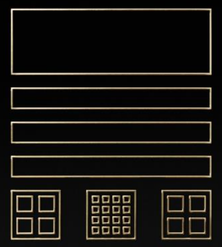
NVIDIA DGX A100/ NVIDIA DGX SUPERPOD最新情報

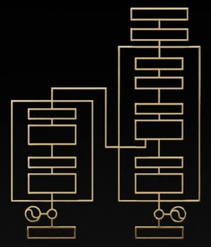
JLUG 2021, 2021/12/10 エヌビディア合同会社, シニアソリューションアーキテクト, 森野 慎也

A NEW COMPUTING LAW

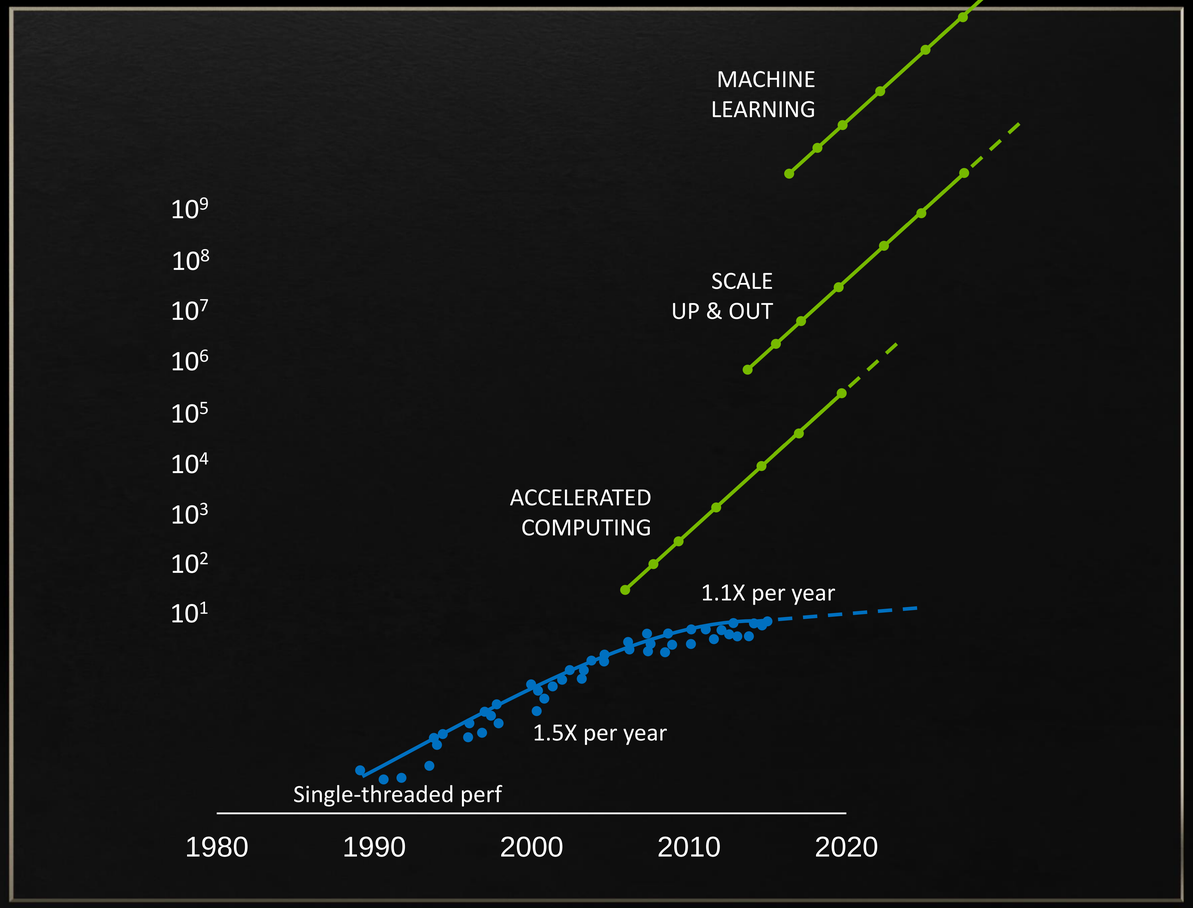
Accelerated Computing



AI



Data Center Scale



GPUDIRECT STORAGE GA AVAILABLE WITH CUDA 11.4

6.6X perf benefit in DL inference

Low latency, high throughput

Multi-node storage acceleration library

Supported by a strong ecosystem

Available on DGX

OS: DGX BASEOS 5.0, Ubuntu 18.04, 20.04, RHEL 8.3

DL frameworks: PyTorch, MXNet

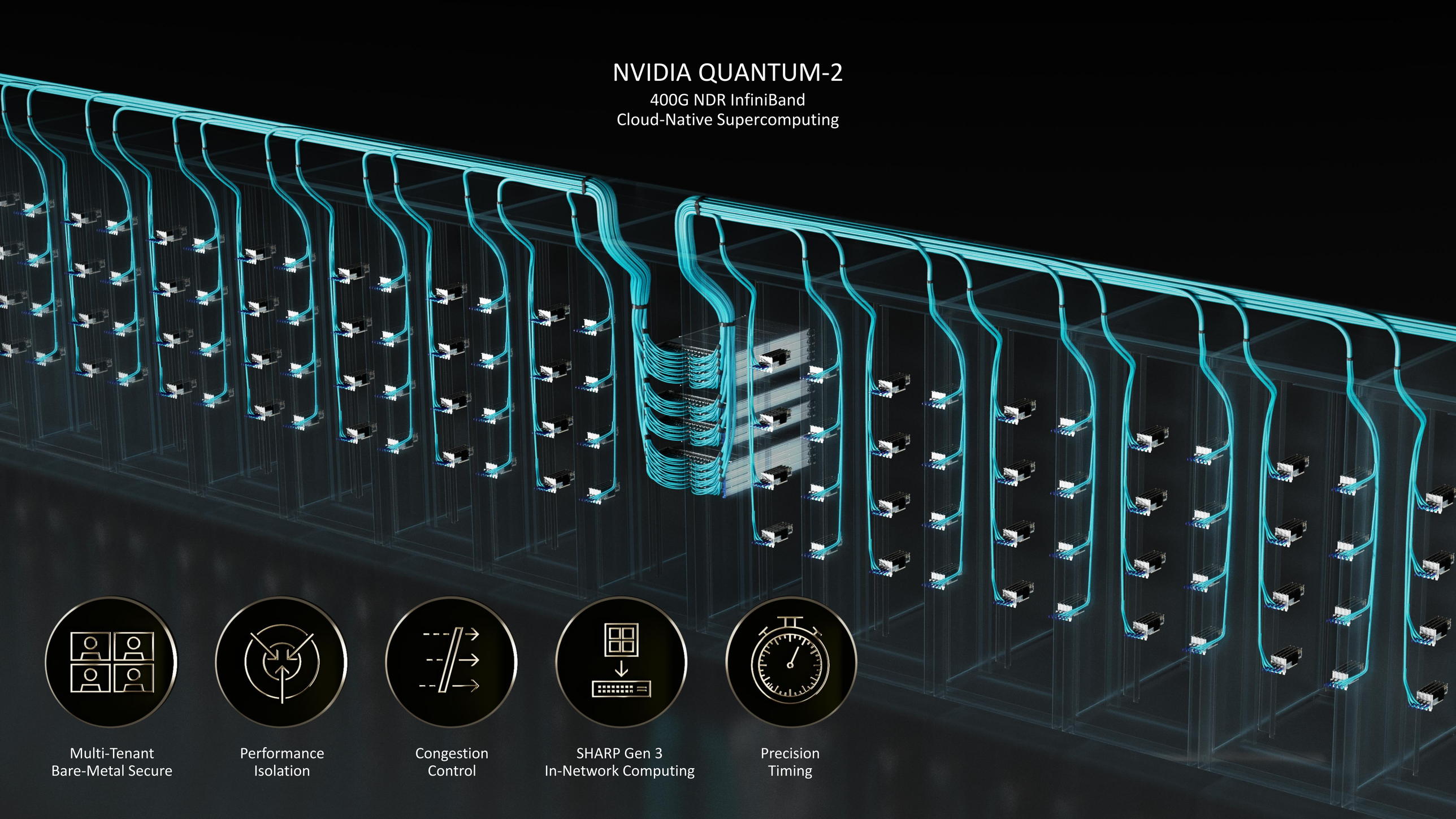
Data analytics frameworks: DALI, RAPIDS cuDF

CUDA 11.4 toolkit integration

Magnum IO GPUDirect Storage | NVIDIA Developer

NVIDIA QUANTUM-2

400G NDR InfiniBand
Cloud-Native Supercomputing



Multi-Tenant
Bare-Metal Secure

Performance
Isolation

Congestion
Control

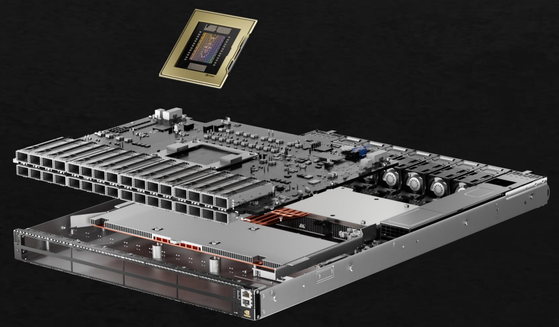
SHARP Gen 3
In-Network Computing

Precision
Timing

NVIDIA QUANTUM-2

400G NDR InfiniBand
Cloud-Native Supercomputing

QUANTUM-2 SWITCH



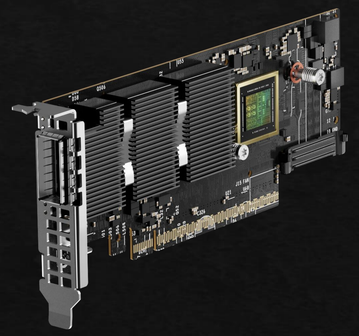
64-Ports of 400 Gbps or 128-Ports of 200 Gbps

32X More AI Acceleration Engines

3X Higher Switching Throughput

Sampling Now

CONNECTX-7 INFINIBAND

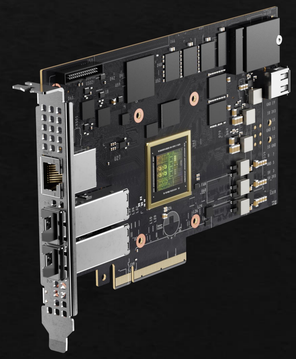


4X In-Network Computing Performance

2X GPUDirect Throughput

Sampling Jan ‘22

BLUEFIELD-3 INFINIBAND



16 Arm 64-Bit Cores | 400 Gbps Crypto Accelerations

4X In-Network Computing Performance

2X GPUDirect Throughput

Sampling May ‘22

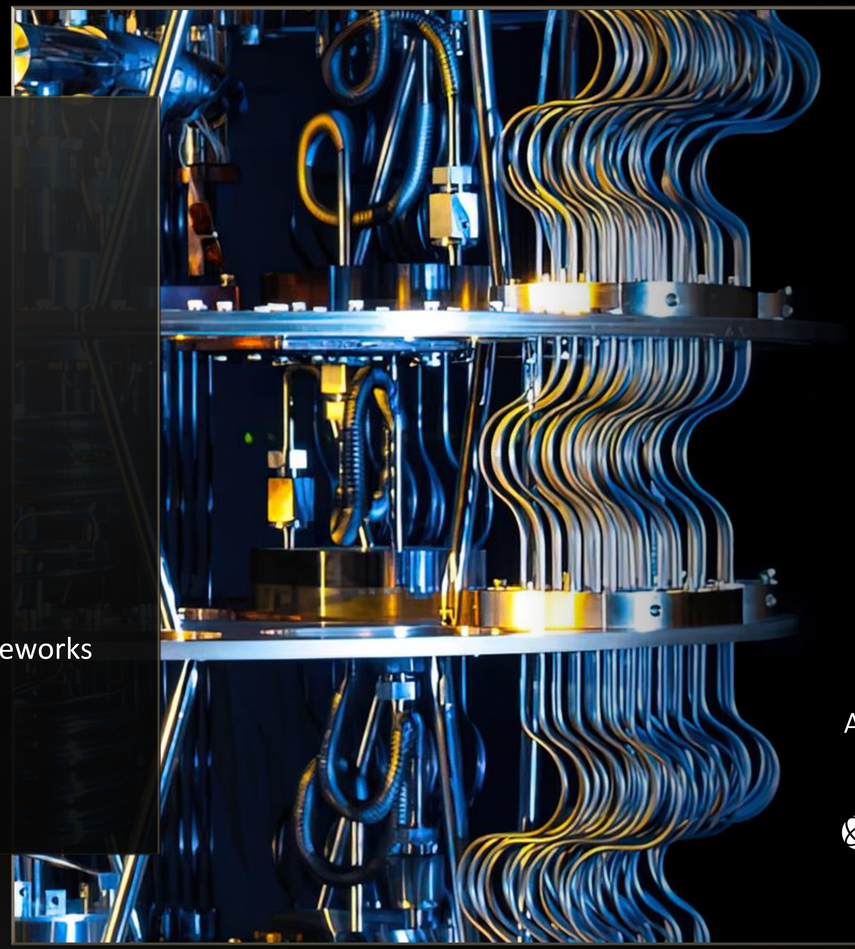
NVIDIA cuQUANTUM BETA

Research the Computer of Tomorrow on the Most Powerful Computer Today

Libraries for State Vector and Tensor Network Simulation

World-Class Performance on Multiple Quantum Algorithms

Integrated In Leading Quantum Computing Frameworks from Google and IBM



LEADING QUANTUM SIMULATORS

Qiskit

Quantum AI

INDUSTRY PARTNERS

IONQ

PASQAL

PsiQuantum

RESEARCH COMMUNITY PARTNERS

Argonne NATIONAL LABORATORY

BERKELEY LAB Lawrence Berkeley National Laboratory

Caltech

JÜLICH Forschungszentrum

Los Alamos NATIONAL LABORATORY

MIT

OAK RIDGE National Laboratory

Pacific Northwest NATIONAL LABORATORY

cuQuantum Available Now for Download
developer.nvidia.com/cuquantum

NVIDIA cuNUMERIC

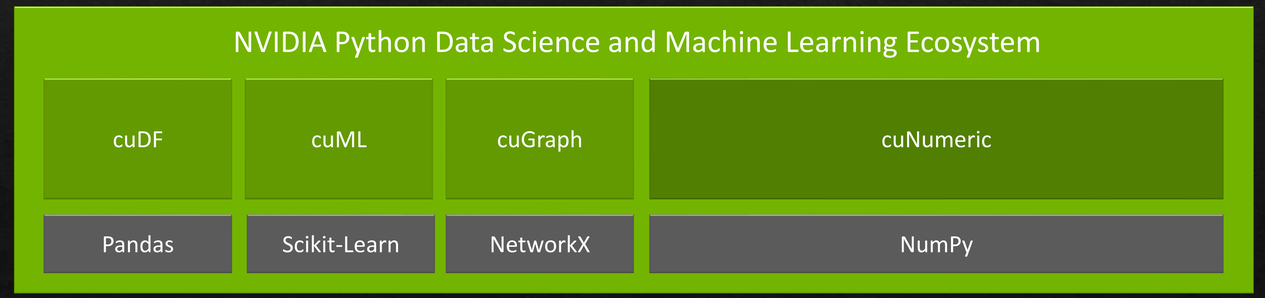
Accelerated Computing At-Scale
for PyData and NumPy Ecosystem

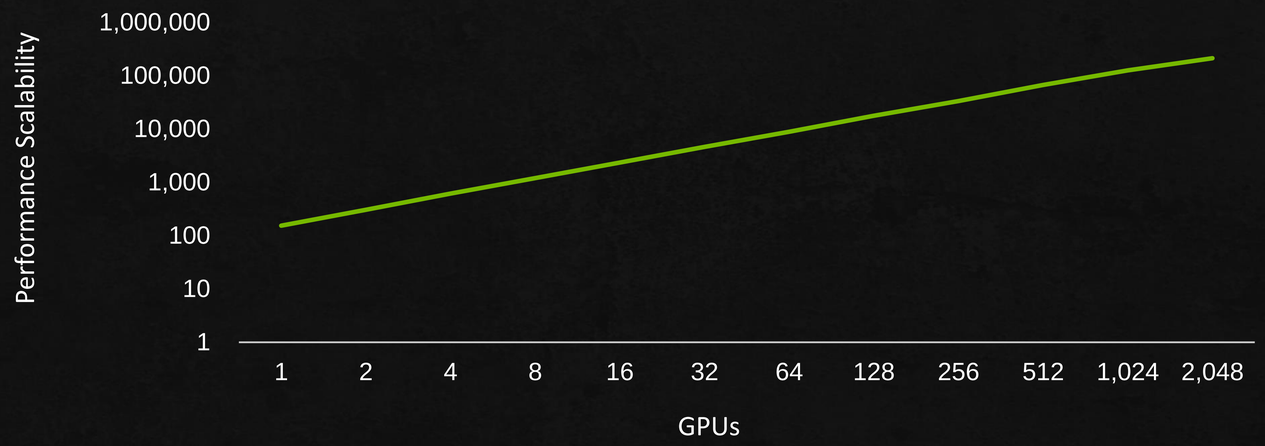
Transparently Accelerates and Scales NumPy Workflows

Zero Code Changes

Automatic Parallelism and Acceleration for Multi-GPU, Multi-Node Systems

Scales to 1,000s of GPUs





Available Now on GitHub and Conda

cuNumeric Library Download | NVIDIA Developer

NVIDIA MODULUS

Physics-ML Neural Simulation Framework

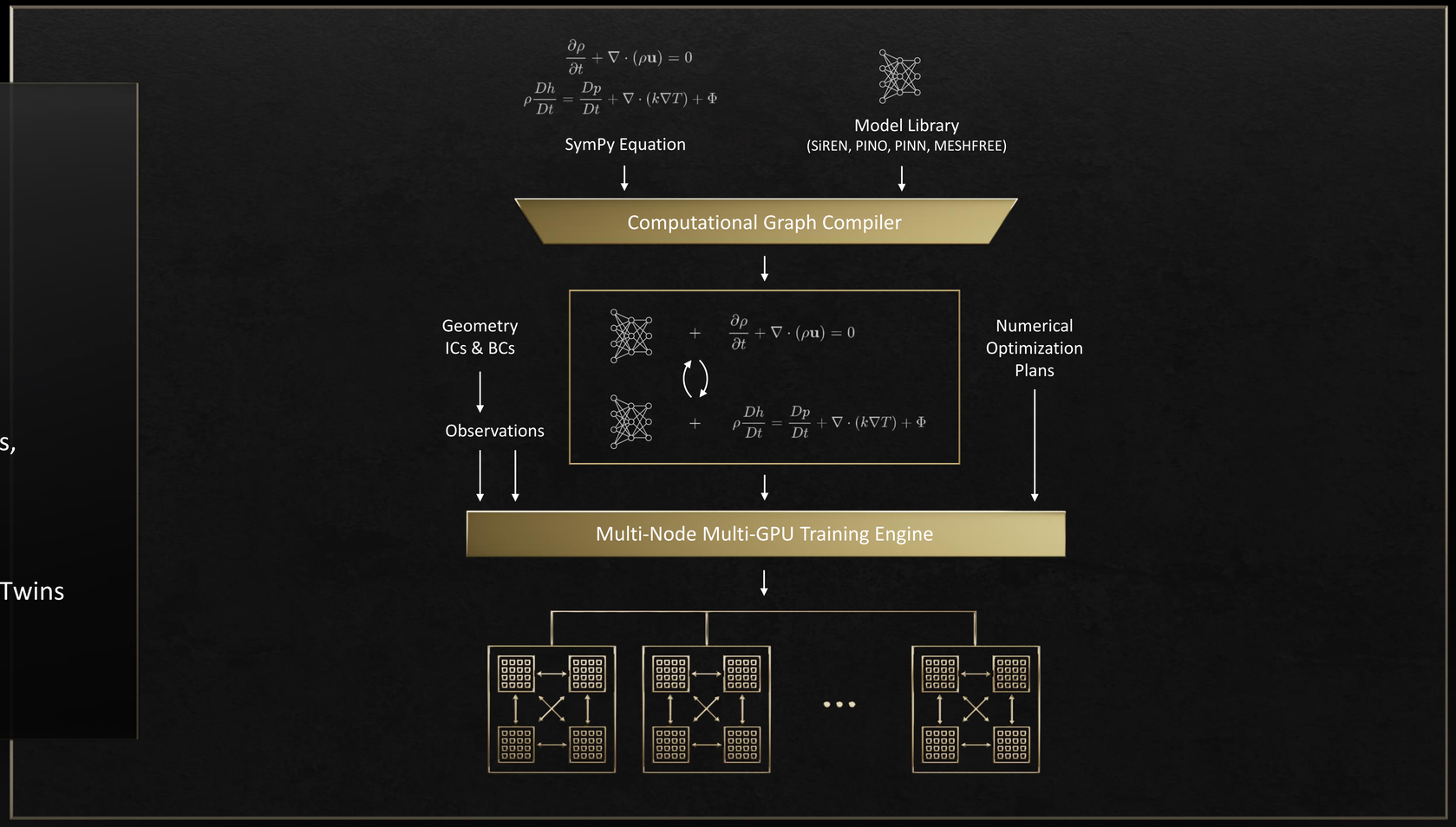
Framework for Developing Physics-ML Models

Train Physics-ML Models Using Governing Physics, Simulation, and Observed Data

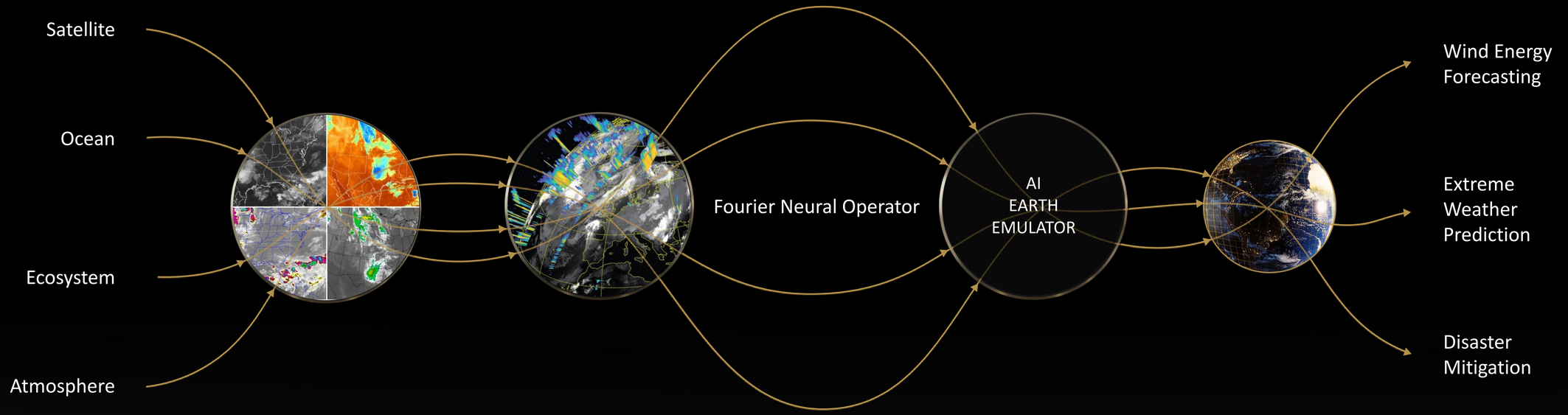
Multi-GPU, Multi-Node Training

1,000-100,000X Speed Models – Ideal for Digital Twins

Available Now
developer.nvidia.com/modulus



EARTH DIGITAL TWIN IN OMNIVERSE



ERA5 ECMWF
Atmospheric Winds & Geopotential
10 TB | 30km | 5 Atmos Layers

100,000X Speed-Up
0.25 Seconds for 7-Day Forecast
Training: 4 Hours on 128 A100 GPUs

RAPIDS — Modulus — Omniverse

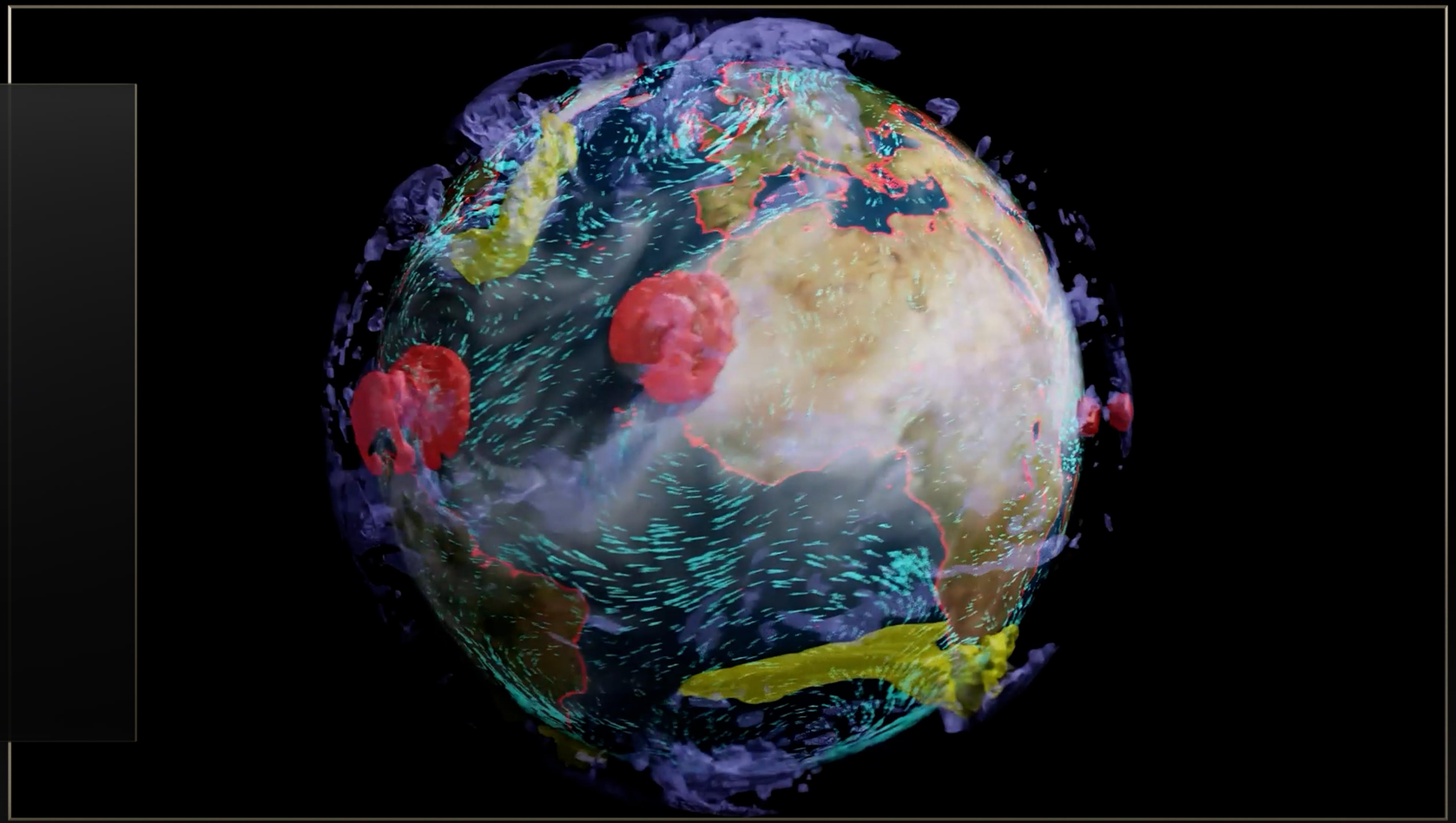
EARTH-2 SUPERCOMPUTER

1 Million X Acceleration for Climate Research

World’s largest digital twin

Running Modulus-created AI physics models

Designed to run Omniverse



More details in 2022


NVIDIA