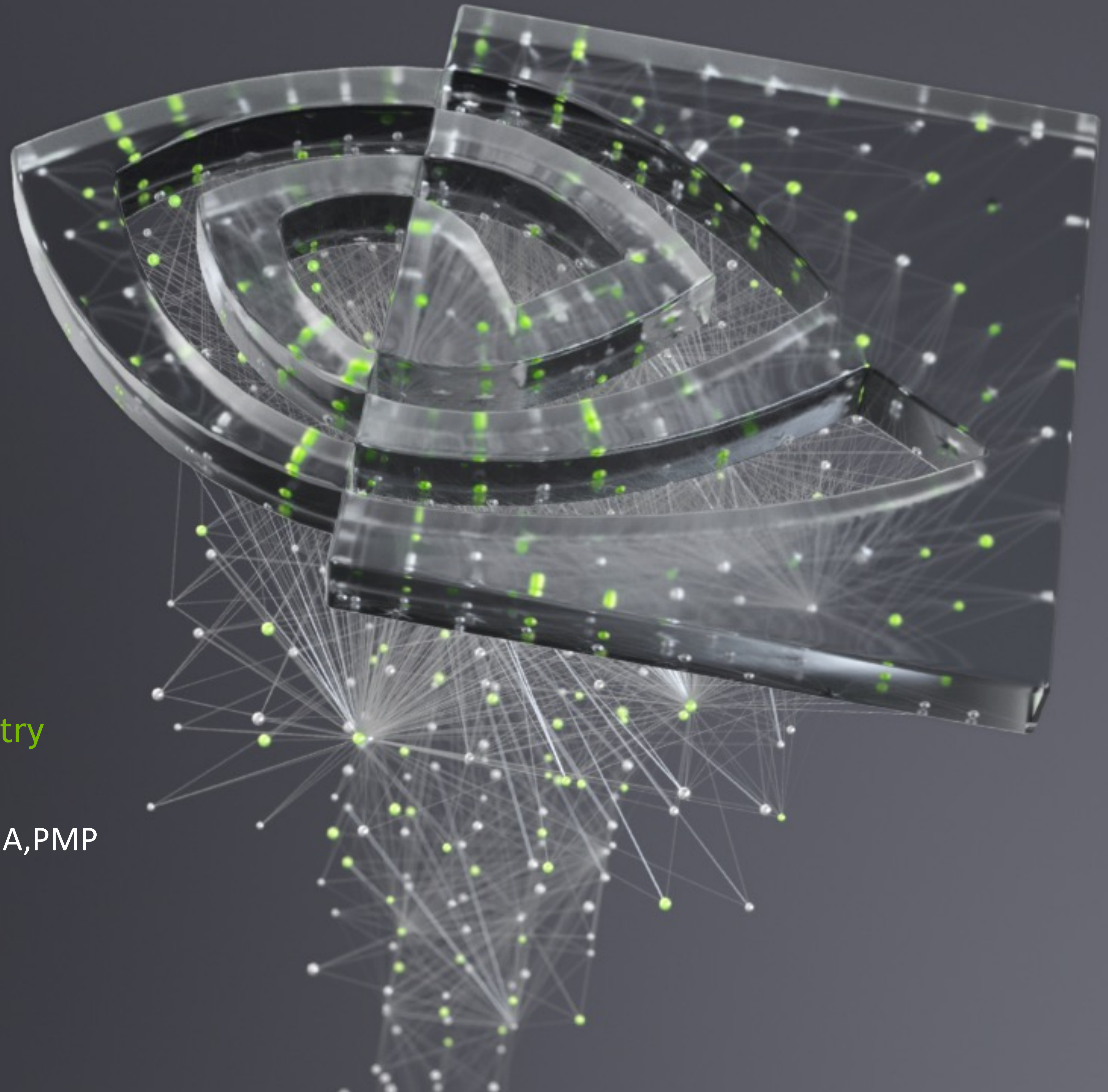
STAC - INNOVATION

Malcolm deMayo, VP Global Financial Services Industry

David Rosen, Director of Sales, US Financial Services
Prabhu Ramamoorthy, Developer Relations, CFA, FRM,CAIA,PMP
Brian Grant, Solutions Architect
Anthony Murphy, Enterprise Account Manager, FSI



FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)

“NVIDIA REINVENTS ITSELF EVERY SINGLE YEAR. WE ARE GOING TO CALL NVIDIA ‘THE GOAT,’ THAT IS, THE GREATEST OF ALL TIME.”

MAD MONEY

NVIDIA pioneered accelerated computing to tackle challenges ordinary computers cannot. We make computers for the da Vincis and Einsteins of our time so that they can see and create the future.

CEO & Founder:	**Jensen H. Huang**
Revenue:	$26.9B 61% YoY
Data Center:	**$11.0B 58% YoY**
R&D Investment:	**$ 5.3B 34% YoY**
Glassdoor best place to work	**#1 (2022) #2 (2021)**

NVIDIA is leading in HPC & AI

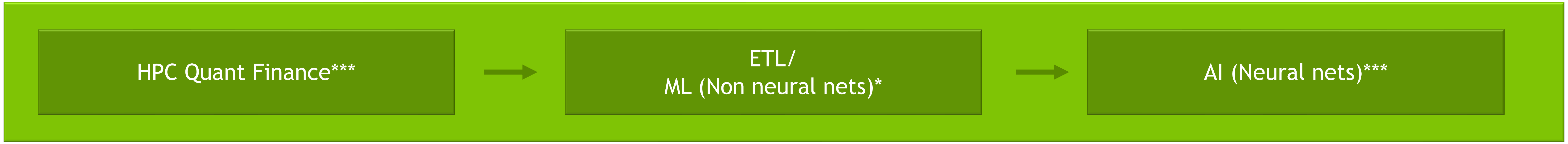
1B Cuda GPUs
250 Cloud ExaFlops
3000 Appl. Accel.
10M Cuda downloads
450+ SDK AI Models
12K startups
3.5M developers

“NVIDIA’s DNA is in every AI solution we evaluated. Its an understatement to say that NVIDIA’s AI Platform are synonymous with AI infrastructure.” *Forrester Wave, AI Infrastructure, Q4 2021*

FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)

BUILDING THE FULL STACK TO ACCELERATED COMPUTE

FSI WORKLOAD TYPES, USE CASES, anmd proof points



- Pricing, Risk (FRTB, CVA, SIMM, XVA) & Simulation
- Monte Carlo Simulation, Other types (VAE, GAN, Bootstrapping)
- Algo Trading (LSTM/RNN FinQuant) & Backtesting
- Use Cases by Framework – CUDA, Iso C++ Parallel Algorithms, Accelerated Python & RAPIDS, OpenACC

Data mining

- Feature Engineering
- Data Prep, ETL & Databases
- ML & Data Science (e.g., XGBOOST)
- Use Cases by Framework – RAPIDS, Spark on GPU

- RNNs – used to process text (Chatbots, sentiment analysis, time series)
- CNNs – Used to process images and text (face recognition)
- GANs – Used in reinforcement learning to train models in real time
- Supervised & Unsupervised Learning
- Aggregation of signals into a strategy
- Reinforcement Learning
- Testing & Evaluation
- Use Cases by Framework – Pytorch, TensorFlow, JAX

AI/ML Use Cases

- Monte Carlo Risk Simulations
- Market Risk (Exotic Derivative pricing, Variable Annuities, Modeling underlying volatilities – e.g., Heston)
- Counterparty Risk (CVA, XVA, FVA, MVA Valuation adjustments)
- Market Generator and Simulator

NVIDIA Platform

- Hardware: Training: DGX/A100
- Software: CUDA, C++, HPC SDK, RAPIDS
- Https://developer.nvidia.cFSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)om/hpc-sdk

Resources

- STAC-A2 blog
- STAC-A3 blog
- JPMC Risk Calculations
- Cohen & Steers GTC session
- Wells Fargo GTC session
- **Citibank NN For Exotic** GTC session
- CBOE Global Markets GTC session
- Bank of America GTC session

FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)

ACCELERATE FINANCIAL MODELING & SIMULATION WITH THE NVIDIA HPC SDK

Graham Lopez, Product Manager HPC Compilers

Big Picture in meeting customer challenges

- Accelerate RT performance
- Streamline time to solution
- Improve Developer productivity
- Improve Portability

Fortran

OpenACC

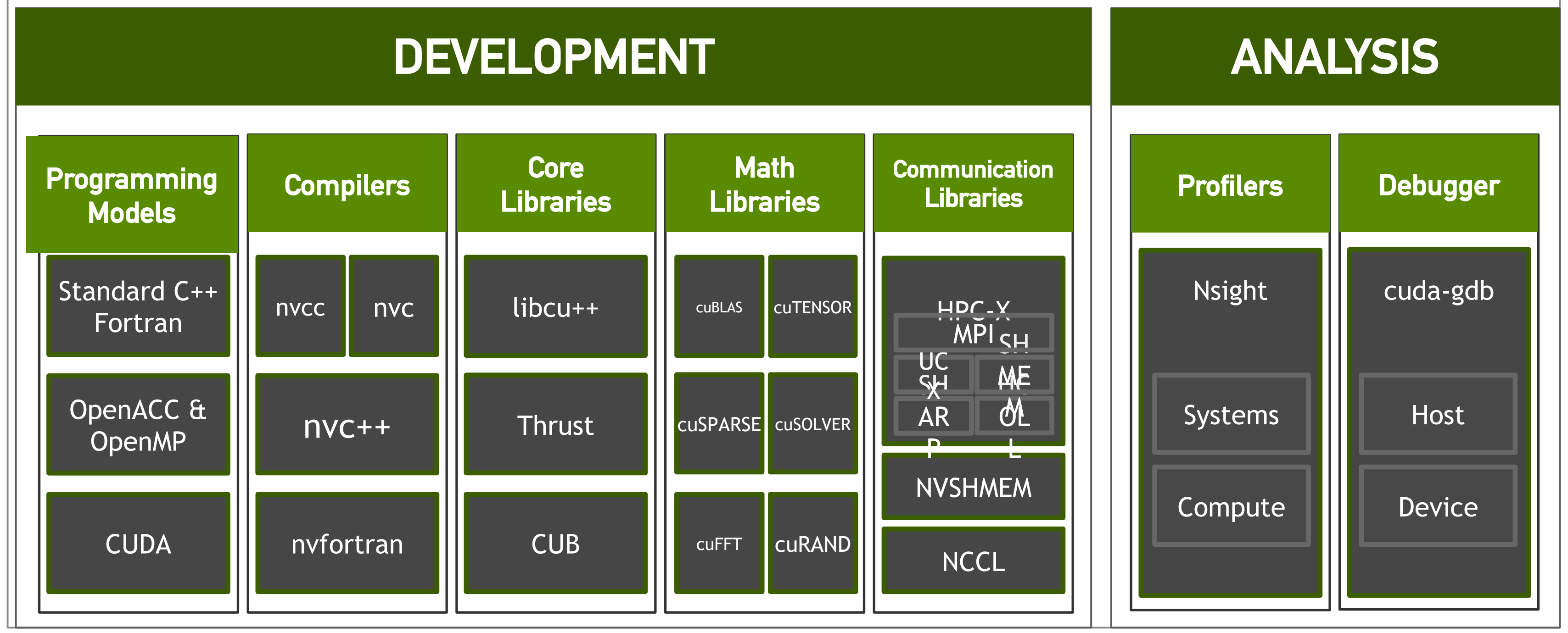
More Science, Less Programming

OpenMP

NVIDIA Solutions:

- HPC CUDA
- Parallelize Algorithms in ISO C++17 & ISO Fortran
- Provide Open-Source solution – OpenACC
- New compilers & libraries to speed adoption

NVIDIA HPC SDK | Available at no charge



DEVELOPMENT

- Programming Models: Standard C++ Fortran; OpenACC & OpenMP; CUDA
- Compilers: nvcc; nvc; nvc++; nvfortran
- Core Libraries: libcu++; Thrust; CUB
- Math Libraries: cuBLAS; cuTENSOR; cuSPARSE; cuSOLVER; cuFFT; cuRAND
- Communication Libraries: HPC-X; MPI; UCX; SHARP; SHMEM; HCOLL; NVSHMEM; NCCL

ANALYSIS

- Profilers: Nsight; Systems; Compute
- Debugger: cuda-gdb; Host; Device

ACCELERATING PYTHON FOR EXOTIC OPTION PRICING

- **Part 1: Use Python to implement Monte Carlo simulation to price the exotic option efficiently**
- **Part 2:** Use Neural Networks and deep learning to approximate the pricing model and speed up inference latency
  - Approximated model calculates option Greeks efficiently
  - TensorRT boosts inference time to state of the art exotic option speed

https://developer.NVIDIA.com/blog/accelerating-python-for-exotic-option-pricing/

Inspired by this the developer (Yi Dong) his blog below NVIDIA Devtech used his case to showcase how far we have come

- CUDA version was 1st ported to a loop-based C++ code
- Includes OpenACC directives for comparative GPU performance
- Three main parts to the algorithm

1. Generate a set of random numbers (cuRAND)
2. Compute the Barrier Option Payoff
3. Sum the Payoffs

Programming | Compute	Speedup
CUDA | A100	87x*
Standard C++ | A100	65x*
OpenACC | A100	37x*

Accelerate Financial Modeling and Simulation with the NVIDIA HPC SDK | NVIDIA On-Demand

*Not STAC Benchmarks

FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)

ACCELERATED COMPUTE FOR DEEP LEARNING

Jacob Holley, PhD & Georgious Papaioannou, PhD, Bank of America

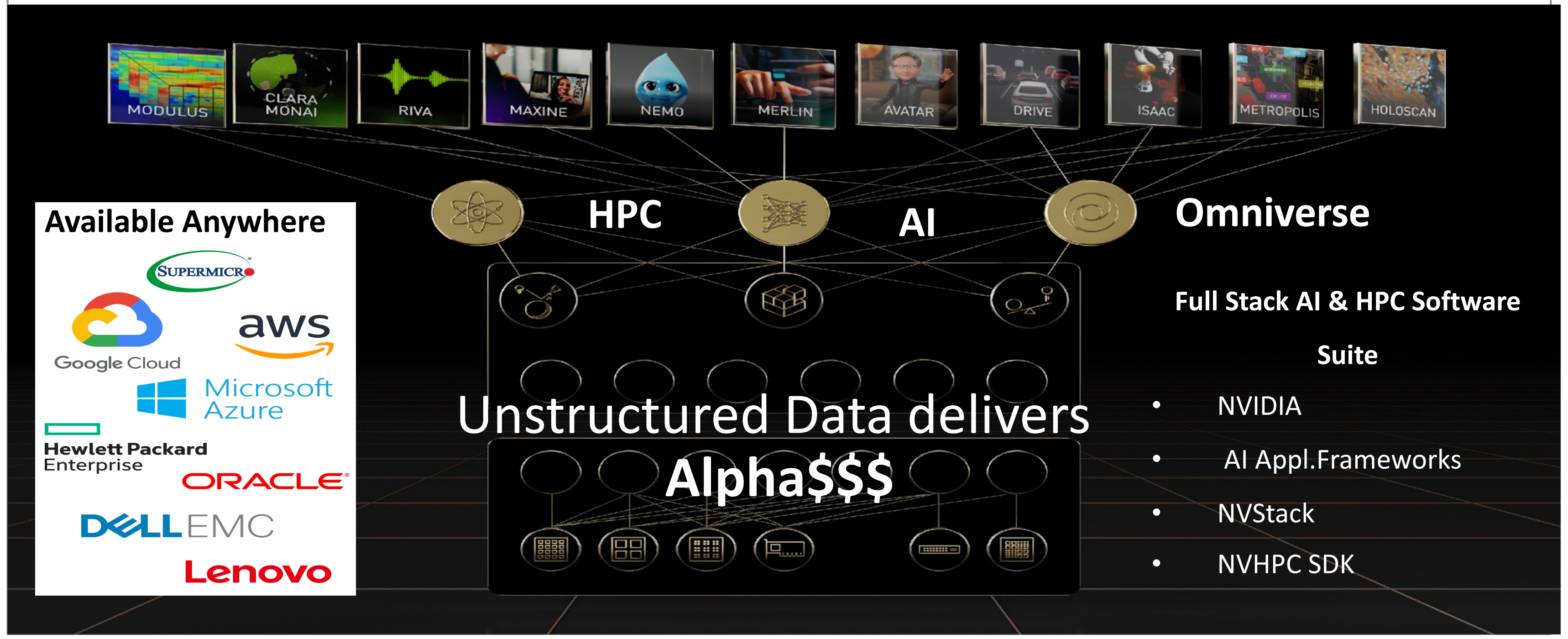
Big Picture in meeting customer challenges

- Democratize AI - Open, Complete, Hardened, & Scalable
- Accelerate HPC, Data Processing, DL, Training & Inference
- Enable FSI to create re-usable capabilities
- Hybrid, Multi-Cloud Portability, Management, Monitor, Govern

NVIDIA Solutions:

- NVIDIA AI Application Frameworks (RIVA, Nemo, AVATAR, Merlin)
- NVIDIA Stack (Triton, RAPDIS, TensorRT, TensorFlow, PyTorch, JAX)
- NVIDIA LaunchPad, NVIDIA Lighthouse,

NVIDIA AI Enterprise| FSI- Ready | Open & Complete



Available Anywhere

Unstructured Data delivers Alpha$$$

Full Stack AI & HPC Software Suite

- NVIDIA
- AI Appl.Frameworks
- NVStack
- NVHPC SDK

Cross-asset risk premia prediction with recurrent GANs and disentangled feature encoding β-VAEs

Challenges:

- Large number of parameters/time lags can lead to overfitting/curse of dimensionality
  - Recurrent neural networks (LSTMs)
  - Interpretable encoding using β-VAEs
- Limited data to reduce variance
  - Synthetic data generation using time-series GAN

S&P 500 06/18/2018 – 08/31/2020

The Prize of Predictability

Scenario	Sharpe Ratio
Buy & Hold the Index	.6
Fully Predictability Model Buy/ Sell Daily	11.0

Sharpe Ratio Rule of thumb

- Good Sharpe b/t 1 & 2
- Very Good Sharpe Ratio b/t 2 & 3
- Excellent Sharpe Ration > 3

Prediction with RNN & GANs, Bank of America GTC Spring 2022, NVIDIA on Demand

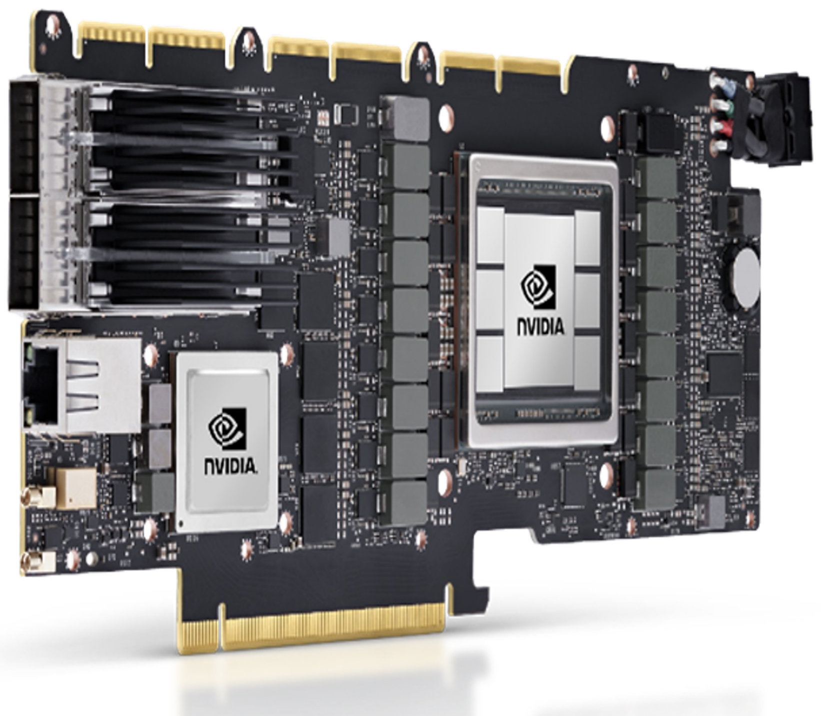
ACCELERATED MARKET DATA PLATFORMS

GPUDirect RDMA + GPU CUDA Kernels + Converged Adapters

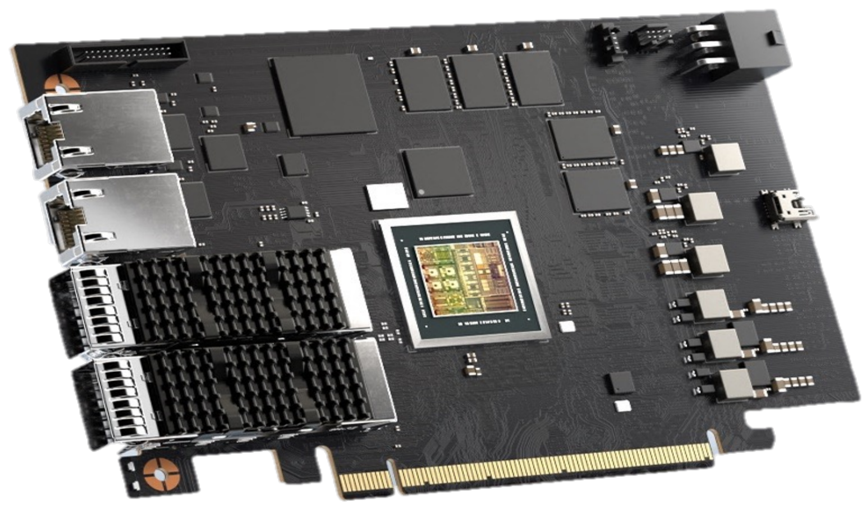
Our Networking solutions support both HPC & AI

- *New low latency applications leveraging GPUs emerging :*
- *Time scale for latency is in microsecond to millisecond domain*

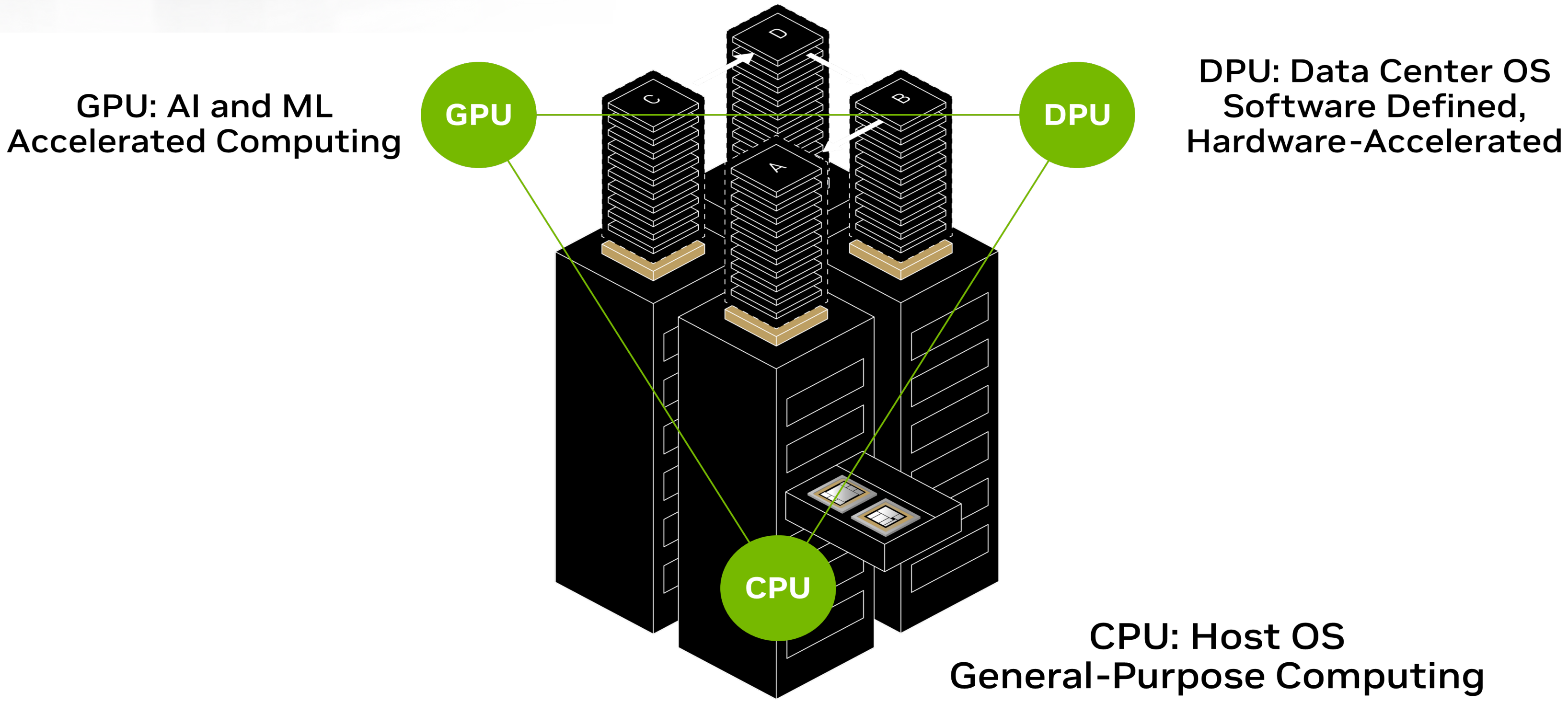
- Market data processing at exchanges / exchange subscribers
  - Enable parallel processing of received packets / index/ETF calculations
  - SIP feed processing (SIP : US securities information processor)
- Semi High frequency trading applications
  - Low latency trade requests (Tick to Analytics to Trade)
- Producer / Consumer applications
  - Accelerating IO Input/Output across Data Center Applications
  - GPU can ingest large amount of network data directly into GPU memory
  - Enables new kinds of analytics in low latency applications



Nvidia Converged A100X



NVIDIA BlueField DPU



FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)

GETTING STARTED WITH NVIDIA AI

NVIDIA AI Enterprise Trial Programs

Test Drive Demo

- Self-directed, remote access demo
  - Predicting NYC Taxi Fares with RAPIDS
  - BERT Question Answer in TensorFlow
- Requires ~1 hour/Access for 48 hours

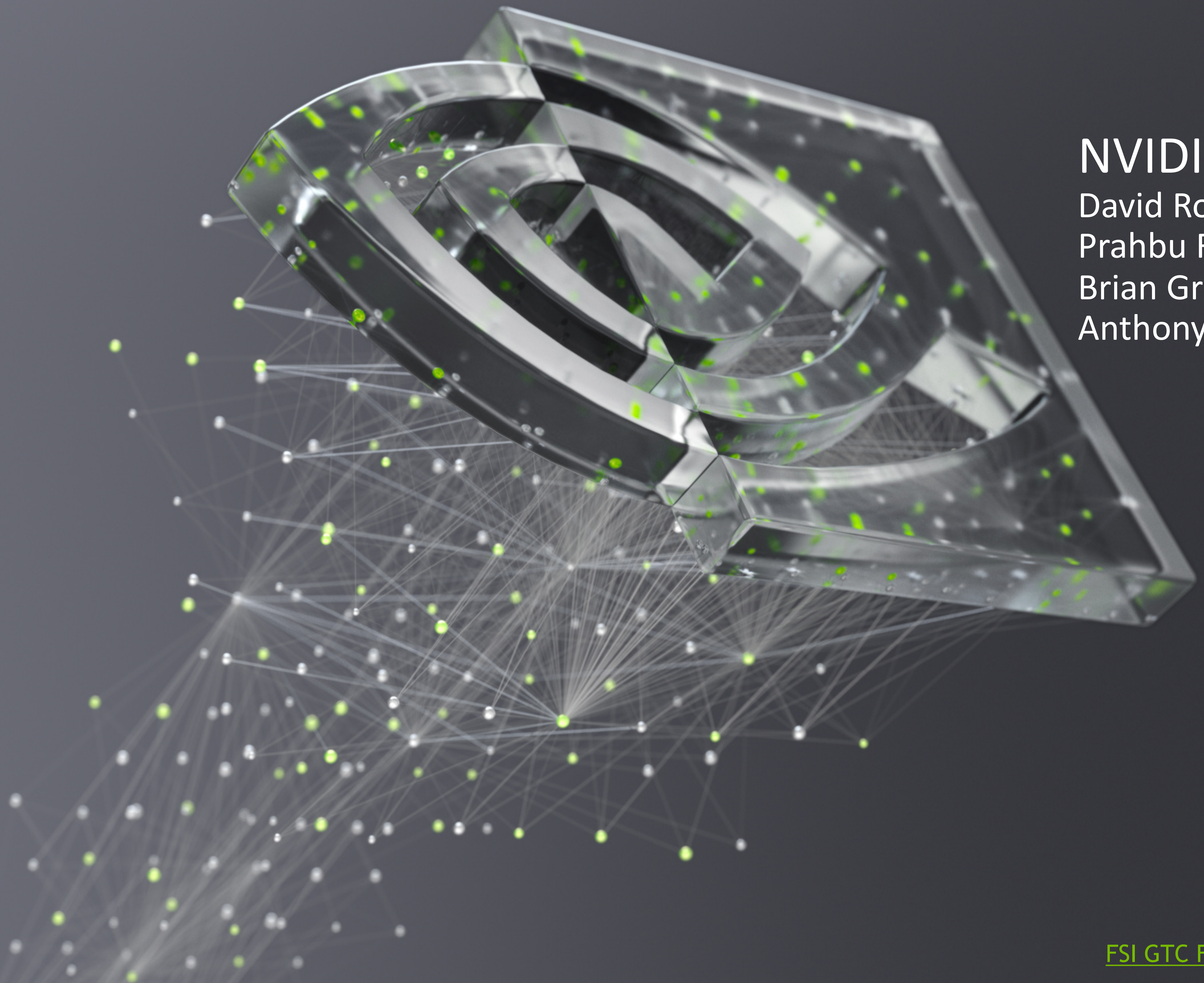
NVIDIA LaunchPad

- AI development and deployment trial program
- Deep dive, hands-on labs for AI practitioners and IT staff
- Requires ~8 hours/Access for 2 weeks

Light House Partner

- C-suite sponsorship
- NVIDIA & Deloitte engagement with customer
- 2-4-week ideation to validation

FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)



Thank you

NVIDIA STAC Team

David Rosen, Director US FSI Sales
Prahbu Ramamoorthy, DevRel, CFA, FRM,CAIA,PMP
Brian Grant, Solutions Architect
Anthony Murphy, Enterprise Account Manager, FSI

Malcolm deMayo
mdemayo@nvidia.com
(203) 984-1168

FSI GTC Fall 2022 Recommended Sessions Slide (highspot.com)