Data Platforms Built for the AI Factory

Bertrand Ounanian
System Engineer Mgr VAST Data
Noureddine Taguelmimt
Sr Solution Architect NVIDIA

VAST

The Data Platform
For The AI Era

1000+
Employees Globally

Cash Flow Positive
A Self-Sustaining Growth Engine

$9.1 Billion
Series E Valuation

Powered By DASE

The First Distributed Systems Architecture Designed for AI

★★★★★
Gartner Peer Insights

84
Net Promoter Score
(validated by OCX Cognition)

Powering The Most Data-Intensive Enterprise and Cloud Providers

VAST Data Platform

Multi-Tenant, Zero-Trust, Global Scale

DataStore

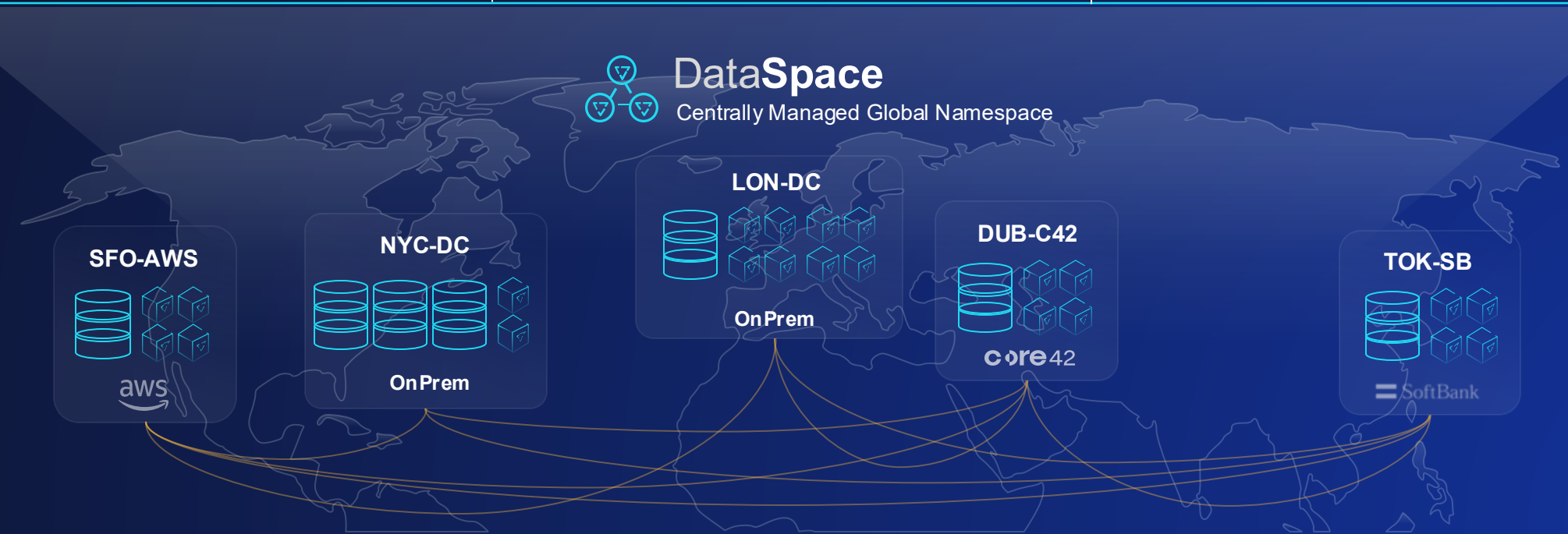
Multi-Protocol File, Object & Block Storage

DataBase

A Transactional Data Warehouse

DataEngine

Parallel, Containerized Computing



World's Leading AI Organizations Run on VAST Data

Model Builders

Powering 100,000+ GPU clusters for the world's leading foundation model builders

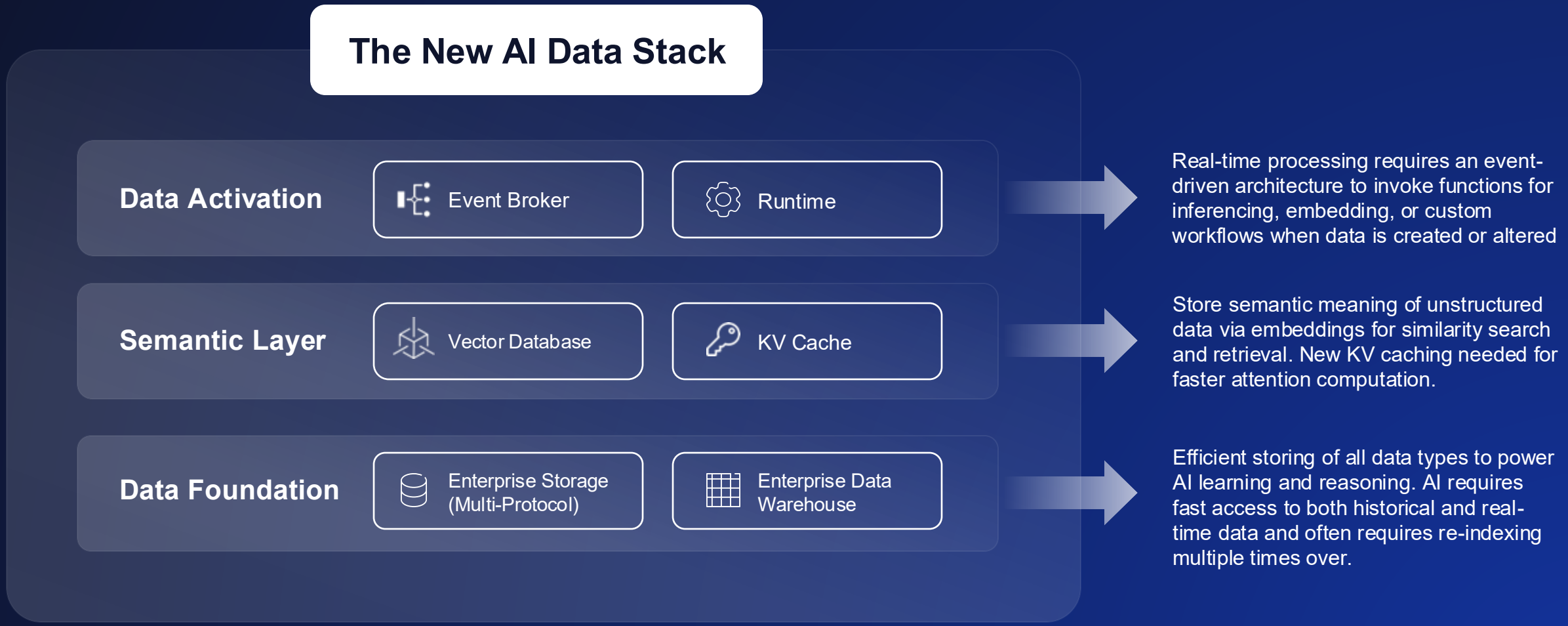
AI CSPs

Providing high-performance data services to 30+ 'neoclouds' all around the world

Enterprise AI

Real-time Data Platform with **NVIDIA NIMs** for unstructured data in the enterprise

The AI Era Needs a New Data Stack



VAST Data Platform

Any Data. Any Workload. No Compromises

Enterprise Workloads

Enterprise Data Corpus

DataStore
DataBase

Automated AI Data Operations

Data Preparation

DataStore
DataBase

Vector Embedding

DataEngine
DataBase

Inference

DataEngine

AI Compliance

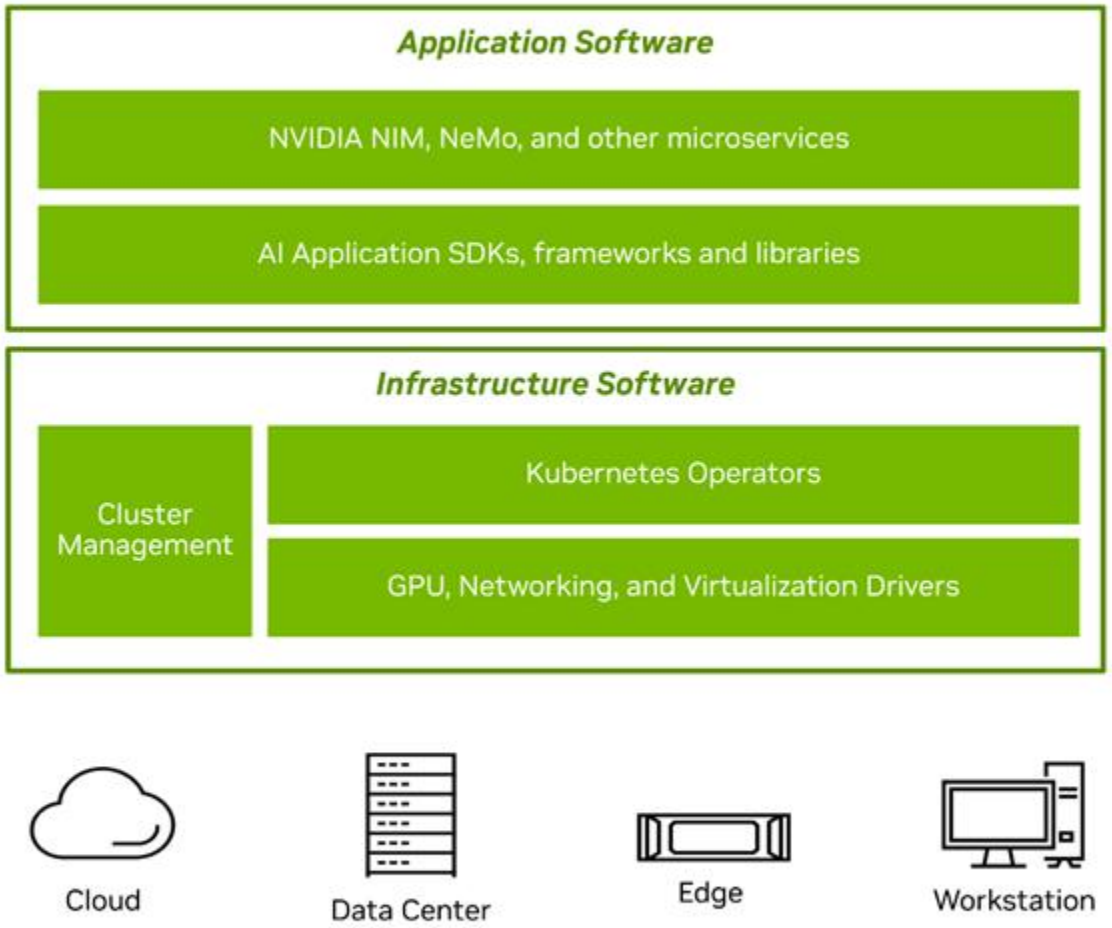
Inference Logging

DataBase

DataSpace
Centrally Managed Global Namespace

NVIDIA AI Enterprise

What's Inside?



Application Software

Optimized NIM and NeMo microservices enhance model performance and speed time to deployment for generative AI

SDKs, Frameworks, and libraries across many domains, including speech AI, route optimization, cybersecurity, and vision AI

Infrastructure Software

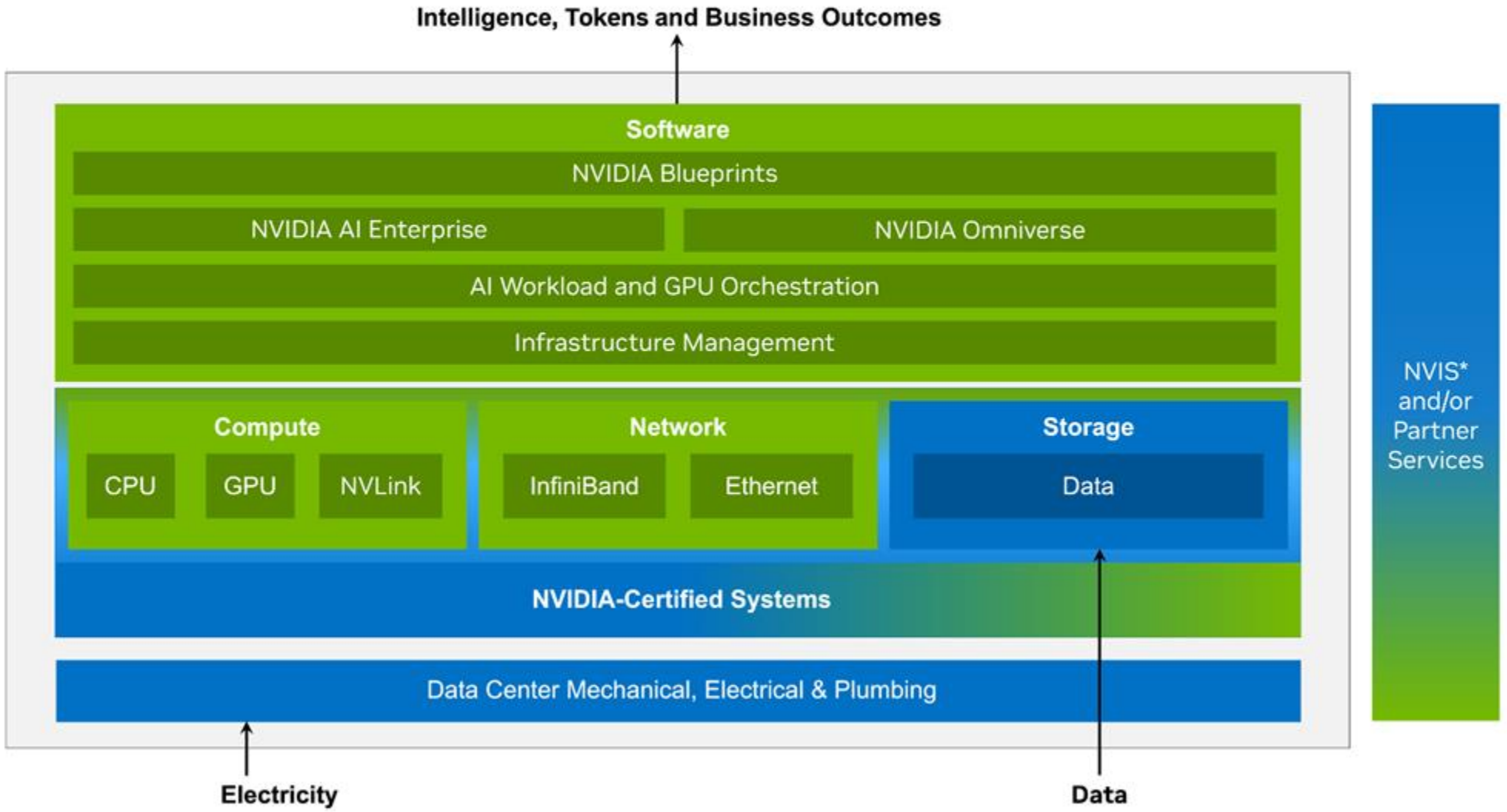
Kubernetes operators for managing GPU and networking in containers and the lifecycle of microservices and AI pipelines

Cluster Management software to provision and monitor servers at scale

Drivers to optimize utilization of NVIDIA GPUs and Networking in bare metal and virtualized environments.

Full Stack approach for AI Factories

Built on customer-validated data center reference architectures



*NVIS = NVIDIA Infrastructure Specialists

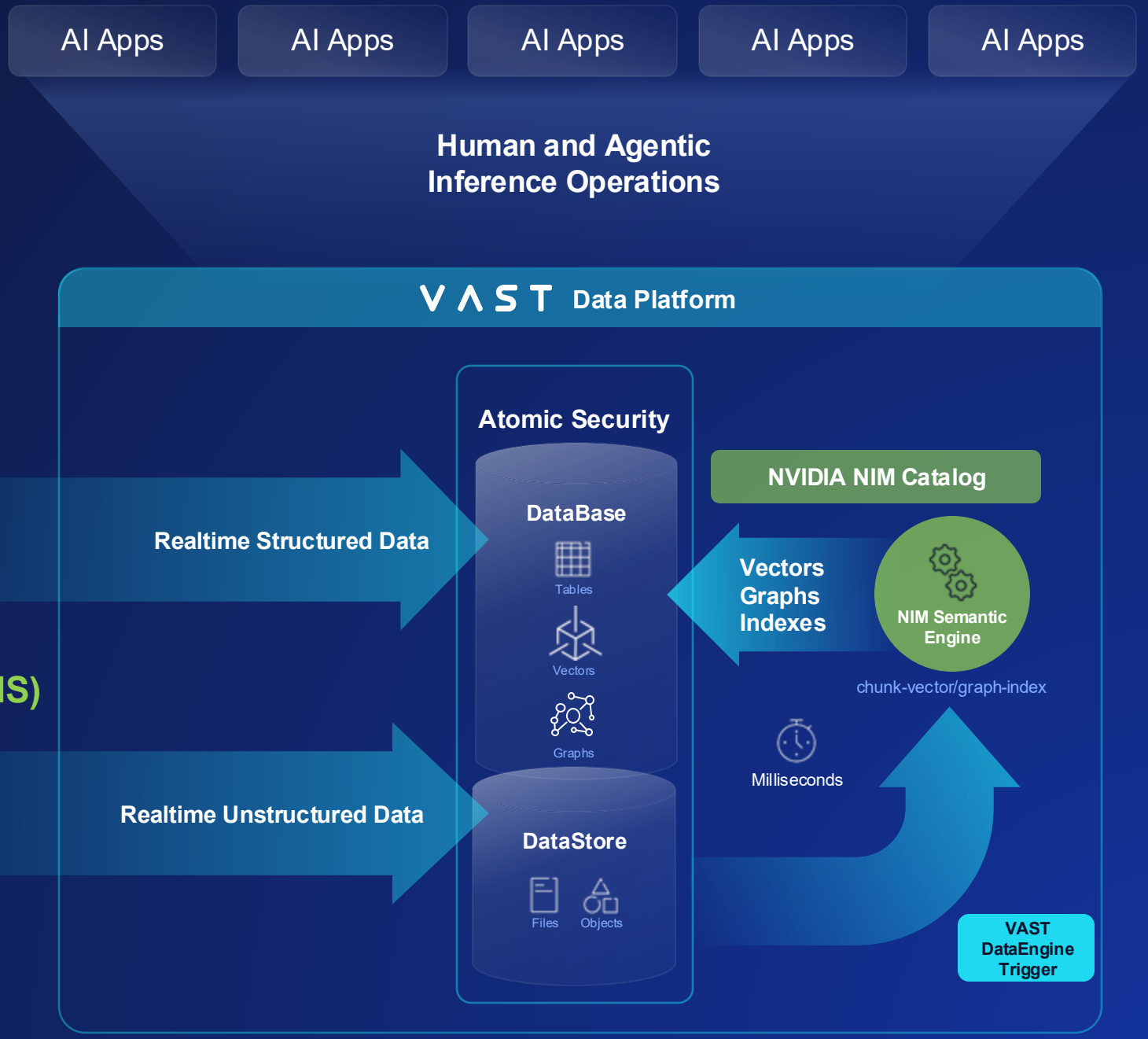
Link:
[Jensen Huang keynote showcases the power of InsightEngine though a demo]

VAST InsightEngine

Real-Time Data Streams

Powered With NVIDIA AI Enterprise (NIMS)

- Multi-Modal PDF Extraction
- Video Search & Summarization



The AI Factory Data Platform

NVIDIA

NVIDIA NIM
Optimized AI inferencing microservice

NVIDIA Blueprints
Reference AI workflow templates

NVIDIA NeMo Platform
Build, customize, and deploy generative AI

VAST

Enterprise Storage (Multi-Protocol)

Enterprise Data Warehouse

Vector Database

Event Broker

Runtime

KV Cache

NVIDIA SPECTRUM-X NETWORKING

NVIDIA

GPUs (H200, B200, etc.)
AI/ML compute acceleration

CPUs (Grace, etc.)
Orchestration and general-purpose workloads

DPUs (BlueField)
Security, networking, and storage tasks

VAST InsightEngine with NVIDIA

Event-Driven Architecture
Invokes NVIDIA NIM in real-time for inferencing and embedding on data creation or modifications

Unifies End-to-End Security
Enforces IT policies through AI semantic layers automatically

Scales to Handle All Your Data
Supports billions of files and trillions of embeddings within one platform

Designed for All Data Types
Store structured and unstructured data via all storage protocols via a single management view

NVIDIA Certified and Validated
NVIDIA DGX SuperPOD, NVIDIA Cloud Partners, and NVIDIA Enterprise Reference Architectures


VAST
+
NVIDIA