



E2 SERVICE ARCHITECTURE

PETER MESSMER, DIRECTOR HPC DEVTECH EMEA

PMESSMER@NVIDIA.COM

ECWMF HPC WORKSHOP

BOLOGNA, OCT 10, 2023

EVE NEEDS 3 MIRACLES

MIRACLE #1

2.5KM Resolution
30,000 SYPY
30 MW

TODAY

40 SYPY on 2,140 CPU at 1 MW
O(750 MW) for 30,000 SYPY

MIRACLE #2

Full State Vector Interactivity
Any Region
Any Time Period

TODAY

Exabytes of Data to Store

MIRACLE #3

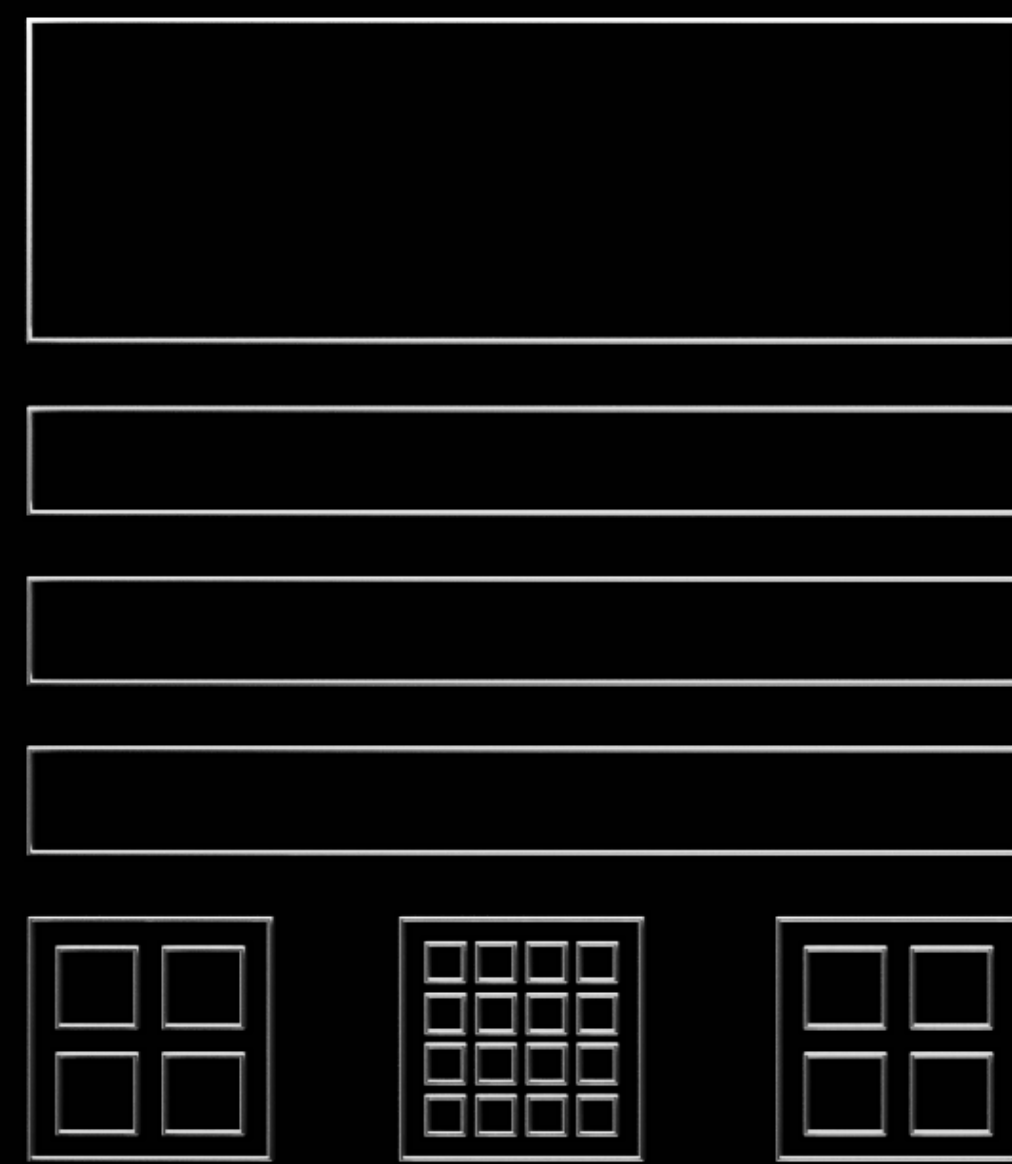
Full State Vector Visualization
From Cloud

TODAY

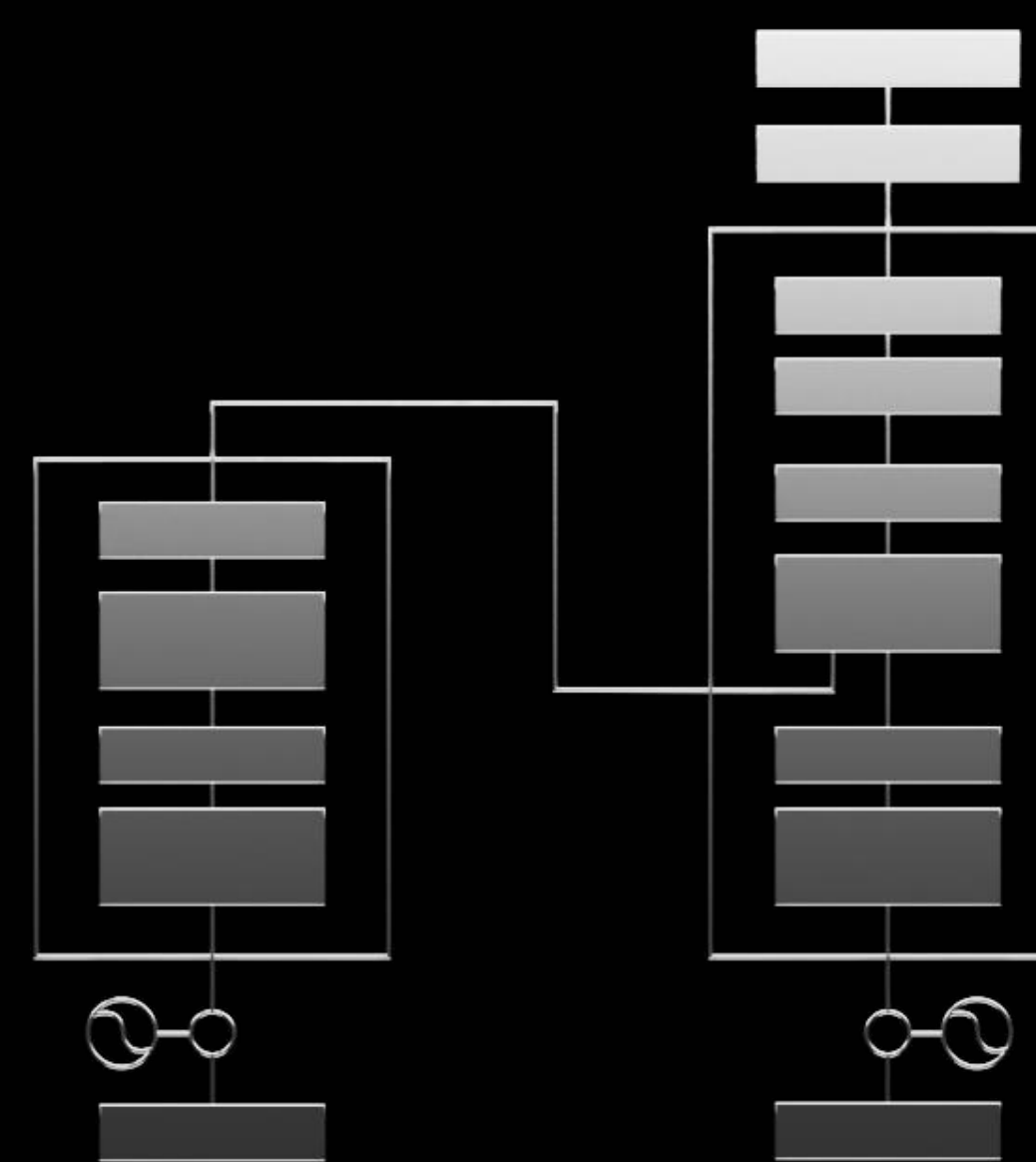
Top-of-Line NVIDIA Visualization
Supercomputer

EARTH-2 — 3 MIRACLES

**NVIDIA
ACCELERATED COMPUTING**



**MODULUS PHYSICS-ML
FOURCASTNET**



**OMNIVERSE
DIGITAL TWINS**



EARTH-2

Earth Climate Digital Twin

This scales (to impact sectors) ... and has plenty of room for AI inside

EU + Africa Japan, UAE, India UK, ZA, AU, Singapore NIS America China

... but the base and mid layers require renewed scientific ambition and a public infrastructure, the need for which we hope to articulate at the Berlin Summit (those interested in contributing are welcome to join)



This simulation was generated from historical data and enables real time visualization of global weather patterns over a 24 hour day. It is accurate enough to reproduce the famed Blue Marble image of the Earth.

Moving forward, we'll be able to scrub global weather patterns different points in time from the past and into the future.

Data source: Max Planck Institute for Meteorology

1:12 PM December 7, 1972

1960 2100

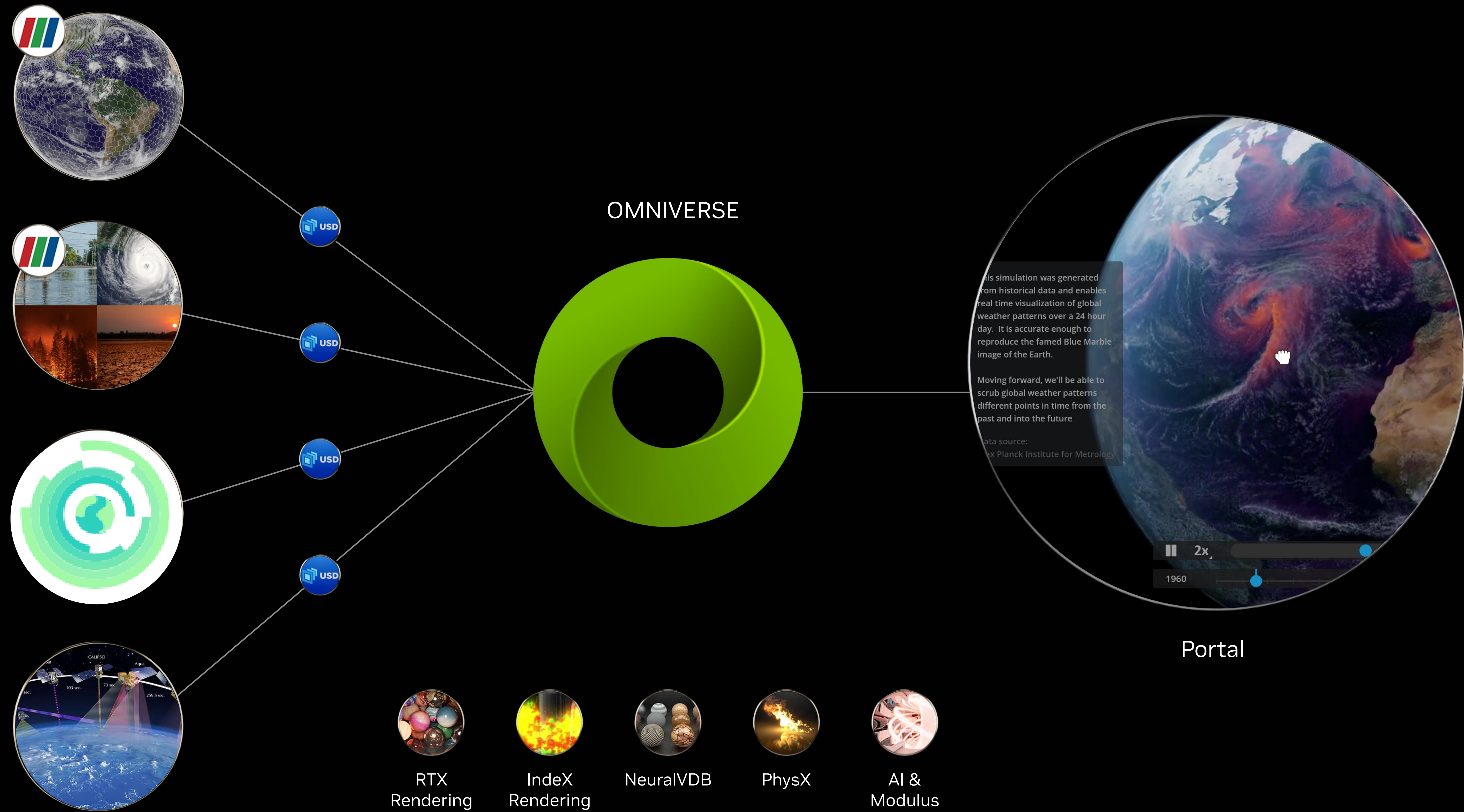
DATA SOURCES / INITIAL CONDITIONS



...

NVIDIA OMNIVERSE FOR INTERACTIVE DIGITAL TWINS

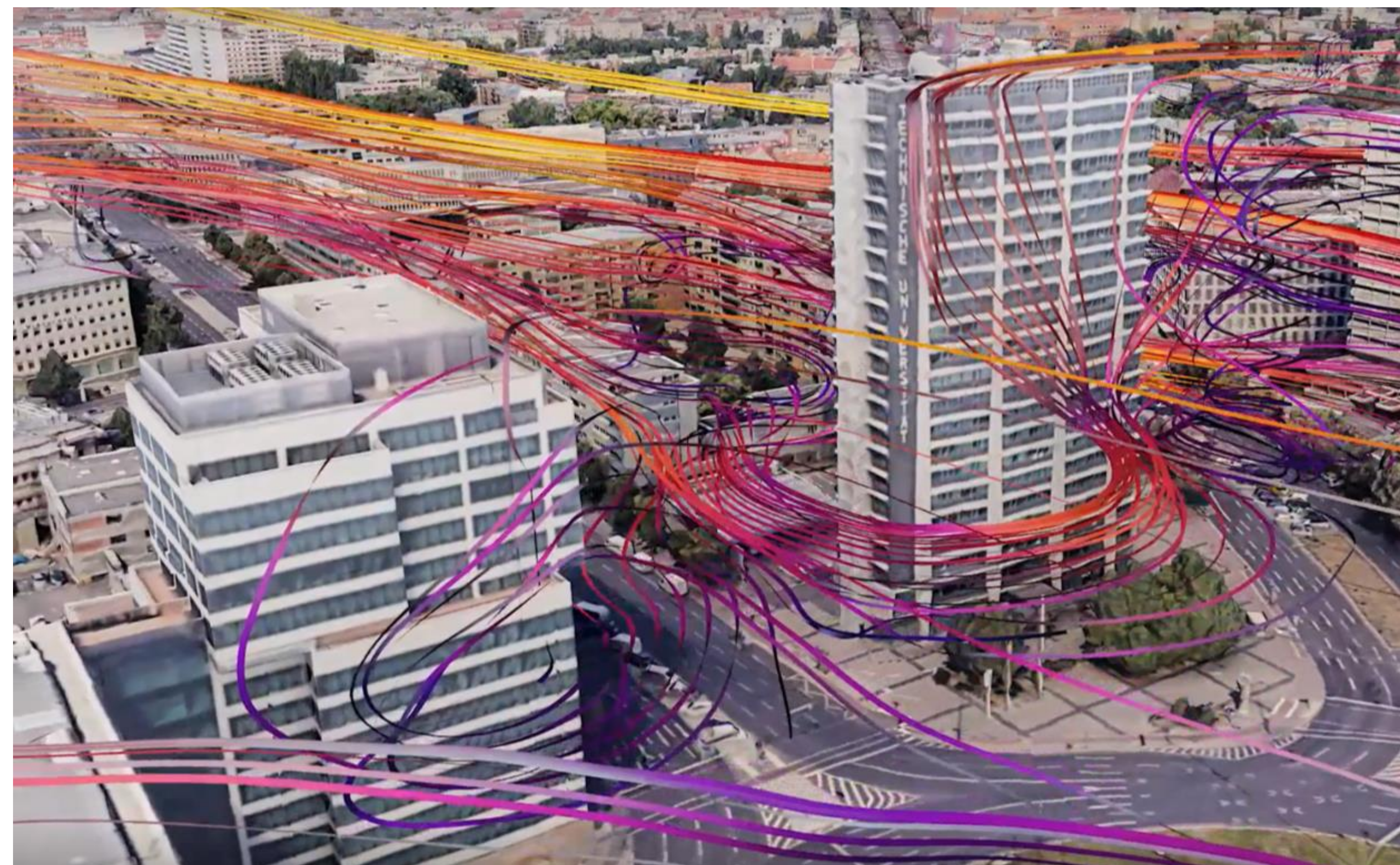
Connecting 3D, Simulation, and Observation Data



INTERACTIVE EXPLORTION OF FUSED MULTI-SOURCE DATA



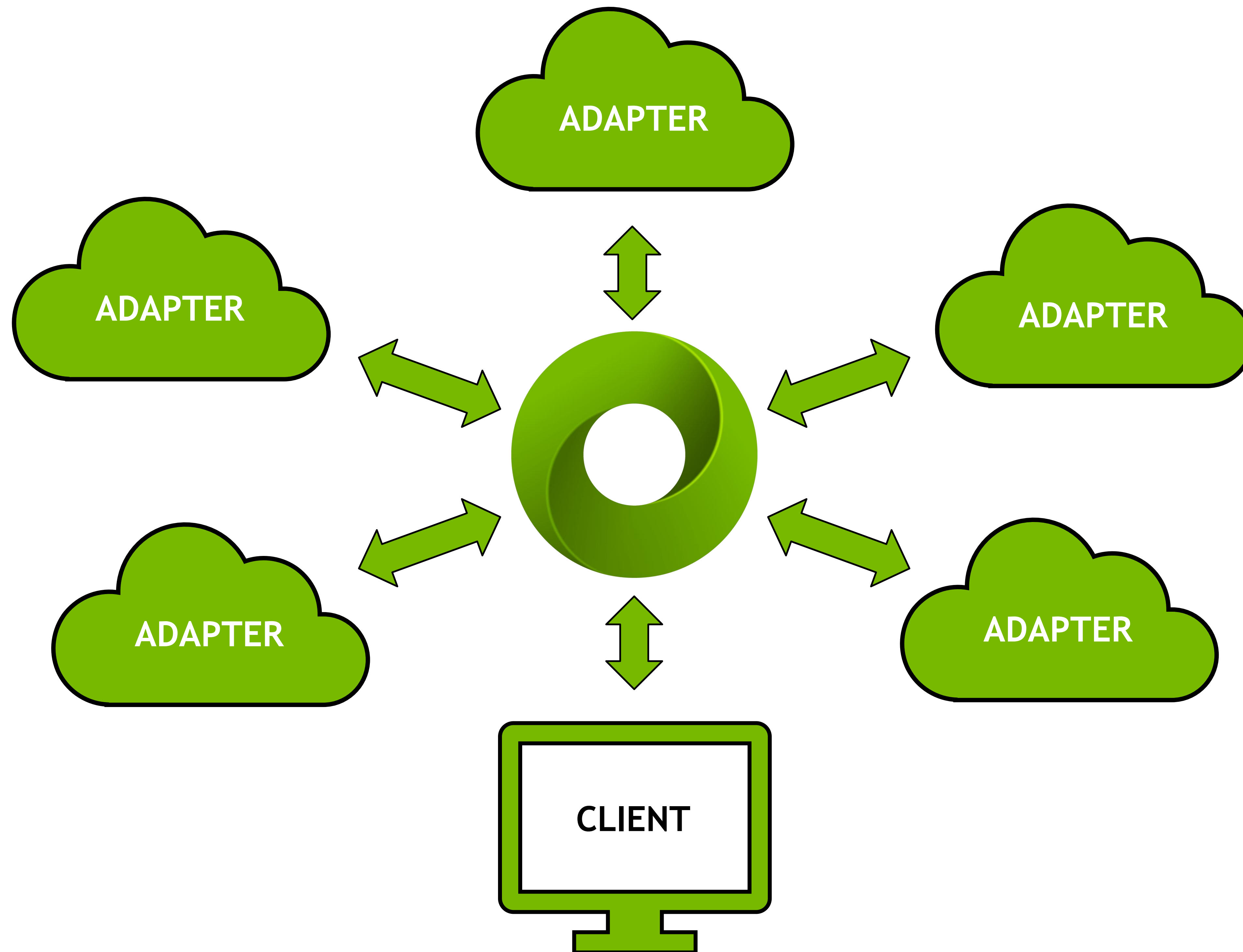
Global View



Local View

DATA FEDERATION - PROVIDE DATA PRODUCTS AT VARIOUS TIMESCALES

ARCHITECTURE



- Spoke and hub of services (client, adapters) connected to Nucleus
 - Scalable message broker via active database
- Client issues data requests
 - Omniverse Client Library (C++ or Python)
 - *Broadcast* JSON formatted messages
- Client/Adapter messaging protocol
 - “Query/Solution”: What is available?
 - “Order/Fulfillment”: Return specific data
- Client:
 - Can operate asynchronously in a Jupyter Notebook or a Python app
- Adapters:
 - Light-weight Python applications
 - Can run anywhere with access to a Nucleus instance
 - Translate data request into external API vocabulary (e.g. CDS)

DATA FEDERATION

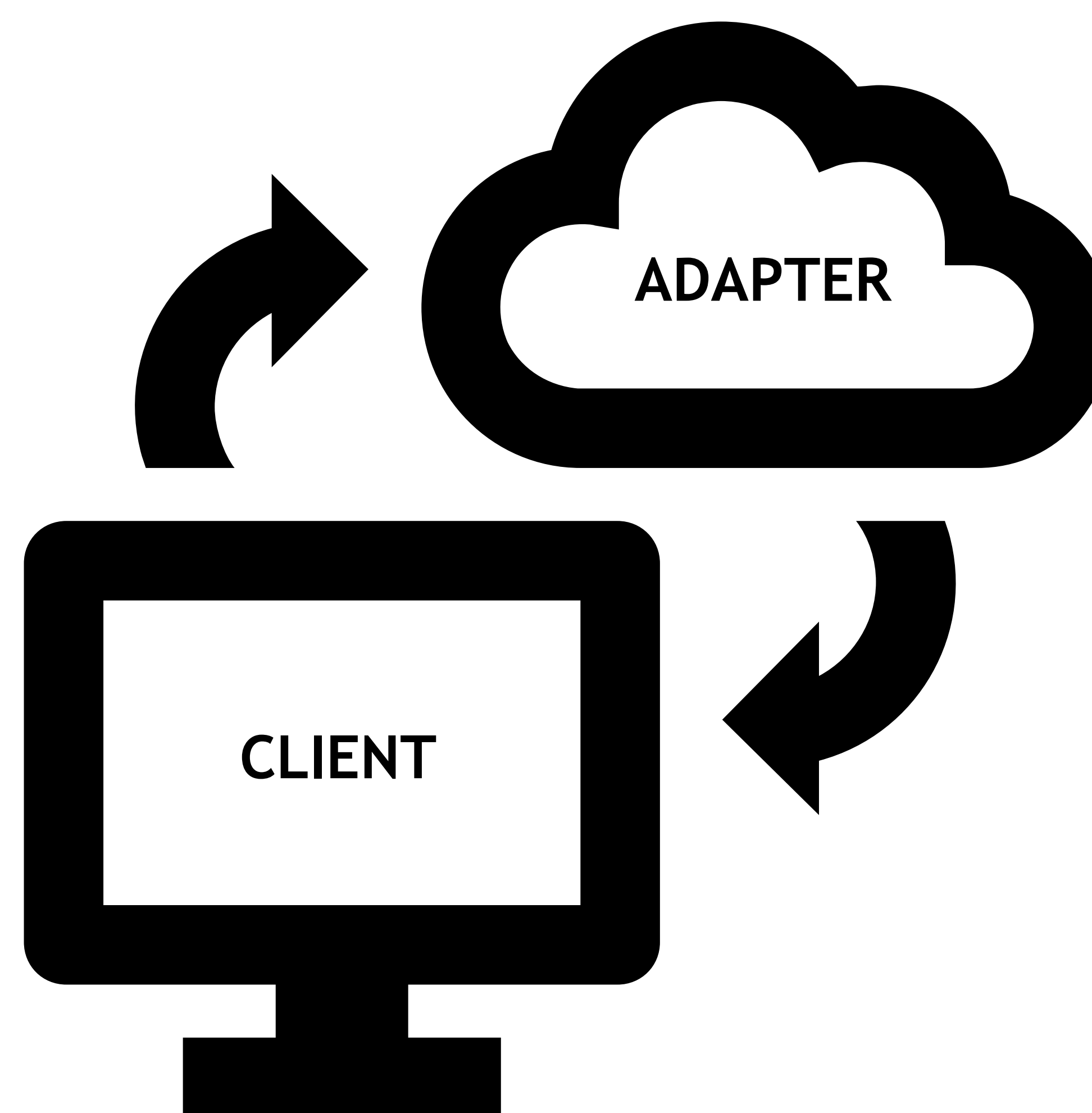
VOCABULARY

QUERY: *"What is available?"*

- Can be empty
 - *"What is available?"*
- Can include specific variables
 - *"Is X available?"*
- Can include specific slices
 - *"What is available here/now?"*
- Can include specific data types
 - *"What textures are available?"*
- *Or any combination of the above*

ORDER: *"Where is this?"*

- Constructed from a specific SOLUTION
- Targeted at a specific adapter
- Contains a complete description of the data being requested



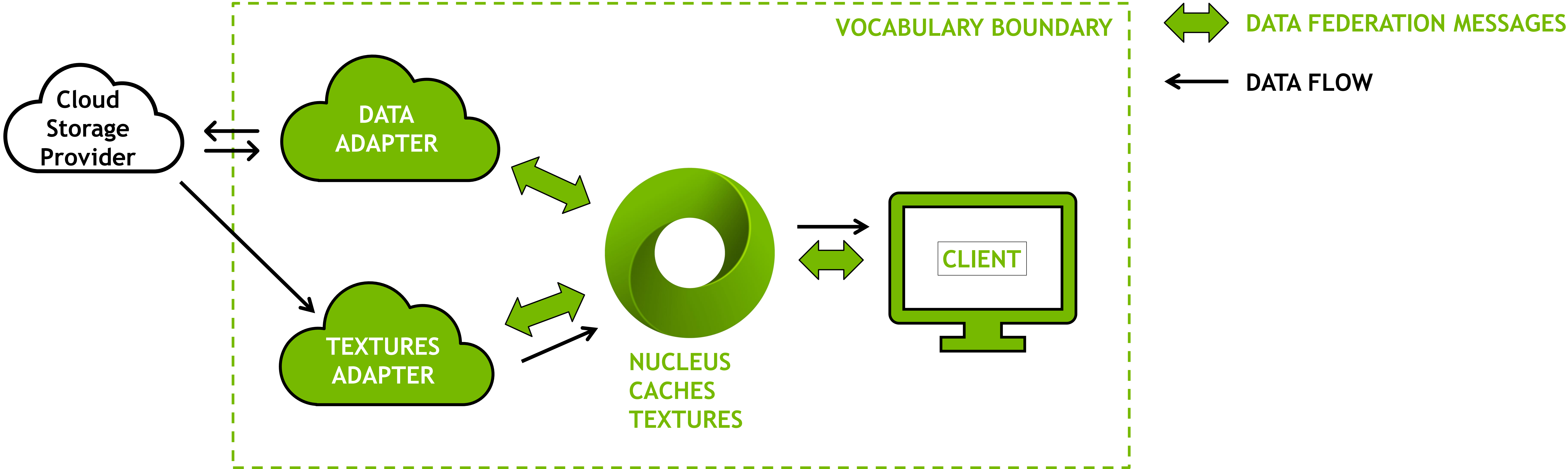
SOLUTION: *"This is available!"*

- One per adapter
- Provides a *complete* description of the data provided by the adapter
 - All coordinate information, including grid information, time coordinates, etc.
- Used by clients to construct ORDERS

FULFILLMENT: *"This is where it is!"*

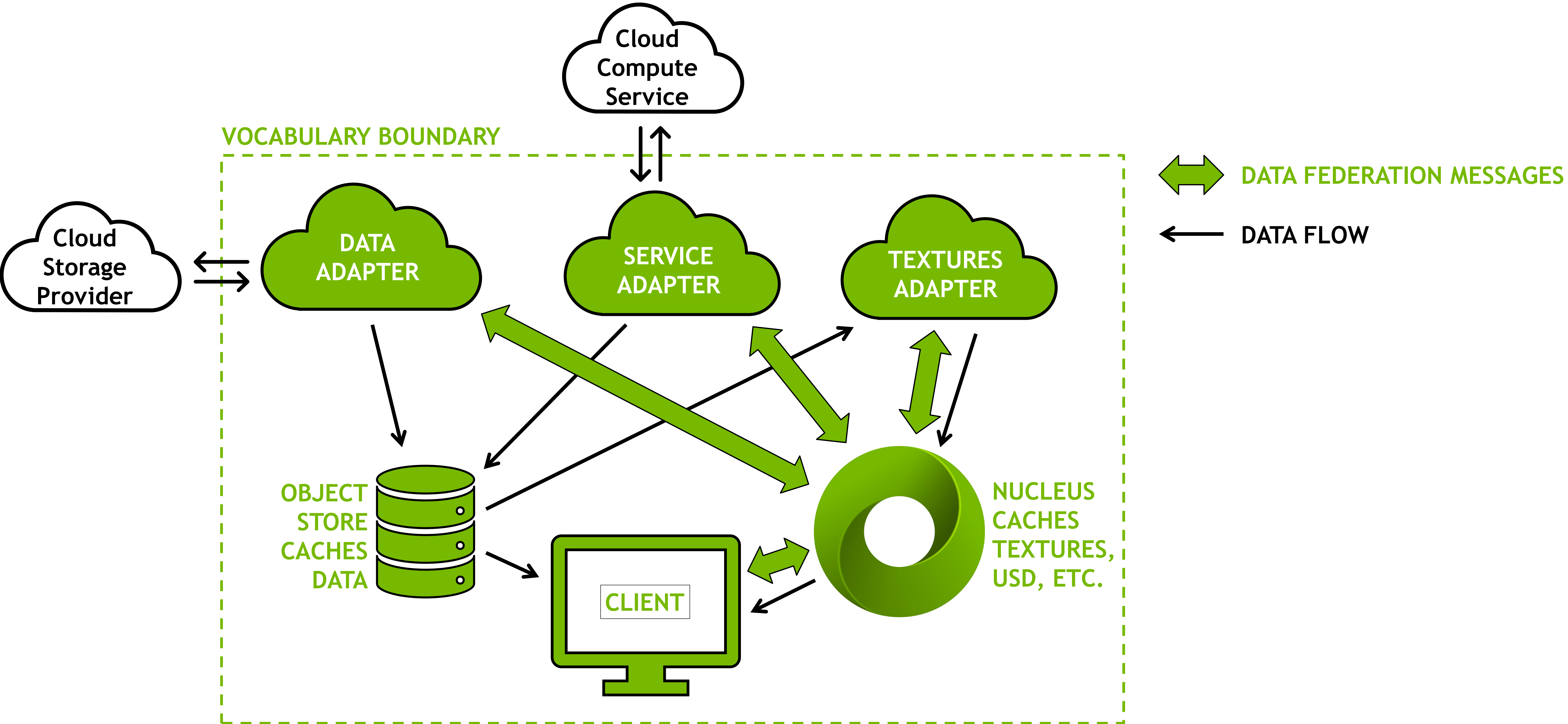
- Can have multiple for a single order
 - i.e., data can be streamed
- Provides URLs and access parameters to the data requested in the ORDER
 - URLs can point to remote data
 - Or data can be cached locally for fast access, and URLs are "local"
- Actual data is pulled from the client (not pushed from adapter)

DATA FEDERATION: BASIC EXAMPLE



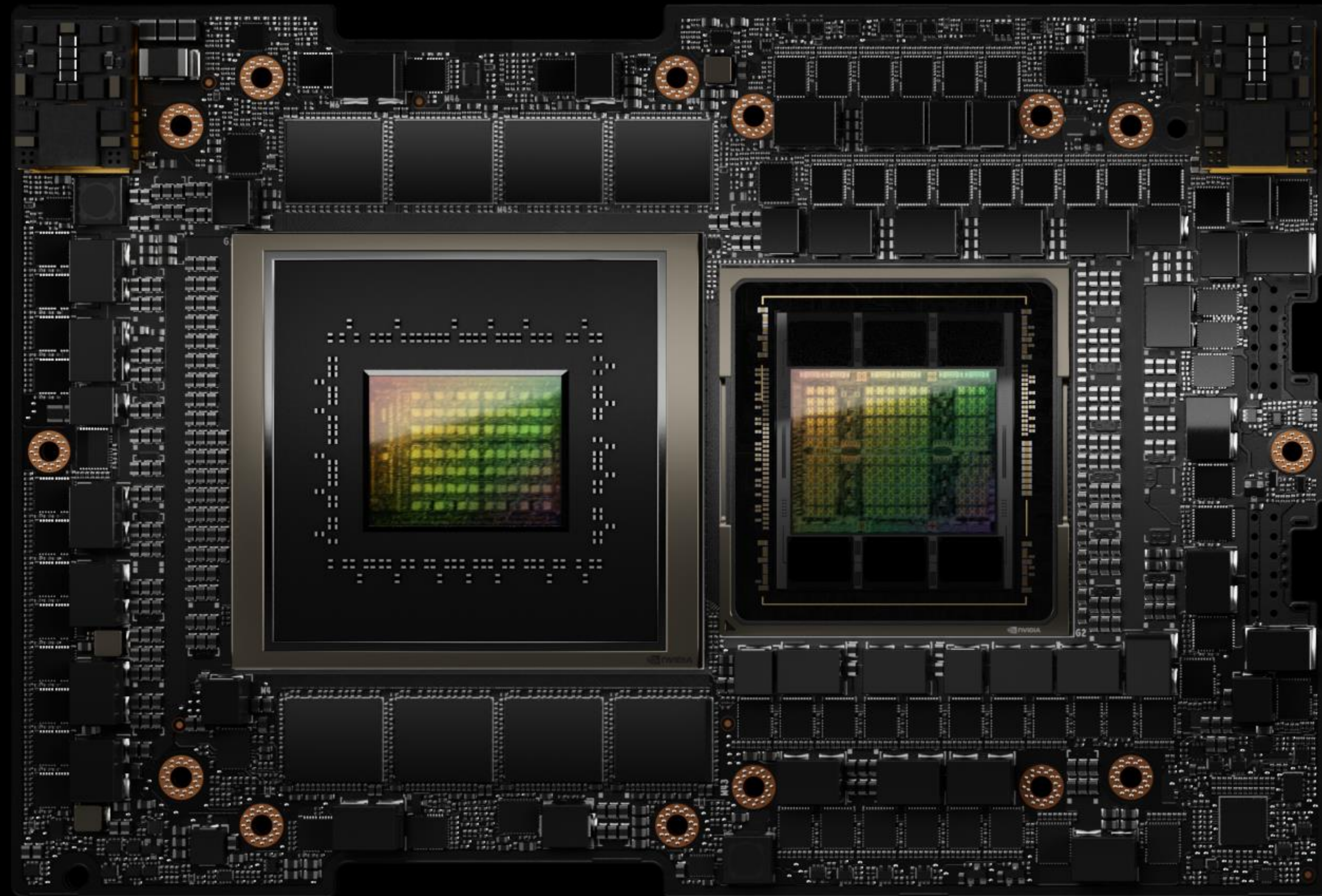
DATA FEDERATION

EXAMPLE ARCHITECTURE



NVIDIA GH200

The Most Advanced Accelerated Computing Platform



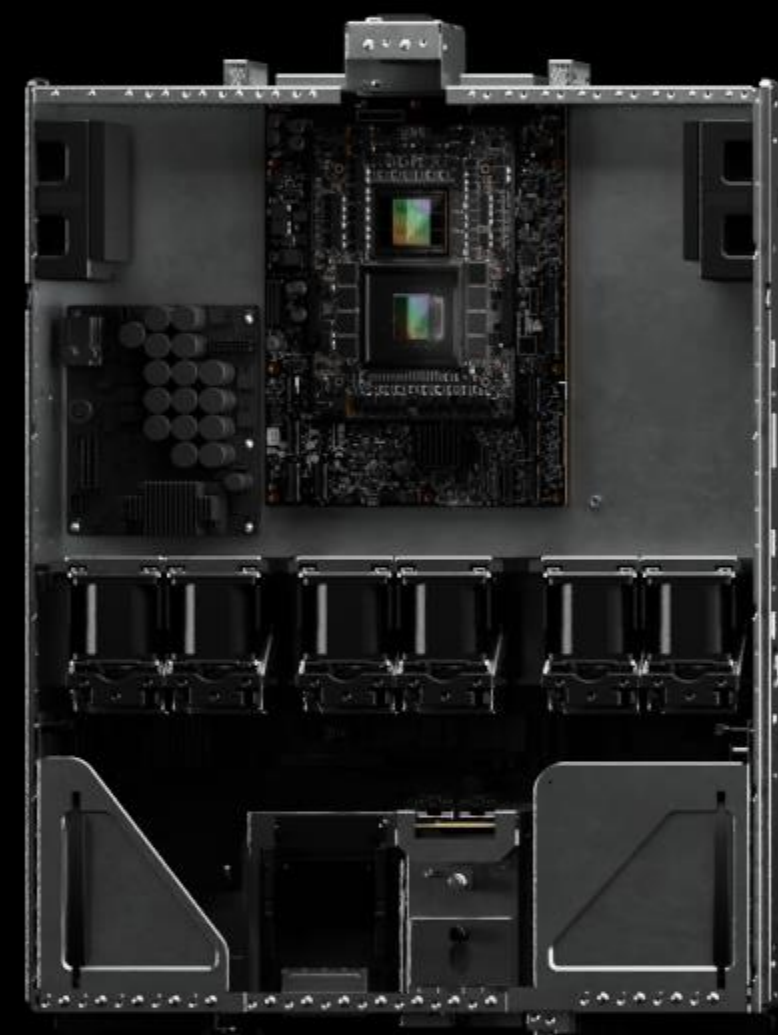
Tightly Coupled CPU and GPU with 900 GB/s NVLink-C2C

Optimized for Memory-Intensive, Accelerated Workloads

Co-Scheduling of CPU and GPU for Maximum Flexibility

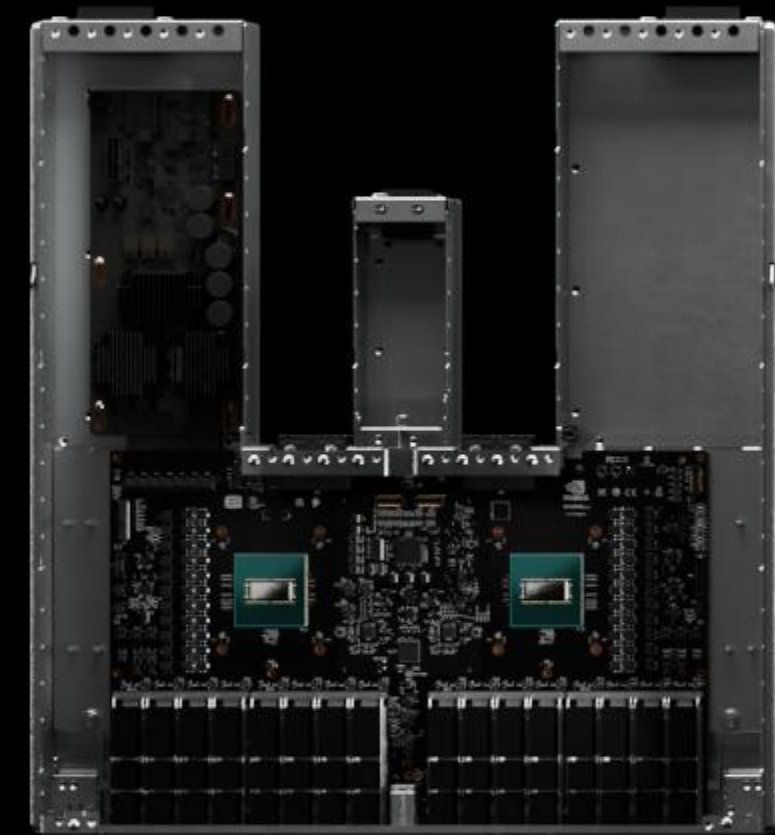
Scale with NVLink Switch System

4 PetaFLOPS TE | 72 Arm Core CPU | 96GB HBM3 | 576GB GPU Memory



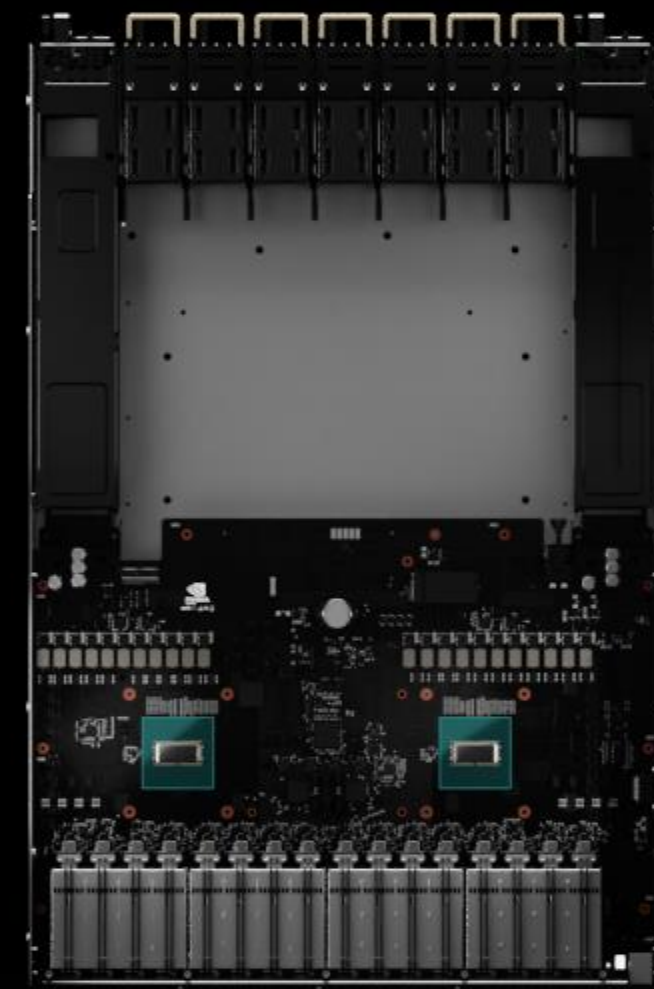
Grace Hopper Superchip

X 8



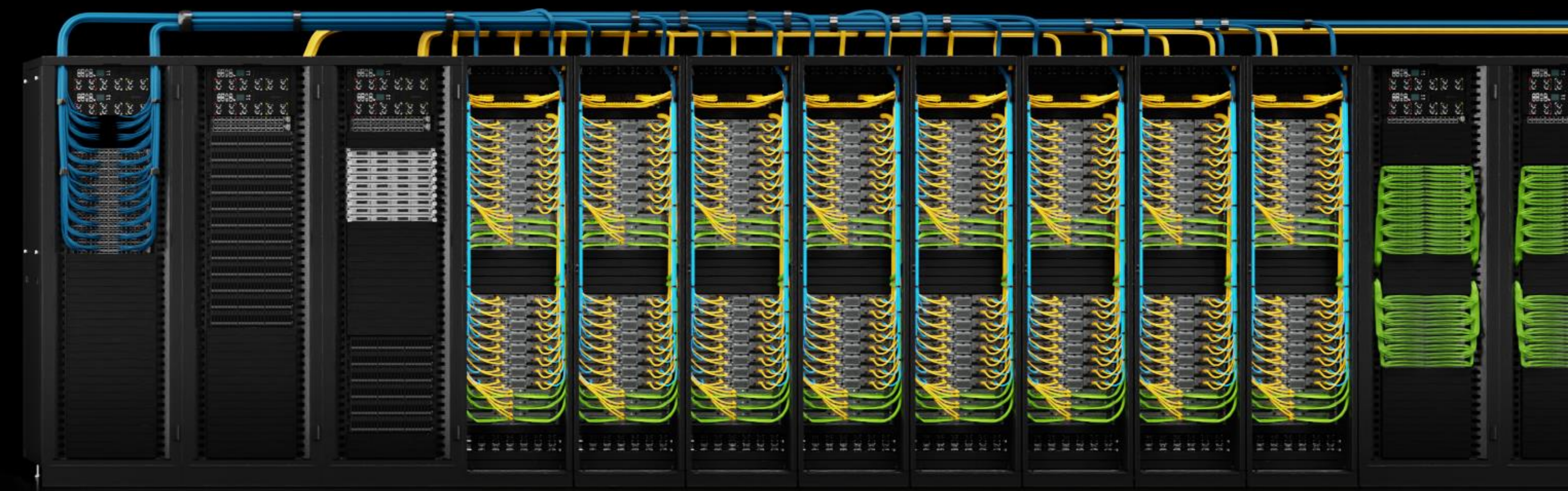
NVLINK Switch x3

X 32



NVLINK Switch x36

=



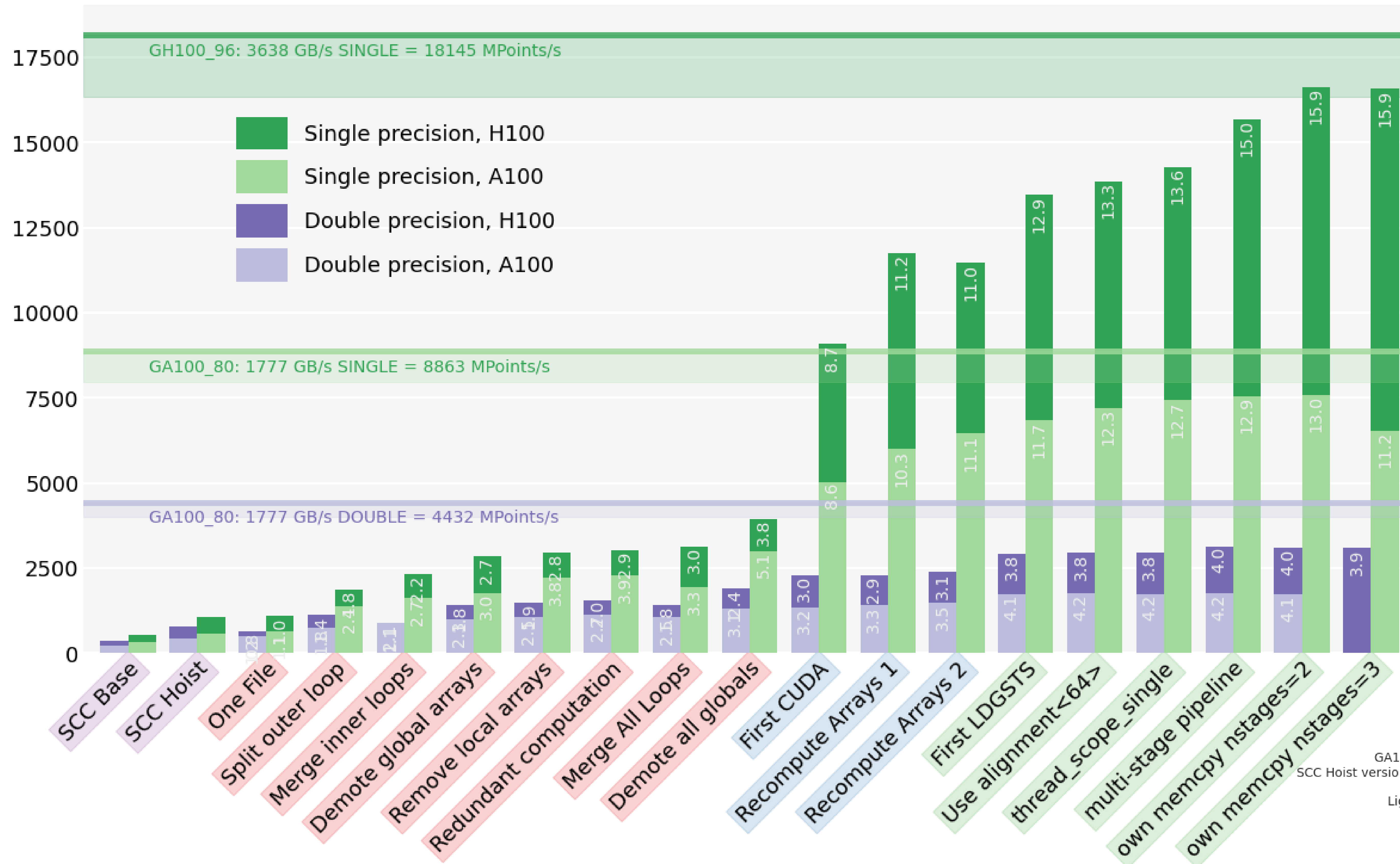
256 Grace Hopper Superchips | 1 ExaFLOPS Transformer Engine
144 TB GPU Memory

IGNORING DIVERSITY HURTS PERFORMANCE TREMENDOUSLY

Leveraging novel hardware features for speed-of-light performance

Perf [Mpoints/s]

Performance Evolution CloudSC

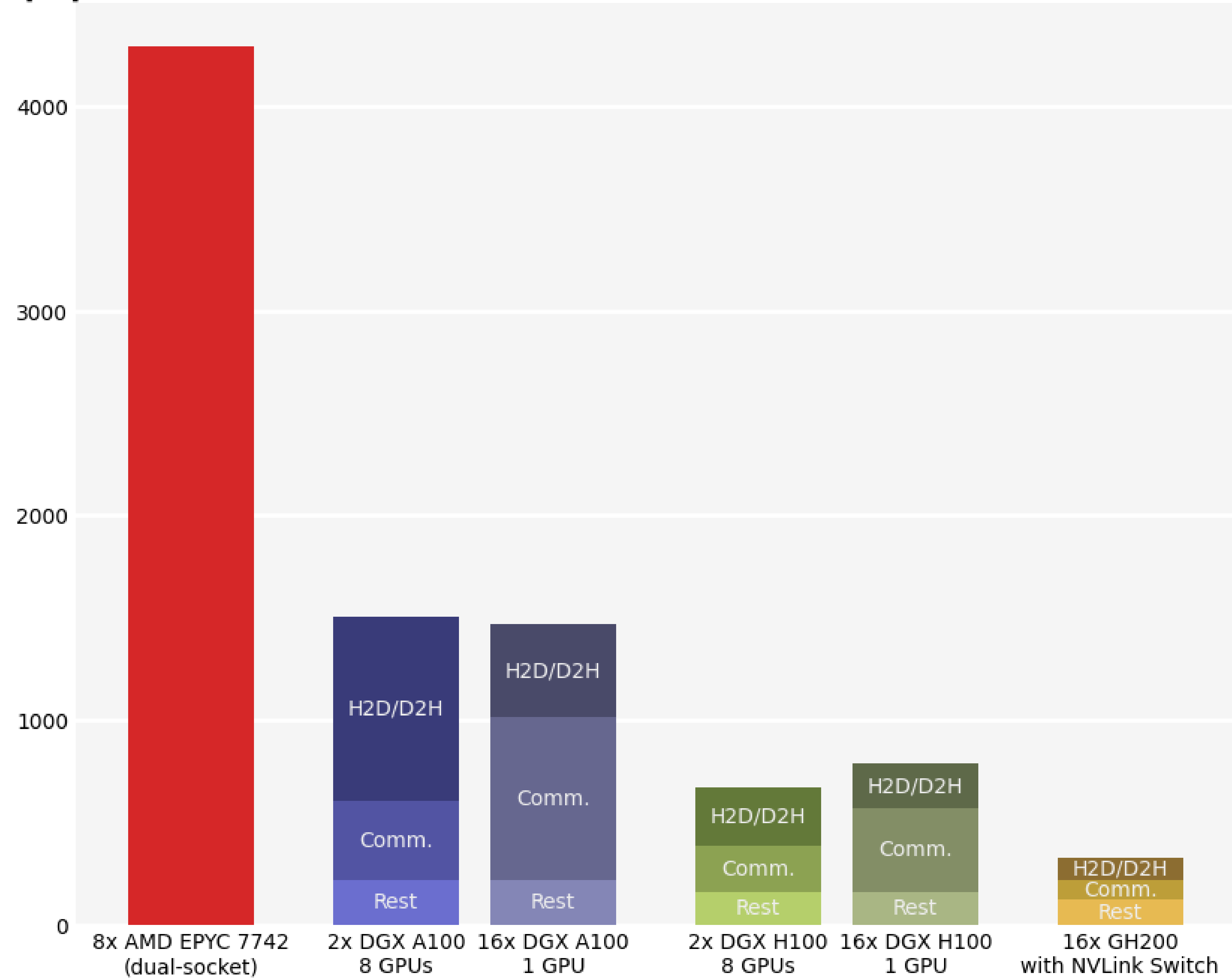


GA100 80 GB vs. GH100 96 GB (HBM3); Speedups against SCC Hoist version with same datatype and machine; NGPTOT=512000
 Peak computed using measured STREAM bandwidth;
 Light colored horizontal span shows 90% range of peak;
 CUDA 11.8; NVHPC nightly; fast-math enabled

UNBLOCKING SPECTRAL TRANSFORM WITH NVLINK SWITCH

- Up to 256 connected GH200 modules
=> Should be able to fit TCc7999
- All-to-all heavy spectral transform
- Including data transfers to/from GPU

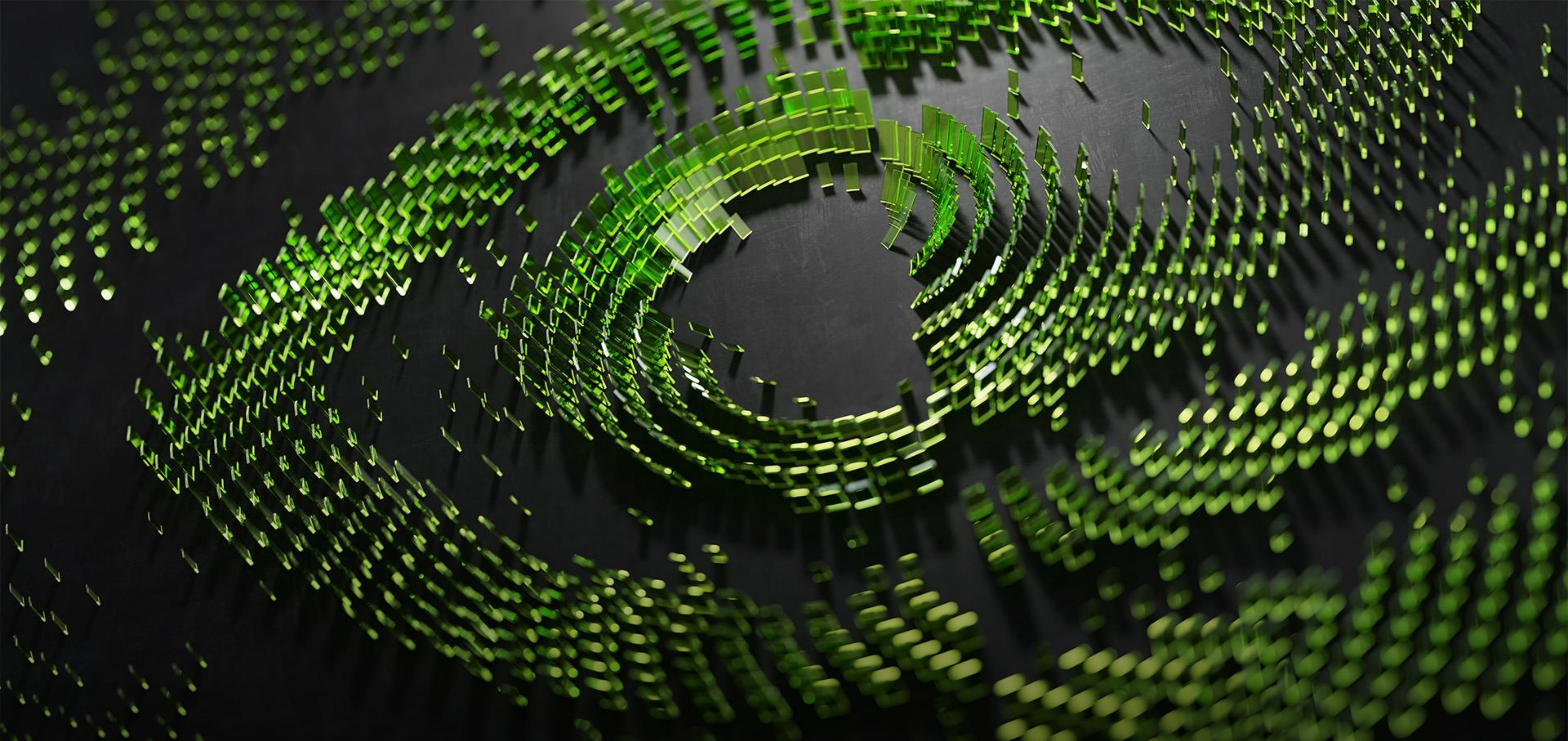
[ms] Time for direct+inverse transform (TCo 2999; including memcopy)





SUMMARY/CONCLUSION

- E2 Platform: Earth Climate Digital Twin
- Interactive access to climate data
 - Archived and live computed
- Data federation
 - Access data form multiple sources
 - Client accepts «best offering», depending on needs
- Simulations packaged as services
 - Tailored/optimized for specific platform
 - Takeing advantage of latest hardware features can have significant performance impact
 - New networking capability can have significant impact
 - Speed-of-light implementations as guiding star, not necessarily for production runs



nVIDIA®