
NVIDIA

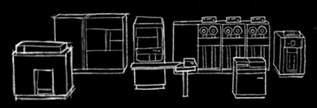
Except for the historical information contained herein, certain matters in this presentation are forward-looking statements. These forward-looking statements and any other forward-looking statements that go beyond historical facts that are made in this presentation are subject to risks and uncertainties that may cause actual results to differ materially. Important factors that could cause actual results to differ materially include: global economic conditions; NVIDIA's reliance on third parties to manufacture, assemble, package and test NVIDIA's products; the impact of technological development and competition; development of new products and technologies or enhancements to NVIDIA's existing product and technologies; market acceptance of NVIDIA's products or NVIDIA's partners' products; design, manufacturing or software defects; changes in consumer preferences and demands; changes in industry standards and interfaces; unexpected loss of performance of NVIDIA's products or technologies when integrated into systems and other factors.

NVIDIA has based these forward-looking statements largely on its current expectations and projections about future events and trends that it believes may affect its financial condition, results of operations, business strategy, short-term and long-term business operations and objectives, and financial needs. These forward-looking statements are subject to a number of risks and uncertainties, and you should not rely upon the forward-looking statements as predictions of future events. The future events and trends discussed in this presentation may not occur and actual results could differ materially and adversely from those anticipated or implied in the forward-looking statements. Although NVIDIA believes that the expectations reflected in the forward-looking statements are reasonable, the company cannot guarantee that future results, levels of activity, performance, achievements or events and circumstances reflected in the forward-looking statements will occur. Except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances. For a complete discussion of factors that could materially affect NVIDIA's financial results and operations, please refer to the reports NVIDIA files from time to time with the SEC, including NVIDIA's most recent Annual Report on Form 10-K, Quarterly Reports on Form 10-Q, and Current Reports on Form 8-K. Copies of reports NVIDIA files with the SEC are posted on NVIDIA's website and are available from NVIDIA without charge.

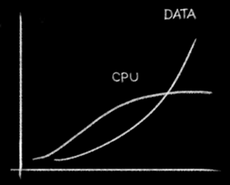

accenture
ADEPT
Adobe
Alibaba Cloud
amazon
Argonne NATIONAL LABORATORY
Ansys
ANTHROPIC
anyscale
AUTODESK
Baidu 百度
BlackRock
BMW GROUP
BNP PARIBAS
Boston Dynamics
Bristol Myers Squibb
c.ai
cadence
Canonical
Canva
CapitalOne
CERN
Chan Zuckerberg Initiative
CLEARML
cohere
CoreWeave
Continental
databricks
DATADOG
data iku
dataloop
DataRobot
ddn
Deepgram
DELL
Deloitte.
Disney Research
DOMINO
ebay
IIElevenLabs
EQUINIX
ExxonMobil
Ford
Foster + Partners
FOXCONN
GE Healthcare
gettyimages
Genentech
GIGABYTE
glean
Google
Goldman Sachs
GSK
H2O.ai
HEXAGON
Honeywell
hp
Hewlett Packard Enterprise
Hugging Face
IBM
imbue
JLR
JOHN DEERE
Johnson&Johnson Innovative Medicine
JPMORGAN CHASE & CO.
Kawasaki
KPF
kt
Lambda
LangChain
Lenovo
Lightning AI
LinkedIn
LOCKHEED MARTIN
LlamaIndex
L'ORÉAL
Los Alamos NATIONAL LABORATORY
LOWE'S
mastercard
MathWorks
MERCK
Meta AI
Medtronic
Microsoft
MISTRAL AI_
MITRE
Modular
moz://a
NASA Langley Research Center
NetApp
NERSC
NETL NATIONAL ENERGY TECHNOLOGY LABORATORY
NETFLIX
NIST National Institute of Standards and Technology
NORTHROP GRUMMAN
NTT Group
OpenAI
ORACLE
orange
PayPal
PEGATRON
perplexity
PETROBRAS
PhotoRoom
PINGAN
Pinterest
PIXAR
QCT
Raytheon Technologies
Recursion
Red Hat
rescale
Rockwell Automation
RSS-Hydro
run:ai
runway
SAP
SAMSUNG ELECTRONICS
aramco
scale
slb
servicenow
Shell
shutterstock
SIEMENS
Snapchat
snowflake
SoftBank
STELLANTIS
SUPERMICRO
synthesia
tabnine
together.ai
Tencent 腾讯
Unilever
Uber
ups
U.S. ARMY
NAVY FORGED BY THE SEA
USPTO
verizon
vmware
Weights & Biases
WEKA
WELLS FARGO
wistron
WPP
XL

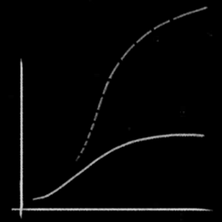

amazon
Argonne NATIONAL LABORATORY
BlackRock
BMW GROUP
BNP PARIBAS
Bristol Myers Squibb
Capital One
CERN
Chan Zuckerberg Initiative
Continental
Disney Research
ebay
ExxonMobil
Ford
Foster + Partners
GE Healthcare
gettyimages
Genentech
Goldman Sachs
GSK
Honeywell
JLR
JOHN DEERE
Johnson&Johnson Innovative Medicine
JPMORGAN CHASE & CO.
Kawasaki
KPF
kt
LinkedIn
LOCKHEED MARTIN
L'ORÉAL
Los Alamos NATIONAL LABORATORY
LOWE'S
mastercard
MERCK
Medtronic
MITRE
NASA Langley Research Center
NERSC
NETL NATIONAL ENERGY TECHNOLOGY LABORATORY
NETFLIX
NIST National Institute of Standards and Technology
NORTHROP GRUMMAN
NTT Group
orange
PayPal
BR PETROBRAS
PING AN
Pinterest
PIXAR
Raytheon Technologies
Recursion
RA Rockwell Automation
RSS-Hydro
SAMSUNG ELECTRONICS
aramco
Shell
shutterstock
SIEMENS
Snapchat
SoftBank
STELLANTIS
Unilever
Uber
ups
U.S. ARMY
AMERICA'S NAVY FORGED BY THE SEA
UNITED STATES PATENT AND TRADEMARK OFFICE uspto
verizon
WELLS FARGO
WPP

1964
IBM S/360



1995
WINDOWS 95
PENTIUM

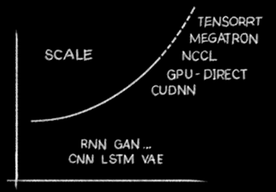




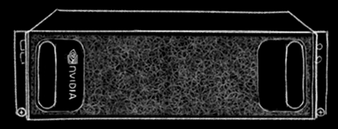
ACCELERATED COMPUTING
1993 UDA → 2003 CG → 2006 CUDA

2012
ALEXNET
"FIRST CONTACT"

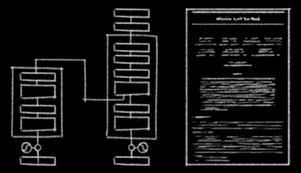
DOUBLING
EVERY 6 MONTHS



2016
DGX-1



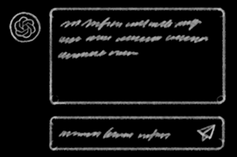
2017
TRANSFORMER



LEARN EVERYTHING

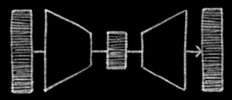
PROTEIN
LANGUAGE
SOUND
PHYSICS
VIDEO
3D
MANIPULATION
IMAGES
GESTURE

2022
OPEN AI CHATGPT

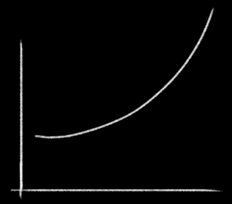


FINE TUNING
GUARDRAILING
ALIGNMENT
PROMPT ENGINEERING
RAG
VECTOR DB
MULTI-MODAL
COT & TOT
AGENTS

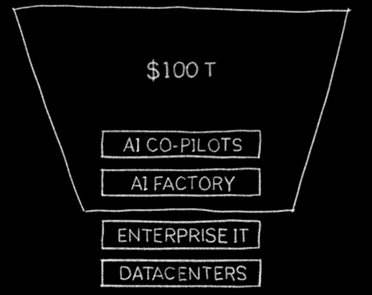
GENERATIVE AI



STARTUPS



A NEW
INDUSTRIAL
REVOLUTION



Ansys, Rescale

Ansys NVIDIA

Building a New Era of CAE with Accelerated Computing & Generative AI

ANSYS SOLVERS OPTIMIZED FOR NVIDIA GPUS

NVIDIA AI ACCELERATED ANSYS SIMULATION

PHYSICS-BASED DIGITAL TWINS

NVIDIA POWERED ANSYS LLMS

TRANSFORM 6G RESEARCH WITH ANSYS PERCEIVE EM INTEGRATION IN OMNIVERSE

ACCELERATED NVIDIA DEVELOPMENT WITH ANSYS

Ansys / Powering Innovation That Drives Human Advancement

SYNOPSYS®

NVIDIA

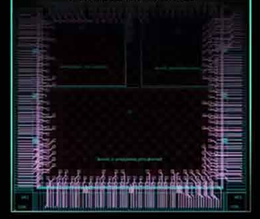
SYNOPSYS: MISSION CRITICAL FOR NVIDIA SILICON SUCCESS

DECADES OF COLLABORATION ACROSS FULL EDA SUITE POWERS ACCELERATED COMPUTING

13X
Verification
Functional Verification

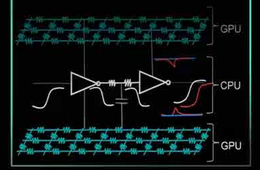
- Synopsys VCS
- NVIDIA L40
- NVIDIA Grace Hopper

10X
Design
Place and Route



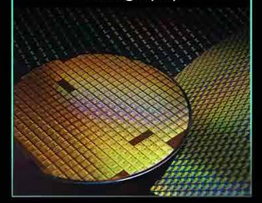
- Synopsys Fusion Compiler
- NVIDIA Grace Hopper

15X
Simulation
SPICE Simulation



- Synopsys PrimeSim
- NVIDIA Hopper
- NVIDIA Grace Hopper

15X
Manufacturing
Computational Lithography



- Synopsys Proteus
- NVIDIA cuLitho
- NVIDIA Grace Hopper

Generative AI
Industry's 1st LLM-Based GenAI EDA Solution

- Synopsys.ai
- NVIDIA NeMo & NIM
- NVIDIA DGX

Systems Software
Testing & Validation of Automotive Software

- Synopsys Electronics Digital Twin, vECU, TPT
- NVIDIA Omniverse

*Performance Speed-Up Based on Projected Results

cadence® + NVIDIA

Deep collaboration on EDA, SDA, Digital Biology and AI

cadence.ai

NVIDIA NeMo

NVIDIA Modulus

NVIDIA BioNeMo

Cerebrus | Virtuoso Studio | Verisium | Optimality | Allegro X AI | Chip NeMo

AI
GenAI Co-Pilots

EDA

Digital | Custom | Verification

SDA

3D-IC | PCB | Multiphysics

OMNIVERSE

Environmental & Data Center Digital Twin

BIO

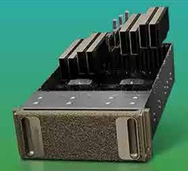
Molecular Sciences

Principled
Simulation +
Optimization

Cadence Palladium & Protium

NVIDIA HGX H100

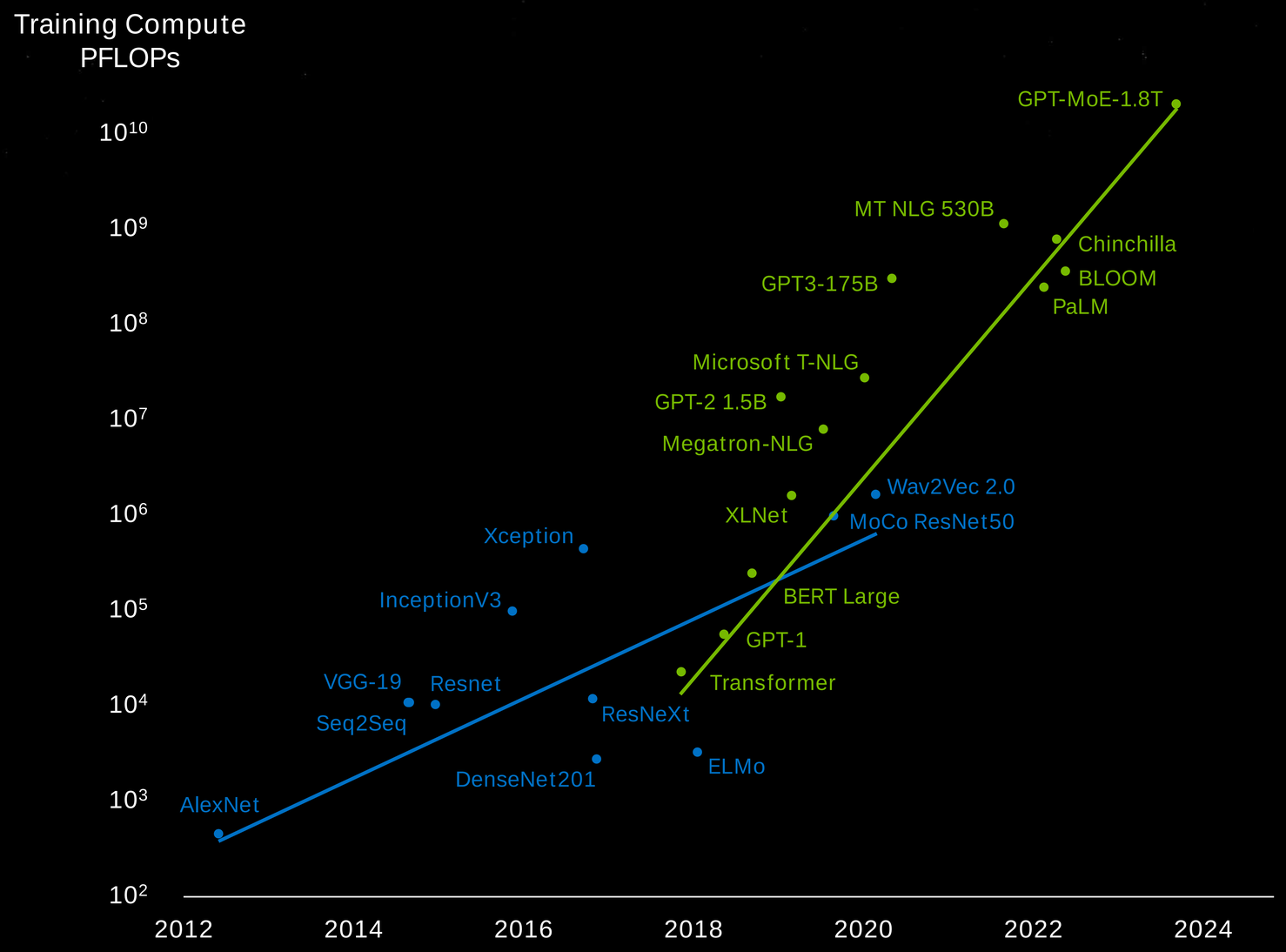
Cadence Millennium

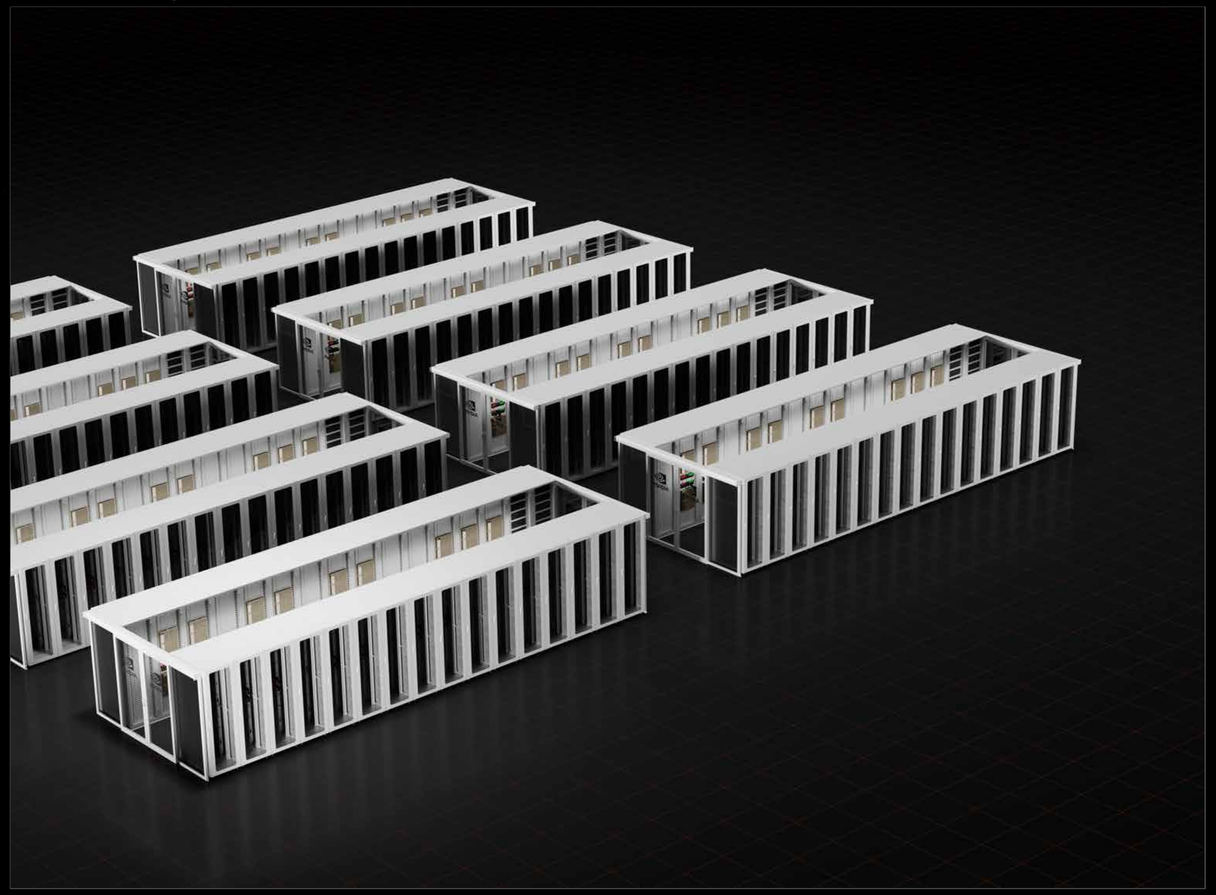


NVIDIA OVX L40

NVIDIA Grace Hopper & Grace Blackwell

Accelerated
Compute


Training Compute
PFLOPs
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
2012
2014
2016
2018
2020
2022
2024
GPT-MoE-1.8T
MT NLG 530B
Chinchilla
BLOOM
PaLM
GPT3-175B
Microsoft T-NLG
GPT-2 1.5B
Megatron-NLG
Wav2Vec 2.0
XLNet
MoCo ResNet50
Xception
BERT Large
InceptionV3
GPT-1
Transformer
VGG-19
Resnet
Seq2Seq
ResNeXt
ELMo
DenseNet201
AlexNet



SELENE
2021

4,480 A100 GPUs

3 EF AI Compute

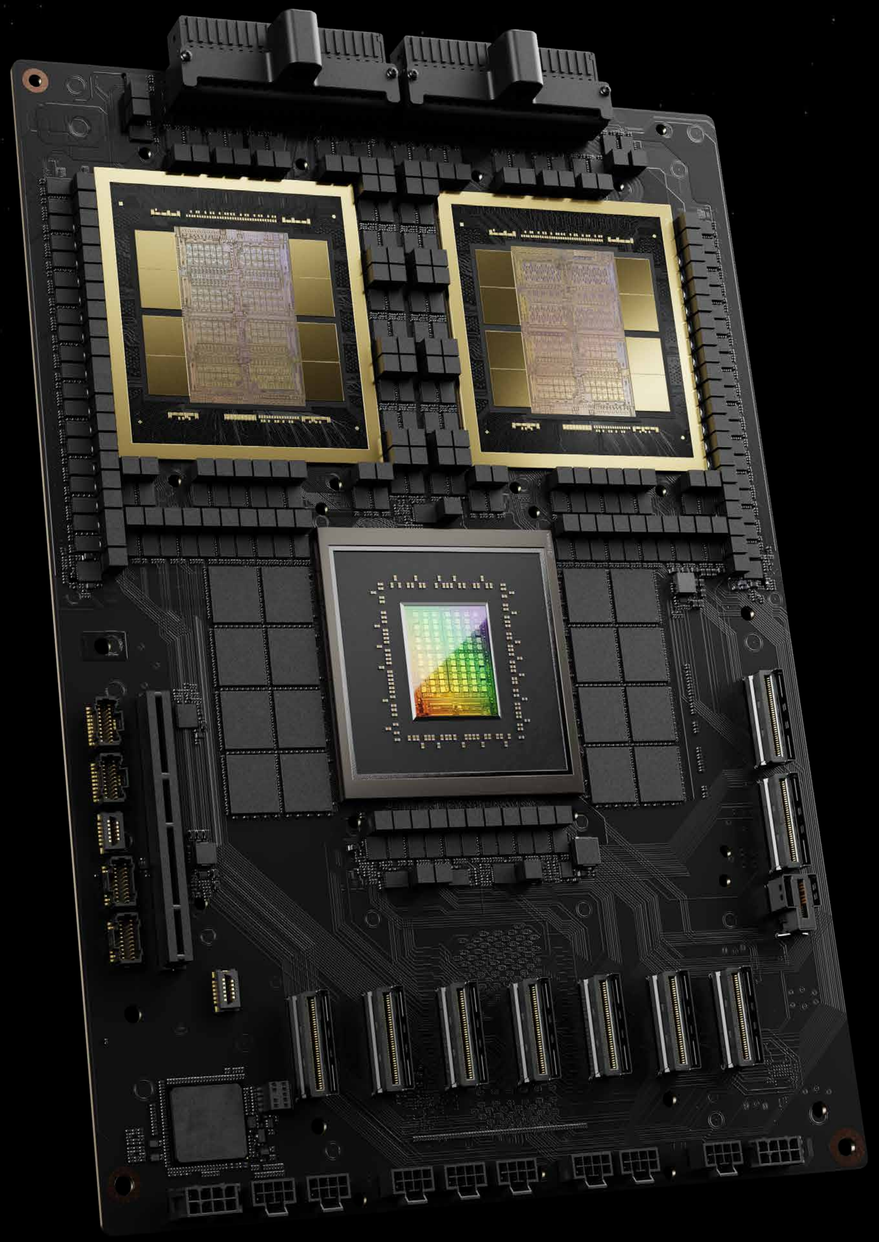
112 TB/s Interconnect BW

EOS
2023

10,752 H100 GPUs

43 EF AI Compute

1,100 TB/s Interconnect BW



ANNOUNCING NVIDIA BLACKWELL PLATFORM FOR TRILLION-PARAMETER SCALE GENERATIVE AI

AI SUPERCHIP
208B Transistors

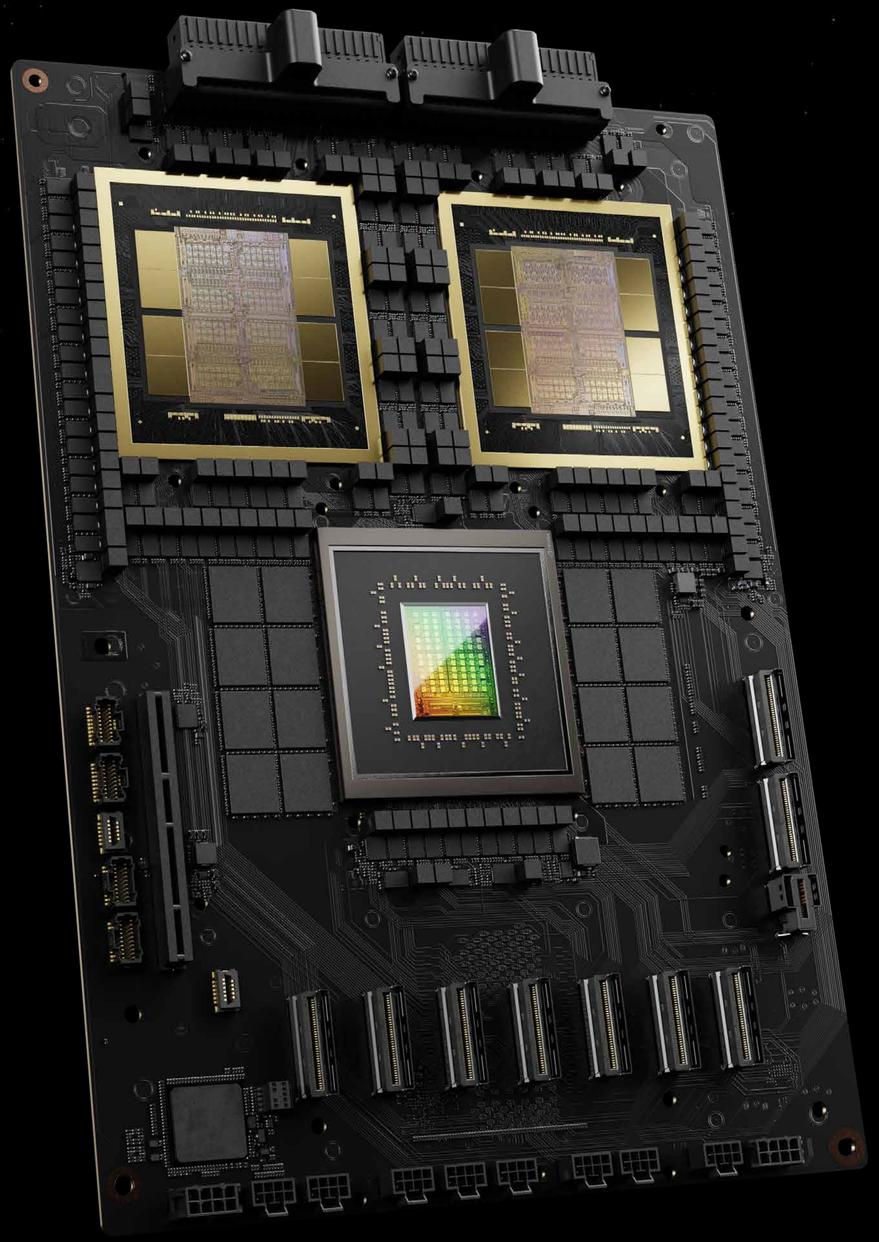
2nd GEN TRANSFORMER ENGINE
FP4/FP6 Tensor Core

5th GENERATION NVLINK
Scales to 576 GPUs

RAS ENGINE
100% In-System Self-Test

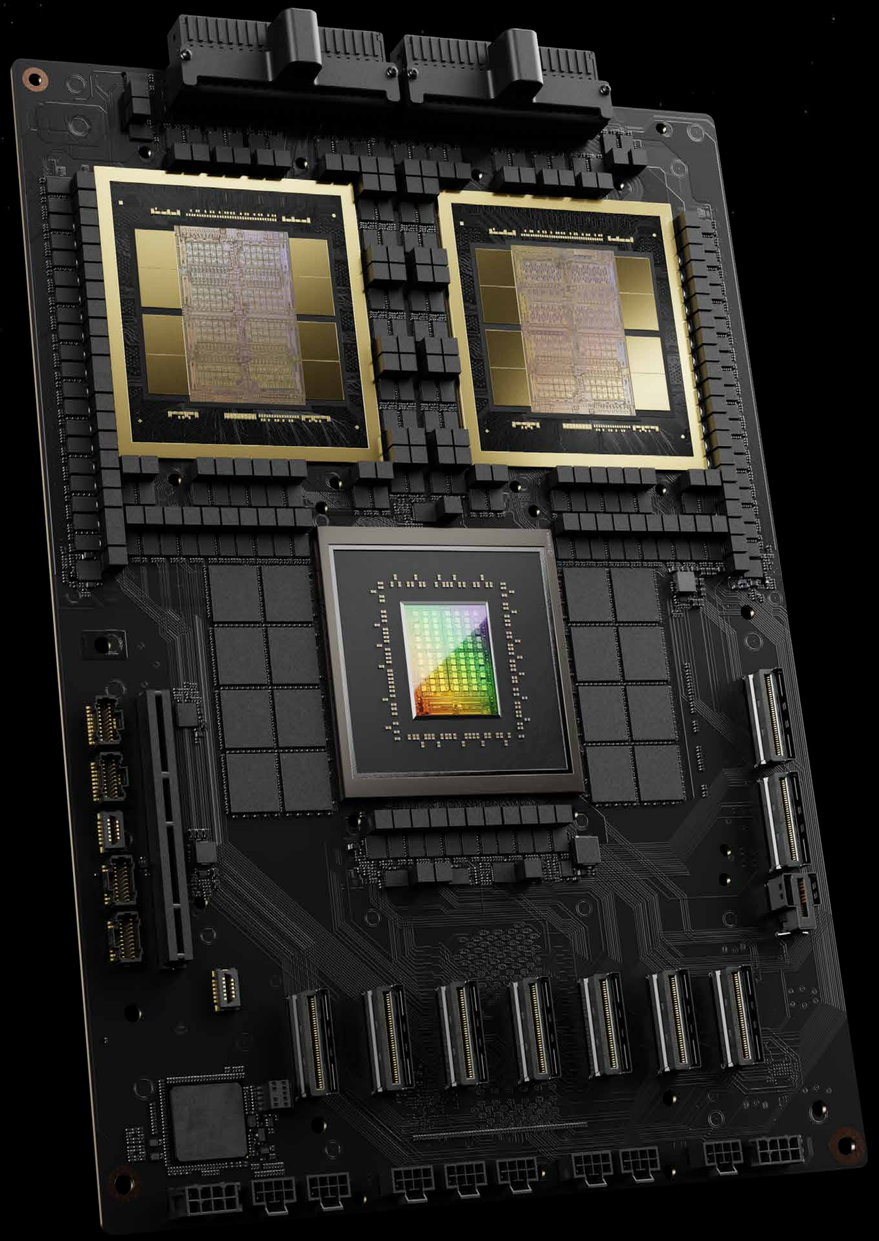
SECURE AI
Full Performance
Encryption & TEE

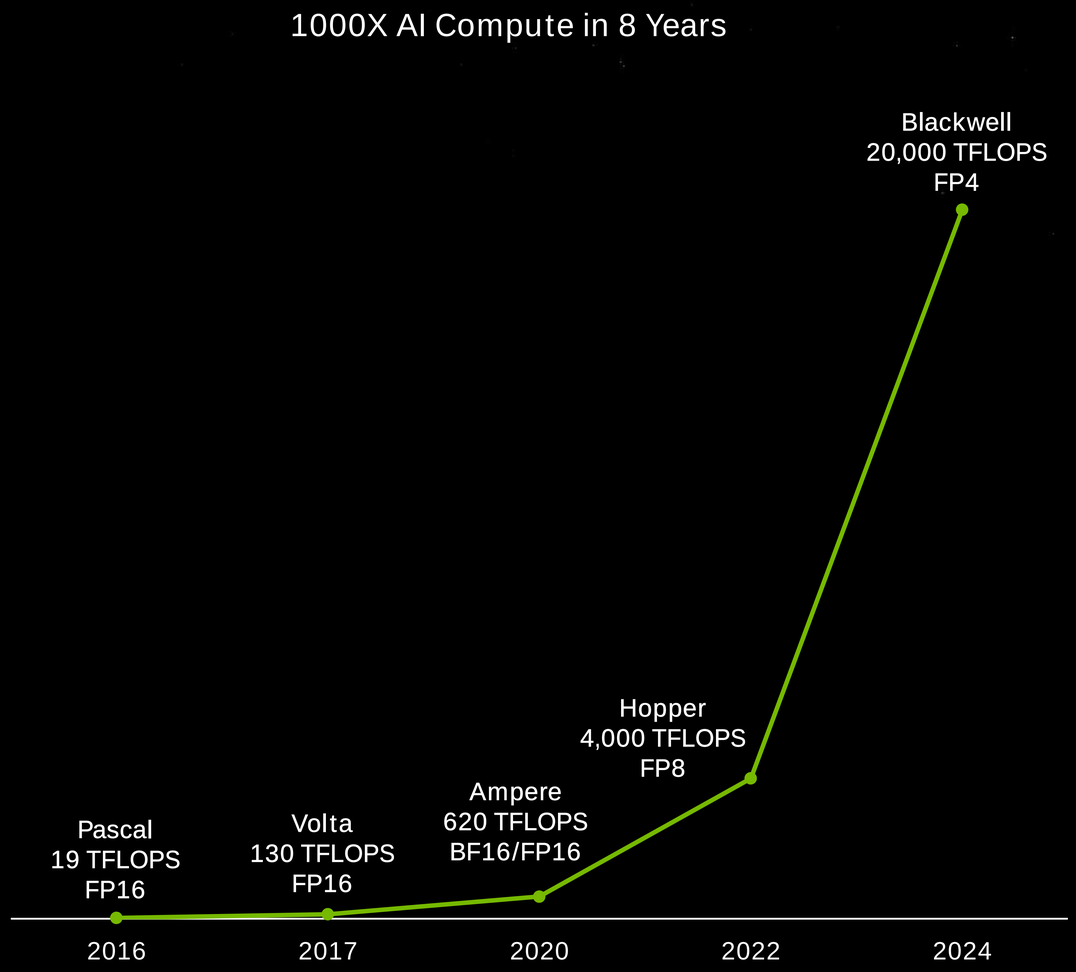
DECOMPRESSION ENGINE
800 GB/sec

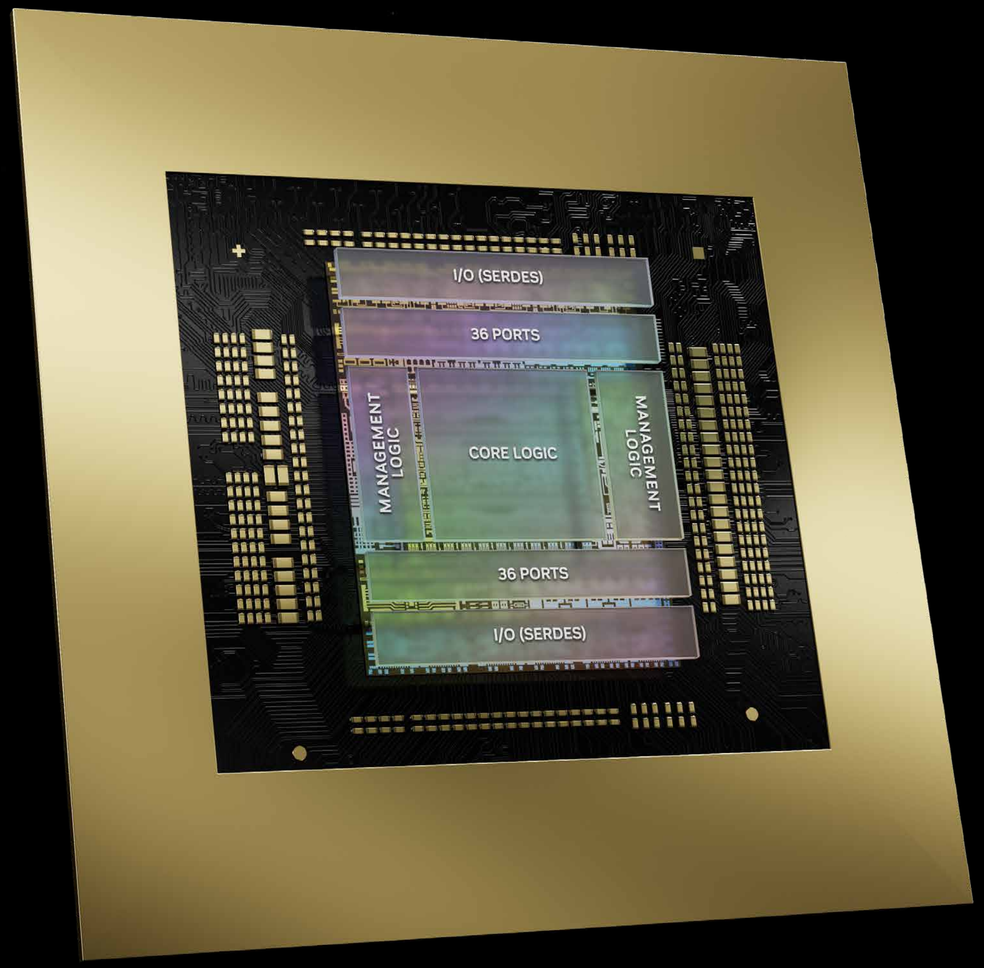


Blackwell GPU

FP8	20 PFLOPS	2.5X Hopper
NEW FP6	20 PFLOPS	2.5X
NEW FP4	**40 PFLOPS**	**5X**
HBM Model Size	740B param	6X
HBM Bandwidth	34T param/sec	5X
NVLINK All-Reduce with SHARP	7.2 TB/s	4X




1000X AI Compute in 8 Years
Blackwell
20,000 TFLOPS
FP4
Hopper
4,000 TFLOPS
FP8
Ampere
620 TFLOPS
BF16/FP16
Volta
130 TFLOPS
FP16
Pascal
19 TFLOPS
FP16
2016
2017
2020
2022
2024



NVLink Switch Chip

50B Transistors in TSMC 4NP

72-Ports Dual 200 Gb/sec SerDes

4 NVLinks at 1.8TB/sec

7.2TB/sec Full-Duplex Bandwidth

SHARP In-Network Compute – 3.6 TFLOPS FP8

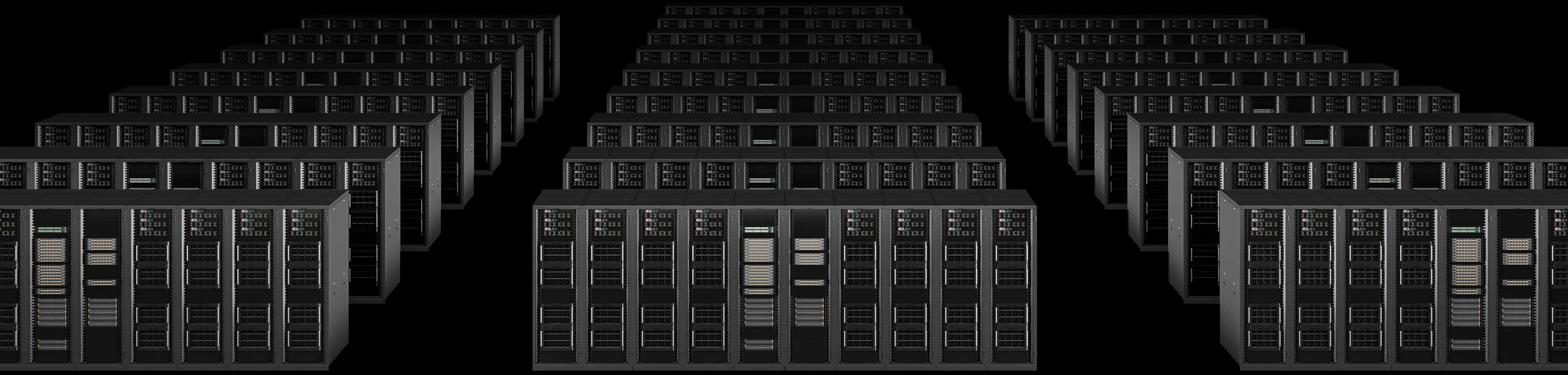
DGX GB200 NVL72
1 Giant GPU

Training FP8	720 PFLOPS	**22X**
Inference FP4	1.44 ExaFLOPS	**45X**
Multi-Node All-to-All	130 TB/sec	**18X**
Multi-Node All-Reduce	260 TB/sec	**36X**


NVIDIA

Train GPT-MoE-1.8T in 90 Days

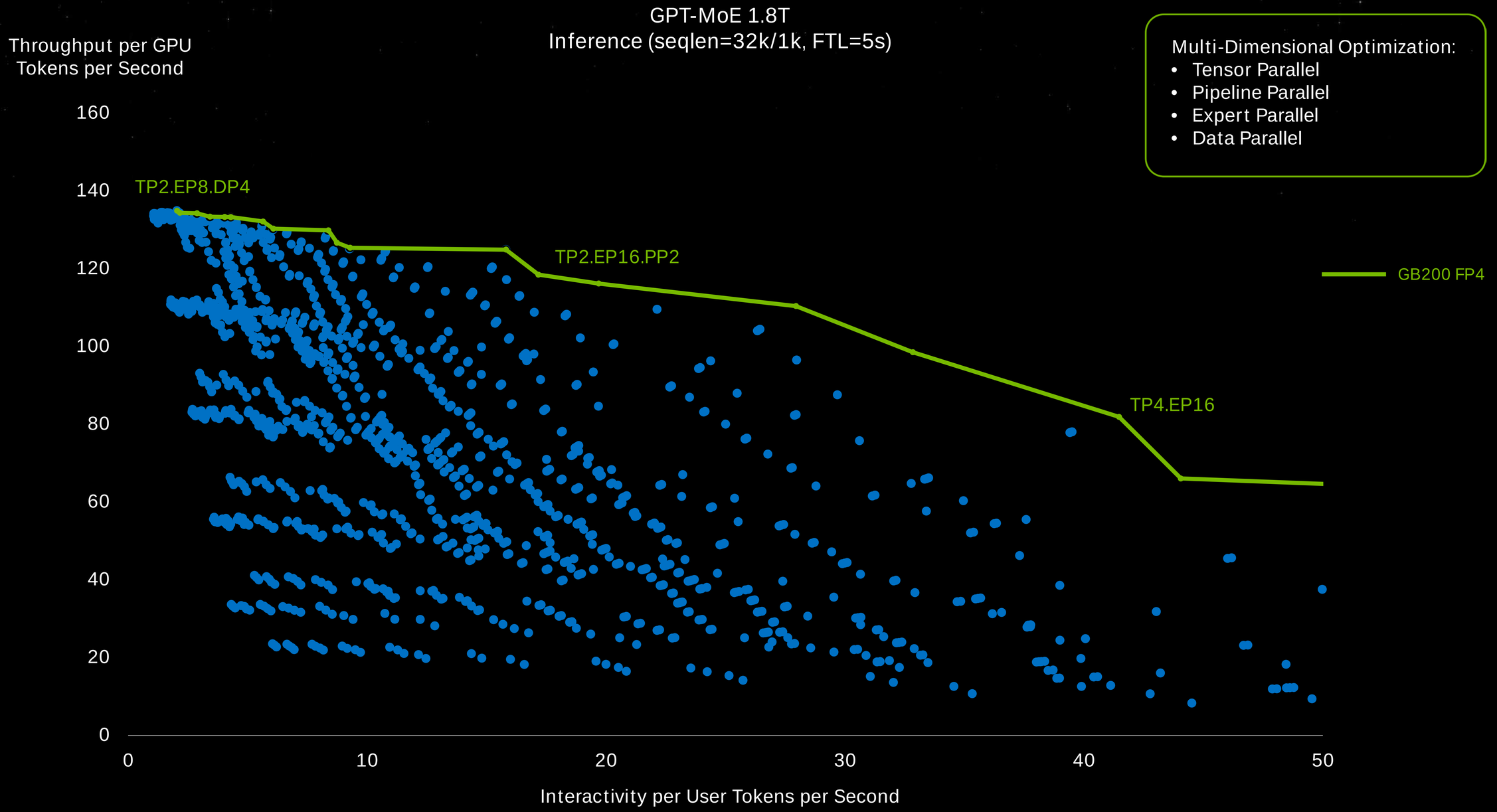
Hopper
8000 GPUs | 15MW

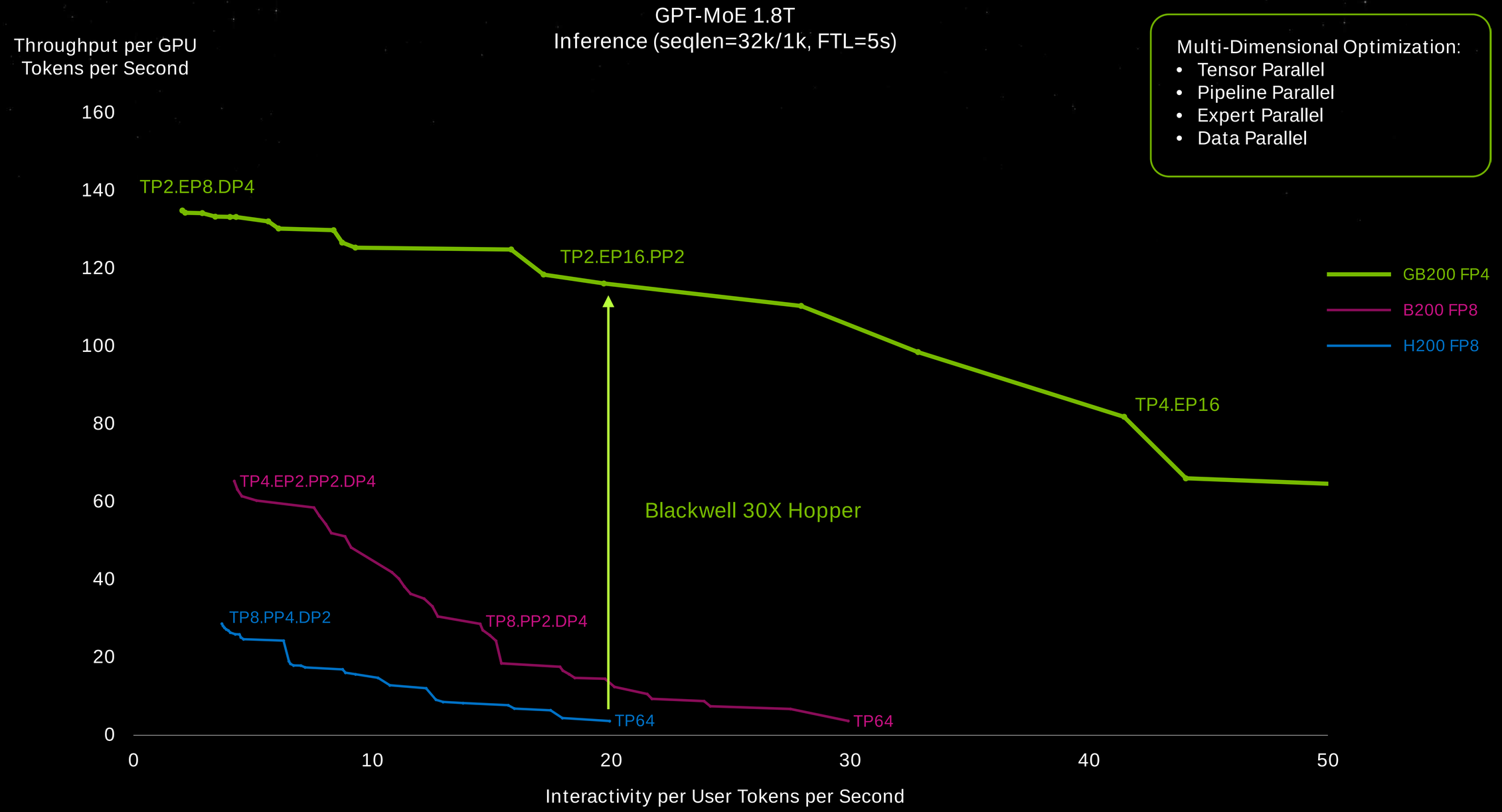


Train GPT-MoE-1.8T in 90 Days

Blackwell GB200 NVL72
2000 GPUs | 4MW

1/4th the Power


GPT-MoE 1.8T
Inference (seqlen=32k/1k, FTL=5s)
Throughput per GPU
Tokens per Second
Multi-Dimensional Optimization:
• Tensor Parallel
• Pipeline Parallel
• Expert Parallel
• Data Parallel
TP2.EP8.DP4
TP2.EP16.PP2
TP4.EP16
GB200 FP4
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
Interactivity per User Tokens per Second


GPT-MoE 1.8T
Inference (seqlen=32k/1k, FTL=5s)
Throughput per GPU
Tokens per Second
Multi-Dimensional Optimization:
• Tensor Parallel
• Pipeline Parallel
• Expert Parallel
• Data Parallel
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
Interactivity per User Tokens per Second
TP2.EP8.DP4
TP2.EP16.PP2
TP4.EP16
TP4.EP2.PP2.DP4
TP8.PP2.DP4
TP8.PP4.DP2
TP64
TP64
Blackwell 30X Hopper
GB200 FP4
B200 FP8
H200 FP8


aws
Google Cloud
Microsoft Azure
ORACLE CLOUD Infrastructure
ADEPT
AI21labs
Character.AI
cohere
essential AI
Hugging Face
Inflection
Meta
MISTRAL AI_
OpenAI
perplexity
Recursion
together.ai
AiVRES
APPLIED DIGITAL
ASRock Rack
ASUS
cisco
CoreWeave
Crusoe
DELL Technologies
EVIDEN
Foxconn HON HAI TECHNOLOGY GROUP
FUJITSU
GIGABYTE
Hewlett Packard Enterprise
IBM Cloud
indosat OOREDOO HUTCHISON
Inventec
Lambda
Lenovo
NORTHERN DATA GROUP
PEGATRON
QCT
Scaleway
Singtel
SoftBank
SUPERMICRO
wistron
wiwynn
YOTTA
YTL COMMUNICATIONS
zt Systems

AWS and NVIDIA expand joint AI offerings

OVER A DECADE OF CO-INNOVATION IN THE CLOUD

AWS AI Security

NVIDIA BLACKWELL COMING TO AWS

AWS NITRO + KMS-ENCRYPTED EFA

NVIDIA DGX Cloud	Amazon EC2
GB200	B100

Project Ceiba

AWS-BUILT AI SUPERCOMPUTER FOR NVIDIA INTERNAL R&D

414 EXAFLOPS | 20,736 GB200 | GEN 4 AWS EFA

Optimizing price performance of FMs running on accelerated computing

Amazon SageMaker

NVIDIA NIM

Transforming industries with the power of AWS and NVIDIA

ADVANCED ROBOTICS SIMULATION

amazon robotics

NVIDIA Omniverse

AI-DRIVEN DRUG DISCOVERY

AWS HealthOmics

NVIDIA BioNeMo

Google NVIDIA

Google Cloud

Data science & analytics

Dataproc | Dataflow | RAPIDS

AI platforms & frameworks

Vertex AI | GKE | DGX Cloud

JAX | XLA | NVIDIA AI Enterprise

TensorRT | Triton | NeMo

Accelerated computing

A3 (H100) | G2 (L4) | Blackwell

Google DeepMind

Gemma

ORACLE | NVIDIA

Sovereign AI	Enterprise Gen AI	Healthcare AI
ORACLE Database	ORACLE Cloud Applications	ORACLE CLOUD Infrastructure
ORACLE CLOUD Infrastructure Generative AI	ORACLE CLOUD Infrastructure Supercluster	ORACLE Distributed Cloud

Microsoft | NVIDIA

Powering the most sophisticated AI innovators

ADEPT

MISTRAL AI_

Inflection

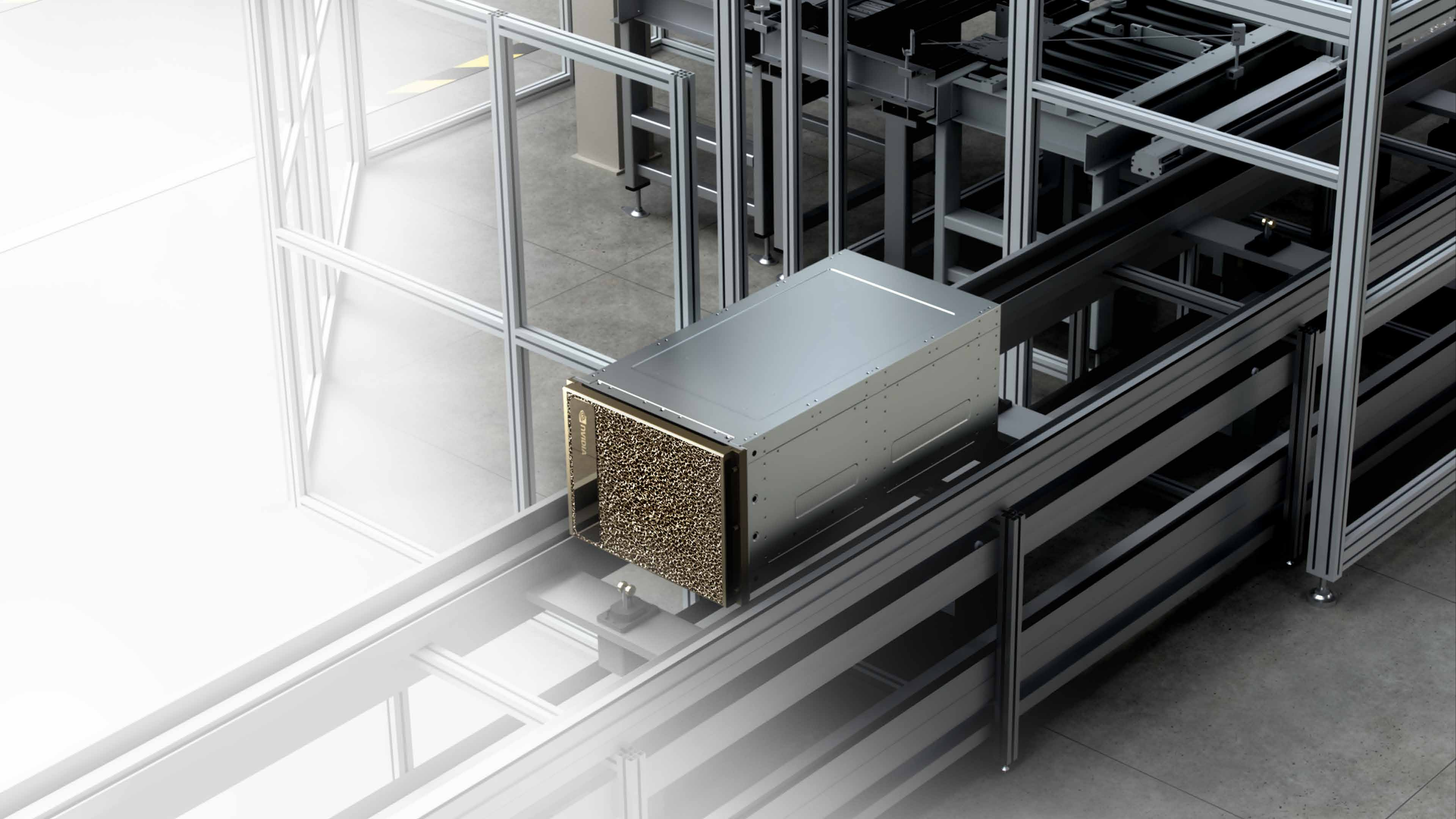
OpenAI

Largest NVIDIA InfiniBand connected supercomputer

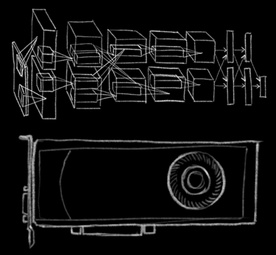
Most deeply integrated cloud for NVIDIA DGX Cloud

Microsoft and NVIDIA Announce

- NVIDIA Clara integration with Azure for healthcare
- NVIDIA NIM available on Azure
- New Azure VM series with NVIDIA H100 NVL
- Microsoft Azure adopts NVIDIA GB200 for customers and AI services
- NVIDIA DGX Cloud Integration with Microsoft Fabric
- NVIDIA Omniverse APIs on Azure



2012
ALEXNET
“FIRST CONTACT”



“CAT”

TEXT

TEXT

IMAGE

VIDEO

TEXT

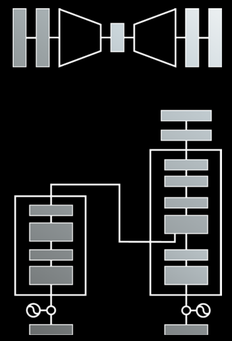
SPEECH

MULTI-MODAL

AMINO ACID

BRAINWAVES

LEARN AND UNDERSTAND EVERYTHING



IMAGE

VIDEO

IMAGE

3D

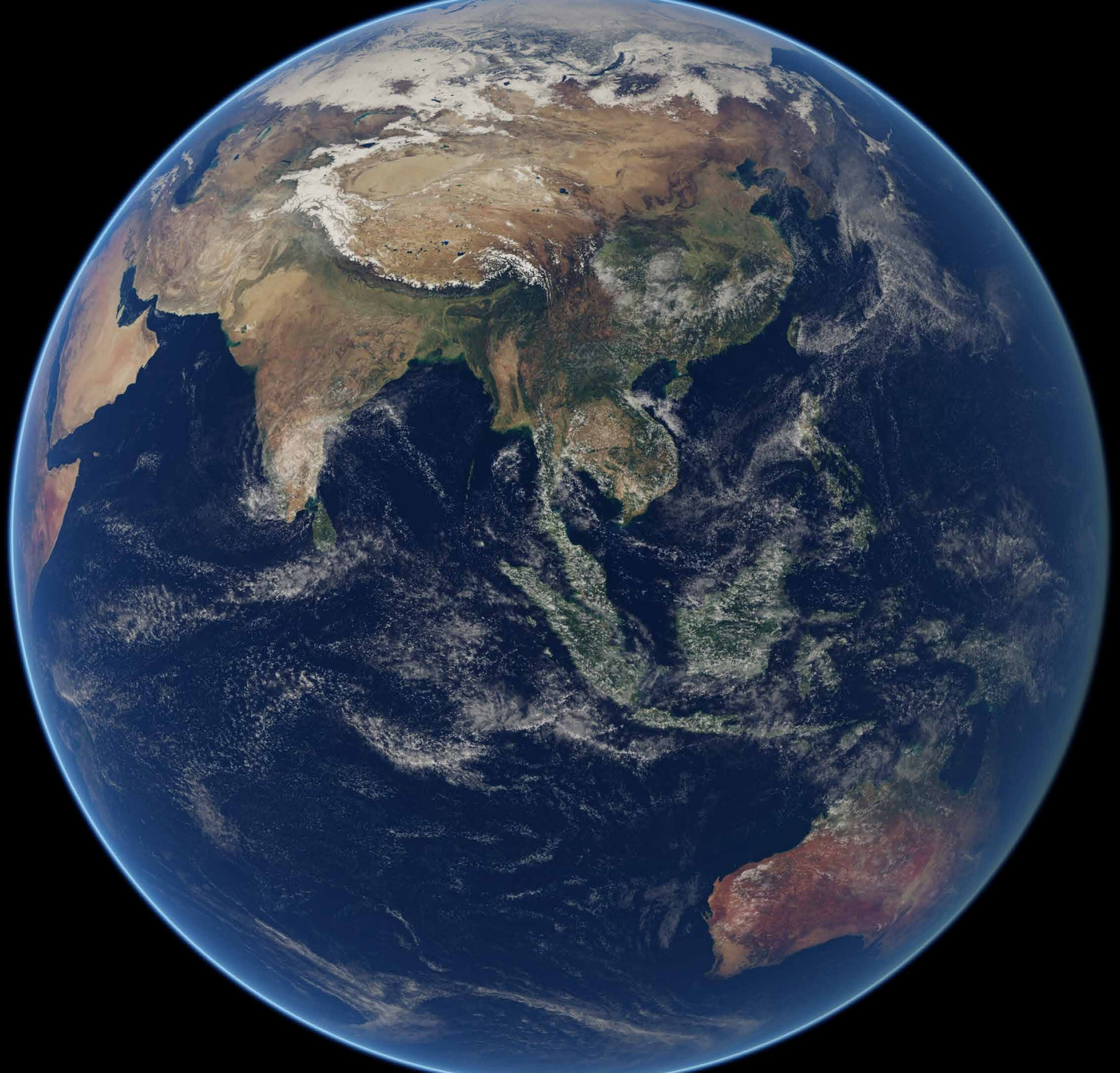
SOUND

ANIMATION

MANIPULATION

PROTEIN

SPEECH



The Weather Company | NVIDIA

TWCo PROCESSING 400 TB OF WEATHER DATA PER DAY

One of the largest collections of global weather data, producing 25B+ forecasts per day for 2.2B locations globally

TWCo GRAF FORECAST MODELLING

Next-generation AI forecast modeling

TWCo WEATHER ENGINE

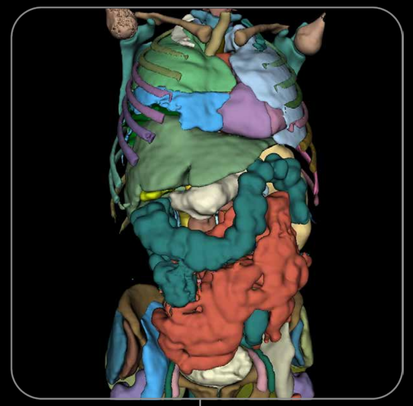
Providing actionable insights to improve social resilience to extreme weather and make weather a competitive advantage for businesses

TWCo WEATHERVERSE

Advanced visualization for better decision making

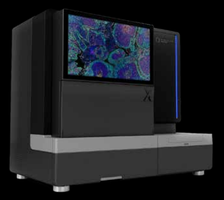
NVIDIA HEALTHCARE

IMAGING & ROBOTICS



MONAI | Holoscan

GENOMICS



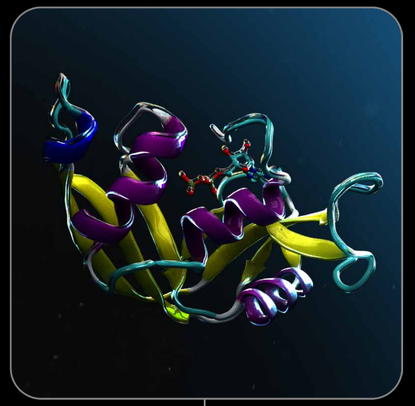
Parabricks

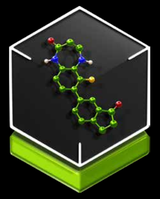
BIOINFORMATICS

pandas

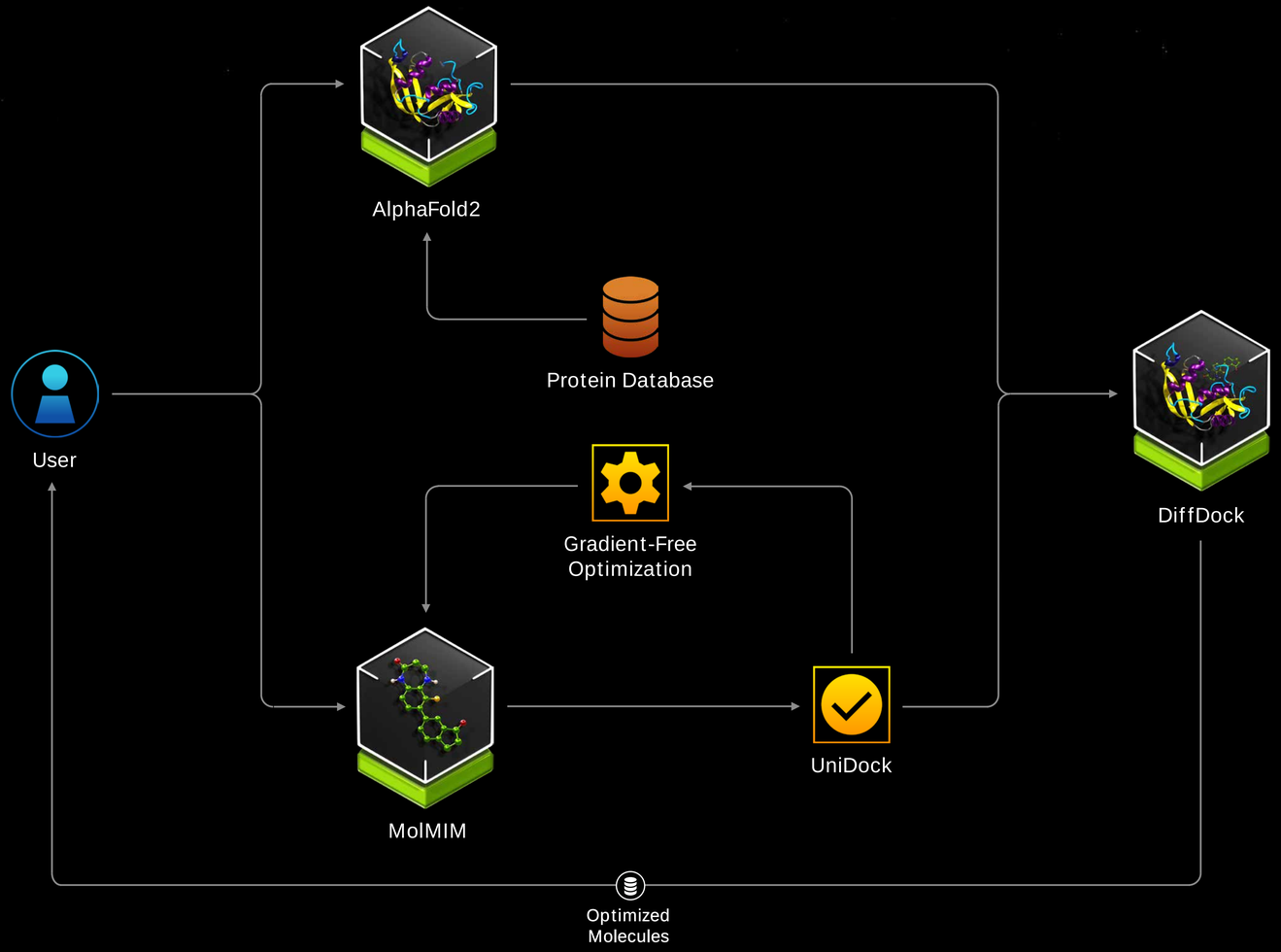
RAPIDS

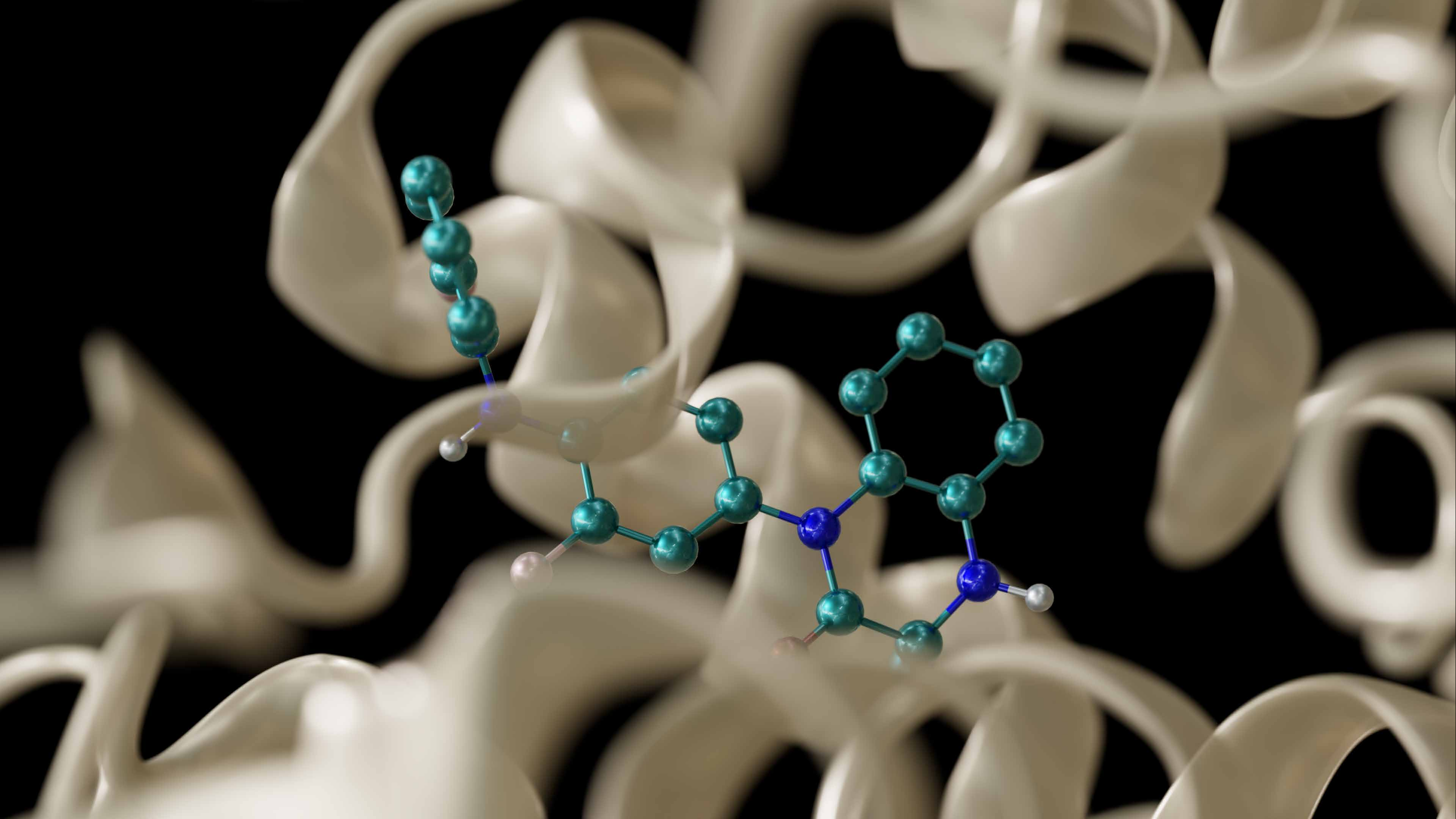
DRUG DISCOVERY

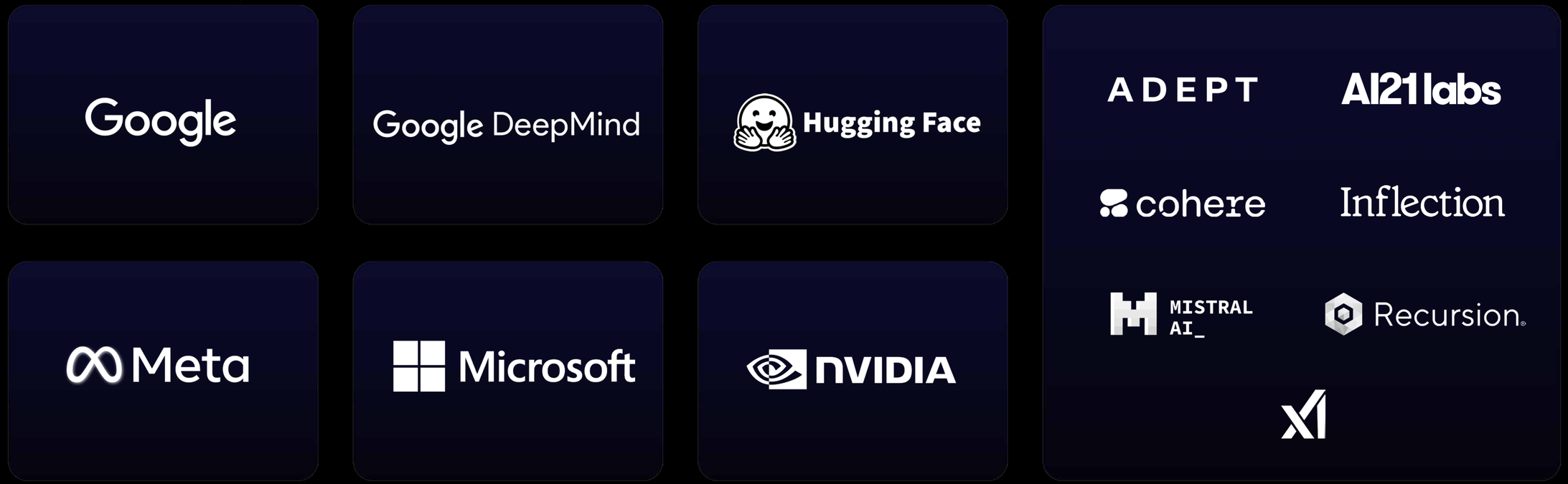


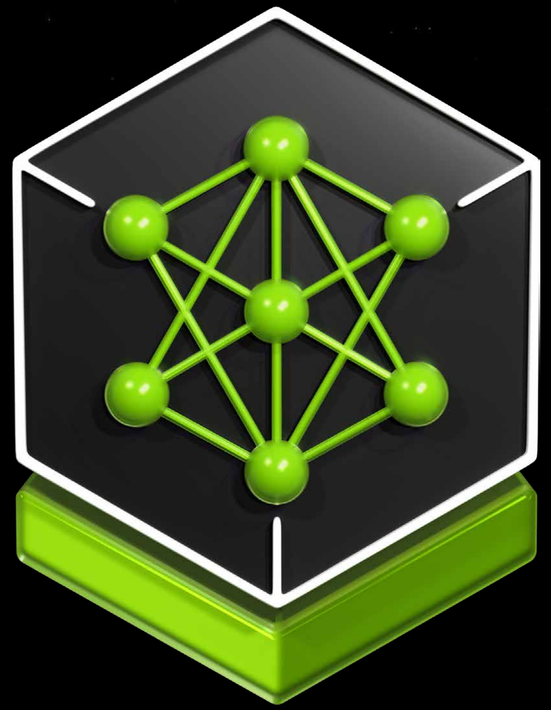


BioNeMo


AlphaFold2
Protein Database
User
Gradient-Free
Optimization
DiffDock
MolMIM
UniDock
Optimized
Molecules




Google
Google DeepMind
Hugging Face
ADEPT
AI21labs
cohere
Inflection
Meta
Microsoft
NVIDIA
MISTRAL AI_
Recursion



NVIDIA INFERENCE MICROSERVICE

Pre-Trained AI Models
Packaged and Optimized to Run Across
CUDA Installed Base

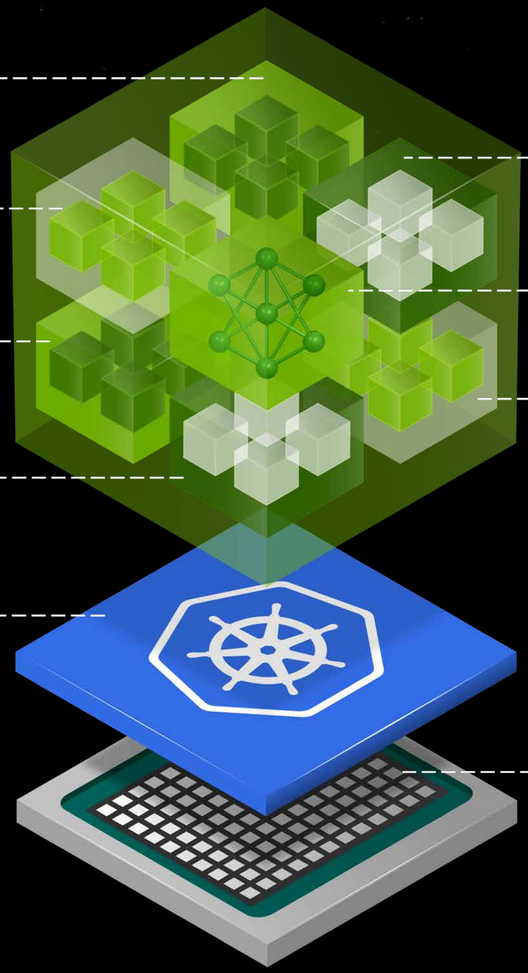
Industry Standard APIs
Text, Speech, Image,
Video, 3D, Biology

Triton Inference Server
cuDF, CV-CUDA, DALI, NCCL,
Post Processing Decoder

Cloud Native Stack
GPU Operator, Network Operator

Enterprise Management
Health Check, Identity, Metrics,
Monitoring, Secrets Management

Kubernetes



TensorRT LLM and Triton
cuBLAS , cuDNN, In-Flight Batching,
Memory Optimization, FP8 Quantization

Optimized Model
Single GPU, Multi-GPU, Multi-Node

Customization Cache
P-Tuning, LORA, Model Weights

NVIDIA CUDA

100’s of Millions of CUDA GPUs Installed Base

NVIDIA

Search NVIDIA AI

Explore Docs Login

Discover

MODELS

Reasoning
Visual Design
Speech & Translation
Video
Biology

DATA PROCESSING

Embedding
Retrieval

OPTIMIZATION

Inference
Route Planning

INDUSTRIES

Gaming
Healthcare
Automotive
Industrial

Top Open Foundation Models

The leading open models built by the community, optimized and accelerated by NVIDIA's enterprise-ready inference runtime

Open Full Page

Input

Try Python Node.js Shell

Input Prompt View Examples

A happy dog hanging out at the park

View Parameters

Reset Parameters Run

Output

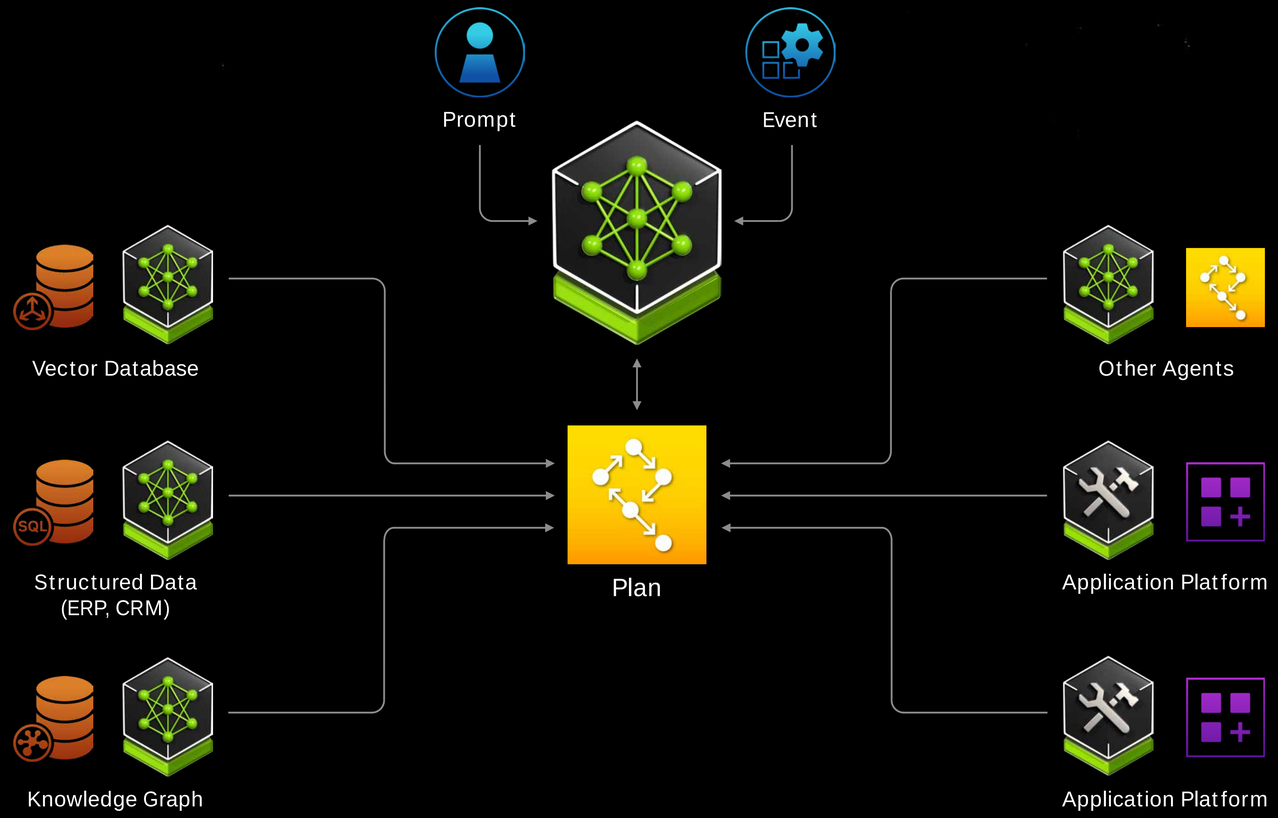
Preview JSON

Trending Now

The latest and most popular additions to the list

Explore by Collection

Discover new use-cases and the right set of APIs to turbocharge your enterprise


Prompt
Event
Vector Database
Structured Data
(ERP, CRM)
Knowledge Graph
Plan
Other Agents
Application Platform
Application Platform

What is CTL? Write an example CTL test.

Prompt

Response

Event

NVIDIA ChipNeMo
Llama 2 70B
Before Fine-Tuning

CTL stands for "Combinational Timing Logic" and it's a type of digital circuit used in ASIC (Application-Specific Integrated Circuit) design. CTL is used to implement finite state machines, counters, and other sequential logic circuits.

Here is an example of a CTL Test for a simple finite machine:

1. Set up the initial state: Set EN to 0 and wait for the state machine to settle into state S0.

2. Test the transition from S0 to S1: Set EN to 1 and wait for the state machine to transition to state S1. Verify that the output signal (Q) is set to 1.

3. Test the transition from S1 to S0: Set EN to 0 and wait for the state machine to transition back to state S0. Verify that the output signal (Q) is set to 0.

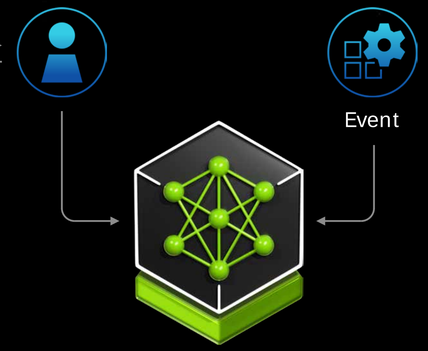
4. Test the reset behavior: Set EN to 0 and wait for the state machine to settle into state S0. Then, assert the reset signal (RST) and verify that the output signal (Q) is set to 0.

5. Test the clock behavior: Set EN to 1 and wait for the state machine to transition to state S1. Then, pause the clock and verify that the output signal (Q) remains set to 1. Resume the clock and verify that the state machine transitions back to state S0 when the clock is restarted.

What is CTL? Write an example CTL test.

Prompt

Response

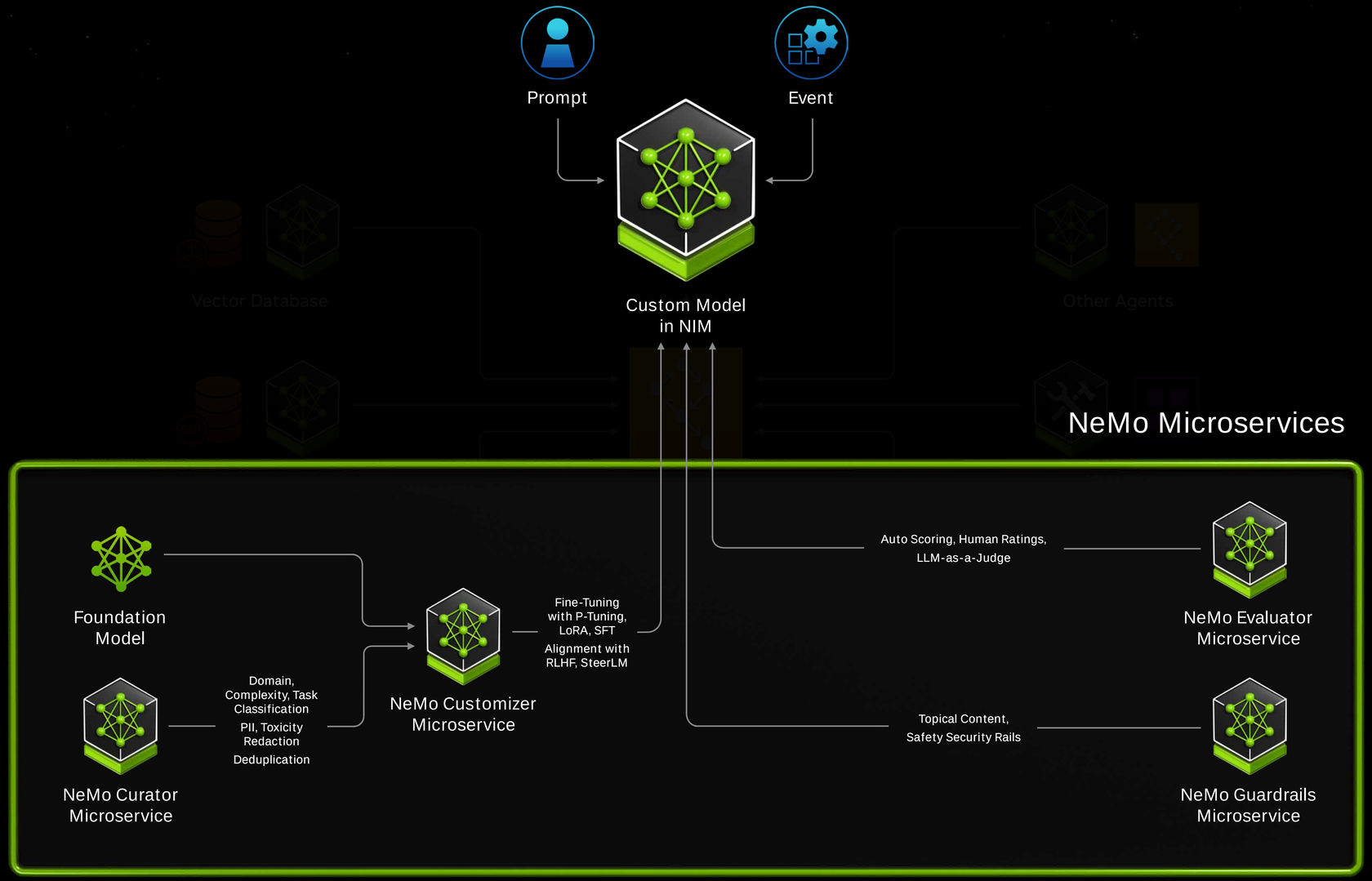


NVIDIA ChipNeMo
Llama 2 70B
With Fine-Tuning

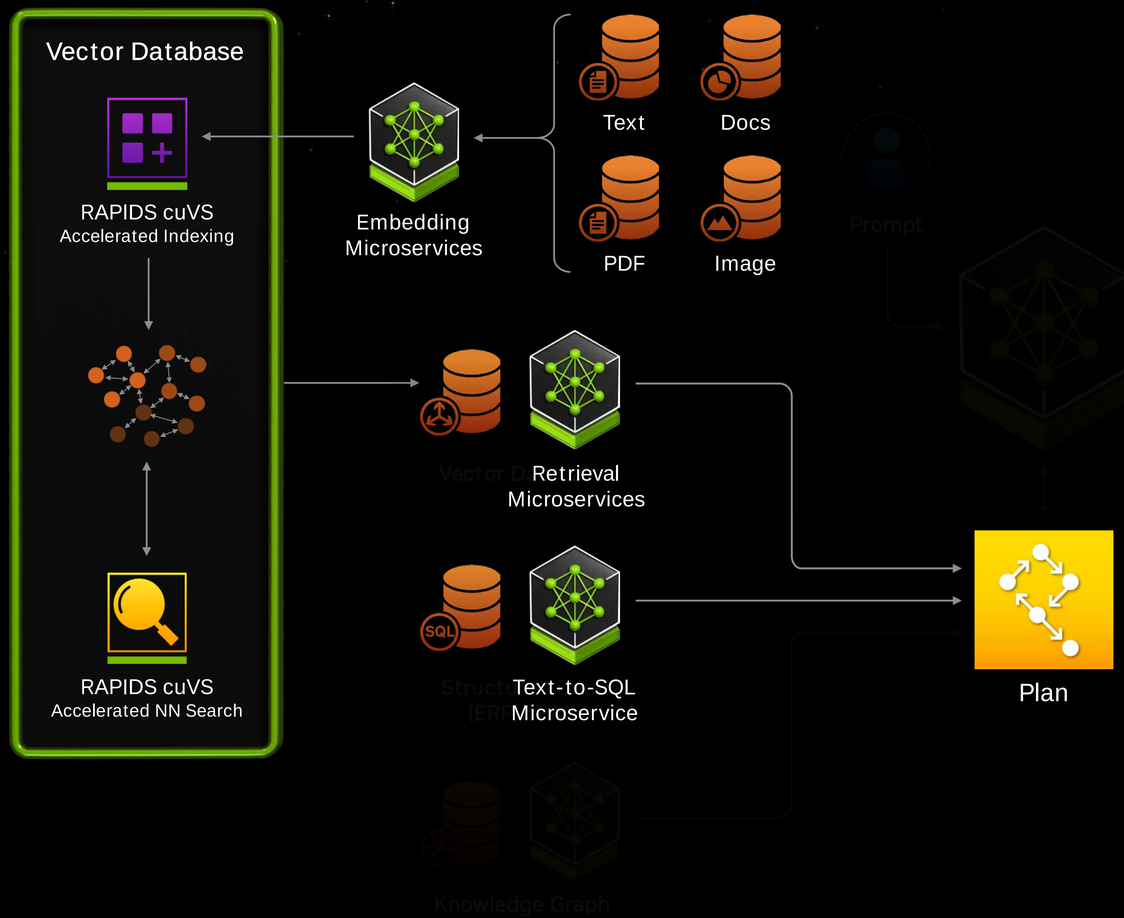
CTL stands for "Compute Trace Library". It is a trace generation library used to create functional and performance tests for the GPU architecture. CTL provides an interface to write tests in C++ that can be compiled into a trace that runs on various simulation models, such as Fmodel, RTL, and Emulation.

Here is an example of a simple CTL test that allocates a buffer, initializes it with some values, and then reads back the values to verify that they are correct:

```
#include "ctl.h"
using namespace ctl;
int main(int argc, char* argv[]) {
// Set up the test
ctlAddTest("simple_buffer", "", "ga100");
// Allocate a buffer and initialize it
with some values
unsigned* buffer;
ctlMalloc((void**)&buffer, 1024, "BUFFER",
CTL_GLOBAL_MEMORY);
ctlMemset(buffer, 0, 1024);
ctlMemcpy(buffer, (void*)0x12345678, 4,
ctlMemcpyHostToDevice);
// Set up the task
ctlCreateTask("TASK", "kernel");
ctlSetQmdCtaRasterDimension("TASK", 1, 1,
1);
ctlSetQmdCtaThreadDimension("TASK", 1, 1,
1);
// Set up the kernel
char program[] =
"!!SPA7.0\n"
".THREAD_TYPE COMPUTE\n"
".option
```


Prompt
Event
Custom Model in NIM
Vector Database
Other Agents
NeMo Microservices
Foundation Model
NeMo Curator Microservice
Domain, Complexity, Task Classification
PII, Toxicity Redaction
Deduplication
NeMo Customizer Microservice
Fine-Tuning with P-Tuning, LoRA, SFT
Alignment with RLHF, SteerLM
Auto Scoring, Human Ratings, LLM-as-a-Judge
NeMo Evaluator Microservice
Topical Content, Safety Security Rails
NeMo Guardrails Microservice

NeMo Retriever



NeMo Retriever

Vector Database

RAPIDS cuVS
Accelerated Indexing

RAPIDS cuVS
Accelerated NN Search

Embedding
Microservices

Text

Docs

PDF

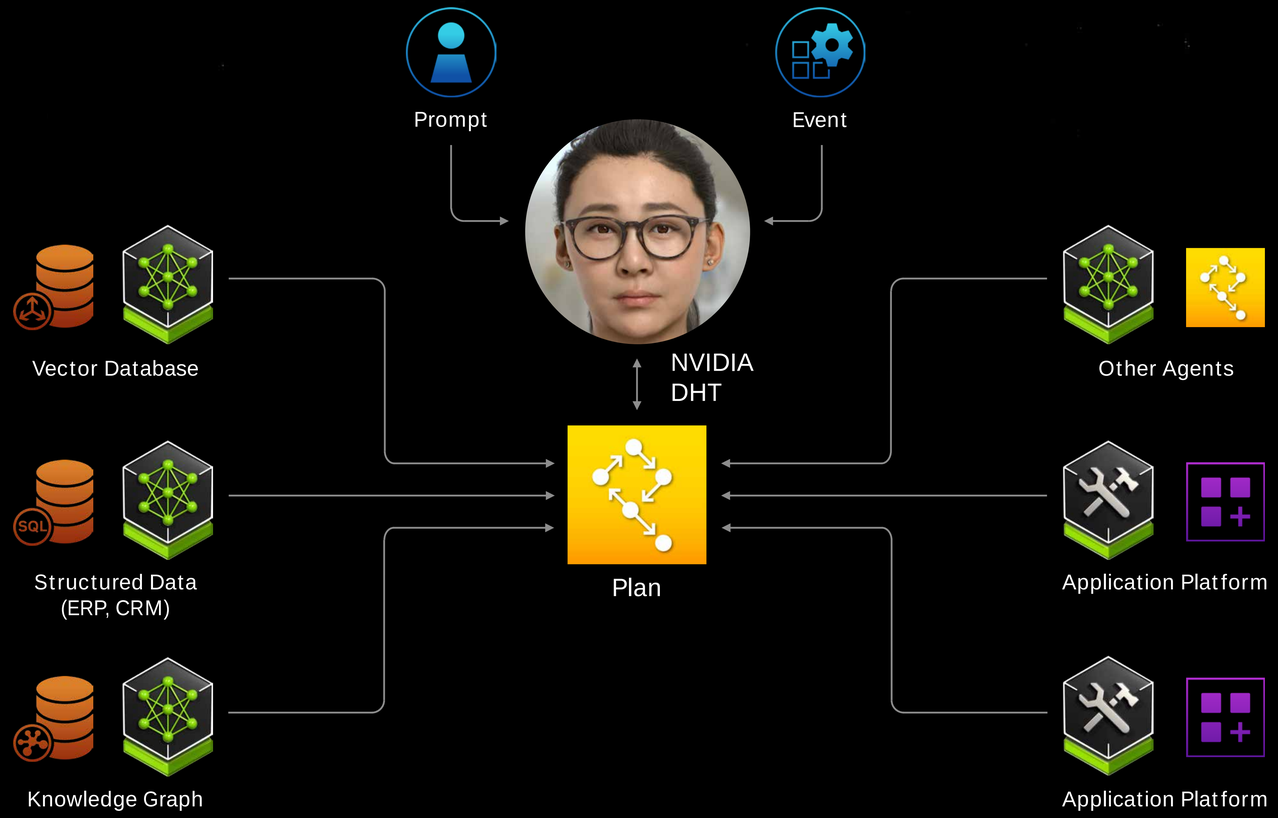
Image

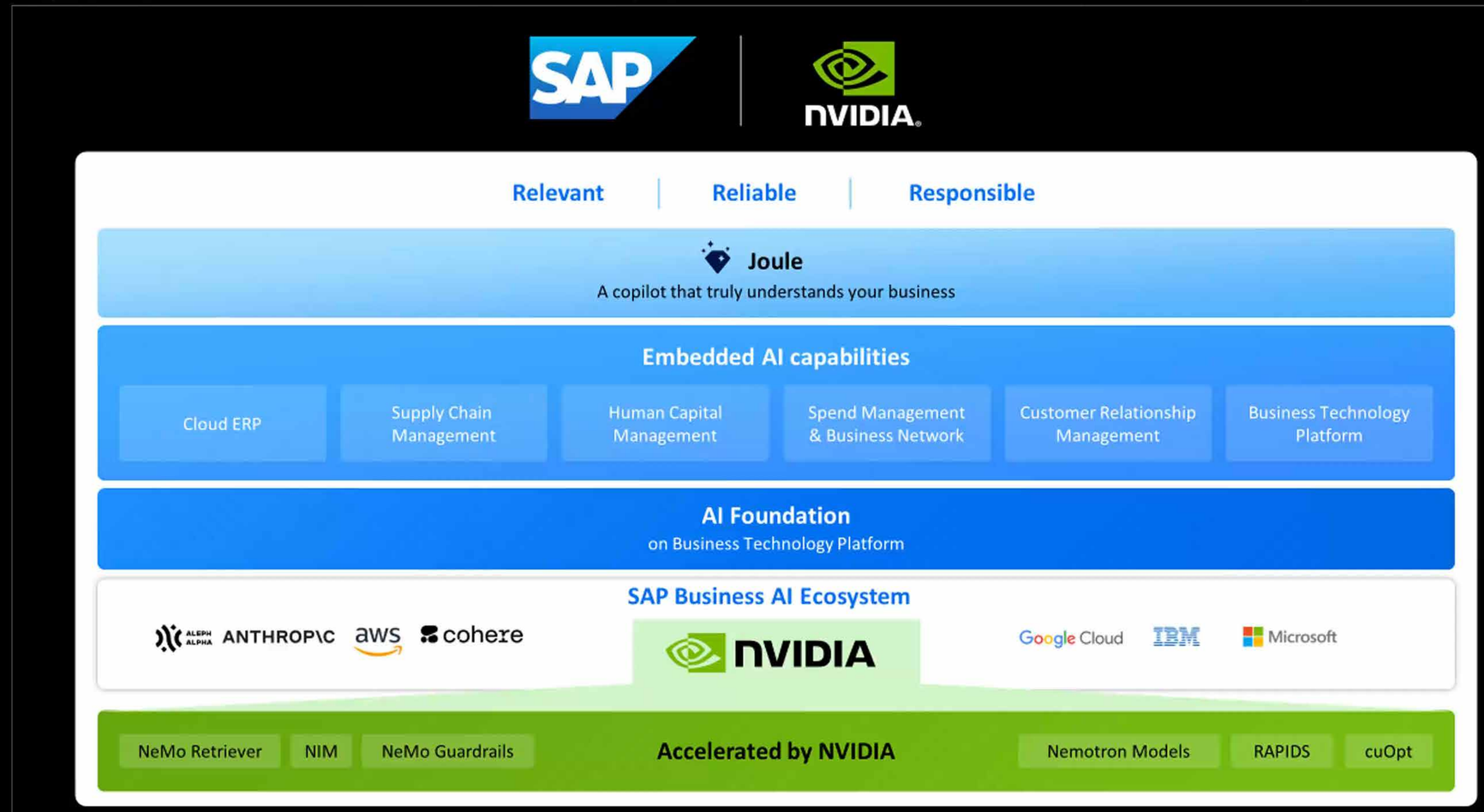
Retrieval
Microservices

Text-to-SQL
Microservice

Plan

Chat with PDF


Prompt
Event
Vector Database
NVIDIA
DHT
Other Agents
Structured Data
(ERP, CRM)
Plan
Application Platform
Knowledge Graph
Application Platform


Relevant
Reliable
Responsible
Joule
A copilot that truly understands your business
Embedded AI capabilities
Cloud ERP
Supply Chain Management
Human Capital Management
Spend Management & Business Network
Customer Relationship Management
Business Technology Platform
AI Foundation
on Business Technology Platform
SAP Business AI Ecosystem
ALEPH ALPHA
ANTHROPIC
aws
cohere
NVIDIA
Google Cloud
IBM
Microsoft
NeMo Retriever
NIM
NeMo Guardrails
Accelerated by NVIDIA
Nemotron Models
RAPIDS
cuOpt

servicenow

THE INTELLIGENT WORKFLOW COMPANY

Now Assist Generative AI Experiences powered by ServiceNow Platform

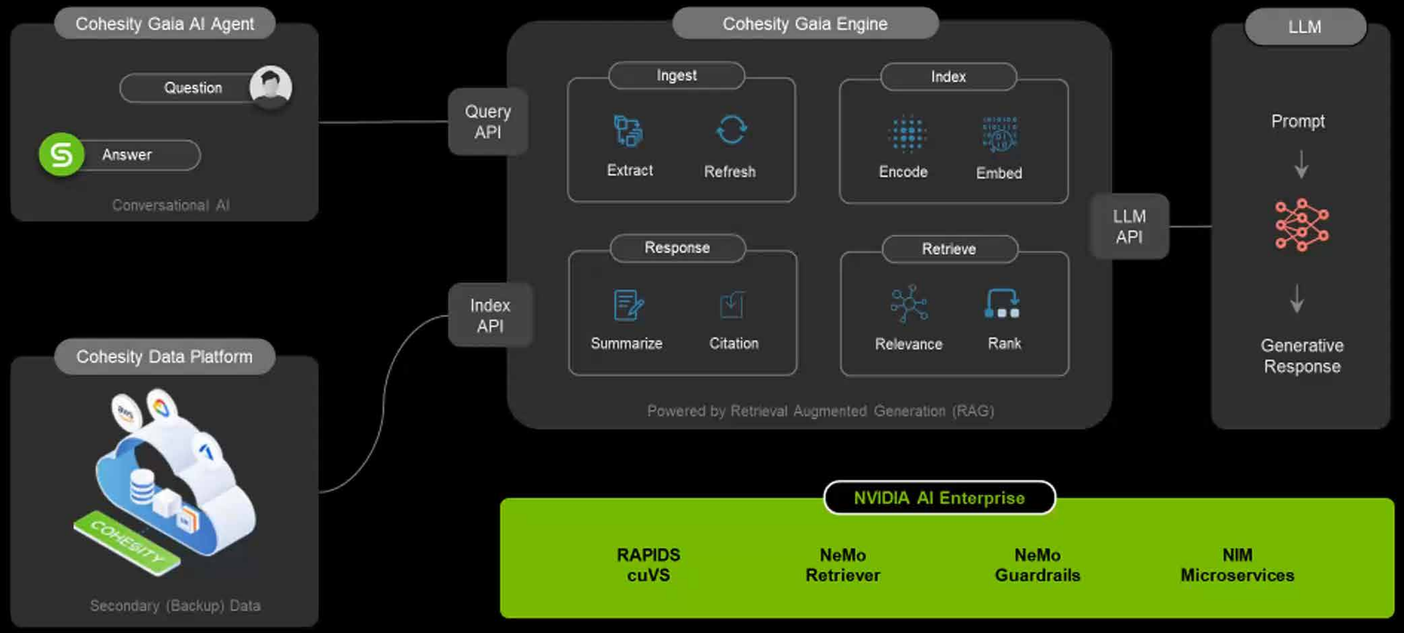
AI Research & Development	Engineering & Applied Science	ServiceNow Global Cloud Services
Enterprise Foundation Models	Domain Specific AI Models	AI Embedded Workflows
AI Trustworthiness	Enterprise Workflows	Privacy and Security
StarCoder		Grounded in Customer Data

POWERED BY NVIDIA SINCE 2018

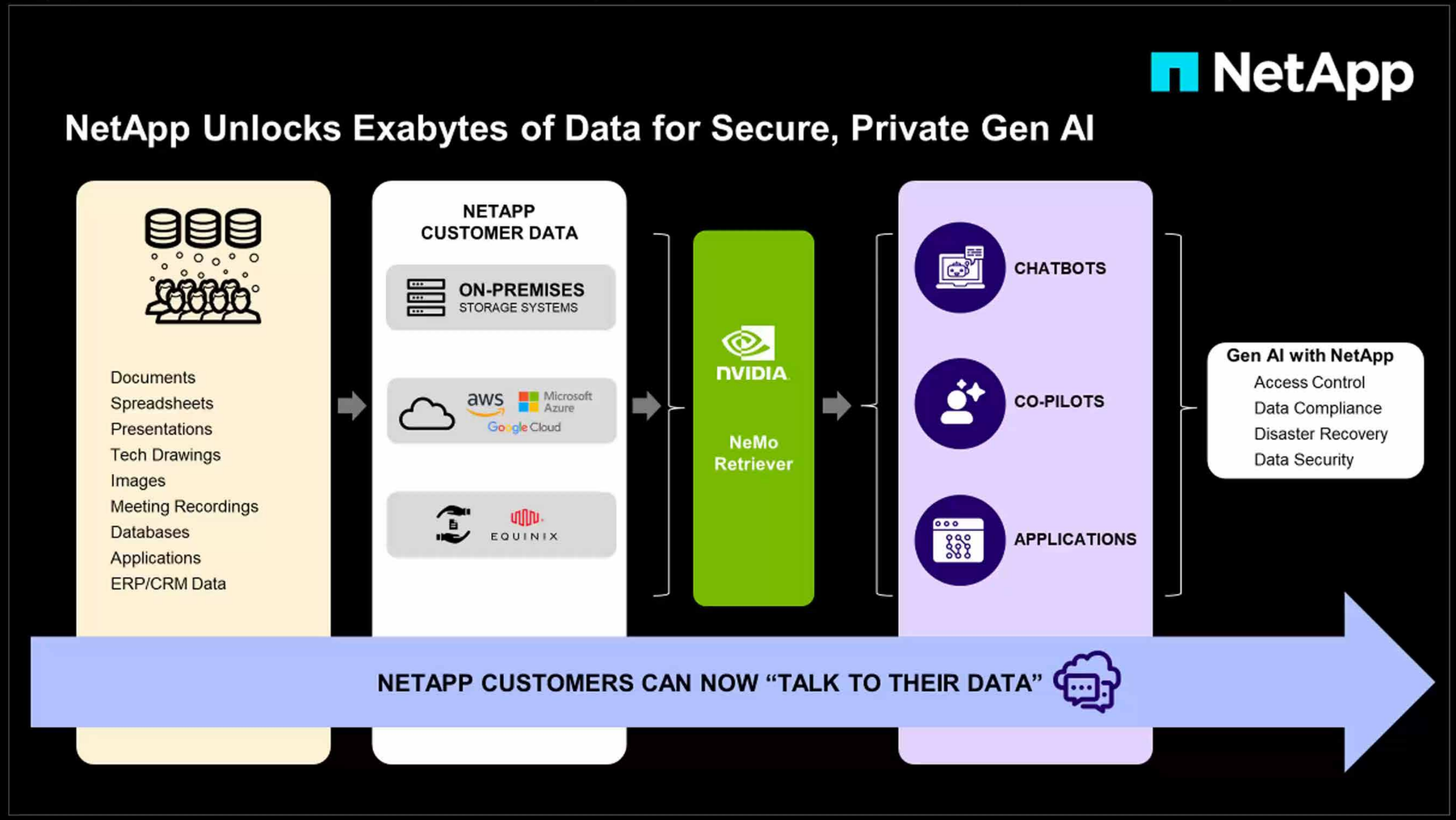
DGX Cloud | DGX SuperPOD | NeMo | TensorRT | NIM Microservices | NVIDIA AI Enterprise

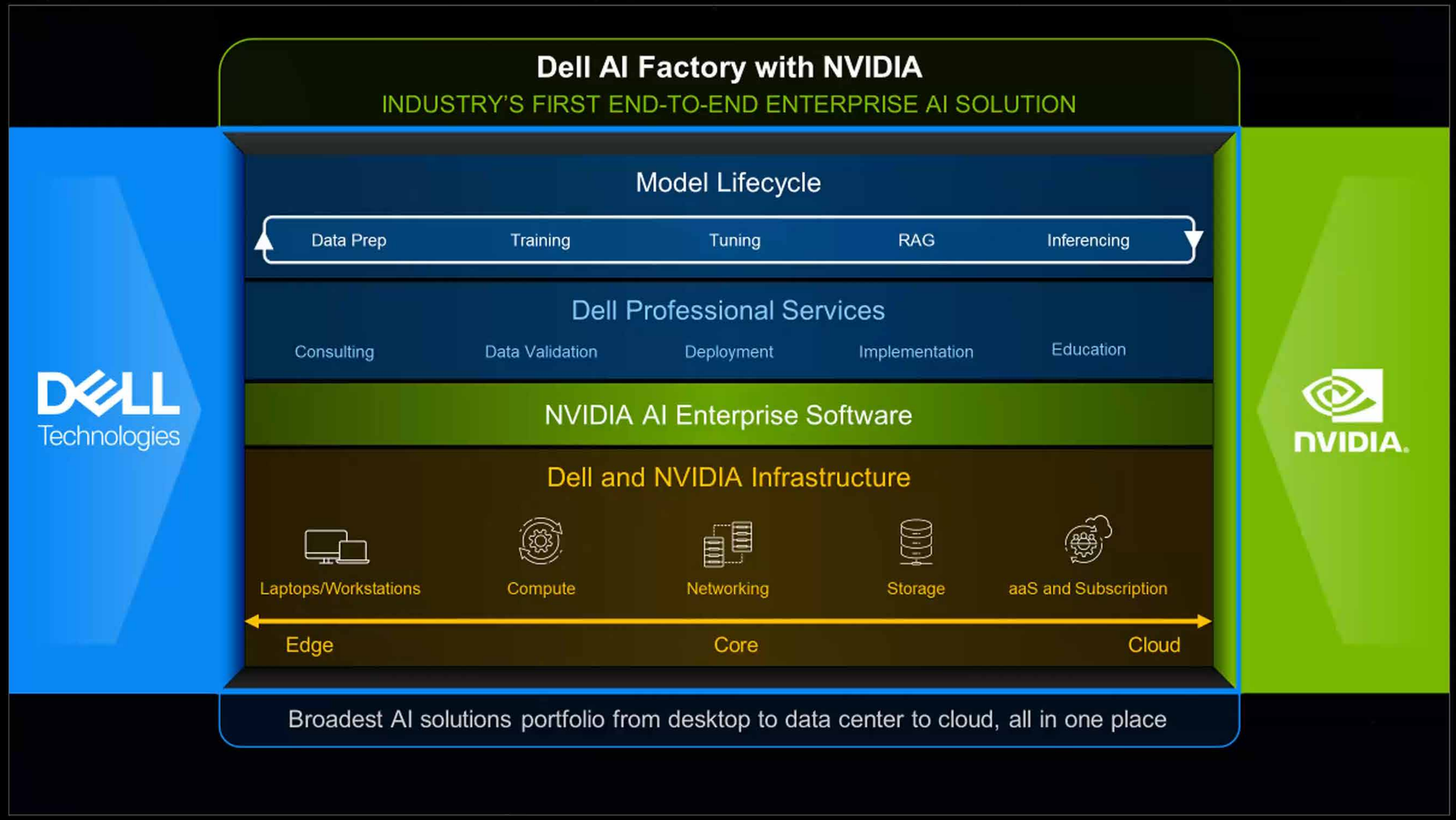
COHESITY | NVIDIA

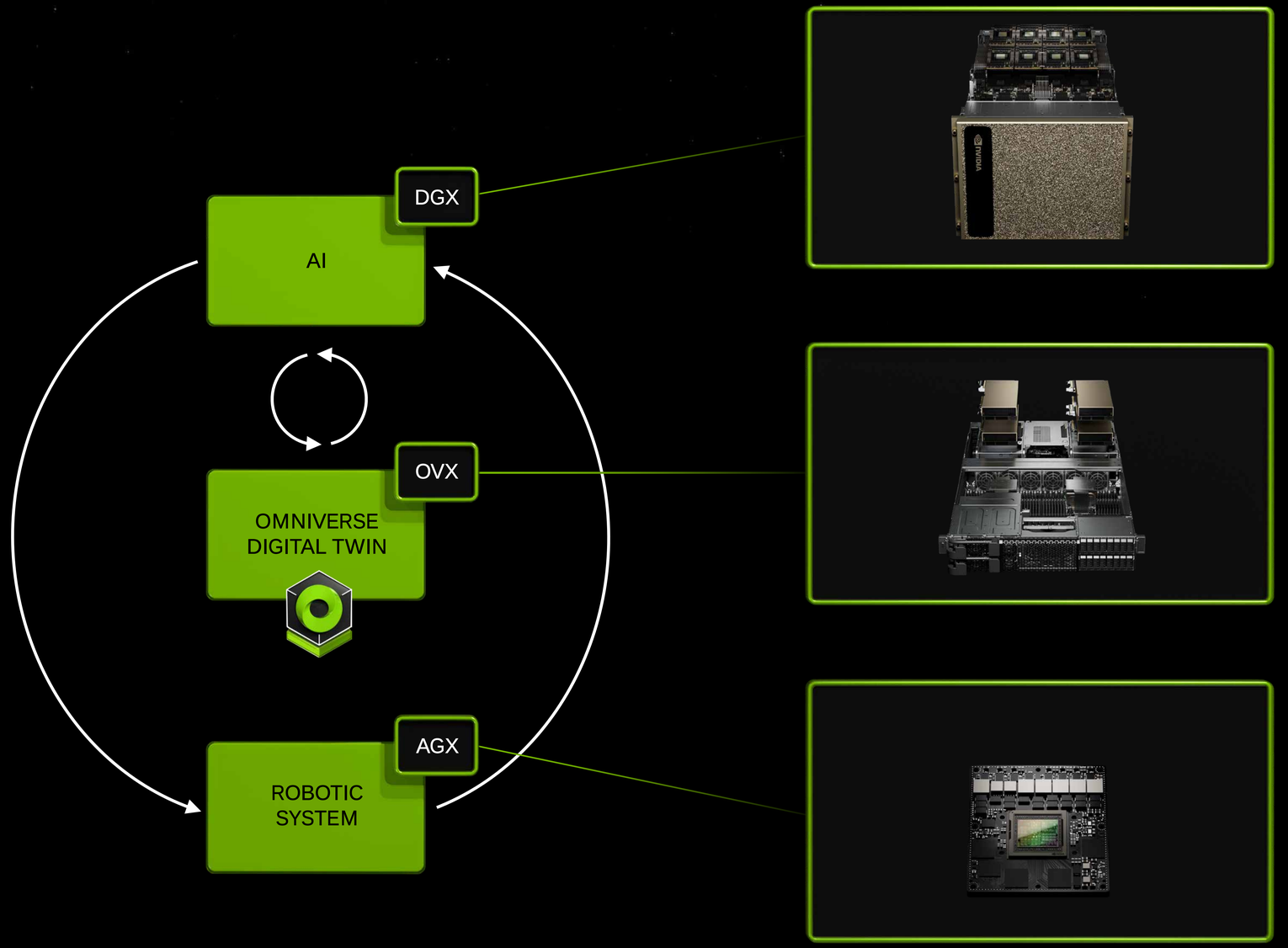
Cohesity Gaia – First to Provide AI-Powered Business Insights from Secondary Data




snowflake
DATA + AI — ALL ON NVIDIA
APPLICATIONS
Snowflake Copilot
Snowflake Document AI
SNOWPARK CONTAINER SERVICES
Model Training, Fine Tuning
SNOWFLAKE CORTEX
Serverless AI, LLMs, Search
ENTERPRISE DATA
ACCELERATED BY NVIDIA
NeMo Retriever
Triton Inference Server
TensorRT


NetApp
NetApp Unlocks Exabytes of Data for Secure, Private Gen AI
Documents
Spreadsheets
Presentations
Tech Drawings
Images
Meeting Recordings
Databases
Applications
ERP/CRM Data
NETAPP
CUSTOMER DATA
ON-PREMISES
STORAGE SYSTEMS
aws
Microsoft Azure
Google Cloud
EQUINIX
NVIDIA
NeMo
Retriever
CHATBOTS
CO-PILOTS
APPLICATIONS
Gen AI with NetApp
Access Control
Data Compliance
Disaster Recovery
Data Security
NETAPP CUSTOMERS CAN NOW “TALK TO THEIR DATA”

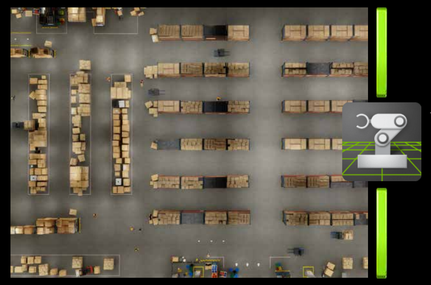

Dell AI Factory with NVIDIA
INDUSTRY'S FIRST END-TO-END ENTERPRISE AI SOLUTION
DELL Technologies
Model Lifecycle
Data Prep
Training
Tuning
RAG
Inferencing
Dell Professional Services
Consulting
Data Validation
Deployment
Implementation
Education
NVIDIA AI Enterprise Software
Dell and NVIDIA Infrastructure
Laptops/Workstations
Compute
Networking
Storage
aaS and Subscription
Edge
Core
Cloud
NVIDIA
Broadest AI solutions portfolio from desktop to data center to cloud, all in one place


DGX
AI
OVX
OMNIVERSE
DIGITAL TWIN
AGX
ROBOTIC
SYSTEM

AMR Robot Digital Twins

Isaac Sim

Warehouse Digital Twin



Isaac Sim

Worker Digital Twins

Isaac Sim

USD APIs

Omniverse Channel APIs

NVIDIA
Omniverse Cloud

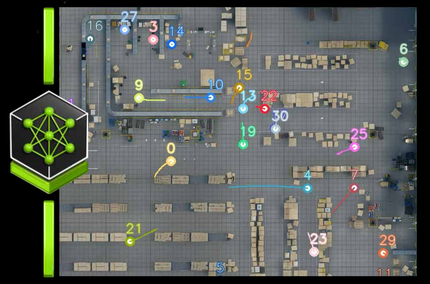
NVIDIA
DGX Cloud

Occupancy APIs

cuOpt APIs

VLM APIs

Sensor Fusion

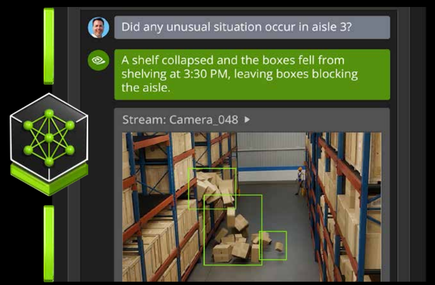


Metropolis

Route Planning

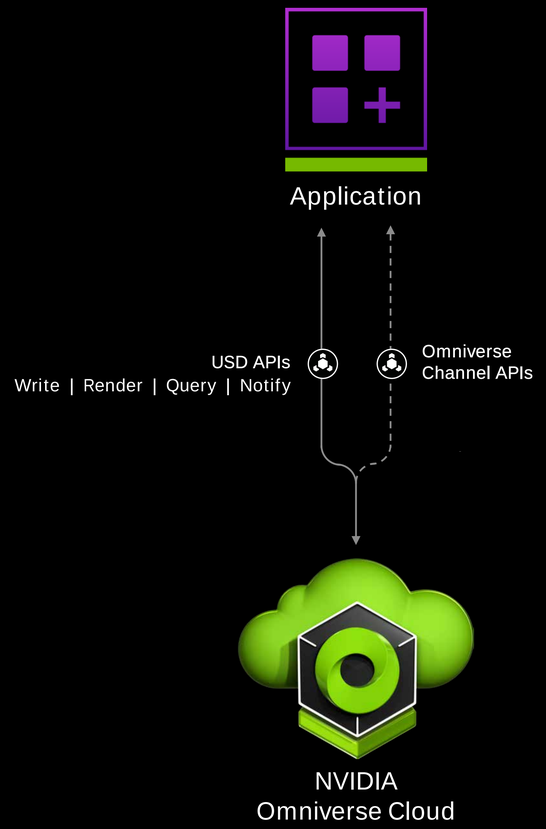
cuOpt

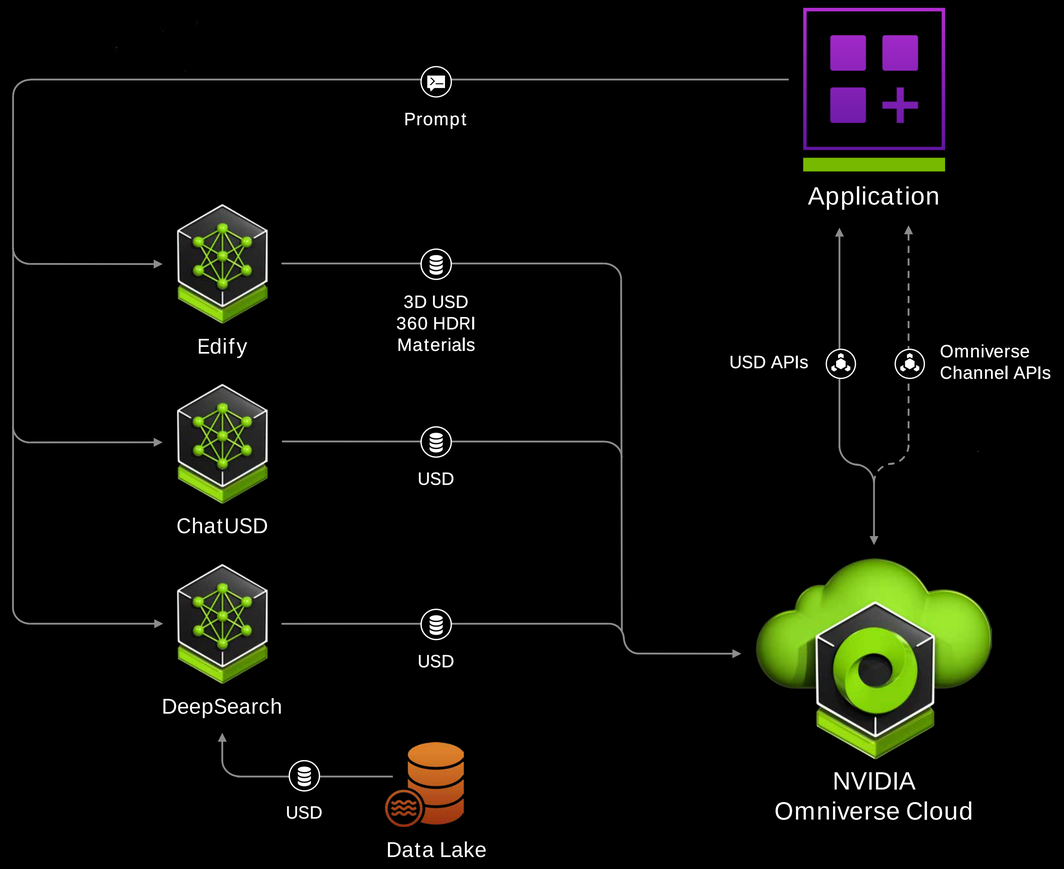
Operator UI



Metropolis


LESTAK-10
DIGITAL TWIN


Application
USD APIs
Write | Render | Query | Notify
Omniverse
Channel APIs
NVIDIA
Omniverse Cloud


Prompt
Application
Edify
3D USD
360 HDRI
Materials
USD APIs
Omniverse
Channel APIs
ChatUSD
USD
DeepSearch
USD
USD
Data Lake
NVIDIA
Omniverse Cloud

SIEMENS | NVIDIA

Building the Next Era of Industrial Digitalization Together

Immersive digital twins

- Siemens Xcelerator
- NVIDIA Omniverse Cloud

Generative AI

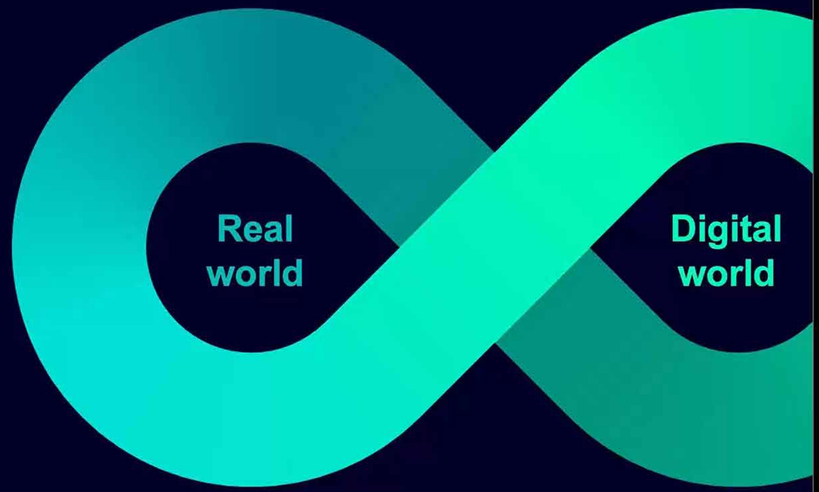
- Siemens Xcelerator
- NVIDIA AI Enterprise

Industrial Edge AI & Robotics

- Siemens Xcelerator
- NVIDIA Robotics

Industrial AI-Physics Simulation

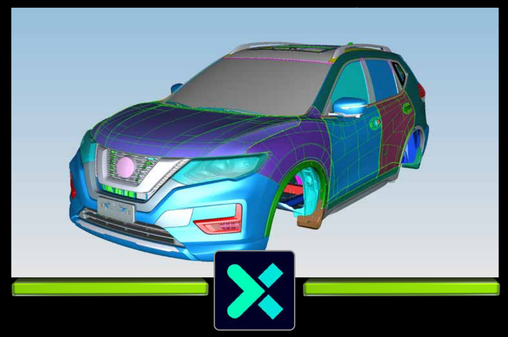
- Siemens Xcelerator
- NVIDIA Accelerated Computing



SIEMENS


HD HYUNDAI
HD HYUNDAI

NISSAN



3DEXCITE

AYA

Omniverse
Channel APIs

USD APIs

NVIDIA
Omniverse Cloud

T

Team
Hyundai

GENERAL

Dashboard

Workspaces

ORDERS

New Asset

My Orders

Cart

PRESETS

Templates

USER MANAGEMENT

Members

Environments

3D | 360° | 2D

REMOVE

SAVE AS PRESET

Vertical Horizon Offset

0

Environment Rotation

0

Prompt

Breathtaking swiss mountain landscape with a large open space in a cinematic afternoon atmosphere.

GENERATE

ADD TO TEMPLATE

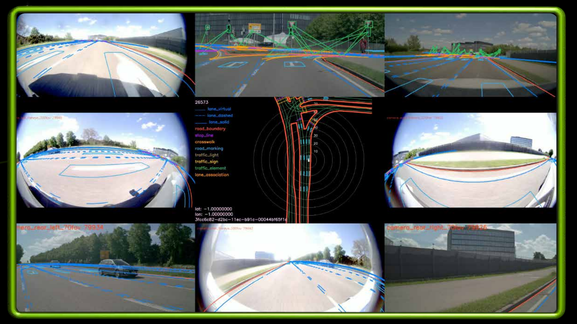
ADD TO CART

Apple Vision Pro

NVIDIA
Omniverse Cloud

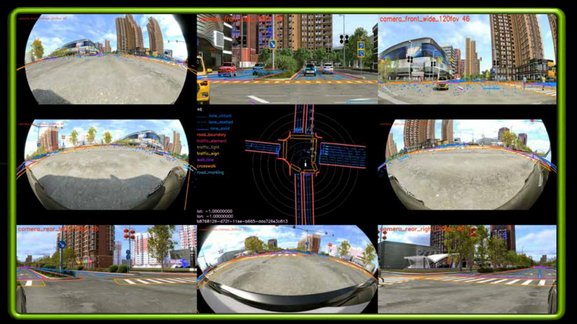
DGX

MSFNET
ROADNET
PLANNET



OVX

DRIVE SIM
IN
OMNIVERSE



AGX

DRIVE AV
STACK

DRIVE THOR
ASIL-D AV COMPUTER & STACK

BYD

NVIDIA Robotics Platform Adoption

1.3M

Robotics Developers

100,000 ROS Developers
6,000 Companies Developing on Orin

DGX

ISAAC PERCEPTOR

OVX

ISAAC SIM IN OMNIVERSE

AGX

ISAAC ROBOTIC STACK

JETSON ORIN
ROBOTIC COMPUTER & STACK

DGX

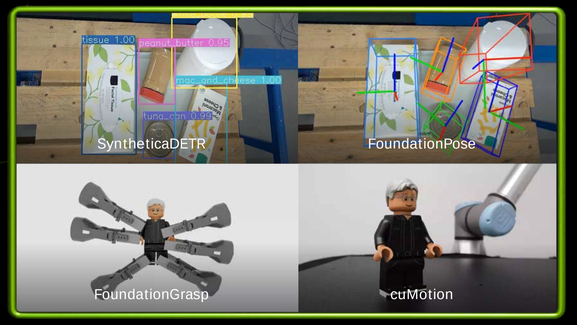
POSE
GRASP
MANIPULATE

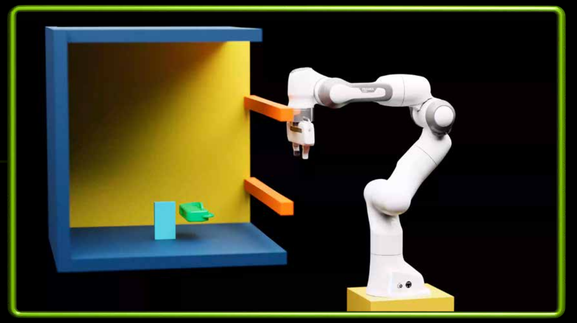
OVX

ISAAC SIM
IN
OMNIVERSE

AGX

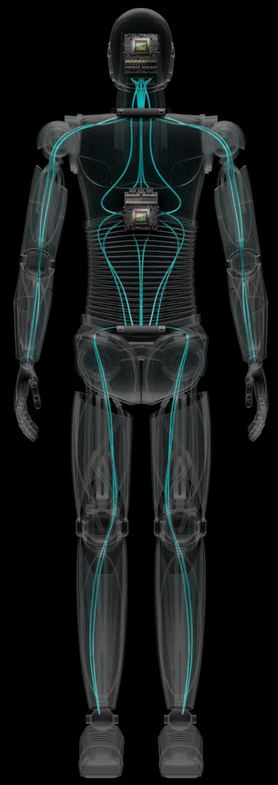
ISAAC
MANIPULATOR
STACK





JETSON ORIN
ROBOTIC COMPUTER & STACK

YASKAWA
UNIVERSAL ROBOTS
PICKNIK
SOLOMON
Vision with Intelligence
READY ROBOTICS
FRANKA ROBOTICS



DGX

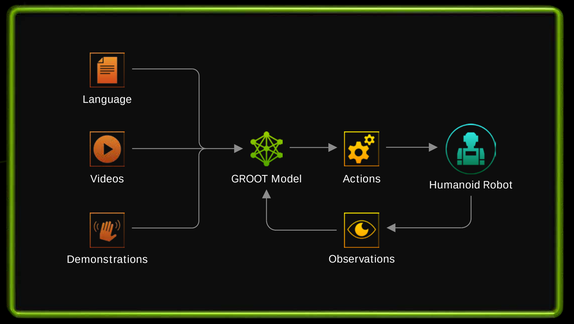
GR00T
FOUNDATION
MODEL

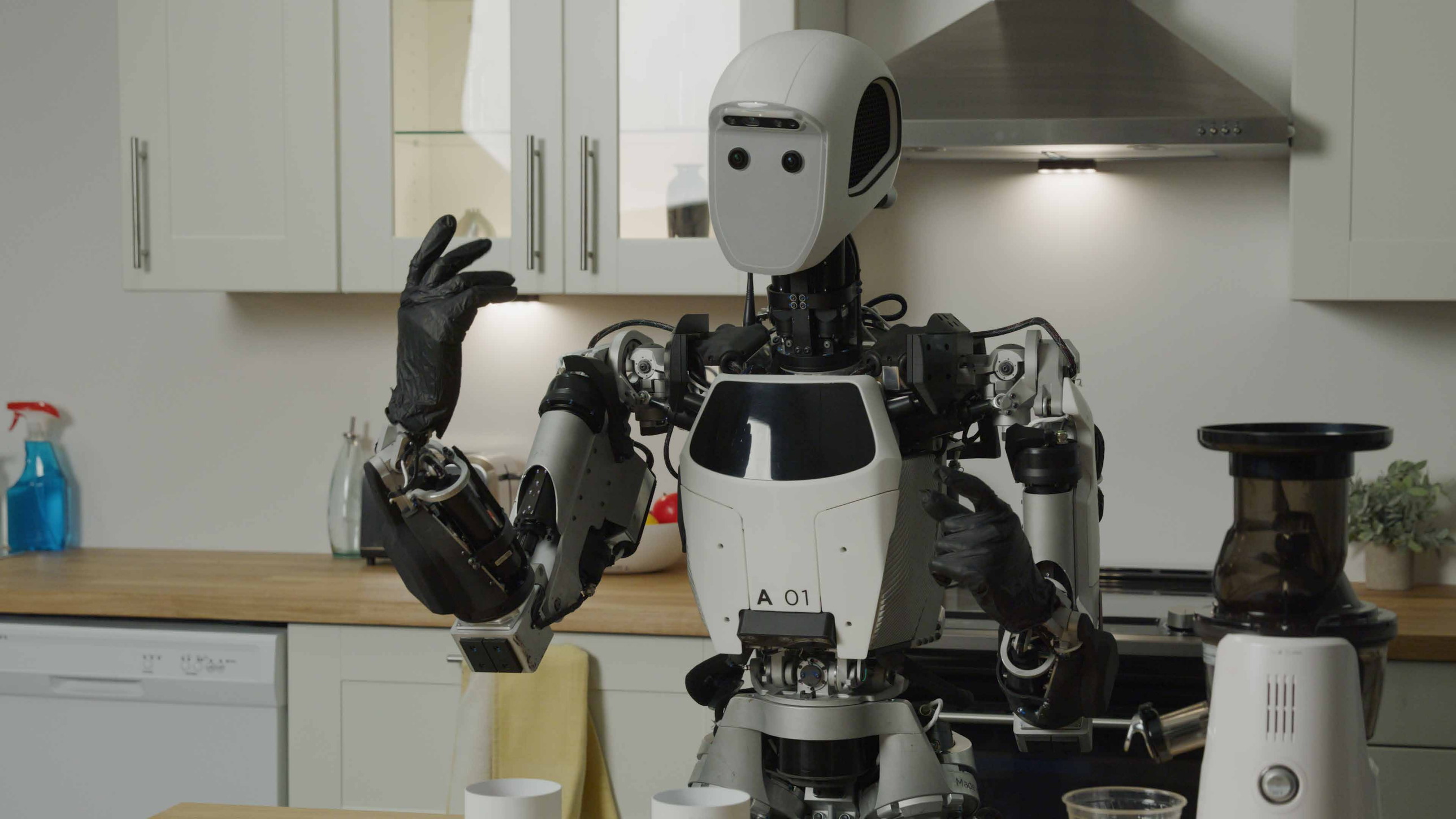
OVX

ISAAC LAB
IN
OMNIVERSE

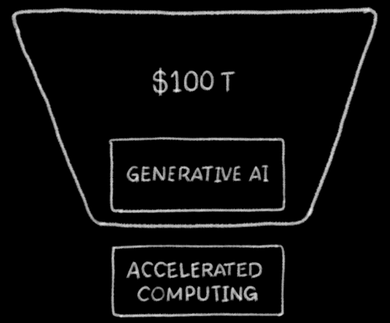
AGX

GR00T
STACK

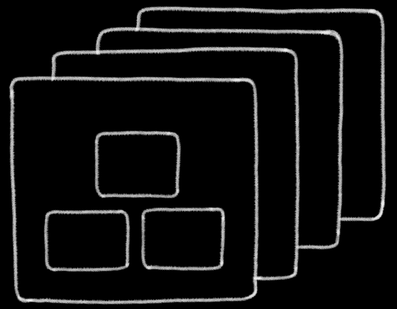



A 01

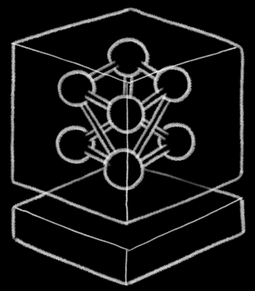
A NEW INDUSTRIAL REVOLUTION



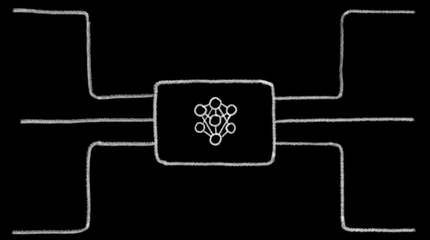
NEW INDUSTRY



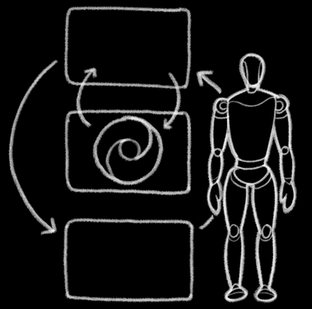
BLACKWELL PLATFORM



NIMs

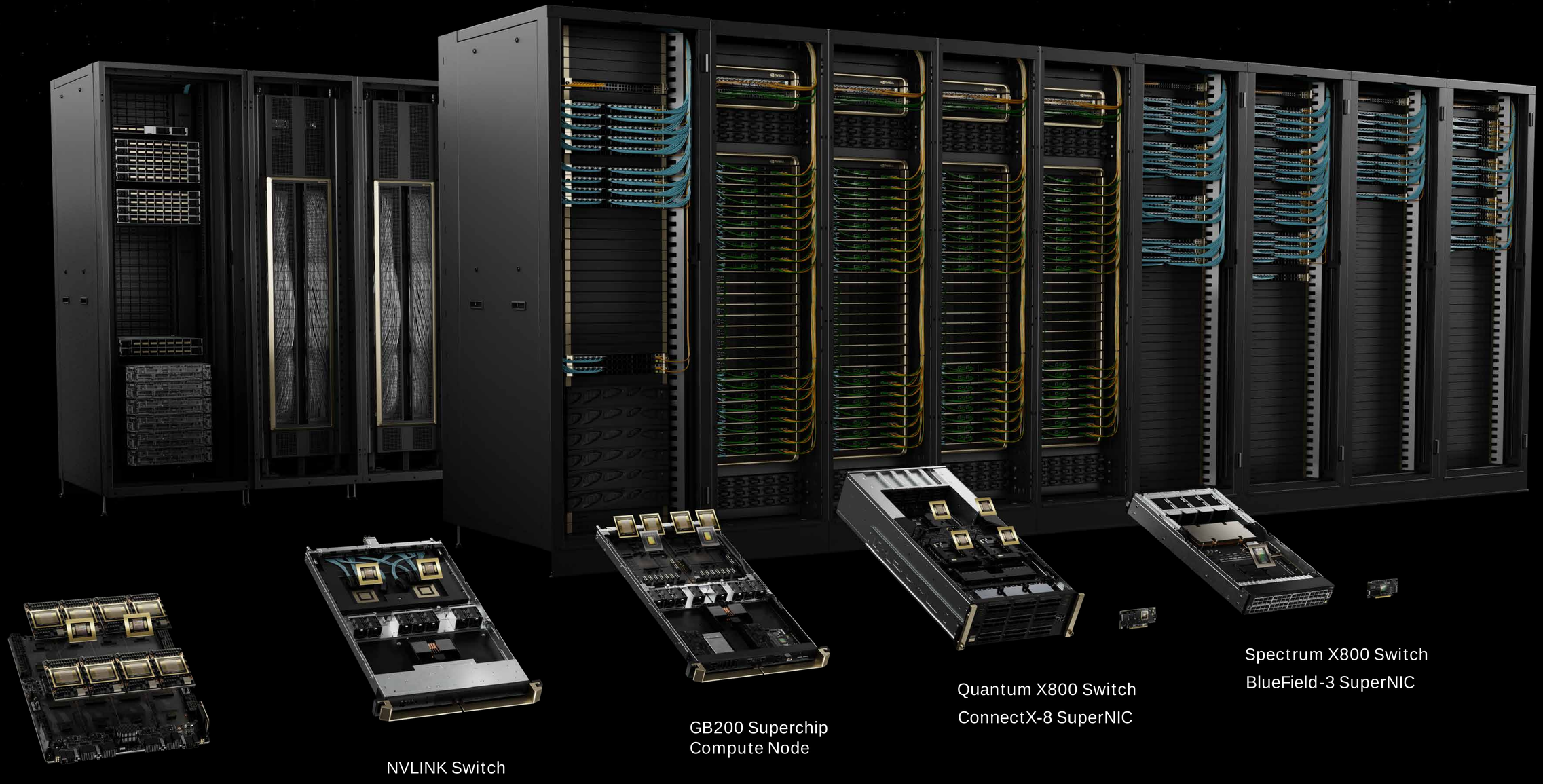


NEMO AND NVIDIA AI FOUNDRY



OMNIVERSE AND ISAAC ROBOTICS

NVIDIA Blackwell Platform



HGX B100

NVLINK Switch

GB200 Superchip
Compute Node

Quantum X800 Switch
ConnectX-8 SuperNIC

Spectrum X800 Switch
BlueField-3 SuperNIC


NVIDIA