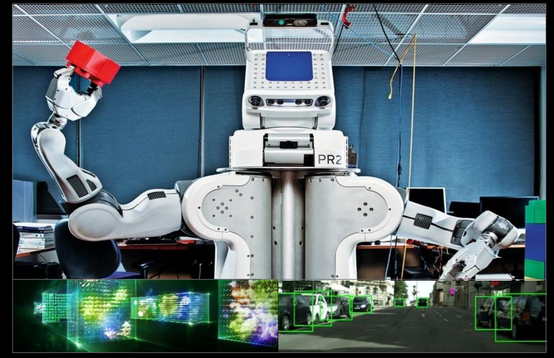
MARC HAMILTON

VP Solutions Architecture and Engineering

NVIDIA
“THE AI COMPUTING COMPANY”

GPU COMPUTING

COMPUTER GRAPHICS



ARTIFICIAL INTELLIGENCE

THE PROMISE OF AI

Increased access to healthcare

Improved patient outcomes

Safer cities

Safer & more efficient transportation

Intelligent manufacturing

16T
Global GDP Boost
by AI by 2030

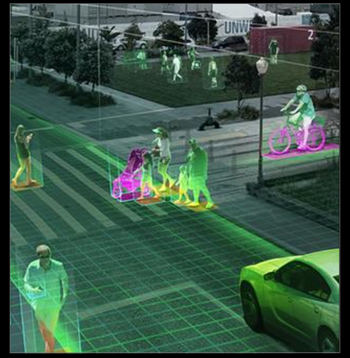
58M
New Jobs Created
by AI by 2022

54%
Jobs Requiring Reskilling
by 2022

SOURCE: AI could contribute up to $15.7 trillion to the global economy in 2030, PwC, "Sizing the prize: What's the real value of AI for your business and how can you capitalise?"
58 million new jobs created by AI by 2022 and 54% of jobs requiring reskilling, World Economic Forum, "The Future of Jobs."

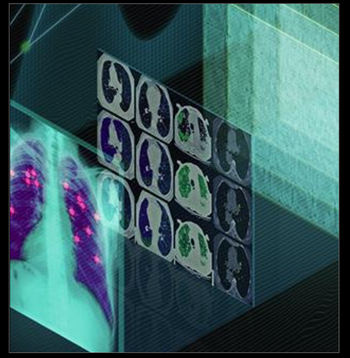
AI IS FUELING GLOBAL INDUSTRIES

Multi-Trillion Dollar Global Industries Turning to AI

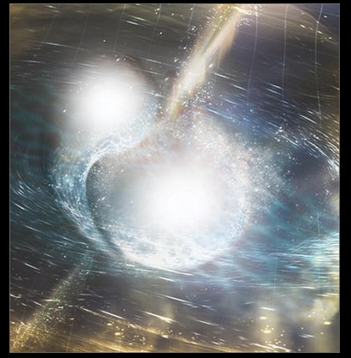


SMART CITIES

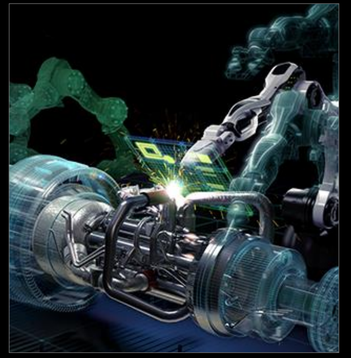
PUBLIC SAFETY



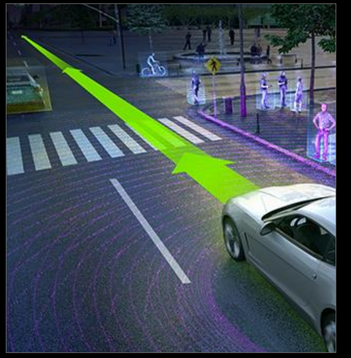
HEALTHCARE



STARTUPS



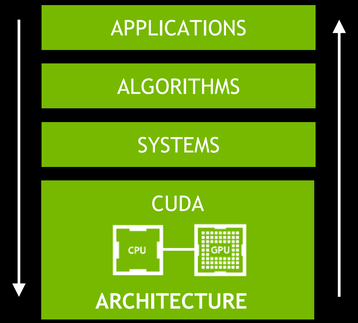
INDUSTRIAL

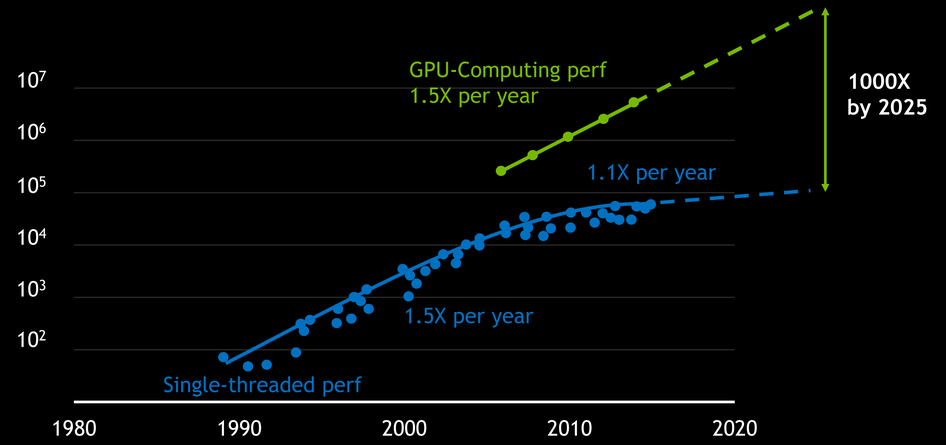


TRANSPORTATION

GPU TECHNOLOGY UPDATE

RISE OF GPU COMPUTING

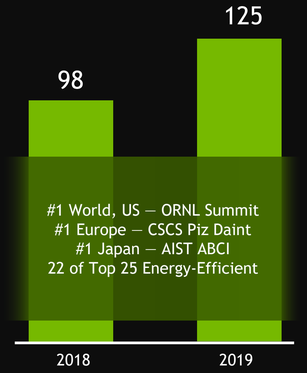




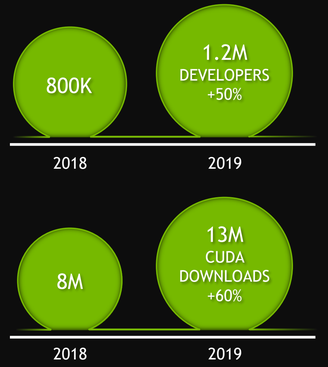
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

A YEAR OF RAPID GROWTH

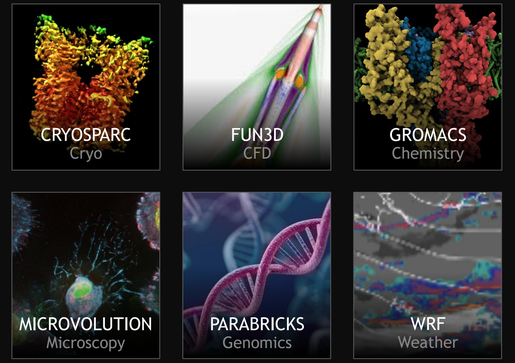
25% MORE TOP500 SUPERCOMPUTERS



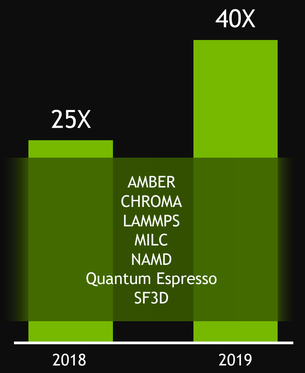
50% GROWTH OF NVIDIA DEVELOPERS



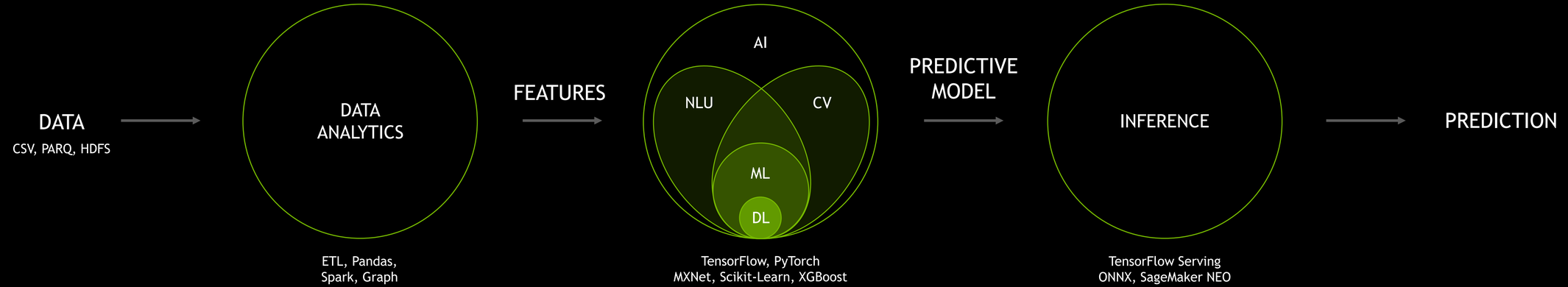
600+ CUDA APPS



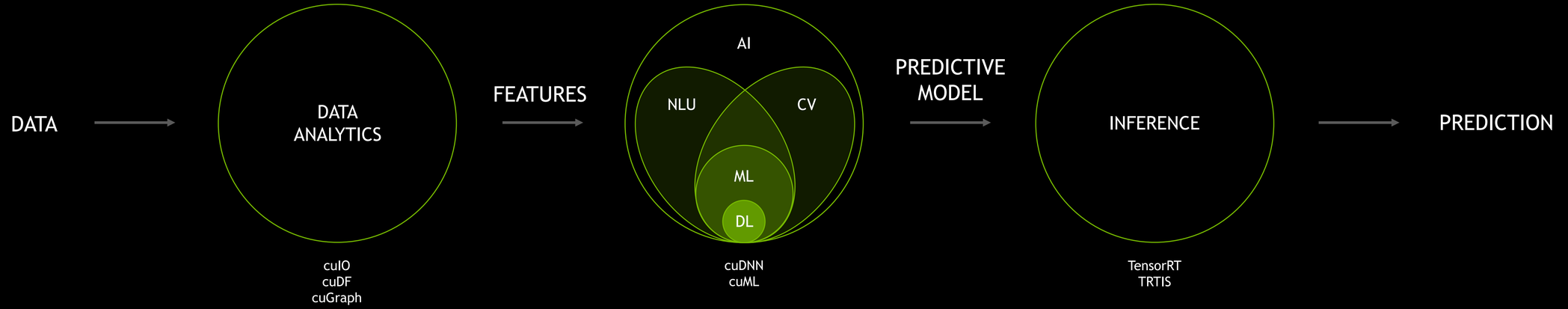
MORE PERF SAME GPU



DATA SCIENCE – A NEW PILLAR OF DISCOVERY

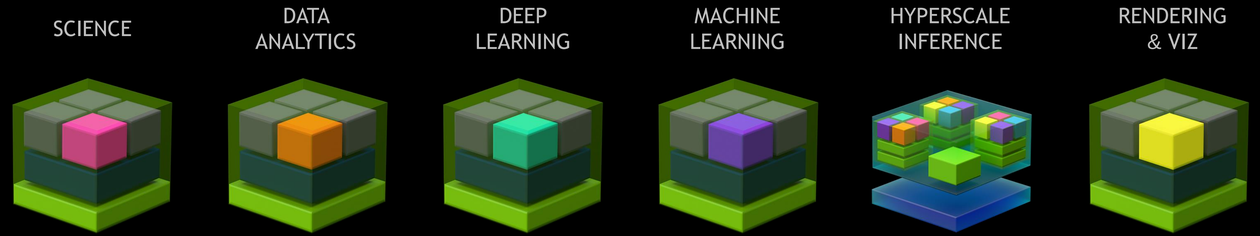


DATA SCIENCE – A NEW PILLAR OF DISCOVERY

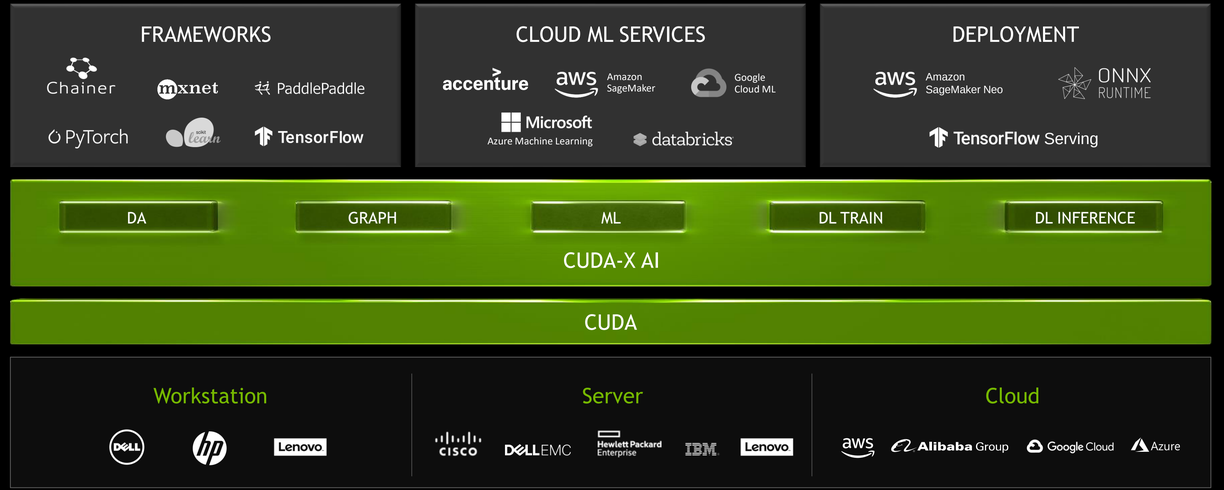


NGC APPLICATION ACCELERATION STACKS

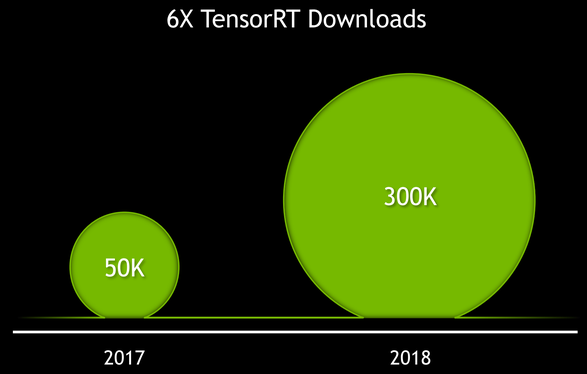
WEALTH OF ACCELERATED APPS MAXIMIZE DATACENTER THROUGHPUT, UTILIZATION, EFFICIENCY



NVIDIA CUDA-X AI ECOSYSTEM

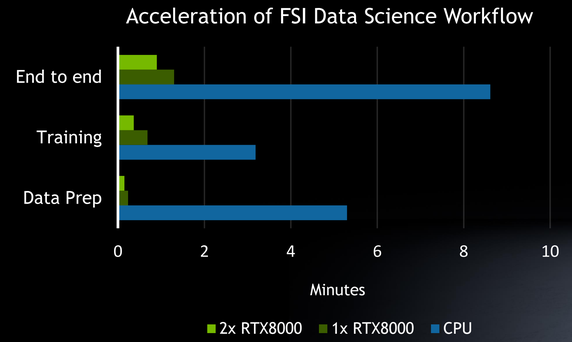


ANNOUNCING WORLD'S LEADING TECH COMPANIES ADOPT CUDA-X AI TO ACCELERATE MODEL DEPLOYMENT



- Voice Search
- Image Search
- Recommendations
- Home Assistant
- News Feed
- Translation
- eCommerce

ANNOUNCING WORLD'S TOP COMPUTER MAKERS OFFER WORKSTATIONS OPTIMIZED FOR DATA SCIENCE



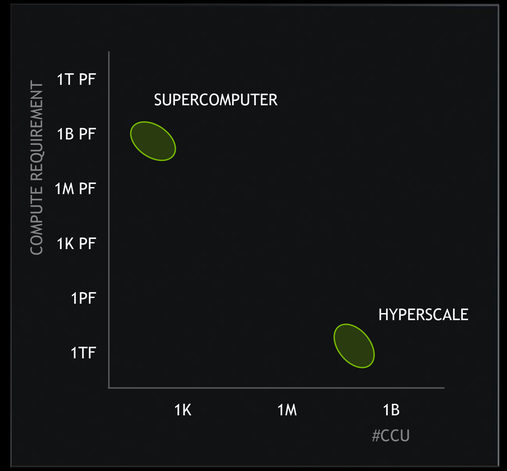
POWERED BY NVIDIA GPU AND CUDA-X AI

Dual Quadro RTX 8000 with 96 GB Memory

Pre-installed for CUDA-X Accelerated Data Science — RAPIDS, TensorFlow, PyTorch, Caffe, Anaconda Distribution

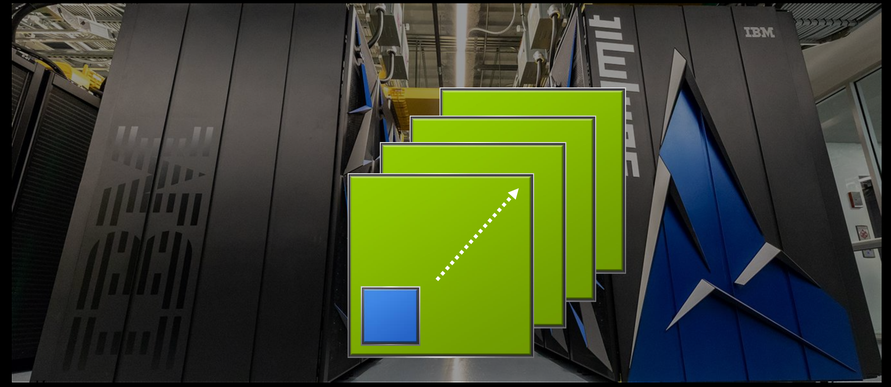
10X Faster

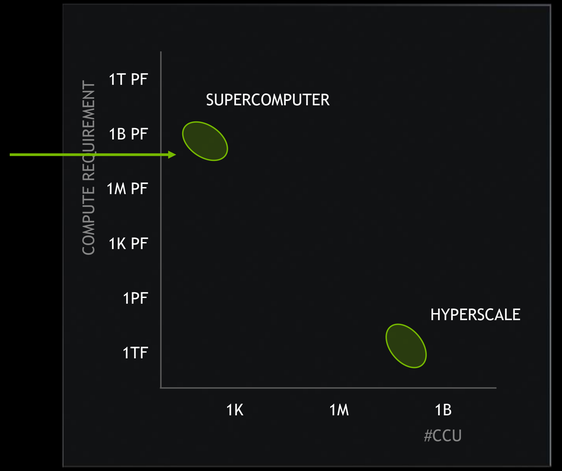
SUPERCOMPUTER vs. HYPERSCALE



SUPERCOMPUTER vs. HYPERSCALE

Supercomputer | Capability Machine | Scale-up Architecture

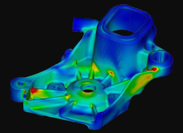




ANNOUNCING 3X PERFORMANCE ON SUMMIT FOR HPL-AI

HPL-AI: A New Approach to Benchmarking AI Supercomputing

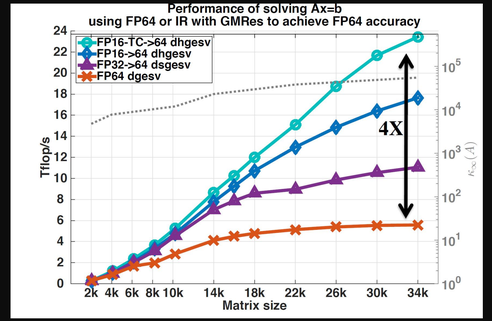
FUSION OF HPC & AI



HPC (Simulation) - FP64

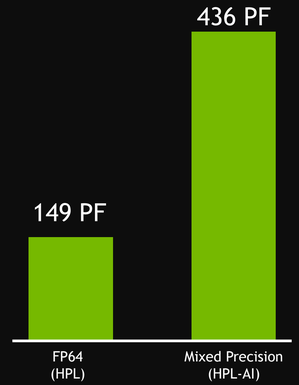
AI (Machine Learning) - FP16, FP32

HPL-AI & ITERATIVE REFINEMENT SOLVERS



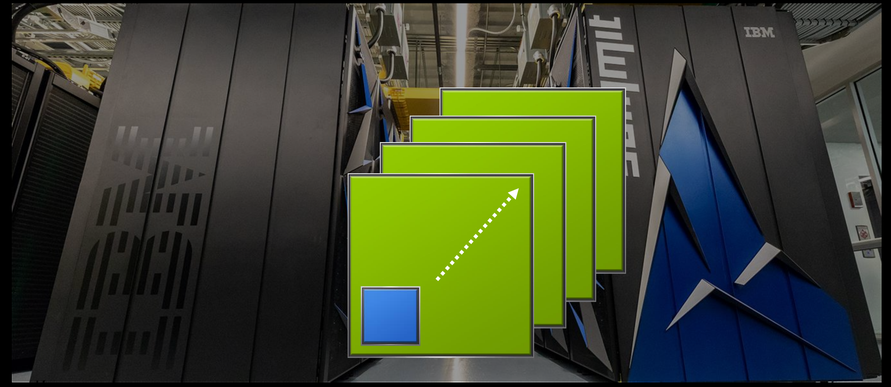
Proposed by Prof Jack Dongarra, et al

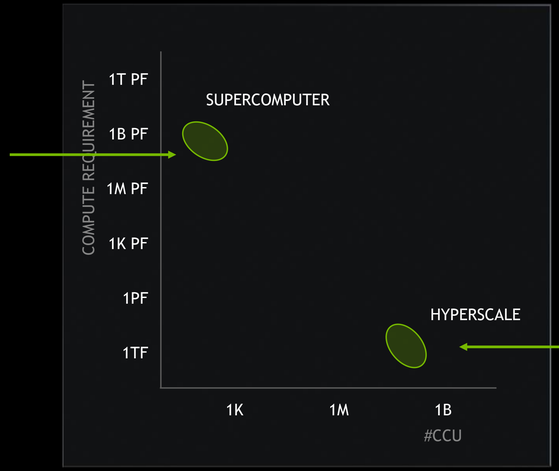
3X MORE PERF ON SUMMIT w/ TENSOR CORE GPUs



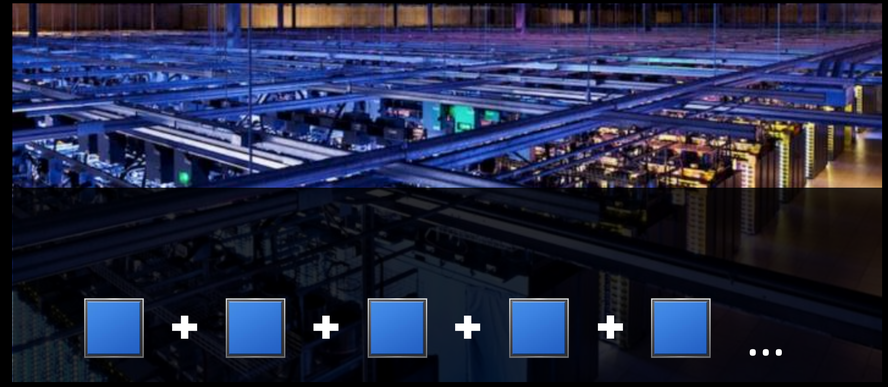
SUPERCOMPUTER vs. HYPERSCALE

Supercomputer | Capability Machine | Scale-up Architecture



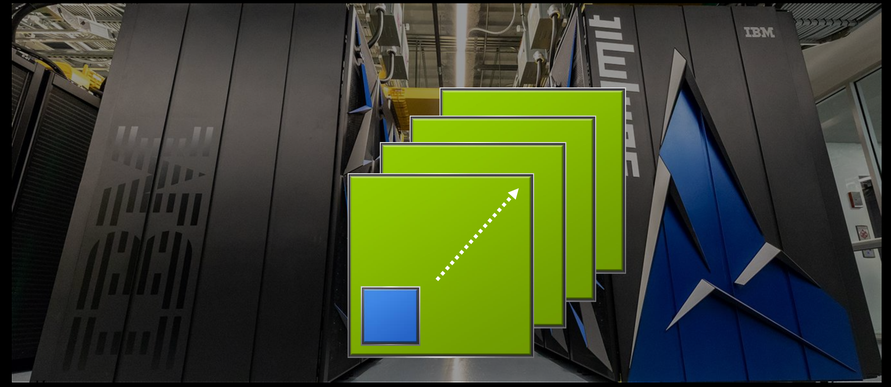


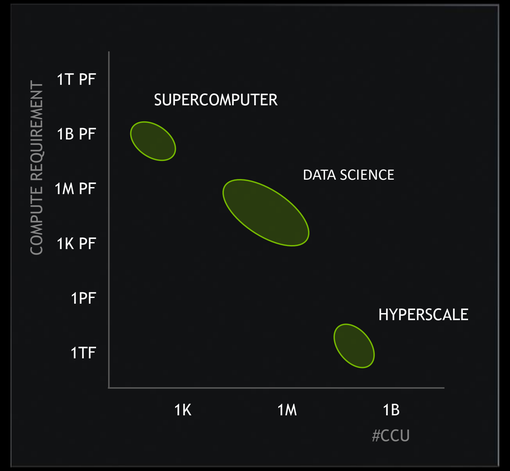
Hyperscale | Capacity Machine | Scale-out Architecture



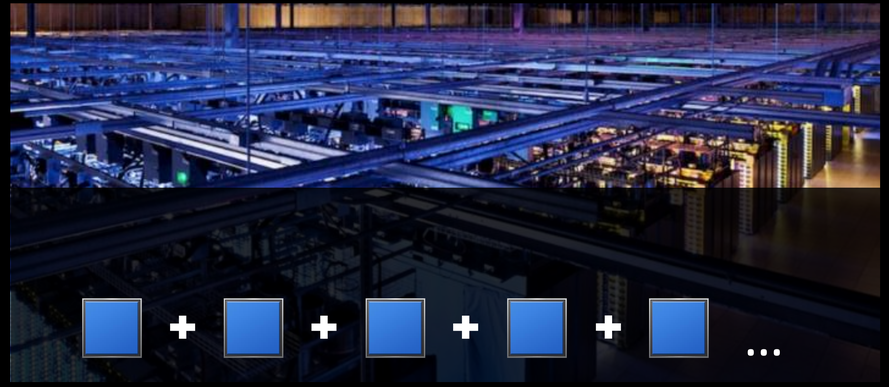
DATA SCIENCE – THE NEW HPC CHALLENGE

Supercomputer | Capability Machine | Scale-up Architecture

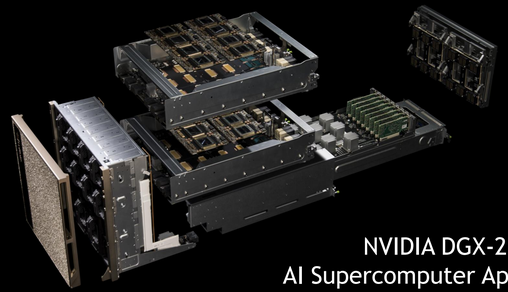




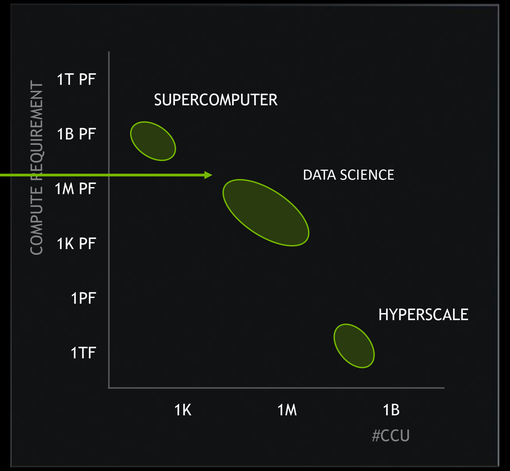
Hyperscale | Capacity Machine | Scale-out Architecture



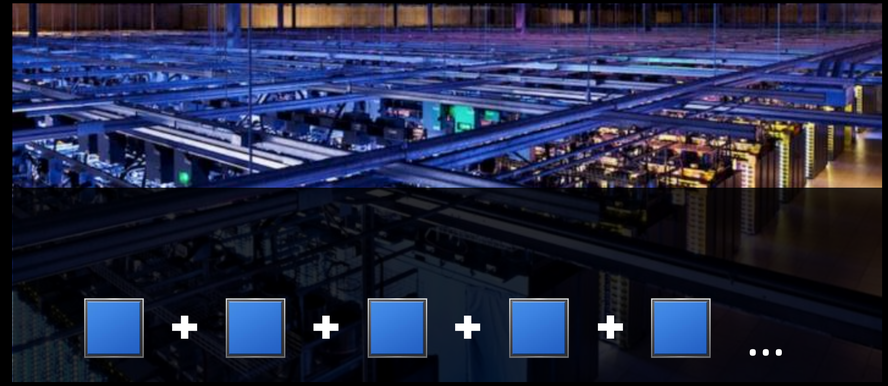
DATA SCIENCE – THE NEW HPC CHALLENGE



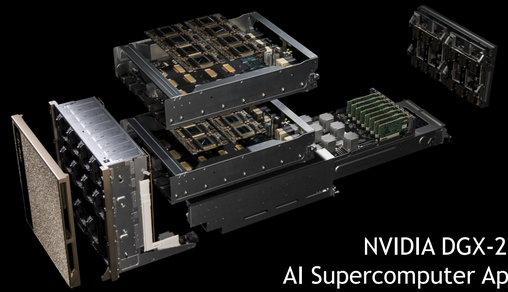
NVIDIA DGX-2
AI Supercomputer Appliance
16x V100 | 2 PF | 512GB HBM2
8x MLNX IB



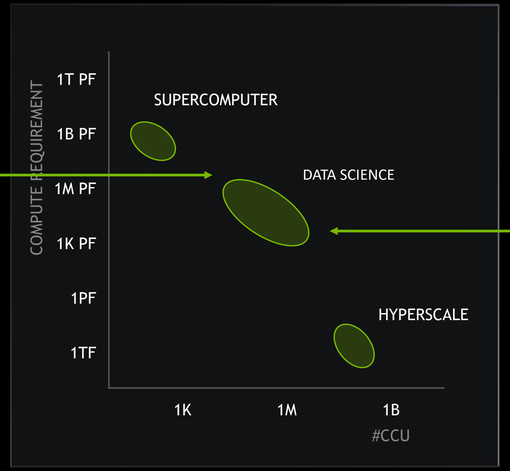
Hyperscale | Capacity Machine | Scale-out Architecture



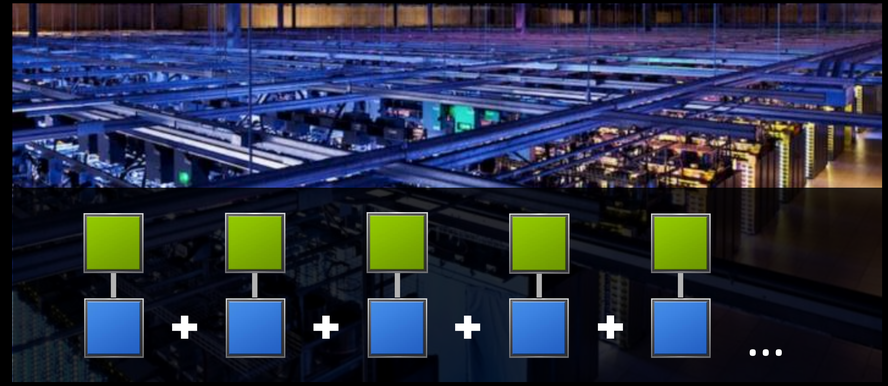
DATA SCIENCE – THE NEW HPC CHALLENGE



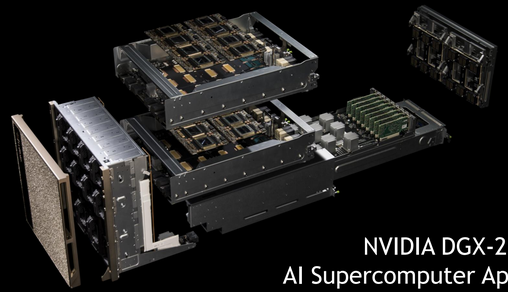
NVIDIA DGX-2
AI Supercomputer Appliance
16x V100 | 2 PF | 512GB HBM2
8x MLNX IB



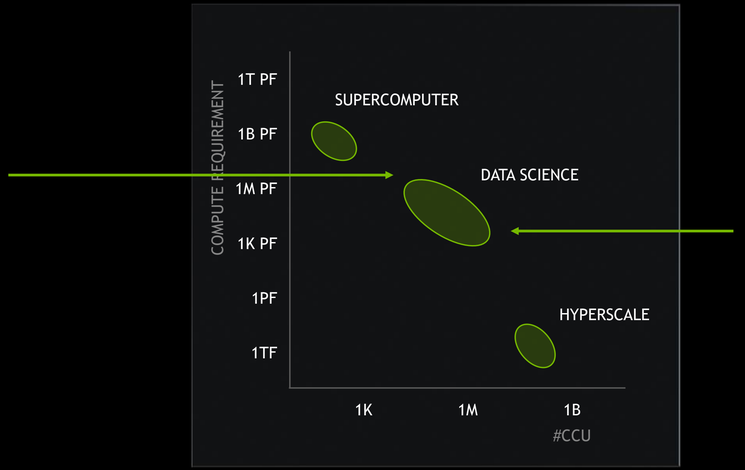
Hyperscale | Capacity Machine | Scale-out Architecture



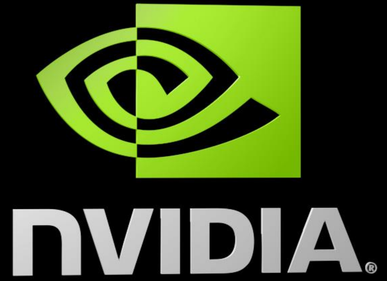
DATA SCIENCE – THE NEW HPC CHALLENGE



NVIDIA DGX-2
AI Supercomputer Appliance
16x V100 | 2 PF | 512GB HBM2
8x MLNX IB



Data Science Server
4x T4 | 260 TF FP16 | 64GB GDDR6
MLNX or BRCM EN


NVIDIA


Mellanox
TECHNOLOGIES

CUDA TO ARM

Energy-Efficient Supercomputing

NVIDIA GPU Accelerated Computing Platform On ARM

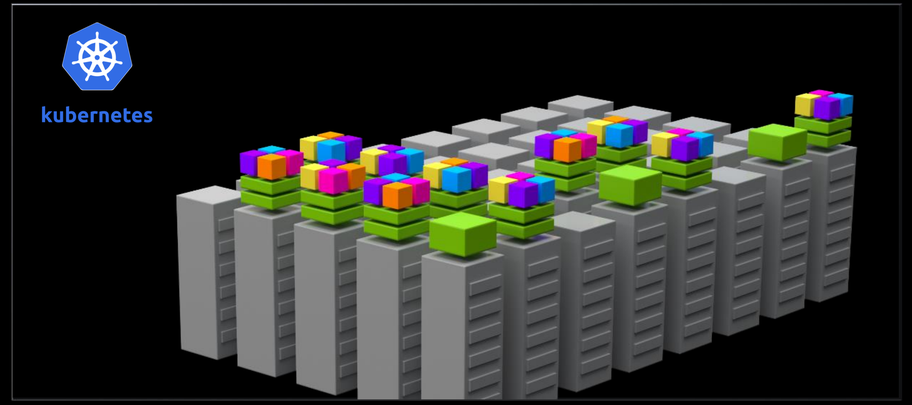
Optimized CUDA-X HPC & AI Software Stack

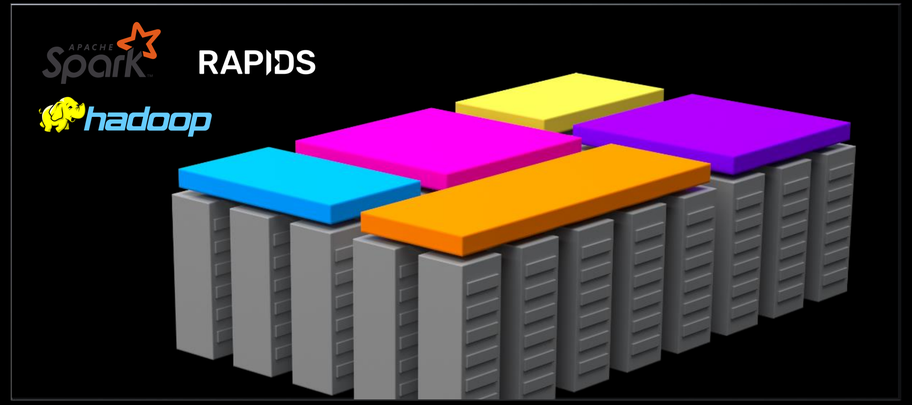
CUDA, Development Tools and Compilers

Available End of 2019

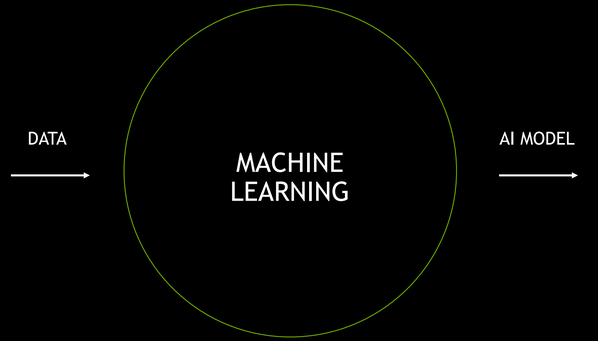
NVIDIA AI INFRASTRUCTURE

DATACENTER BECOMES A COMPUTE ENGINE





AI LEADERSHIP NEEDS AI INFRASTRUCTURE LEADERSHIP



SATURNV

The Worlds Largest Enterprise
AI Infrastructure Buildout

1500 DGX Nodes

12,600 GPUs

1.5 ExaFLOPs

5MW Average Power

NVIDIA DGX SUPERPOD

AI Leadership Requires
AI Infrastructure Leadership

Test Bed for Highest Performance Scale-Up Systems
9.4 PF on HPL | ~200 AI PF | #22 on Top500 list
<2 mins To Train RN-50

Modular & Scalable GPU SuperPOD Architecture
Built in 3 Weeks
Optimized For Compute, Networking, Storage & Software

Integrates Fully Optimized Software Stacks
Freely Available Through NGC

Autonomous Vehicles | Speech AI |
Healthcare | Graphics | HPC

96 DGX-2H

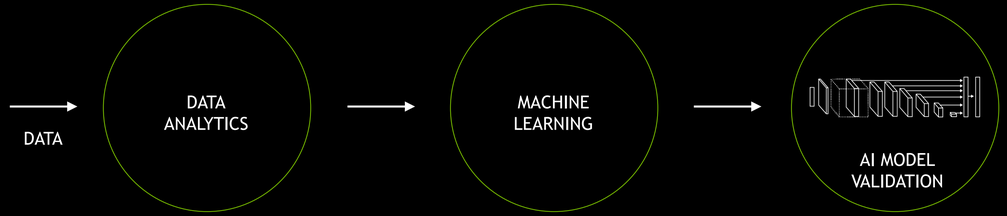
10 Mellanox EDR IB per node

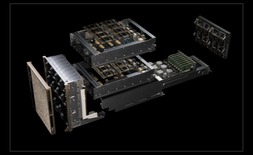
1,536 V100 Tensor Core GPUs

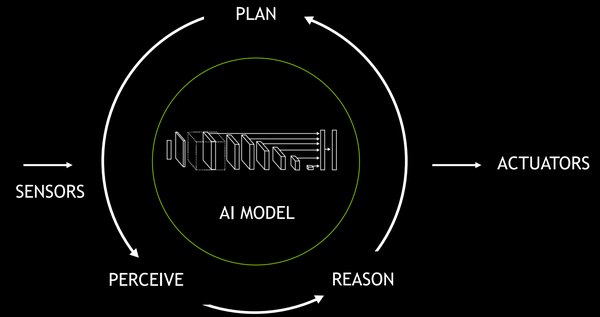
1 megawatt of power

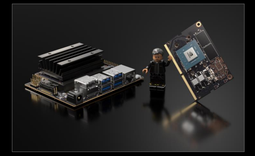
AI IN INDUSTRY

BUILDING AI & DEPLOYING AI



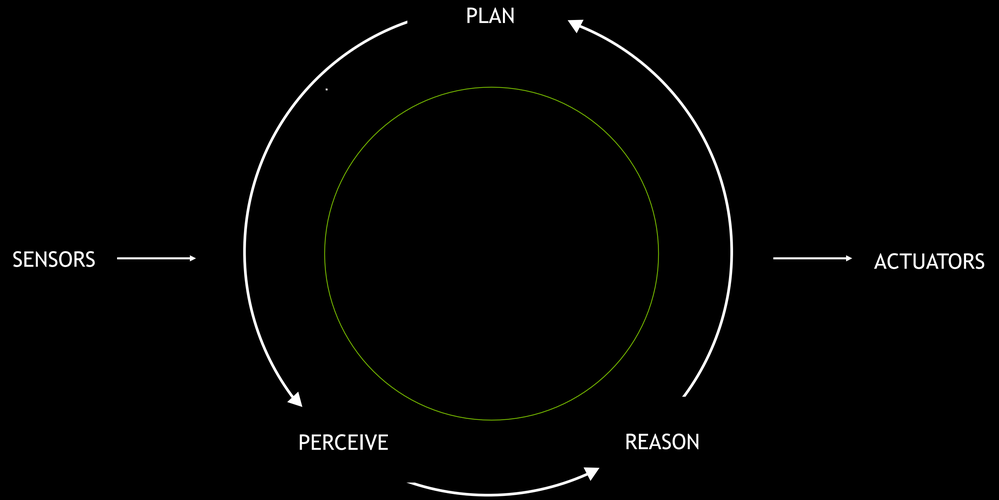




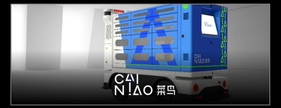


THIS IS AI

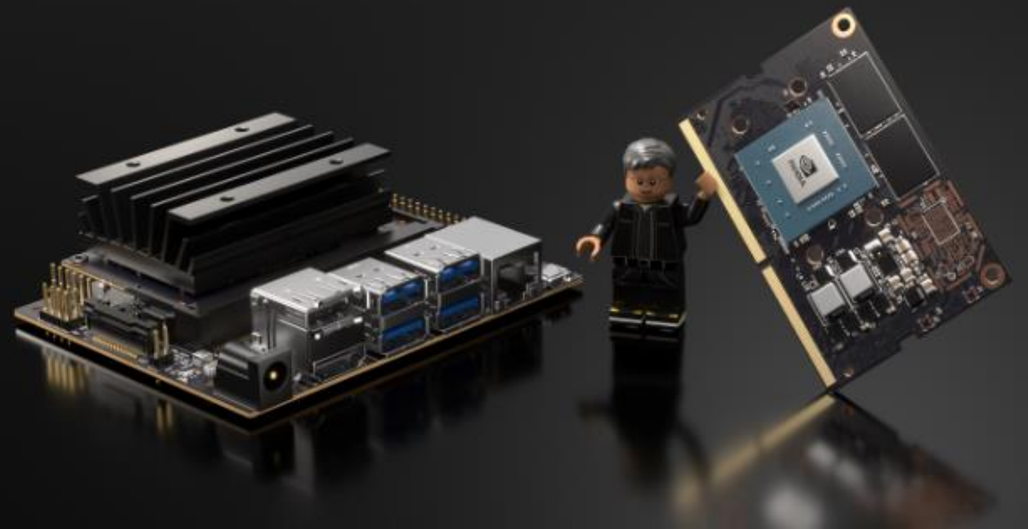
THIS IS AI



JETSON POWERING AUTONOMOUS MACHINES

WAREHOUSE	DELIVERY	AGRICULTURE	RETAIL	INDUSTRIAL
				
				
				
				

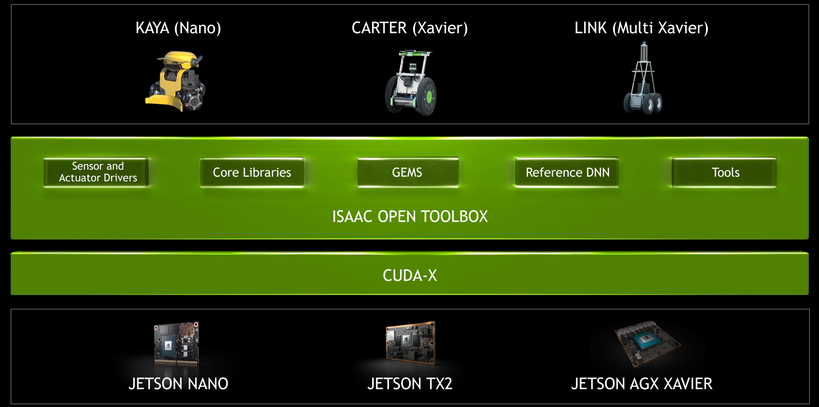
ANNOUNCING JETSON NANO



$99 NVIDIA CUDA-X AI COMPUTER

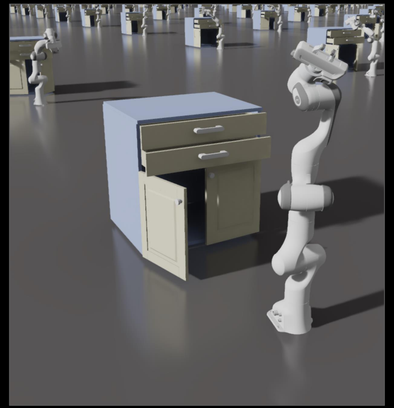
CUDA-X acceleration stack
High-resolution sensor support
Runs all CUDA-X AI models

ANNOUNCING ISAAC OPEN SDK



Isaac Robot Engine

Isaac Sim

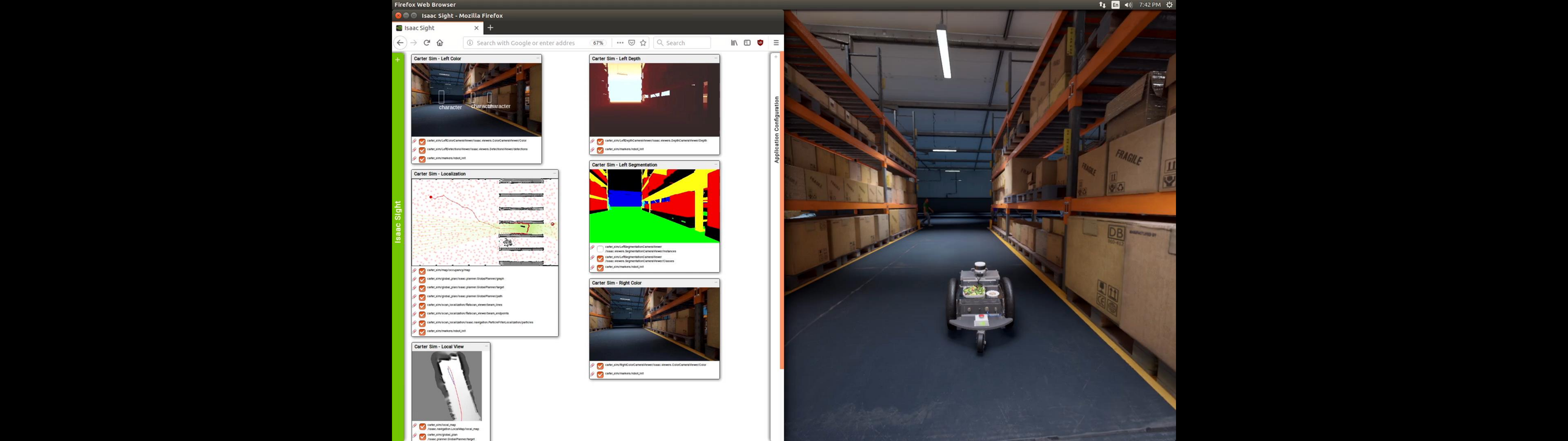


Isaac Gym

Isaac Robot Engine - Modular robot framework
Isaac Sim - Virtual robotics laboratory
Isaac Gym - Reinforcement learning simulator
Isaac Robot Apps - Kaya, Carter and Link

Available at developer.nvidia.com/isaac-sdk


Kaya


Firefox Web Browser
7:42 PM
Isaac Sight - Mozilla Firefox
Isaac Sight
Search with Google or enter addres
67%
Search
Isaac Sight
Carter Sim - Left Color
character
Carter Sim - Left Depth
Carter Sim - Left Segmentation
Carter Sim - Localization
Carter Sim - Right Color
Carter Sim - Local View
Application Configuration
FRAGILE

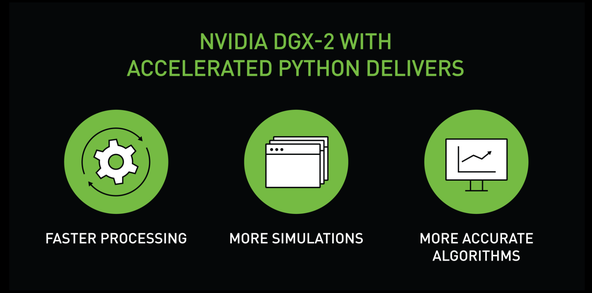
NVIDIA DELIVERS MORE THAN 6000X SPEEDUP ON BENCHMARK ALGORITHM FOR HEDGE FUNDS IN PYTHON

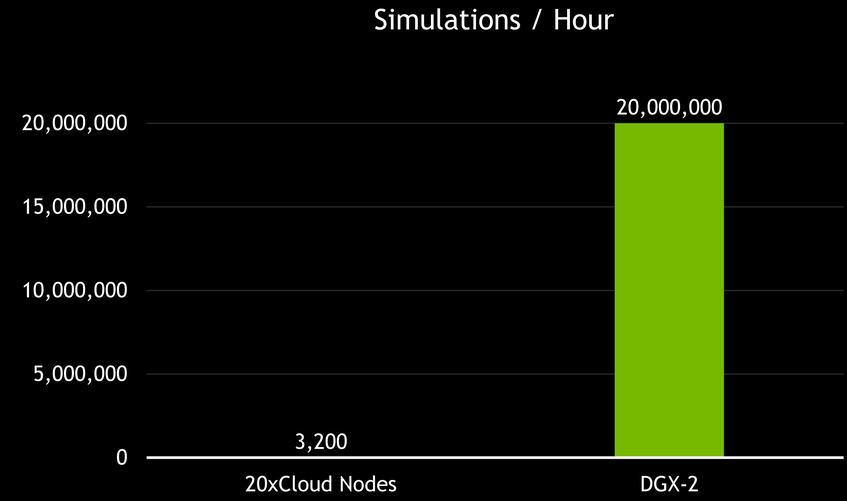
20,000,000 simulations per hour on an industry defined benchmark, compared to the prior record of 3,200.
Over 6,000 times faster on DGX-2

Powered by NVIDIA DGX-2, using all 16x V100 GPUs

Accelerated Python via the RAPIDS packages and Numba - replicate this performance without needing in-depth knowledge of GPU programming

How will you use that power - faster time to market?
More complex models? Test more scenarios? Some of each?





https://blogs.nvidia.com/blog/2019/05/13/accelerated-backtesting-hedge-funds/
www.STACresearch.com/news/2019/05/13/NVDA190425

REAL-TIME FRAUD DETECTION

Recently, PayPal was looking to deploy a new fraud detection system. The team working on it set a high bar: this system had to operate worldwide 24/7, and work in real-time to protect customer transactions from potential fraud. In spec'ing the system, it became evident that CPU-only servers couldn't meet these requirements.

Using NVIDIA T4 GPUs, PayPal delivered a new level of service, using GPU inference to improve real-time fraud detection by 10% while lowering server capacity by nearly 8x.

NVIDIA METROPOLIS

Smart City Platform

Billions of IoT sensors

The data lifeblood of a modern city

The fuel for AI software:

- Reducing traffic congestion
- Energy grid management
- Finding lost children
- Other new services

500M+

Sensors
Worldwide Today

$158B

Smart Cities
Funding by 2022

SOURCE: $158B, Smart Cities Initiative funding by 2022, IDC, "Worldwide Semiannual Smart Cities Spending Guide."

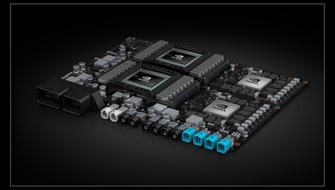
THIS IS AI

THE DRIVE INITIATIVE

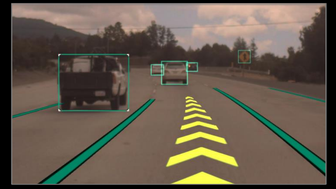
DGX Saturn V

Constellation

Xavier



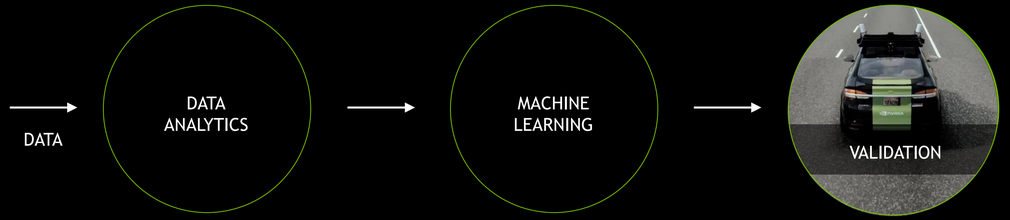
DRIVE AV

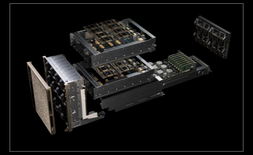


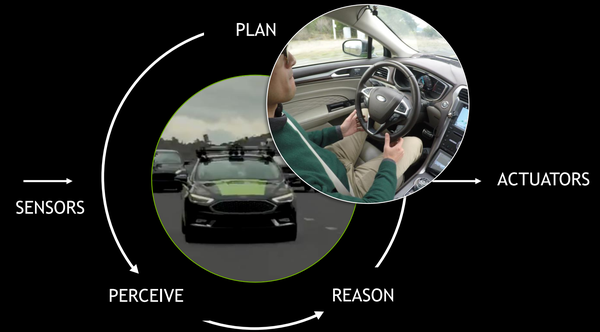
DRIVE IX

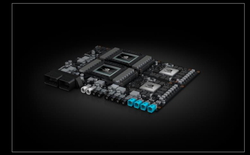
KITT Resim

NVIDIA SELF-DRIVING CAR



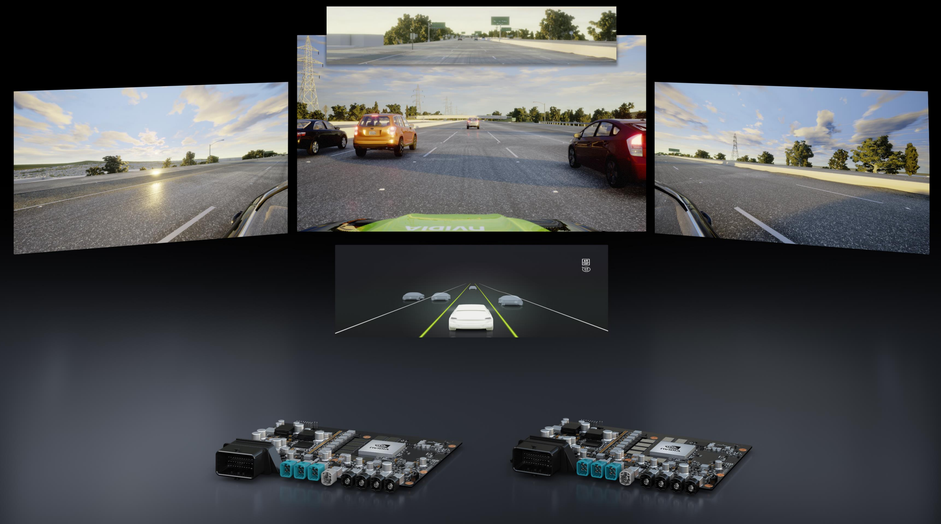






AI FOR TRANSPORTATION

AD is revolutionizing transportation
Saving lives
Reducing shipping costs
Reduced insurance costs
Vehicle of future is software defined
NVIDIA DRIVE – an open platform for research and production



1.5B
Vehicles in the World

$10T
Transportation Industry

AI FOR TRANSPORTATION: NVIDIA DRIVE

NVIDIA DRIVE
Highway Loop to NVIDIA
DECEMBER 20, 2018

77 MILES

6 HIGHWAY INTERCHANGES

34 LANE CHANGES

0 DISENGAGEMENTS

HEALTHCARE DATA IS ENORMOUS

The Perfect Fuel for AI

Genomics Data
2x/7Months

Instrument Data
3+ TB/day

Hospital Data
50 PB/Year

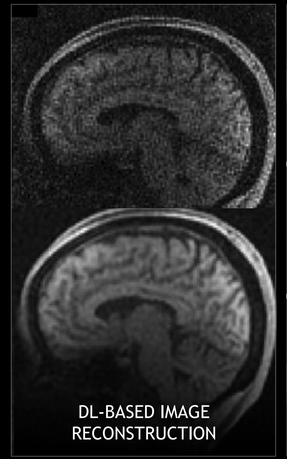
MEDICAL IMAGING

Essential tool of early detection and disease management

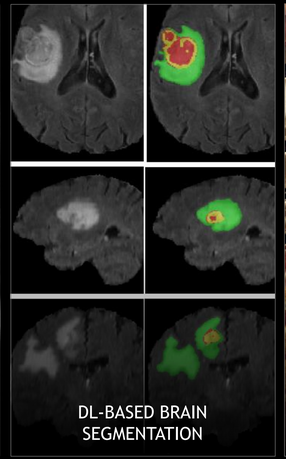
Demand outpacing supply of world's radiologists

Imaging field enormously complex

Perfect application for AI



DL-BASED IMAGE RECONSTRUCTION



DL-BASED BRAIN SEGMENTATION

CINEMATIC RENDERING

70%

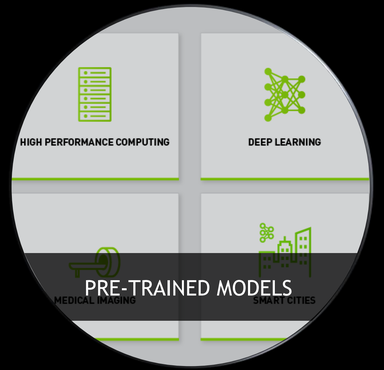
Medical Imaging Research based on DL Today

$8.6B

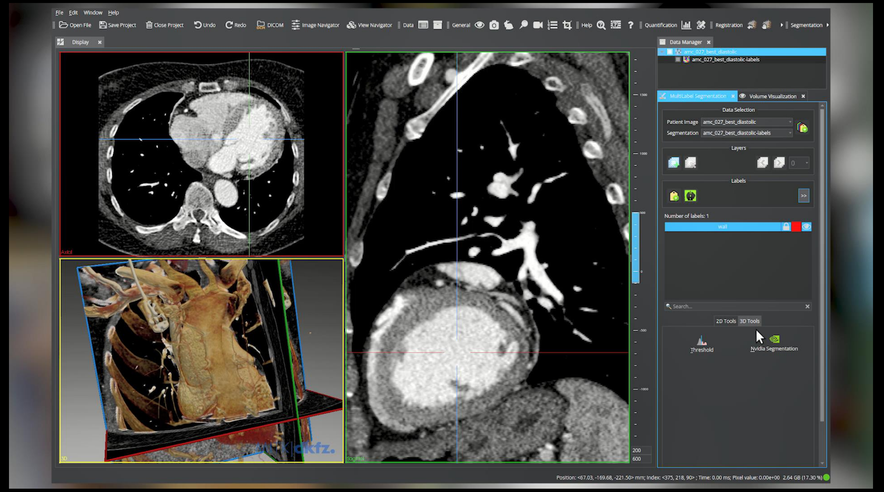
Annual Software Revenue for AI Use Cases by 2025

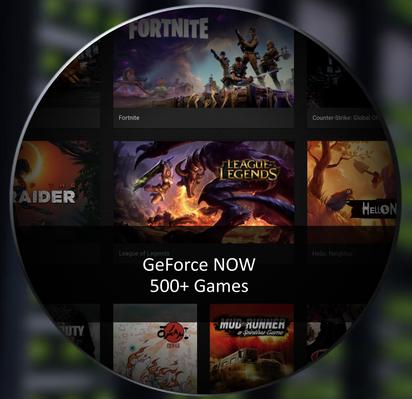
SOURCE: Global software revenue from 22 key healthcare AI use cases will grow to $8.6 billion annually by 2025, Tractica, "Artificial Intelligence for Healthcare Applications."

CLARA AI TOOLKIT



AI-ASSISTED ANNOTATION





GeForce NOW
500+ Games

15 Datacenters
300K Players
1M on Waiting List

SoftBank

LG U+

Announcing
GFN Alliance

ANNOUNCING RTX SERVER

Datacenter Graphics Server Design

40 Turing GPUs in 8U

Virtualize graphics apps up to 320 CCU

Optimized end-to-end stack for rendering, remote workstation, and cloud gaming

ANNOUNCING RTX SERVER POD

Modular Designs for Enterprise & Cloud Edge Datacenters

Pods scale to 32 RTX servers

1,280 GPUs in 10 racks

High-speed storage connected with MLNX IB

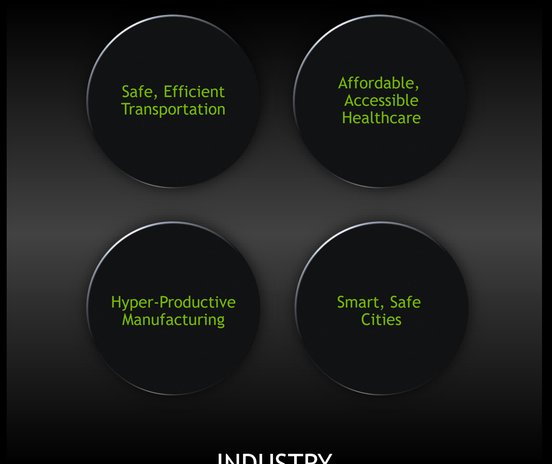
Up to 10,000 concurrent users per RTX Pod

NVIDIA RTX SERVER

PILLARS OF NATIONAL AI INITIATIVES

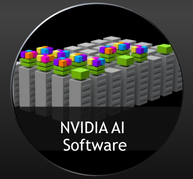
AI CENTERS OF EXCELLENCE

SKILLING AND RESKILLING



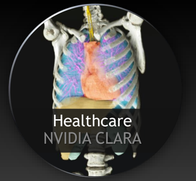
INDUSTRY SOLUTIONS

NVIDIA PARTNERSHIP FRAMEWORK



TECHNOLOGY &
ECOSYSTEM

EXPERTISE &
INVESTMENT



INDUSTRY SOLUTION
PLATFORMS


NVIDIA