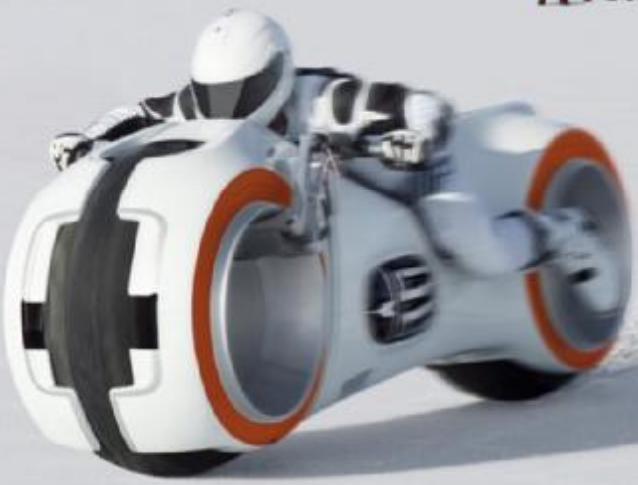
AI, Big Data and Education

Brent Franich Regional Director AP FlashBlade



PURESTORAGE®

TOP EDUCATION CHALLENGES

- Deliver relevant skills for 21st century demands
- Provide flexible learning platform – AI course maps
- Accelerate innovation in computer-assisted learning & communications
- Support AI by launching new, fast platforms for big data science & analytics
- Establish & leverage Partnerships in Government & Industry
- Curriculum for those that affected by automation
- Look at & plan for “what’s next”

© 2017 PURE STORAGE INC.

ARTIFICIAL NTELLIGENCE

Google to invest in Taiwan's AI talent evolution2018-03-22
Alphabet Inc's Google yesterday announced a large investment in Taiwan's artificial intelligence (AI) development, saying it would hire at least 300 AI engineers, cultivate 5,000 talents and train more than 50,000 digital marketing personnel this year.

Judicial officials assess new technology2018-03-22
The Ministry of Justice and the Judicial Yuan yesterday presented assessments on proposed measures and programs for judicial reform, with officials agreeing to look into the application of artificial intelligence (AI) and other new technologies to ameliorate deficiencies in the nation's judicial system. Legal professionals have started to use AI to .

New UK program offers R&D links2018-03-22
A Taiwan-UK innovative industries program was launched in Taipei yesterday to improve research and development (R&D) cooperation between the two nations. The British government is to spend £200,000 (US$281,348) in the program's first year to bolster collaboration in key areas including biotechnology, artificial intelligence, robotics, clean energy and autonomous vehicles. ..

Microsoft looking to hire Taiwanese with artificial intelligence knowledge2018-03-19
Microsoft Corp is focusing on Taiwanese specializing in artificial intelligence (AI) during its current recruitment season, a Microsoft Taiwan official said on Wednesday. Michael Chang (張仁炯), chief executive of the new Microsoft AI research and development center in Taiwan, said the company wants to recruit Taiwanese students because of the ...

Forget phonetics, the future is AI2018-03-17
Democratic Progressive Party (DPP) Legislator Yeh Yi-jin (葉宜津) on Thursday last week threw her hat into the ring as a candidate in the party's primary for Tainan mayor. So far, Yeh has advocated doing away with the Zhuyin Fuhao (注音符號, commonly known as Bopomofo) phonetic system — the primary method ..

© 2017 PURE STORAGE

AI REVOLUTIONIZING EVERY INDUSTRY

Automotive- Volvo

Volvo and its subsidiary, Zenuity, are racing to build fully autonomous cars by 2021

Farming- Blue River

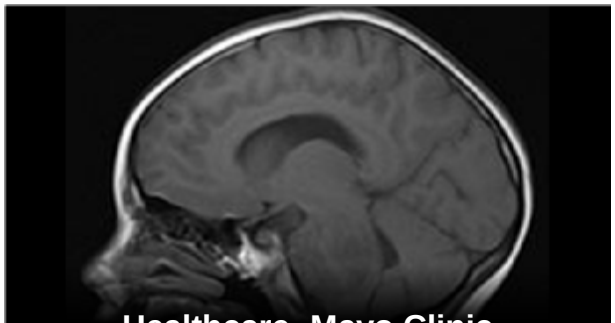
10% of lettuce in the US is harvested by LettuceBot, using AI to maximize crop yield & minimize chemicals

Kitchen- LG

LG Instaview refrigerator with AI-powered Alexa helps owners shop for groceries with their voice

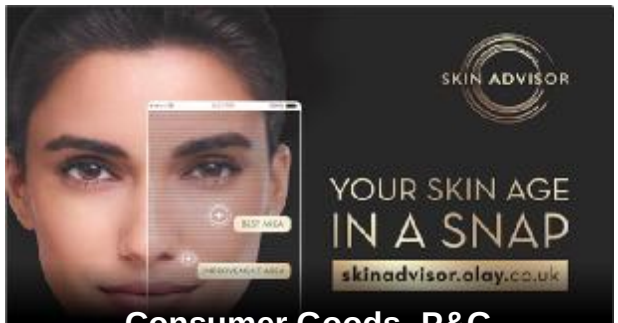
Marketing- SAP

SAP's Brand Impact tool measures ROI of brand sponsorships using video analytics and AI



Healthcare- Mayo Clinic

Finds genetic markers in images to avoid surgery for tumor samples & recommend treatments



Consumer Goods- P&G

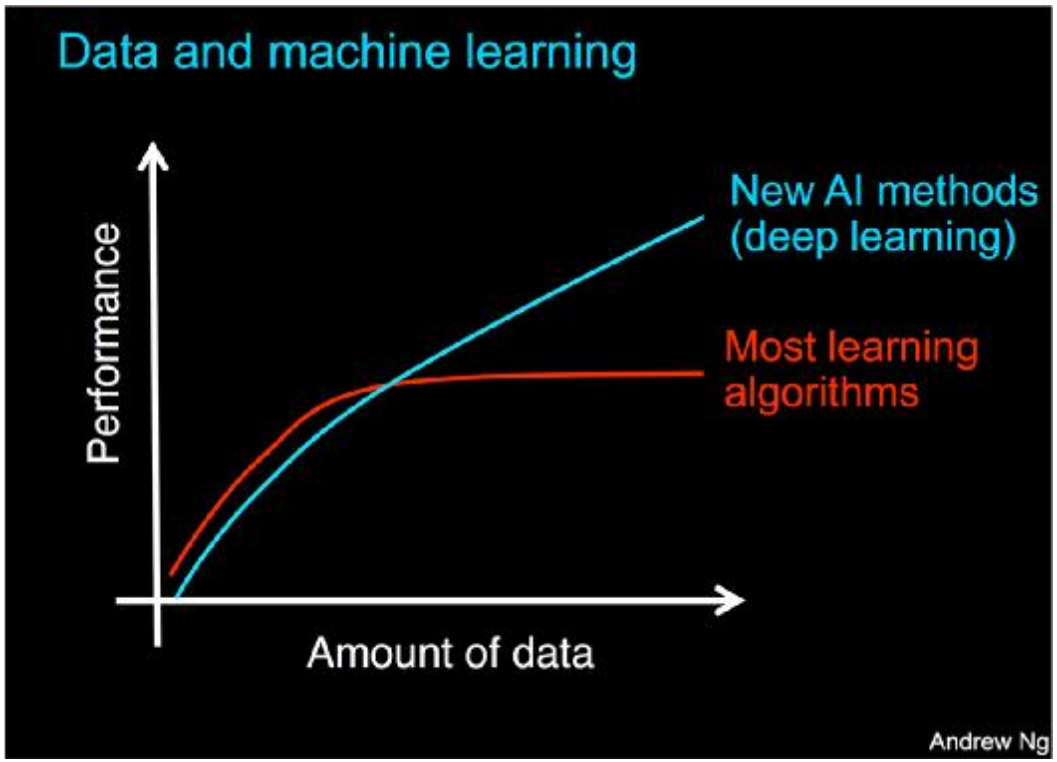
Proctor & Gamble's Olay is using AI to inspect skin & improve trouble spots, all with a mobile phone

© 2017 PURE STORAGE INC.

© 2017 PURE STORAGE INC.

DATA IS VITAL TO MACHINE LEARNING

OBSERVATION BY PROF. ANDREW NG, AI LUMINARY



© 2017 PURE STORAGE INC.

MODERN DATA CHANGES EVERYTHING

DECADE AGO

TODAY

DATA IS NOW DIFFERENT

Small to Large Files

Random to Sequential Access

Real-time or Batched

Apps & Data Evolve Quickly

Elastic Infrastructure

© 2017 PURE STORAGE INC.

STORAGE FOR AI:
BOTH A TRUCK AND A RACE CAR

CAPACITY
LARGE FILES
THROUGHPUT
SEQUENTIAL ACCESS

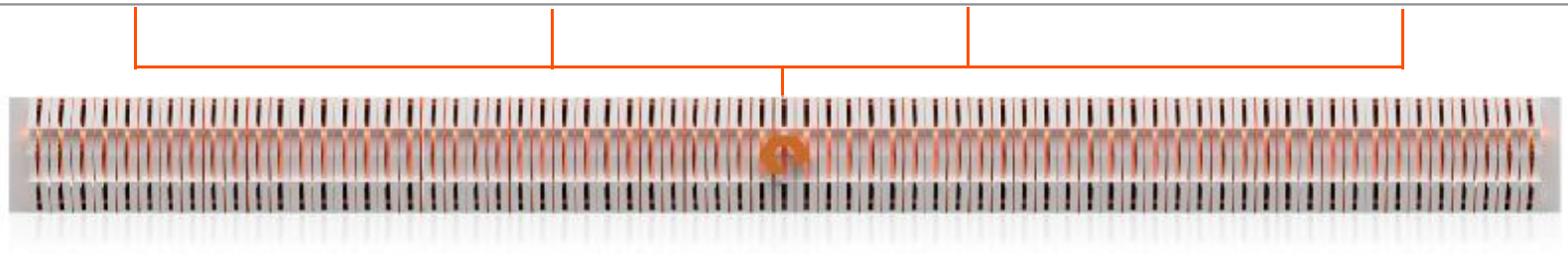
CONCURRENCY
SMALL FILES
LATENCY
RANDOM ACCESS

© 2017 PURE STORAGE INC.

WIDE RANGE OF NEEDS IN AI PIPELINE

SIGNIFICANT CHALLENGE TO LEGACY STORAGE, NOT FOR FLASHBLADE

	INGEST	CLEAN & TRANSFORM	EXPLORE	TRAIN
	From sensors & machines	CPU Servers	GPU Server	GPU Production Cluster
Access Pattern	sequential	sequential or random	random	random
Access Type	write	read & write	read	read
File Size	metadata is small data is small to large	small to large	small to large	small to large
Concurrency	depends on # of sources	high	low	high



REAL-WORLD AI

BUILDING A DATA PIPELINE IS CRITICAL

> “Building and managing data pipelines is typically one of the most costly pieces of a complete machine learning solution. ”

Jeremy Hermann & Mike Del Balso
Uber Machine Learning Platform
https://eng.uber.com/michelangelo/

> “If your boss asks you, tell them that I said [to] build a unified data warehouse. ”

Andrew Ng
Professor at Stanford University, Luminary of AI
Nuts and Bolts of Applying Deep Learning

© 2017 PURE STORAGE INC.

COMPLEXITIES OF AI IN PRODUCTION

INGEST

From sensors, machines, & user generated

CLEAN & TRANSFORM

Label, anomaly detection, ETL, prep, stage

CPU Servers

EXPLORE

Quickly iterate to converge on models

GPU Server

TRAIN

Run for hours to days in production cluster

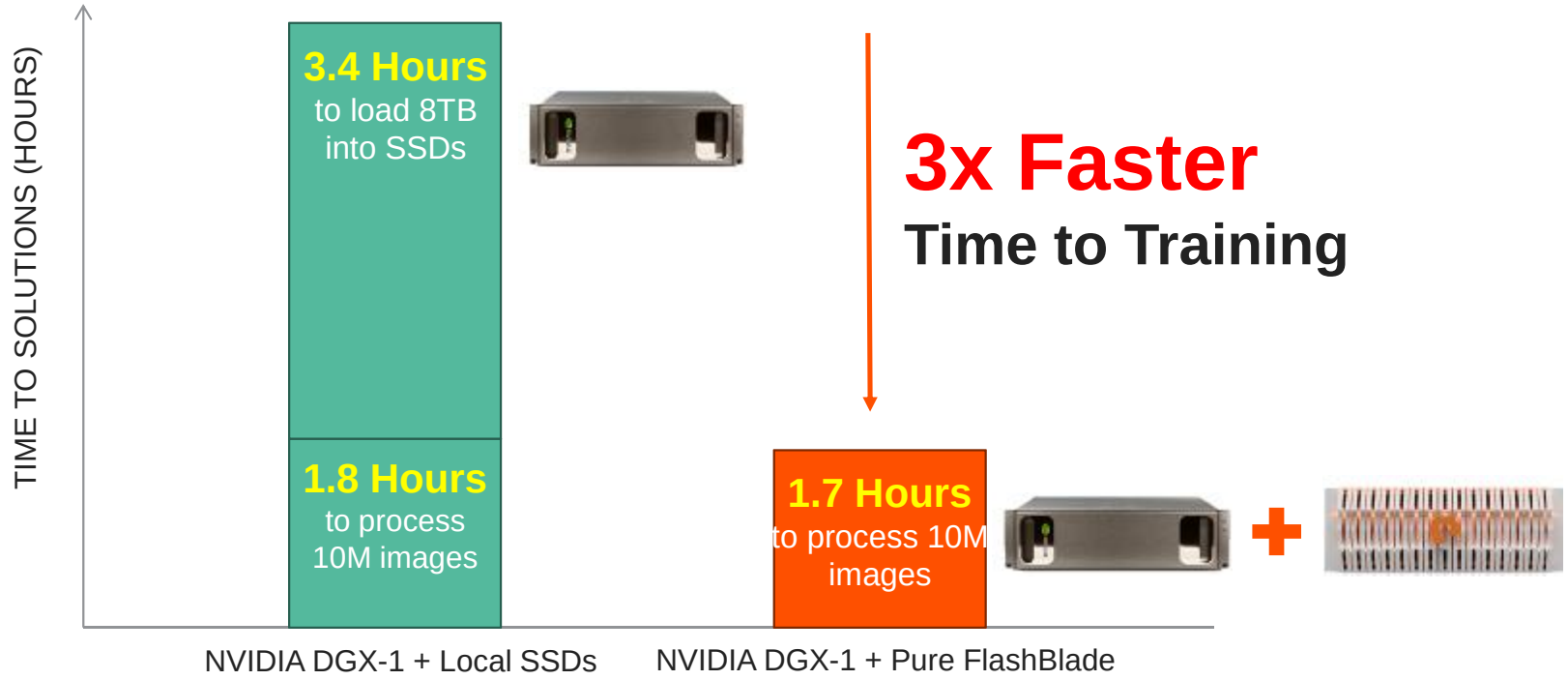
GPU Production Cluster

© 2017 PURE STORAGE INC.

PURESTORAGE

3X FASTER END-TO-END

TENSORFLOW TRAINING BENCHMARK WITH RESNET-50



© 2017 PURE STORAGE INC.

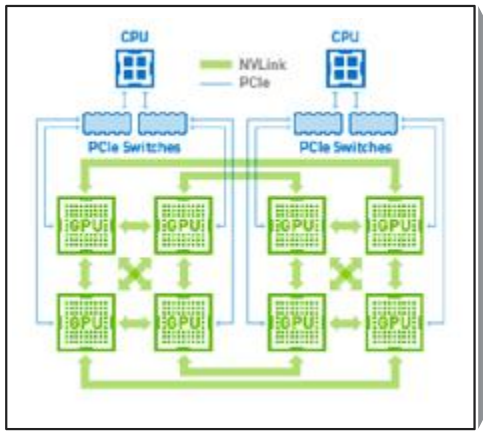
DELIVERING DATA THROUGHPUT FOR AI

Deep Learning Needs Maximum Read Performance, Mostly Small Files, To Keep Training Computers Busy

DGX-1

13K Images/Sec for each DGX-1

Assume 115KB on average for images



FLASHBLADE

Delivers 15 GB/s to keep

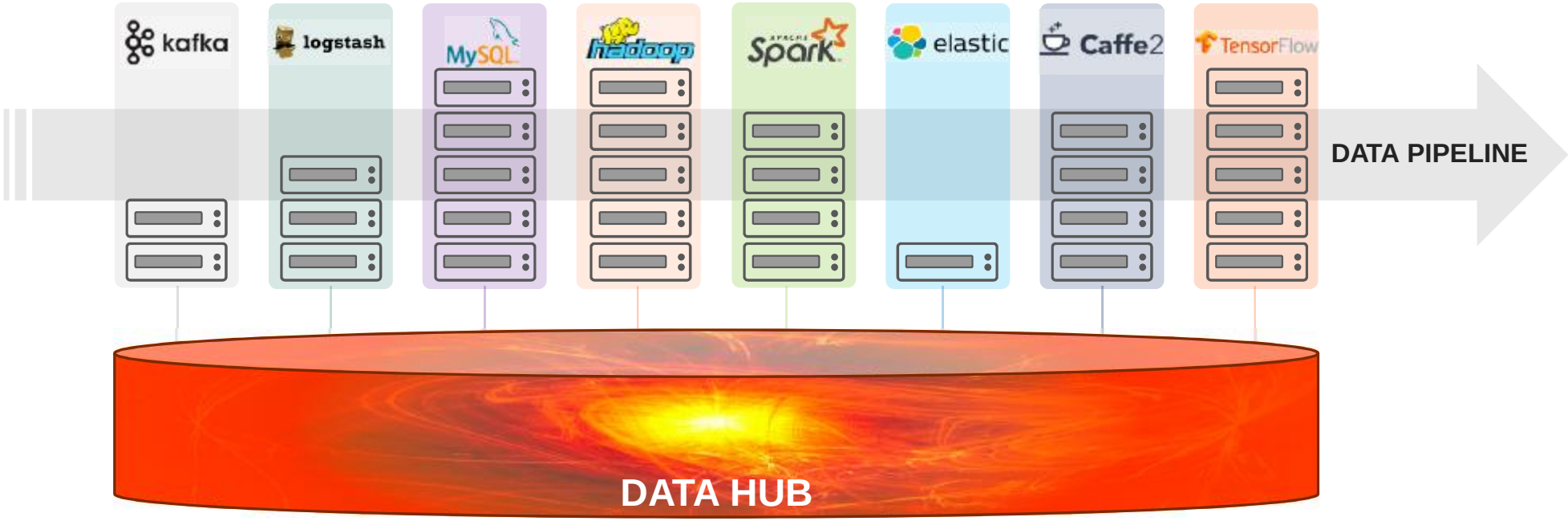
all 10 DGXs in rack with training data

© 2017 PURE STORAGE INC.

For DGX-1 13K images per second performance: http://files.shareholder.com/downloads/AMDA-1XAJD4/4389242263x0x918093/50C3BC56-468D-4A02-941B-C0599570915A/JHH_SC16_FINAL_PUBLISHED.pdf

IDEAL PLATFORM FOR MODERN ERA

DYNAMIC DATA HUB ARCHITECTED FOR REAL-TIME & ELASTIC DATA



"TUNED FOR EVERYTHING"
Small, Random to Large, Seq.
Architected for the Unknown

REAL-TIME
Low Latency Performance for Instant Response

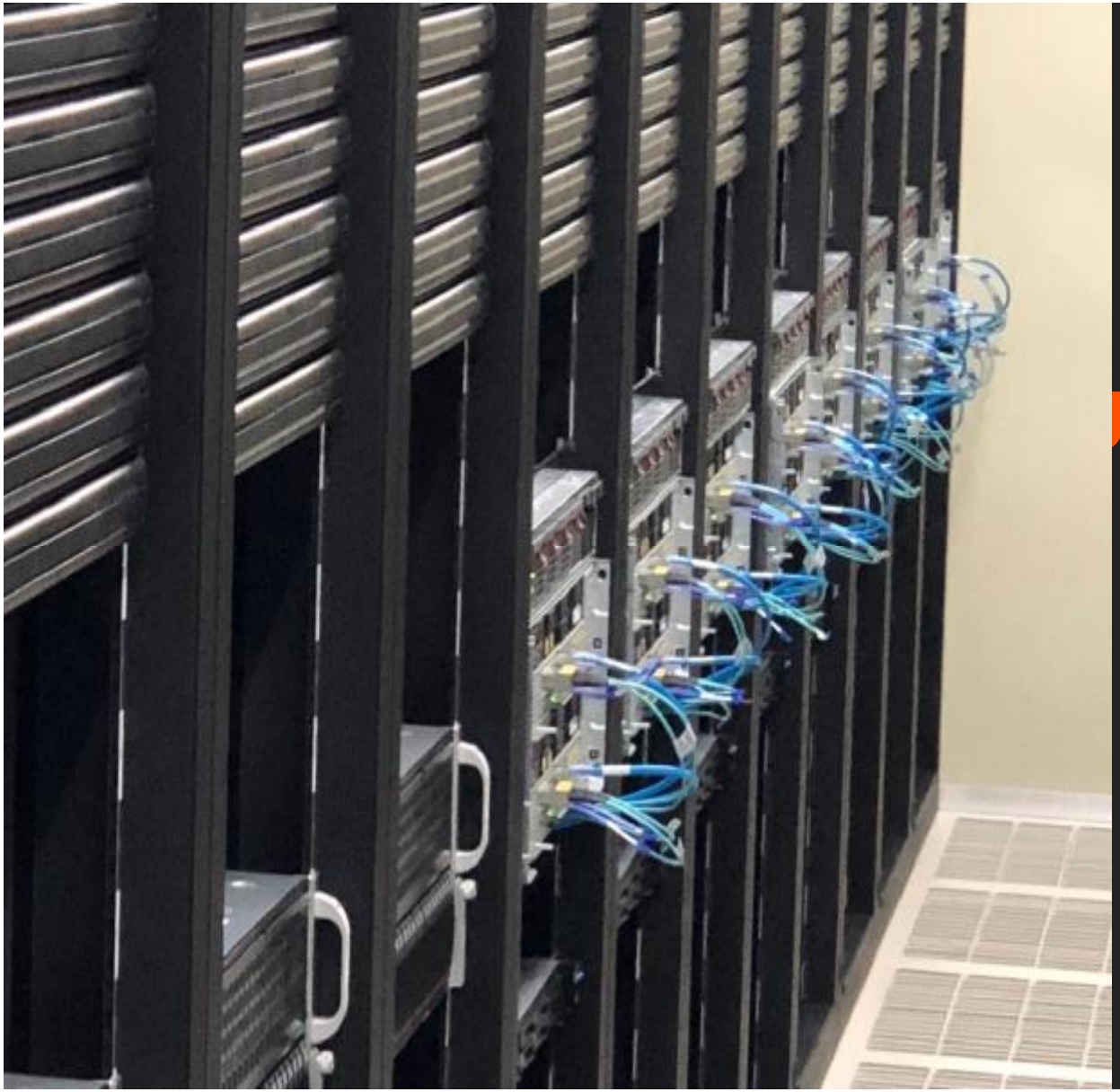
ALL-FLASH
Modern, Ultra-Fast Technology

PARALLEL
No Serial Bottlenecks for Max Throughput

ELASTIC
Grow Non-Disruptively with More App Clusters

SIMPLE
Focus More on Insights, Not Infrastructure

SOCIAL MEDIA – Be more than a social network. Know what brings people together. Be their safe place.	**HOSPITALITY** – Deliver one of a kind guest experiences.	**SAAS** – Unparalleled uptime to the millions of businesses whose data calls their cloud home.	**AUTONOMOUS MACHINES** – Pioneer new innovations in 30 seconds or less.	**TRAVEL** – Customers know best, but aren't all created equal.
FORMULA 1 RACING – Lap 1: Crunch the numbers Lap 2: Adjust Lap 3: Win	**RETAIL** – Data driven product consumption. It pays.	**FINANCIAL SERVICES** – Predict market volatility, identify trading patterns and ultimately improve investment strategy.	**LIFE SCIENCES** – Sequence more. Guess less.	**INVESTMENT MANAGEMENT** – Let quants be quants.



USING AI TO DRIVE BUSINESS GROWTH

Delivering AI research for Kakao, Korea's leading web company to develop and provide more intelligent services to millions of users

A LEADER FOUR YEARS IN A ROW!

Figure 1. Magic Quadrant for Solid-State Arrays

Source: Gartner (July 2017)

Gartner

A MAGIC QUADRANT LEADER

Source: Gartner Magic Quadrant for Solid State Arrays, July 2017.

This graphic was published by Gartner, Inc. as a part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from Pure Storage. Gartner does not endorse any vendor, product, or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including and warranties of merchantability or fitness for a particular purpose.

GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally, and is used herein with permission. All rights reserved.

© 2017 PURE STORAGE INC.

PURESTORAGE®

WE'VE HELPED > 3000 CUSTOMERS

ACCELERATE POSSIBLE

ACCELERATE DIGITAL BUSINESS

TRANSFORM IT TO THE CLOUD MODEL

HARNESS VALUE OF DATA

© 2017 PURE STORAGE INC.

FOUR ESSENTIAL THINGS FOR AI

FRAMEWORKS & APPLICATIONS

Caffe2

mxnet

PYTORCH

TensorFlow

COMPUTE

FROM CPU TO GPU SERVERS

NVIDIA

STORAGE

POWER ENTIRE AI PIPELINE

PURESTORAGE

SKILLS POWER AI

© 2017 PURE STORAGE INC.

PURESTORAGE

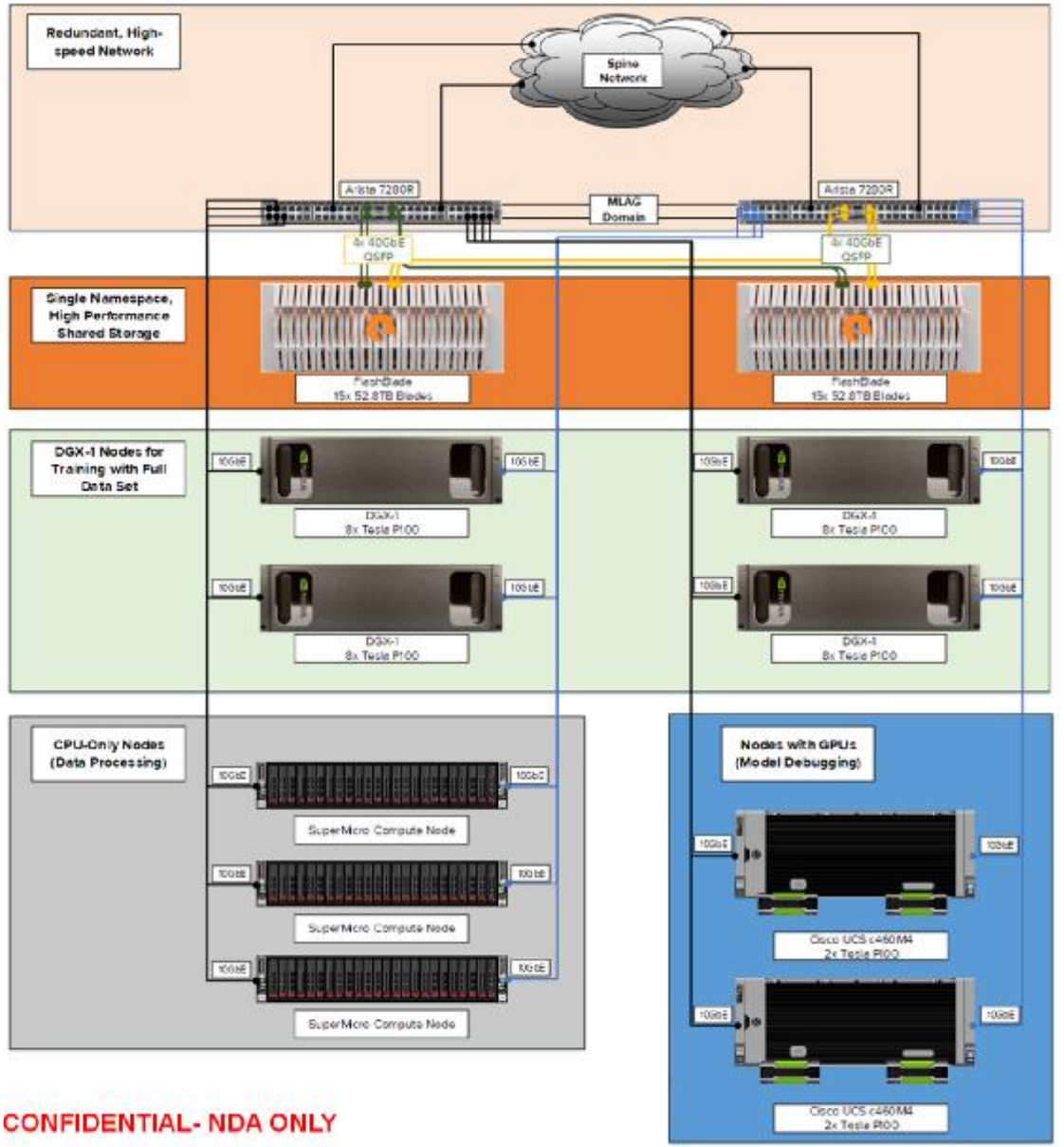
ALL OF THE ABOVE IS IRRELEVANT IF THE TALENT IS NOT THERE TO CREATE THE FUTURE

ZENUITY AI PIPELINE FUELED BY FLASHBLADE + DGX-1

4 NVIDIA DGX-1
DL Training Cluster

2 PURE FLASHBLADE
Over 1 PB of training data w/ performance headroom

ENTIRE AI PIPELINE ON A SINGLE HUB
Preprocessing, Exploring, and Training on FlashBlade



© 2017 PURE STORAGE INC.

PURE STORAGE CONFIDENTIAL- NDA ONLY