NVIDIA DGX STATION™ A100
A WORKGROUP APPLIANCE FOR THE AGE OF AI

DATA CENTER PERFORMANCE WITHOUT THE DATA CENTER

BIGGER MODELS, FASTER ANSWERS (UNPARALLELED PERFORMANCE)

TRAINING

BERT Large Pre-Training Phase 1 (Relative Performance)
DGX Station A100 320GB; Batch Size=64; Mixed Precision; With AMP; Real Data; Sequence Length=128

OVER 4X FASTER

OVER 3X FASTER

INFERENCE

BERT Large Inference (Relative Performance)
DGX Station A100 320GB; Batch Size=256; INT8 Precision; Synthetic Data; Sequence Length=128, cuDNN 8.0.4

3.17X

LINEAR SCALABILITY

4 GPUs

2 GPUs

1 GPU

3,975

2,066

7,666

MULTI/GPU SCALABILITY

ResNet-50 V1.5 Training (Images per Second)
DGX Station A100 320GB; Batch Size=192; Mixed Precision; Real Data; cuDNN Version=8.0.4; NCCL Version=2.7.8; NGC MXNet 20.10 Container

4.35X

3X FASTER

3X FASTER

A POWERFUL TOOL FOR DATA SCIENCE TEAMS
A SHARED SYSTEM WITHOUT LIMITS
TRAINING, INFERENCE, DATA ANALYTICS

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

Wool Coat

CUSTOMERS ALSO BOUGHT