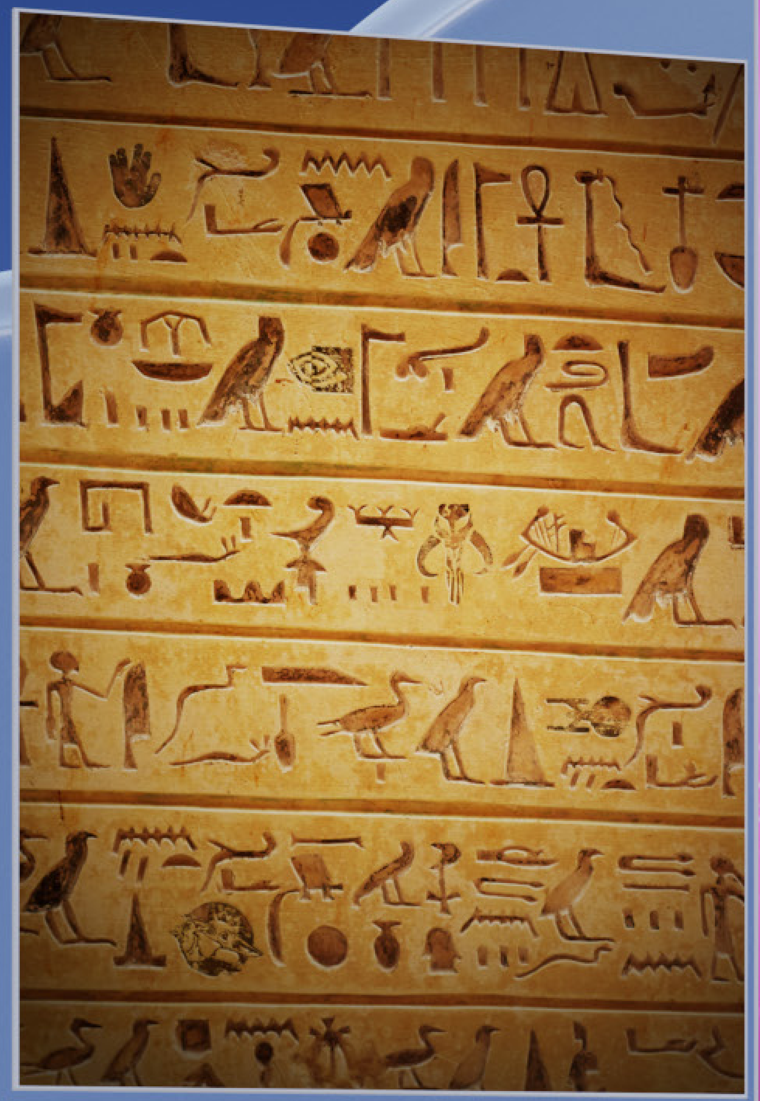
NVIDIA AI FOUNDATIONS

Google Cloud

Microsoft Azure

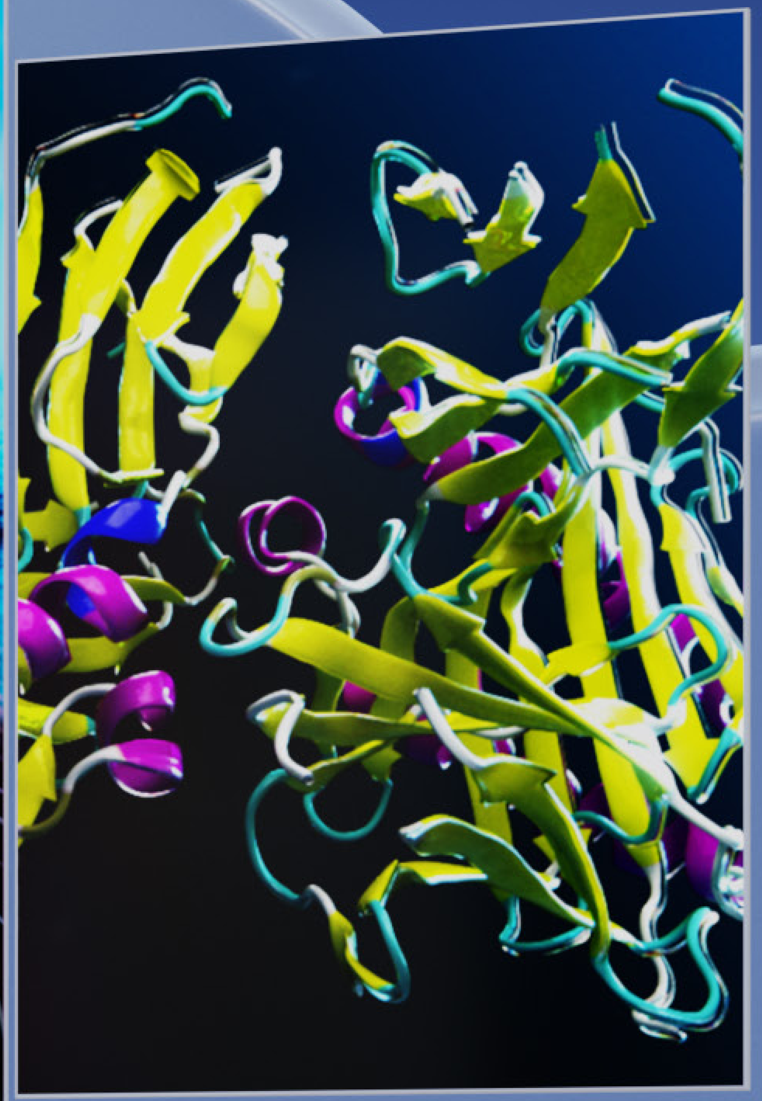
ORACLE CLOUD Infrastructure


NVIDIA NeMo Service
In-Application API
Millions of Queries
In: Query Zero-Shot
or Prompt Examples
Out: Skilled Response
Guardrail Service
NeMo Customization
and Inference Service
Inform
8B
43B
530B
Vector DB
Enterprise Data
Developer API
Customization Data
P-Tuning
Reinforcement Learning
From Human Feedback
NVIDIA DGX Cloud



NEMO

PICASSO



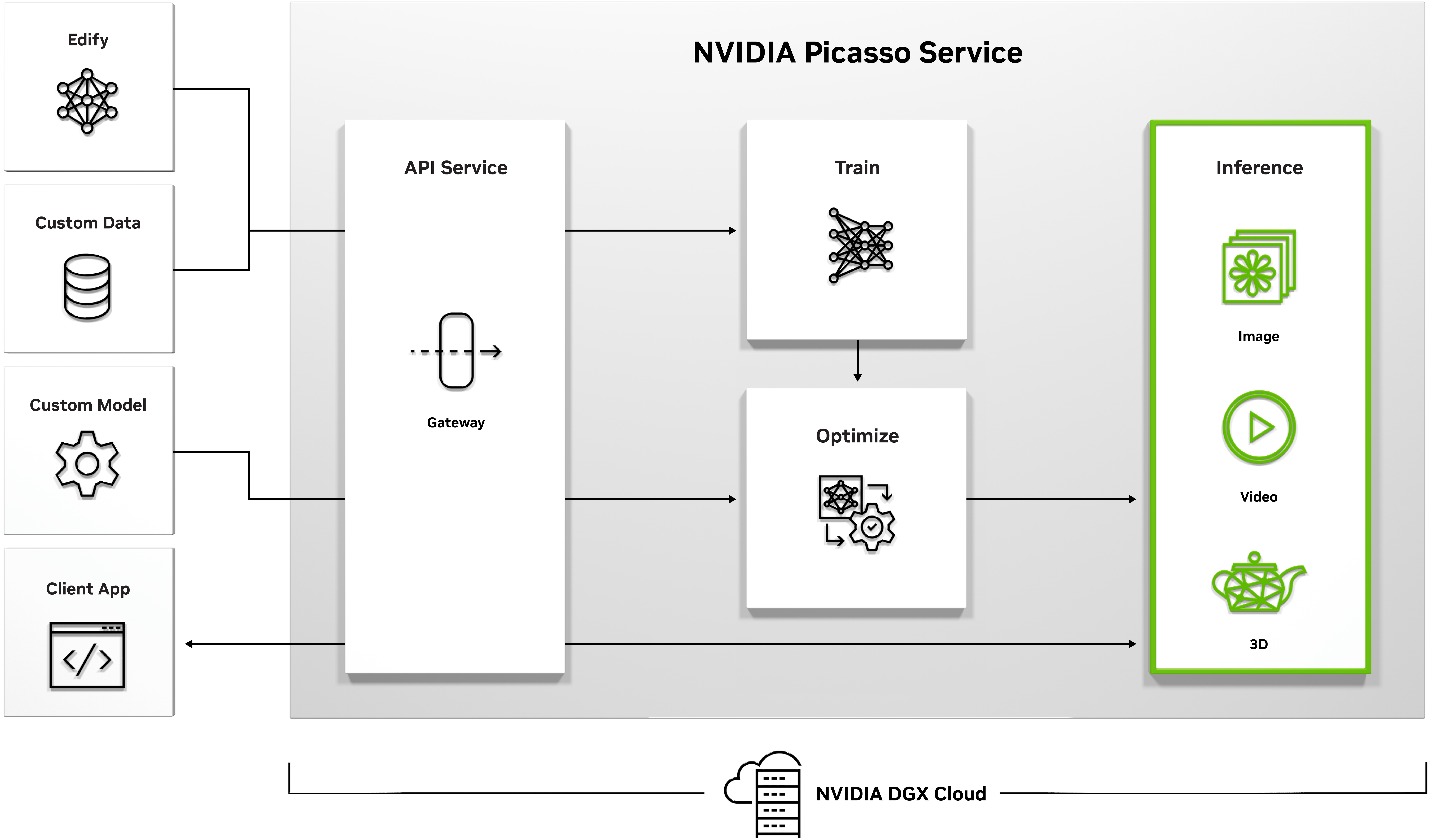
BIONEMO

NVIDIA AI FOUNDATIONS

Google Cloud

Microsoft Azure

ORACLE CLOUD Infrastructure


Edify
Custom Data
Custom Model
Client App
NVIDIA Picasso Service
API Service
Gateway
Train
Optimize
Inference
Image
Video
3D
NVIDIA DGX Cloud

gettyimages

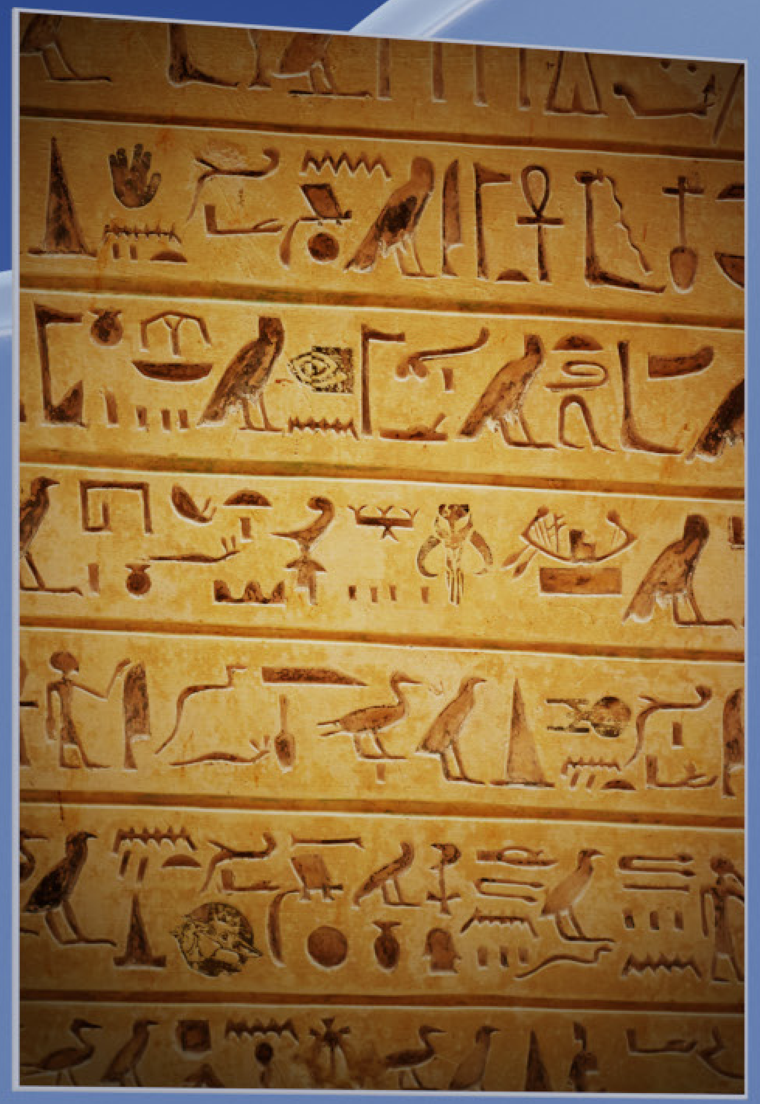
NVIDIA

shutterstock®

NVIDIA

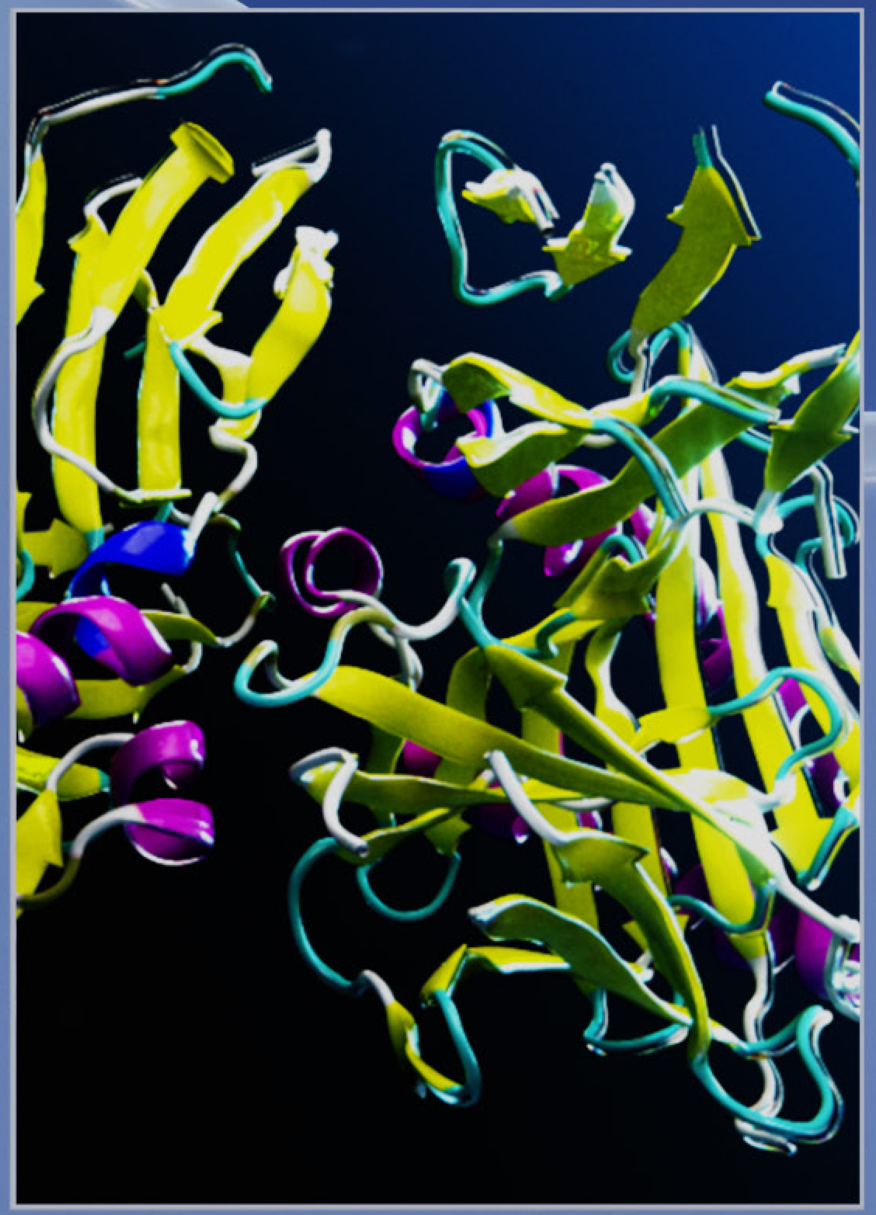
Adobe

NVIDIA



NEMO

PICASSO



BIONEMO

NVIDIA AI FOUNDATIONS

Google Cloud

Microsoft Azure

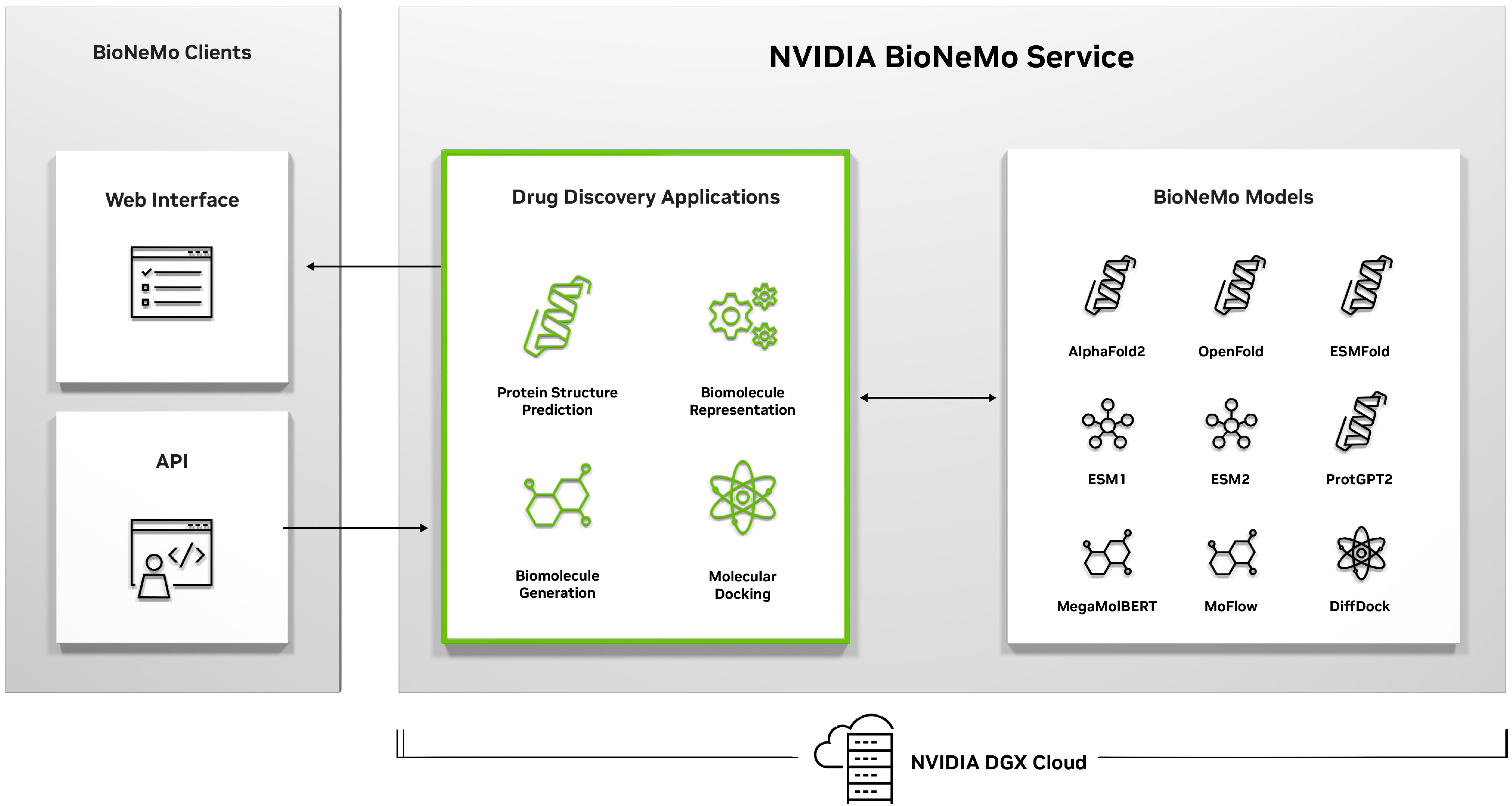
ORACLE CLOUD Infrastructure

absci

EVOZYNE
NATURAL MACHINES

Exscientia

Insilico
Medicine


BioNeMo Clients
Web Interface
API
NVIDIA BioNeMo Service
Drug Discovery Applications
Protein Structure Prediction
Biomolecule Representation
Biomolecule Generation
Molecular Docking
BioNeMo Models
AlphaFold2
OpenFold
ESMFold
ESM1
ESM2
ProtGPT2
MegaMolBERT
MoFlow
DiffDock
NVIDIA DGX Cloud

ALCHEMAB
THERAPEUTICS

AMGEN

AstraZeneca

EVOZYNE
NATURAL MACHINES

Exscientia

Innophore

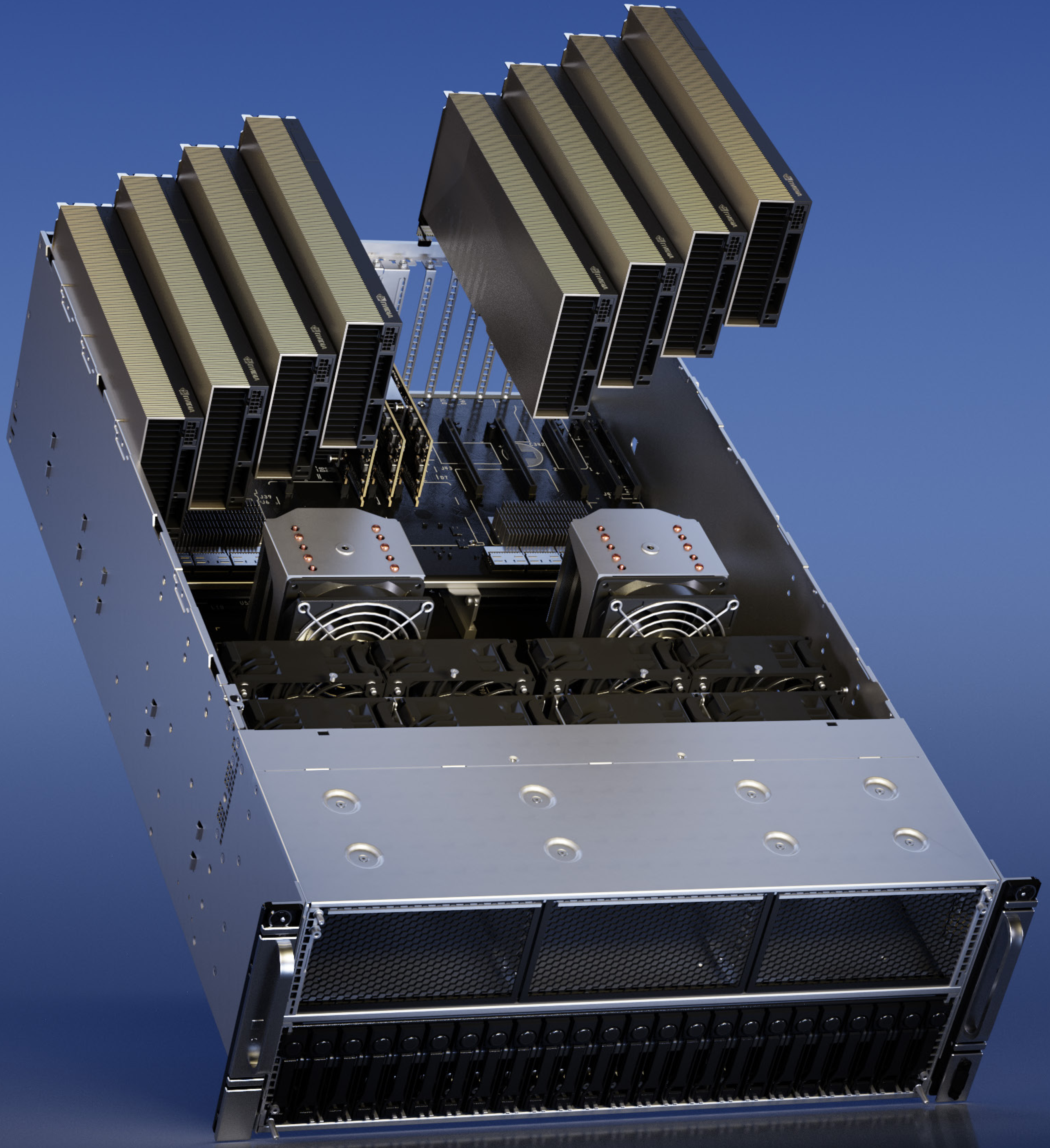
Insilico
Medicine


L4
L40
H100 NVL
GRACE-HOPPER

L4 - 120X
vs. CPU for AI Video

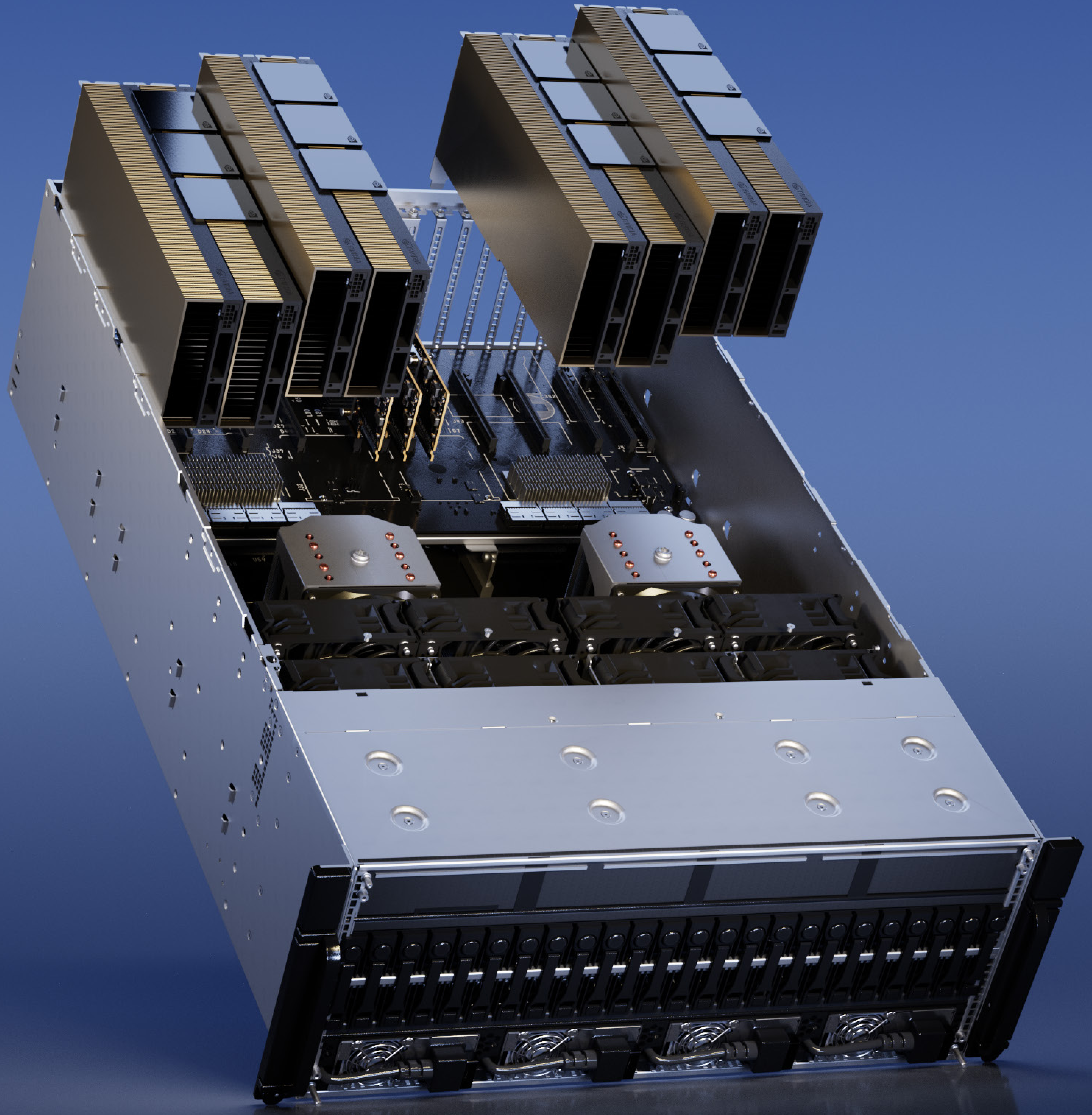
Google Cloud

NVIDIA



L40 - 10X
vs. T4 for GPT-3 20B


DEMO: RUNWAYML – GREEN SCREEN



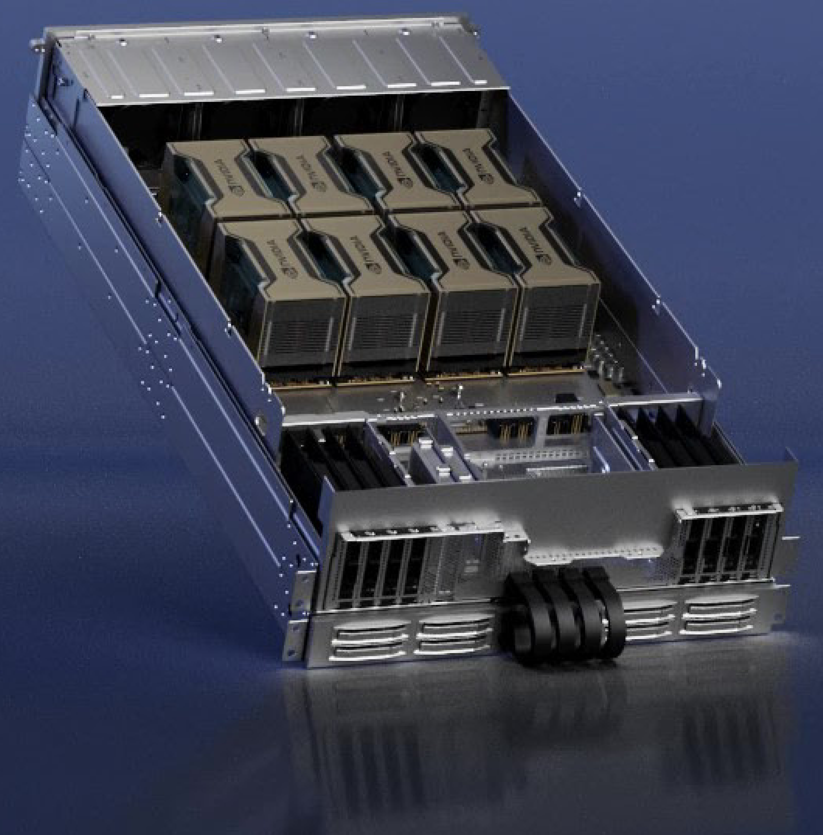
H100 NVL - 10X

vs. HGX A100 for GPT-3 175B

Query : How can my sign printing business help a local restaurant expand their business to more locations?

GPT 3: Your sign printing business can be a valuable partner for a local restaurant looking to expand to more locations. By providing high-quality signage and promotional materials, you can help the restaurant establish a strong brand presence, attract customers, and maintain a consistent image across all locations. Here are some ways your sign printing business can help:

1. Get to know your customer before you begin designing content for them. Review their existing signage and printed materials. Consider asking these questions: What is their current brand identity?



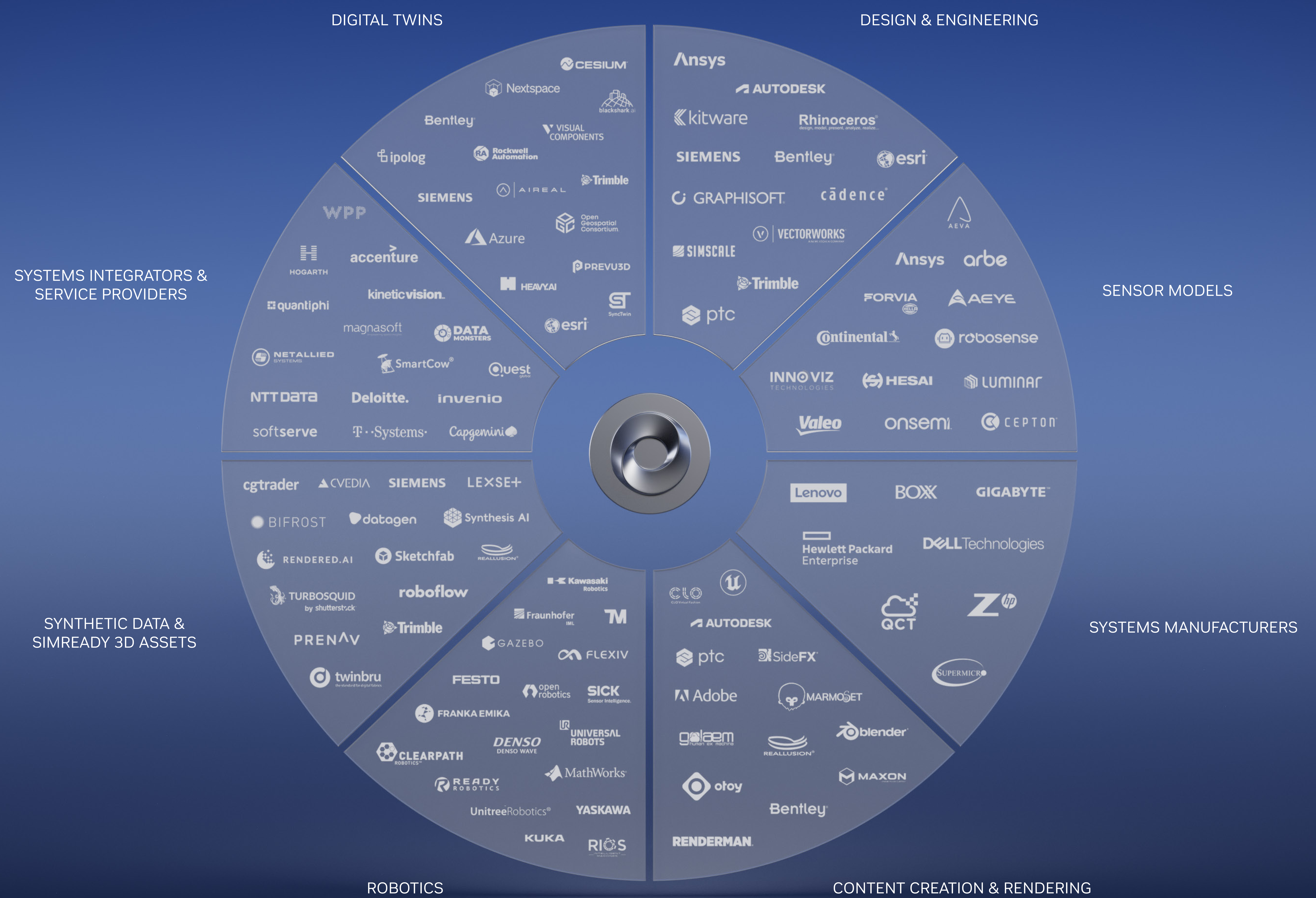
HGX A100

GRACE-HOPPER - 7X

vs. PCIe for CPU to GPU Bandwidth


L4
L40
H100 NVL
GRACE-HOPPER


NVIDIA AI
NVIDIA Omniverse


DIGITAL TWINS
CESIUM
Nextspace
blackshark.ai
Bentley
VISUAL COMPONENTS
ipolog
Rockwell Automation
Trimble
SIEMENS
AIREAL
Open Geospatial Consortium
Azure
PREVU3D
HEAVY.AI
SyncTwin
esri
DESIGN & ENGINEERING
Ansys
AUTODESK
kitware
Rhinoceros
SIEMENS
Bentley
esri
GRAPHISOFT
cadence
VECTORWORKS
SIMSCALE
Trimble
ptc
SENSOR MODELS
AEVA
Ansys
arbe
FORVIA
AEYE
Continental
robosense
INNOVIZ TECHNOLOGIES
HESAI
LUMINAR
Valeo
onsemi
CEPTON
SYSTEMS INTEGRATORS & SERVICE PROVIDERS
WPP
HOGARTH
accenture
kineticvision
quantiphi
magnasoft
DATA MONSTERS
NETALLIED SYSTEMS
SmartCow
Quest
NTT Data
Deloitte.
invenio
softserve
T··Systems·
Capgemini
SYNTHETIC DATA & SIMREADY 3D ASSETS
cgtrader
CVEDIA
SIEMENS
LEXSET
BIFROST
datagen
Synthesis AI
RENDERED.AI
Sketchfab
REALLUSION
TURBOSQUID by shutterstock
roboflow
PRENAV
Trimble
twinbru
ROBOTICS
Kawasaki Robotics
Fraunhofer IML
GAZEBO
FLEXIV
FESTO
open robotics
SICK Sensor Intelligence.
FRANKA EMIKA
UNIVERSAL ROBOTS
DENSO DENSO WAVE
CLEARPATH ROBOTICS
MathWorks
READY ROBOTICS
UnitreeRobotics
YASKAWA
KUKA
RIOS
SYSTEMS MANUFACTURERS
Lenovo
BOXX
GIGABYTE
Hewlett Packard Enterprise
DELL Technologies
QCT
Z hp
SUPERMICRO
CONTENT CREATION & RENDERING
CLO
AUTODESK
ptc
SideFX
Adobe
MARMOSET
golaem
REALLUSION
blender
otoy
MAXON
Bentley
RENDERMAN


NVIDIA AI
NVIDIA Omniverse
CONCEPT & STYLING
DESIGN & ENGINEERING
SOFTWARE & ELECTRONICS
AUTONOMOUS DRIVING
RETAIL EXPERIENCE
SMART FACTORY


ThinkStation
NVIDIA OMNIVERSE
ThinkStation
BOXX
DELL Technologies
Lenovo


DELL Technologies
GIGABYTE
Hewlett Packard Enterprise
Lenovo
QCT
SUPERMICRO


NVIDIA Omniverse Cloud
Omniverse Launcher
Custom Launcher
3rd Party Tools
Enterprise Nucleus
Service API
API
Resource Management
Service Catalog
USD Composer
Replicator
DRIVE Sim
Isaac Sim
Custom
NVIDIA OVX


USD COMPOSER
USD GDN PUBLISHER
DRIVE SIM
ISAAC SIM
REPLICATOR
KIT
NVIDIA OMNIVERSE CLOUD
Microsoft
Azure


Spark
RAFT
cuOpt
Holoscan
cuLitho
Triton
TMS
NVIDIA AI
NVIDIA Omniverse
Google Cloud
Microsoft Azure
ORACLE CLOUD Infrastructure

NVIDIA