AIRI

AI-READY INFRASTRUCTURE

From **PURE**STORAGE®

Powered by **NVIDIA**

NVIDIA PREFERRED PARTNER

PURESTORAGE®

The Importance of Data in AI

Matt Oostveen

CTO Pure Storage APJ

PURESTORAGE®

MATTHEW OOSTVEEN

Chief Technology Officer: APJ PURE STORAGE

Working on future systems

Career: IBM, Microsoft, IDC, DellEMC

WHAT YOU'LL LEARN TODAY

The Role of Data in AI

Our Technical motivation

AI Tuned Infrastructure

PURESTORAGE®

THE
NEW WORLD
OF DATA

ZB   CORE/EDGE   THINGS   REAL-TIME   INTELLIGENCE

PURESTORAGE®

What is the difference between AI & ML?
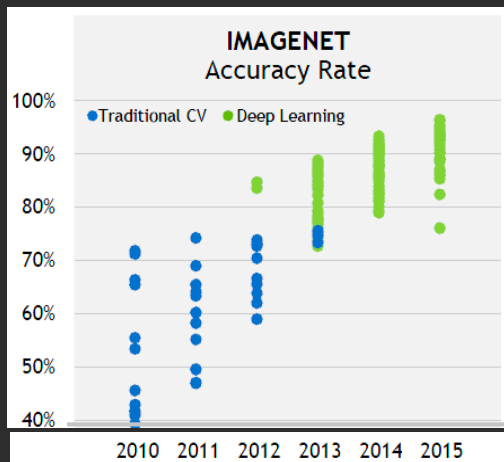
If it's written in <u>Python</u> it's probably <u>ML</u>

If it's written in <u>PowerPoint</u> it's <u>AI</u>

PURESTORAGE®

THE BIG BANG OF INTELLIGENCE
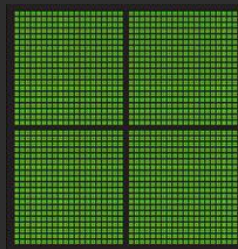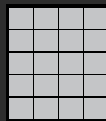
FUELED BY PARALLEL COMPUTE, NEW ALGORITHMS, AND BIG DATA

NEW ALGORITHMS
Massively Parallel Delivering
Superhuman Accuracy

MODERN COMPUTE
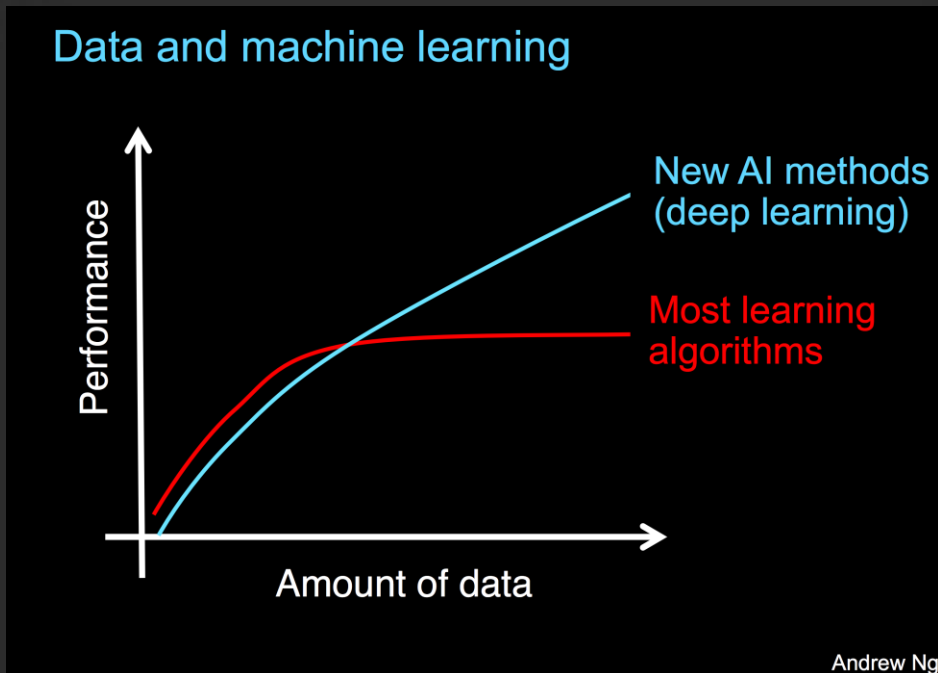Massively Parallel Architecture
Driving Performance

BIG DATA
Data is the New Oil
163 Zettabytes Created in 2025

GPU- *THOUSANDS OF CORES*

PURESTORAGE®

DATA IS VITAL TO MACHINE LEARNING

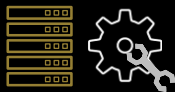OBSERVATION BY PROF. ANDREW NG, AI LUMINARY

DO IT YOURSELF
IS OFTEN THE ONLY OPTION

Never-ending cycles of compiling and tuning open source software

Months of system building and tuning, constant maintenance

Yet legacy solutions full of data bottlenecks, from storage to GPU to apps

DEEP LEARNING = COMPLEXITY

AI complexity: how many skilled researchers are required and available to build high-performance models?

Infrastructure complexity: how many people do I need to deploy, manage, and scale systems for deep learning?

Performance complexity: how much time is spent tuning and configuring pipelines to keep GPUs fed with data?

PURESTORAGE®

COMPLEXITY IS NOT SUSTAINABLE

MONTHS OF SYSTEM BUILDING AND TUNING, CONSTANT MAINTENANCE

NEVER-ENDING CYCLES OF TUNING MODELS & CODE

AIRI IN DEPTH

PURESTORAGE®

AIRI
THE INDUSTRY'S FIRST

COMPLETE AI-READY INFRASTRUCTURE

HARDWARE

NVIDIA® DGX-1™ | 4x DGX-1 Systems | 4 PFLOPS of DL Performance
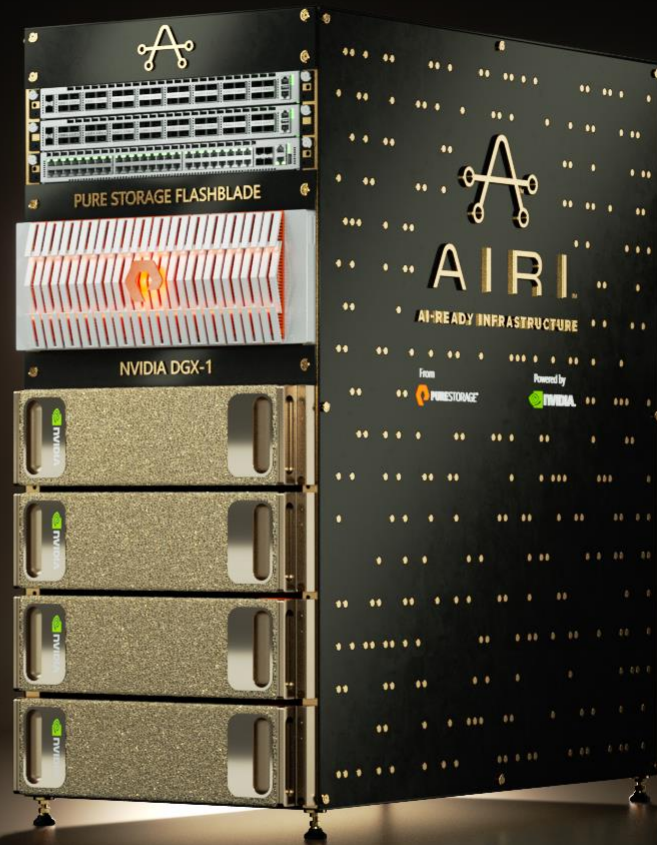
PURE FLASHBLADE™ | 15x 17TB Blades | 1.5M IOPS

ARISTA | 2x 100Gb Ethernet Switches with RDMA

SOFTWARE

NVIDIA GPU CLOUD DEEP LEARNING STACK | NVIDIA Optimized Frameworks

AIRI SCALING TOOLKIT | Multi-node Training Made Simple

LINKING DATA WITH OUTCOMES

PURE STORAGE FLASHBLADE

A I R I

AI-READY INFRASTRUCTURE

PURESTORAGE

Powered by NVIDIA.

NVIDIA DGX-1

Final rendering TBD

UNIFIED ETHERNET FABRIC
Delivers the performance of RDMA, while simplifying integration into existing data centers

HIGH-PERFORMANCE DATA PLATFORM
Keeps GPUs fed with data for efficient scaling of storage and compute resources

CONFIGURATION GUIDE & SCALING TOOLKIT
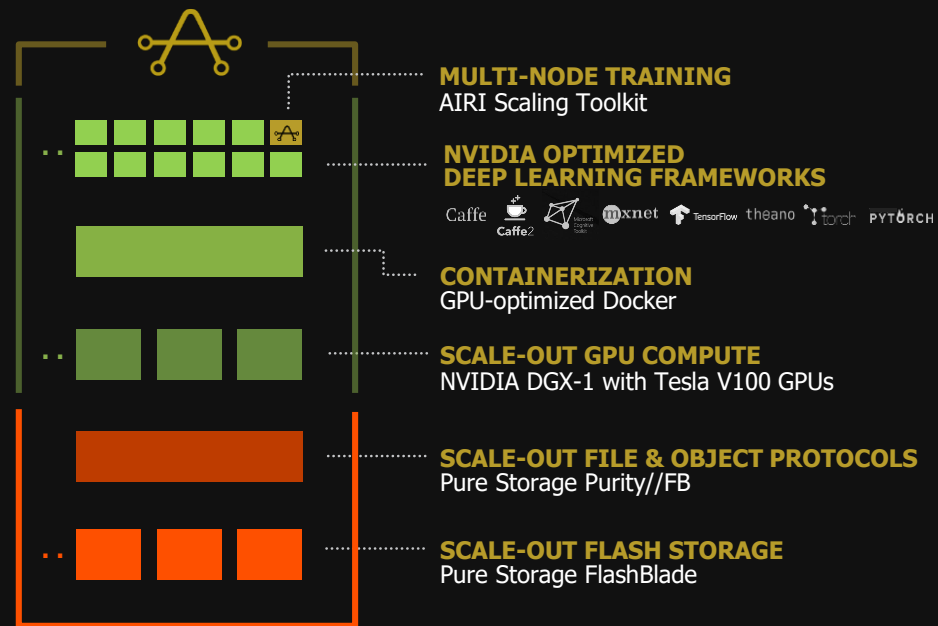Simplify the deployment and validation of high-performance infrastructure for deep learning

PURESTORAGE®

DATA BOTTLENECK ELIMINATED

AIRI™ SLASHES TRAINING TIME BY 4X, BOOST DATA SCIENTIST PRODUCTIVITY

RESNET-50

10244 i/s

4870 i/s

2660 i/s
2540 i/s

1 DGX-1
1 DGX-1 2 DGX-1 4 DGX-1

Synthetic Data Mode
*(Theoretical Maximum
for DGX-1 Performance)*

Multi-Node Training

INCEPTION3

6440 i/s

3160 i/s

1700 i/s
1600 i/s

1 DGX-1
1 DGX-1 2 DGX-1 4 DGX-1

Synthetic Data Mode   Multi-Node Training

VGG16

6300 i/s

3110 i/s

1660 i/s
1640 i/s

1 DGX-1
1 DGX-1 2 DGX-1 4 DGX-1

Synthetic Data Mode   Multi-Node Training
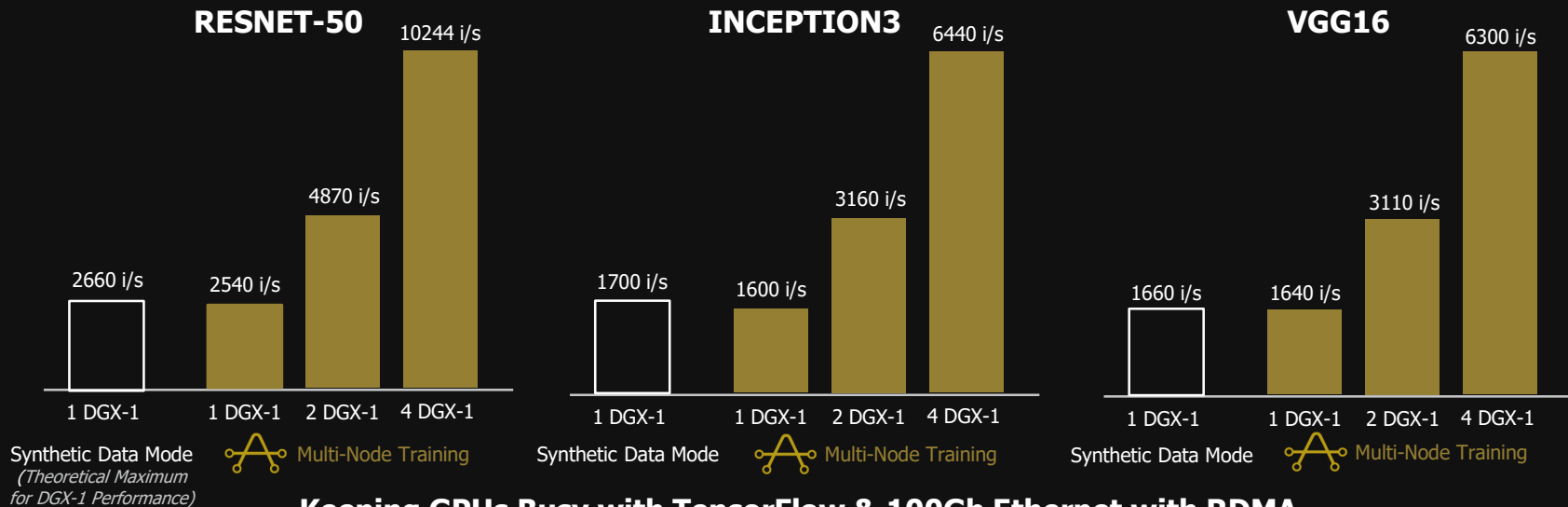
Keeping GPUs Busy with TensorFlow & 100Gb Ethernet with RDMA

AI-AT-SCALE FOR EVERY ORGANISATION

AIRI™ TO EXTEND THE POWER OF NVIDIA® DGX-1™ SYSTEMS

PURE STORAGE FLASHBLADE

NVIDIA DGX-1

INDUSTRY'S FIRST TO SIMPLIFY AI-AT-SCALE
Data scientist teams can focus on algorithms, not infrastructure

50 RACKS UNDER 50 INCHES
Performance of entire data center for each data team

SLASH TRAINING FROM MONTH TO WEEK
Only a few experts can run multi-node training, AIRI makes it simple

AIRI SCALING TOOLKIT

CONFIGURATION AND DEPLOYMENT GUIDE
Guide to enabling RDMA over Converged Ethernet and end-to-end best practices for configuration of storage, networking, and compute.

SCALING TOOLKIT
Tools to reproduce benchmark results and validate deployment of end-to-end AIRI environment.

PURESTORAGE®

PURE STORAGE FLASHBLADE

A

AIRI

AI-READY INFRASTRUCTURE

PURESTORAGE®

Powered by NVIDIA.

NVIDIA DGX-1

Final rendering TBD

KEY TAKEAWAYS

I. Architect for data acquisition, cleaning, exploration, training, and model validation

II. Design infrastructure to scale with the sophistication of data pipelines and models

III. Serve models at scale using best-of-breed tools that support operations

PURESTORAGE®

AIRI

AI-READY INFRASTRUCTURE

From **PURE**STORAGE®

Powered by **nVIDIA.**