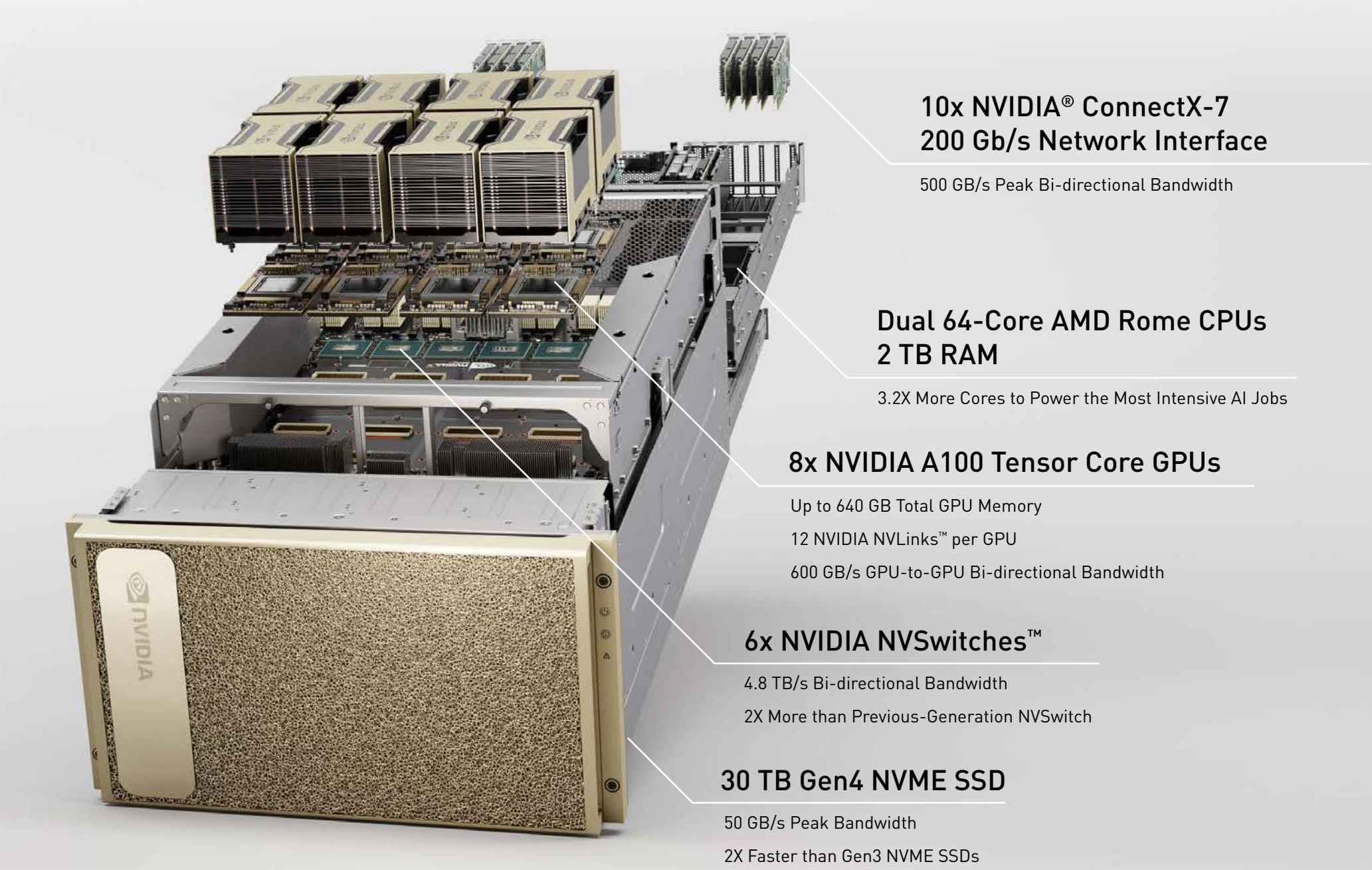# NVIDIA DGX A100
THE UNIVERSAL SYSTEM FOR AI INFRASTRUCTURE

THE BUILDING BLOCK OF THE AI DATA CENTER
AI INFRASTRUCTURE REIMAGINED, OPTIMIZED, AND READY FOR ENTERPRISE AI

10x NVIDIA® ConnectX-7
200 Gb/s Network Interface
500 GB/s Peak Bi-directional Bandwidth

Dual 64-Core AMD Rome CPUs
2 TB RAM
3.2X More Cores to Power the Most Intensive AI Jobs

8x NVIDIA A100 Tensor Core GPUs
Up to 640 GB Total GPU Memory
12 NVIDIA NVLinks™ per GPU
600 GB/s GPU-to-GPU Bi-directional Bandwidth

6x NVIDIA NVSwitches™
4.8 TB/s Bi-directional Bandwidth
2X More than Previous-Generation NVSwitch

30 TB Gen4 NVME SSD
50 GB/s Peak Bandwidth
2X Faster than Gen3 NVME SSDs

1 TRILLION	1 KILOMETER	1 MILLION	30 THOUSAND	50 THOUSAND	5 PETAFLOPS
Transistors	of Traces	Drill Holes	Components	Connector Pins	of AI Performance

GAME-CHANGING PERFORMANCE FOR INNOVATORS
NVIDIA DGX SUPERPOD: UNLOCKING AI AT SCALE FOR EVERY ENTERPRISE

- NVIDIA DGX™ A100 Systems
- Modular, with 20 Node Scalable Units
- NVIDIA Quantum-2 InfiniBand Switches
- Optimized Software from NVIDIA NGC™
- Petabytes of All-Flash Storage
- Deployable in Weeks, Instead of Months

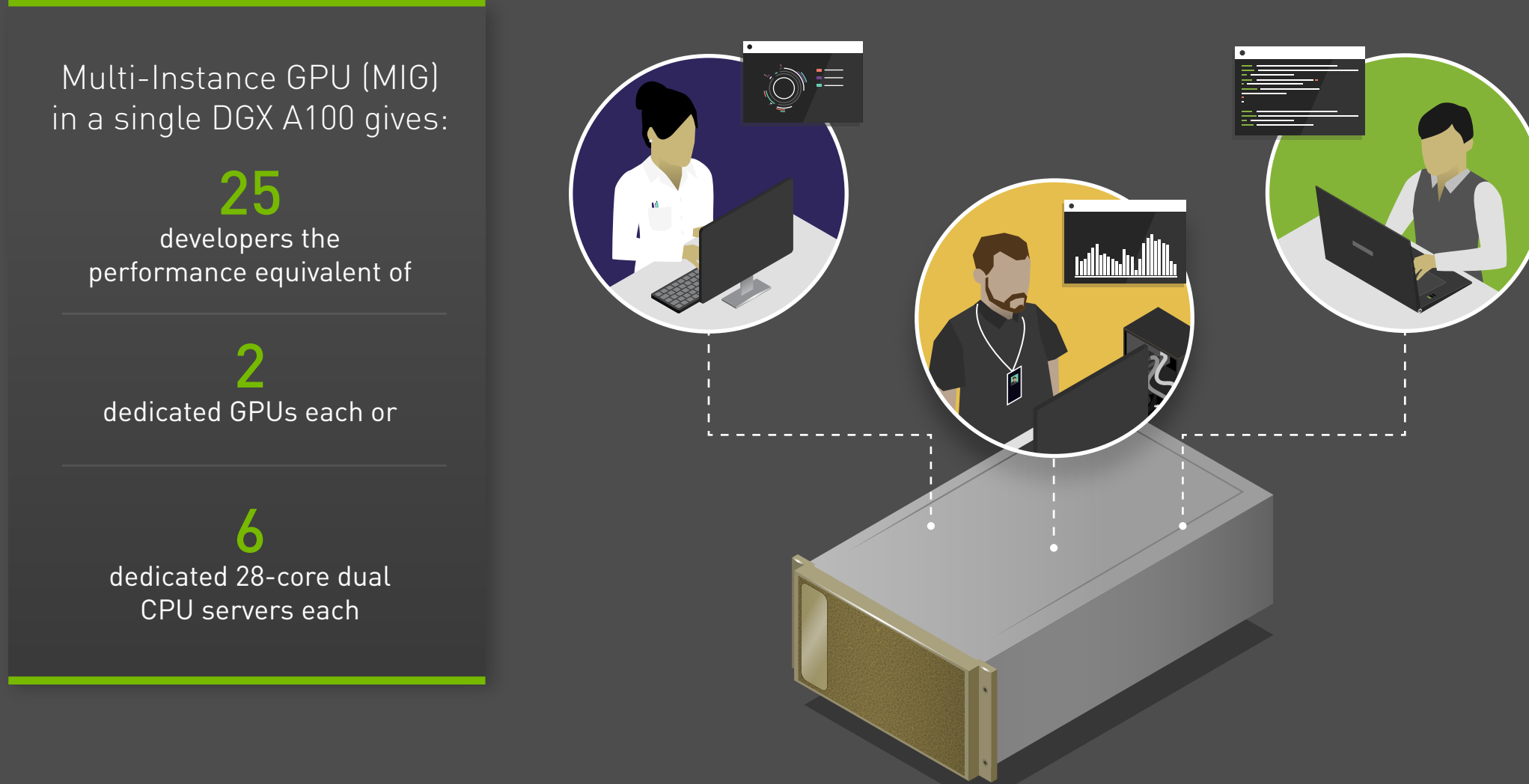## ACCELERATION FOR EVERY AI WORKLOAD
RUN DATA ANALYTICS, TRAINING, AND INFERENCE WORKLOADS ON THE SAME SYSTEM

TRADITIONAL INFRASTRUCTURE IS CONSTRAINED
Infrastructure silos starve AI workloads or waste capacity.

CLUSTER 1: TRAINING | CLUSTER 2: ANALYTICS | CLUSTER 3: INFERENCE

DGX A100 INFRASTRUCTURE IS AGILE
Run any workload on any system for maximized utilization.

UNIFIED AI INFRASTRUCTURE

Today's Workload | Tomorrow's Workload | Next Week's Workload
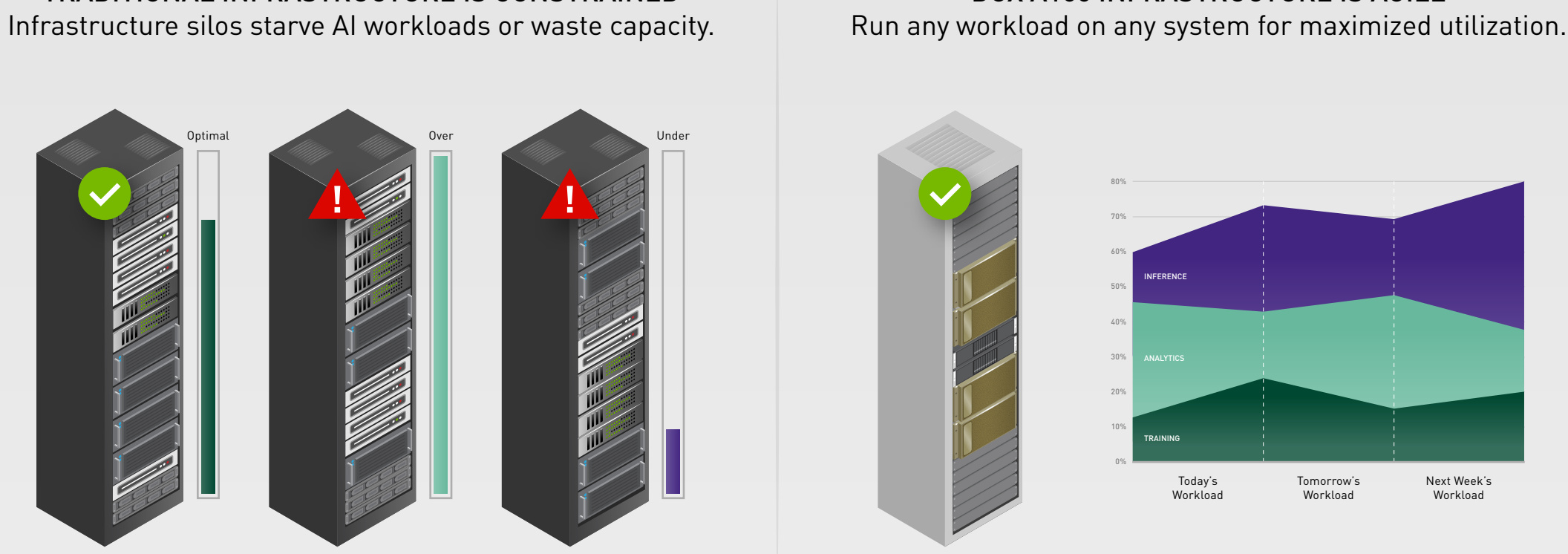
THE MOST POWERFUL TOOL FOR DATA SCIENCE TEAMS
GIVE EVERY DEVELOPER THE POWER TO EXPLORE

Multi-Instance GPU (MIG) in a single DGX A100 gives:

25 developers the performance equivalent of

2 dedicated GPUs each or

6 dedicated 28-core dual CPU servers each

DIRECT ACCESS TO A GLOBAL TEAM OF NVIDIA DGXPERTS
GET UNMATCHED AI EXPERTISE WITH EVERY DGX SYSTEM

INSPIRATION → ? → INSIGHTS

NVIDIA With You Every Step of the Way
DESIGN / PLAN / BUILD / TEST / DEPLOY / OPERATE / MONITOR

10+	100+	20,000+
Years of AI Innovation	GPU-Optimized Software and Tools on NGC	of AI-Fluent Practitioners Around the Globe

ACCELERATE AI, ACCELERATE ROI
DEPLOY FLEXIBLE AI INFRASTRUCTURE THAT REINS IN CAPEX

1/10th THE COST
1/20th THE POWER
1/25th THE SPACE

TODAY'S AI DATA CENTER
$11 Million | 25 Racks | 630 kW

DGX A100 DATA CENTER
$1 Million | 1 Rack | 32.5 kW
5 DGX A100 Systems for AI Training and Inference
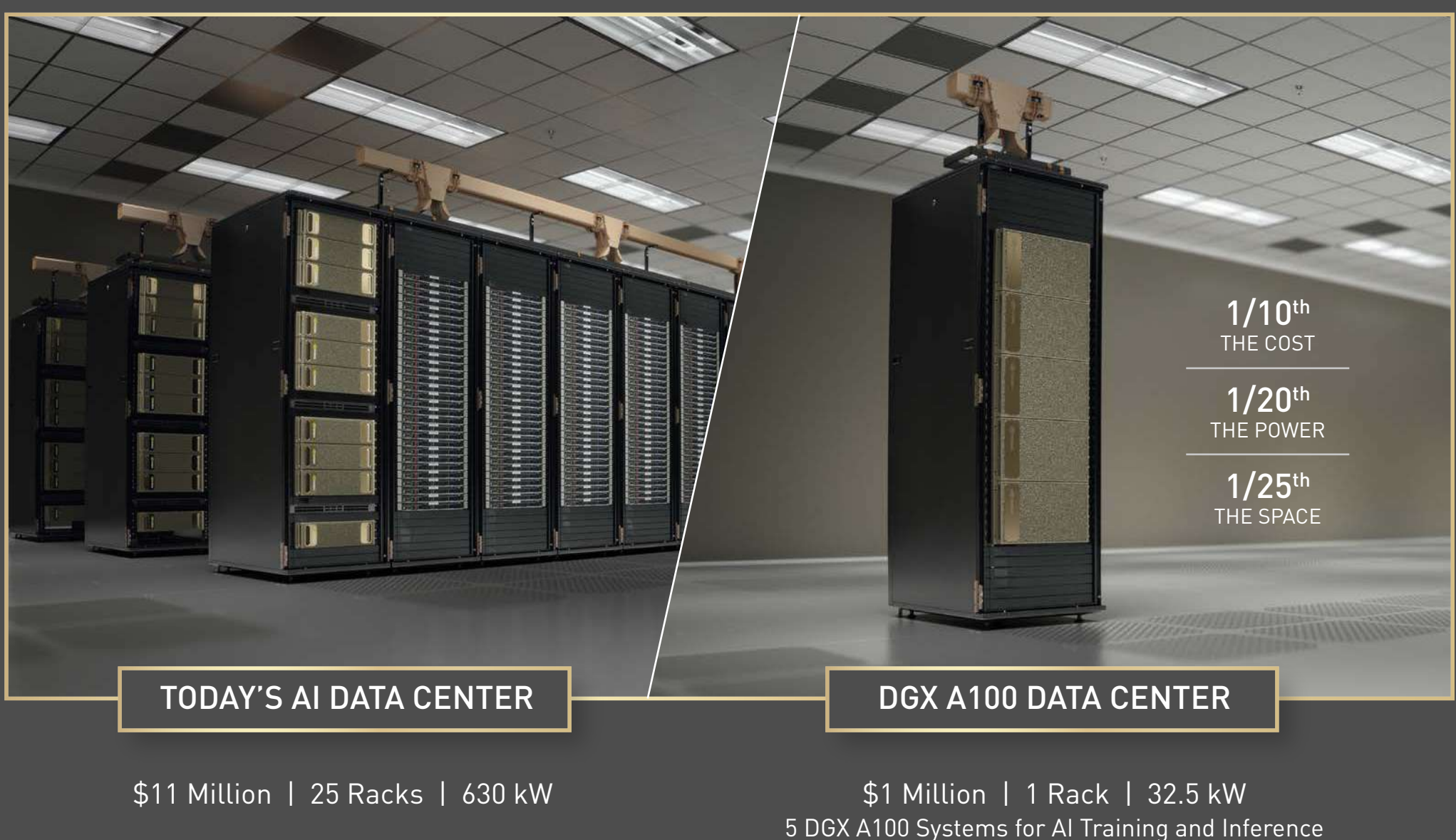
Unlock AI in Your Organization Today
www.nvidia.com/DGXA100

◢ NVIDIA