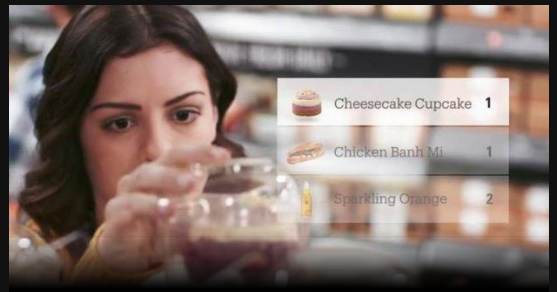
AIに最適な統合インフラストラクチャ「AIRI」

2018年10月31日
ピュア・ストレージ・ジャパン株式会社
FLASHBLADEセールスリード
大浦譲太郎

PURE STORAGE



食料品店- Tesco社 （英国）
精算なしのスーパーマーケット

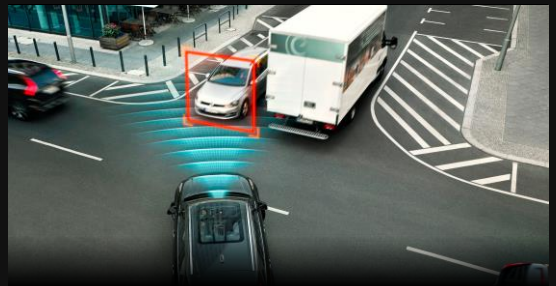
$1.2兆

AIドリブン企業が情報量の少ない同業他社から奪う年間収益
Forrester社調べ

医療- GE社
子供の神経学的遅延を特定

「私たちは、過去何十年もSFの世界の話だったAIを使って問題を解決しています。基本的には、機械学習で改善できない組織など存在しません。」

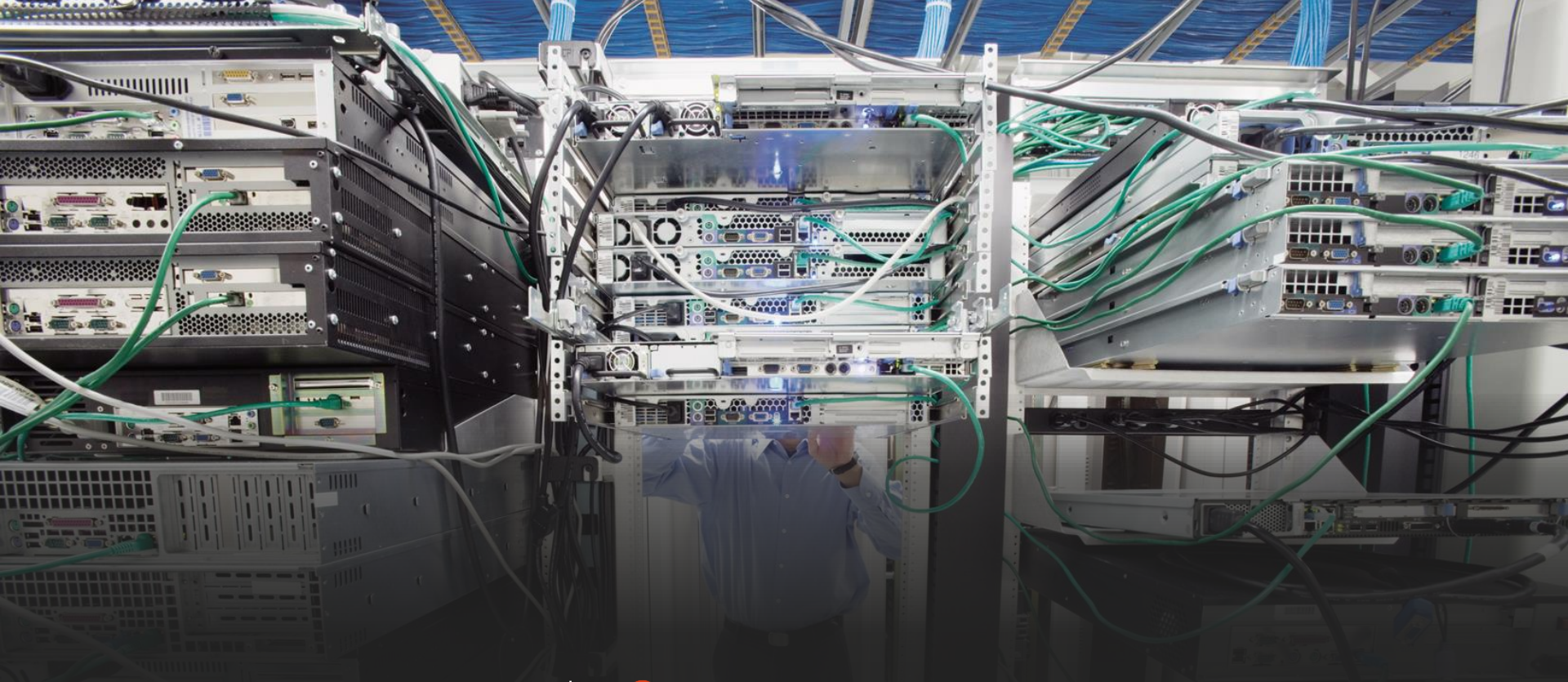
インテリジェンスはあらゆる企業に不可欠
Amazon 社CEO ジェフ・ベゾス



自動車- Zenuity社
2021年までの自律走行車実現を目指す

80%

2020年までにAIを導入する企業の割合
Gartner社調べ



“独自の構築”がほぼ唯一の方法

終わりのないオープンソースソフトウェアのコンパイルおよびチューニングサイクル

何ヶ月もかかるシステム構築とチューニング、絶え間のないメンテナンス

依然として、ストレージ、GPU、アプリケーション間にデータボトルネックを多数含むレガシーソリューション

GTCでのFacebookによる発表（引用）

http://on-demand.gputechconf.com/gtc/2017/presentation/s7815-soumith-chintala-building-scale-out-deep-learning-infrastructure-lessons-learned-facebook-ai-research.pdf

Infra Requirements for DNN Training

Models with **Millions of Parameters**

Training on **Multi-TB datasets**

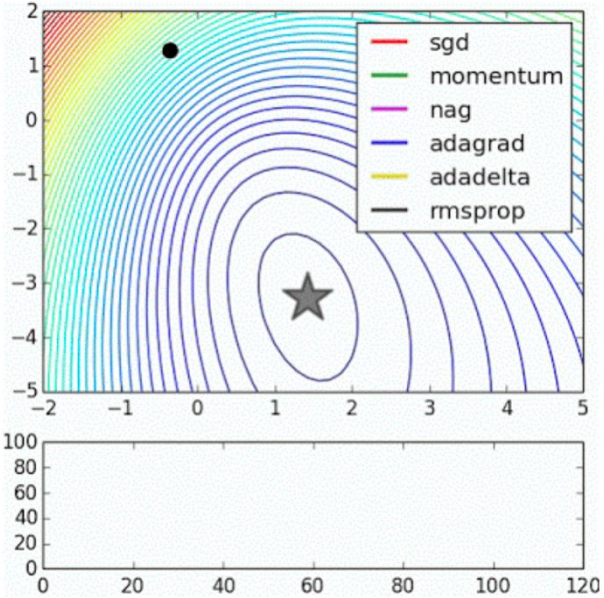
Gradient Descent **algorithms are sequential**

Computer Vision model requires:

- **5-100+ ExaFLOPs** of compute
- **Billions of IOPS**

Many **communication bottlenecks**

Needs a **well-balanced system**

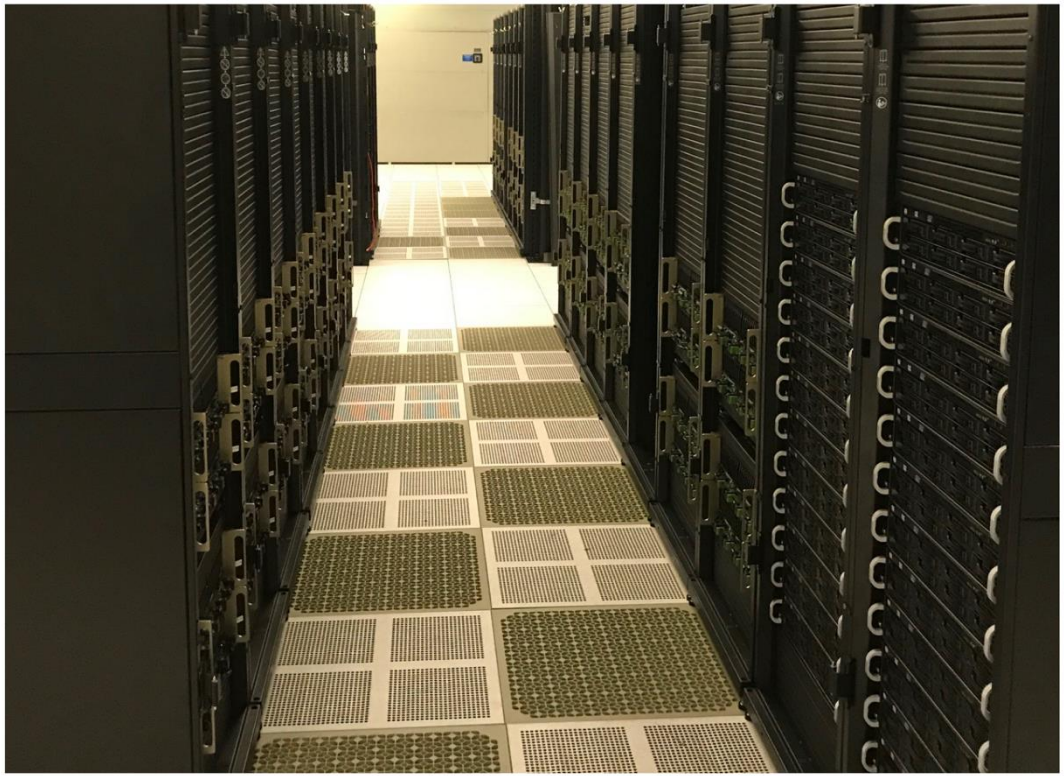


(animation by Alec Radford)

http://on-demand.gputechconf.com/gtc/2017/presentation/s7815-soumith-chintala-building-scale-out-deep-learning-infrastructure-lessons-learned-facebook-ai-research.pdf

FAIR Cluster Hardware - Compute

128 * DGX-1
10.5 PFLOPS total FP32
21 PFLOPS total FP16
Non-blocking IB fabric



http://on-demand.gputechconf.com/gtc/2017/presentation/s7815-soumith-chintala-building-scale-out-deep-learning-infrastructure-lessons-learned-facebook-ai-research.pdf

FAIR Cluster Hardware

Dev servers:

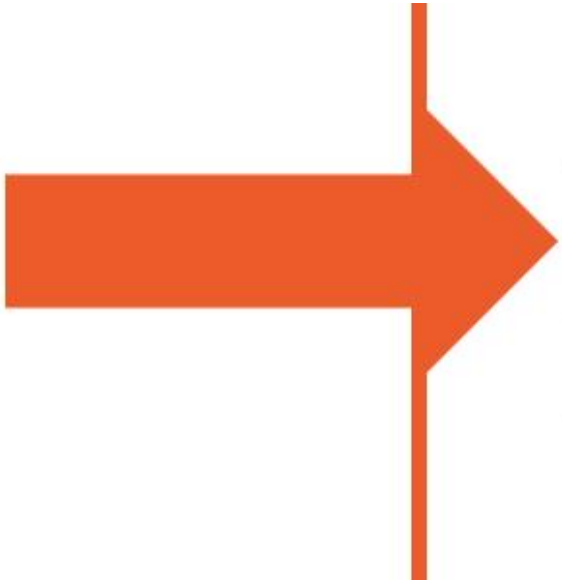
– 2 * GP100 with NVLink

Shared storage system:

- All flash
- Supports ~150,000 50KB files/sec per 100TB
- Sharded datasets

http://on-demand.gputechconf.com/gtc/2017/presentation/s7815-soumith-chintala-building-scale-out-deep-learning-infrastructure-lessons-learned-facebook-ai-research.pdf

拡大します



Shared storage system:

- All flash
- Supports ~150,000 50KB files/sec per 100TB
- Sharded datasets

GPU数とトレーニング処理の相関

Facebookの論文

Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour

https://research.fb.com/wp-content/uploads/2017/06/imagenet1kin1h5.pdf?

データフィードのボトルネックを取り除くと時間あたりのトレーニング処理数はリニアにスケールする。

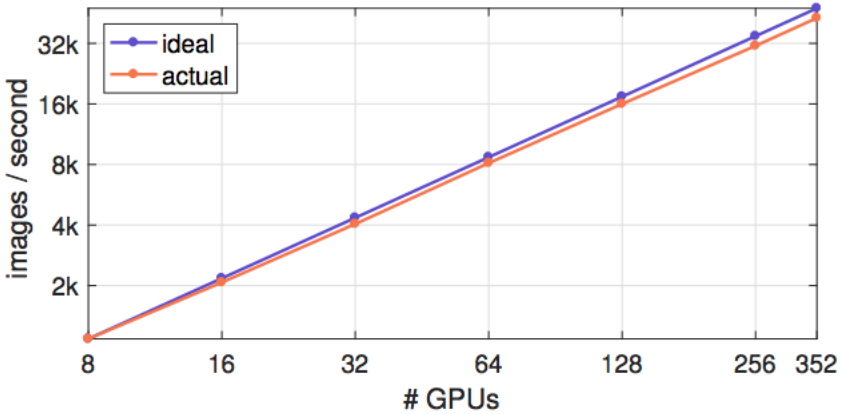


Figure 8. **Distributed synchronous SGD throughput.** The small overhead when moving from a single server with 8 GPUs to multi-server distributed training (Figure 7, blue curve) results in linear throughput scaling that is marginally below ideal scaling (~90% efficiency). Most of the allreduce communication time is hidden by pipelining allreduce operations with gradient computation. Moreover, this is achieved with commodity Ethernet hardware.

FacebookのAIへの取り組み

CNN.co.jp | Featured | World | USA | Business | Tech | Entertainment | Travel | Photo | Video | Odd News | Style | Login

Tech

AIを使って自殺予防、フェイスブックが取り組み拡大

2017.11.28 Tue posted at 16:43 JST

シェア 150 ツイート

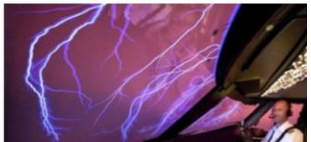
PR

- 【特集】これぞ新たな着想！仮想化を応用したセキュリティ対策！
- 生産性向上は企業の至上命題--明日の働き方を考え、デジタルで未来を切り開く
- CNN.co.jpメルマガ購読者募集中！

PR注目情報

CNN.co.jp読者アンケート実施中
回答いただいた読者の中から
抽選でAmazonギフト券が当たります

編集部セレクト

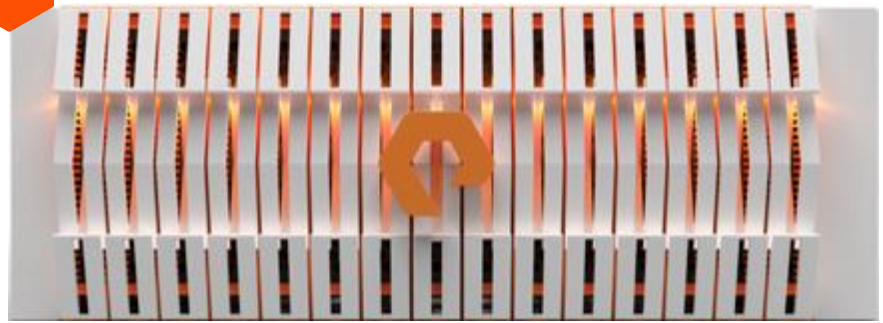


PURESTORAGE®

FLASHBLADE

将来を見据えたクラウドスケールのデータプラットフォーム

NEW



大規模、高速、シンプル

スケールアウトの独特なハードウェア/ソフトウェア設計
4Uシャーシに最大1.6PB、500K IOPS、15 GB/秒
75台のブレードにまで拡張可能予定

ファイル + オブジェクト

NFS + S3全体でデータを共有
ファイル、オブジェクト、コンテナに最適

17TB
または
52TB
の非圧縮容量
スケールアウトブレード

PURITY FB

ELASTIC FABRIC

100TB未満から
数十PBまで
柔軟な
拡張

PURESTORAGE®

AI モデルトレーニングのワークロード

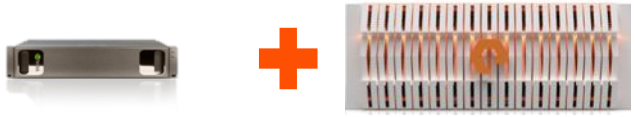
トレーニングの
ワークフロー

BENCHMARK
環境

GPUのみの場合

Setup #1: Synthetic Data from
System RAM into GPUs

I/O + CPU + GPU

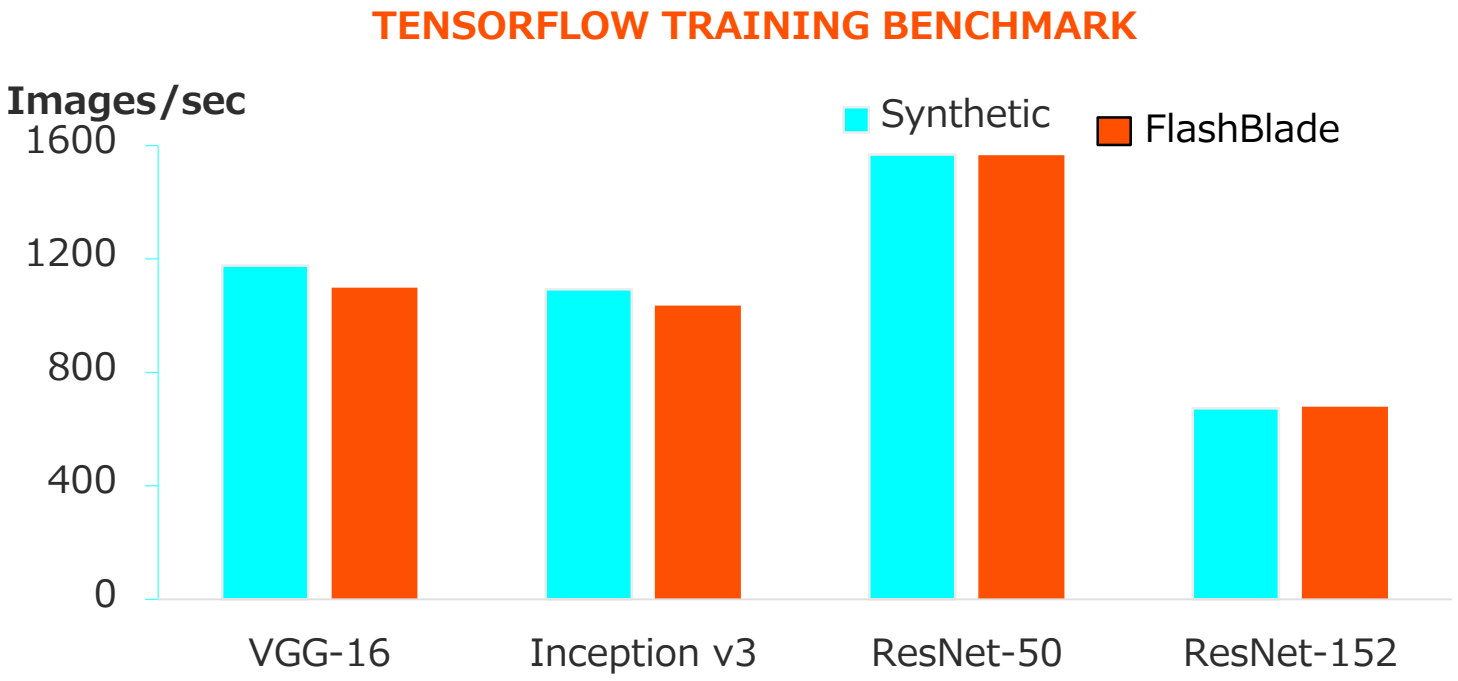


Setup #2: Real Image Data from
FlashBlade into DGX-1

PURESTORAGE®

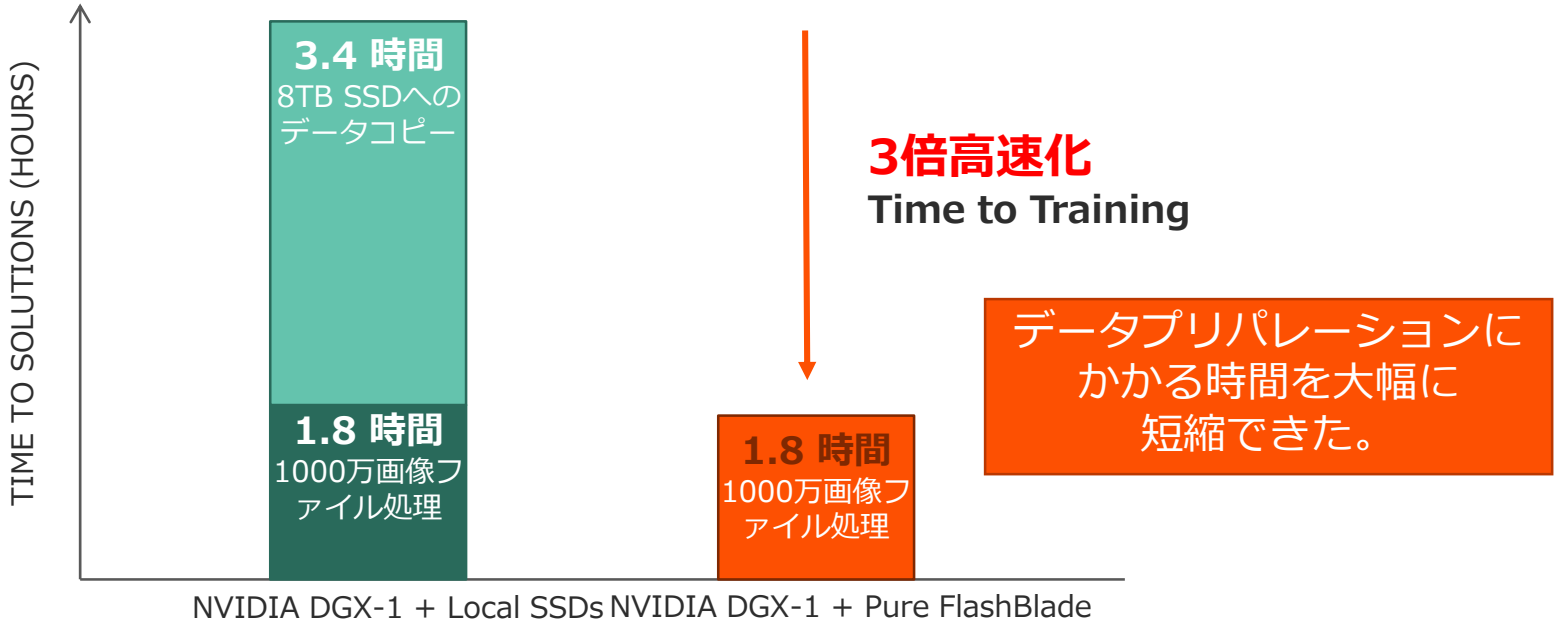
ローカルデバイスに匹敵する性能

FLASHBLADE はGPUに絶え間なくデータを供給



END-TO-ENDで３倍高速化

TENSORFLOW TRAINING BENCHMARK WITH RESNET-50



お客様事例

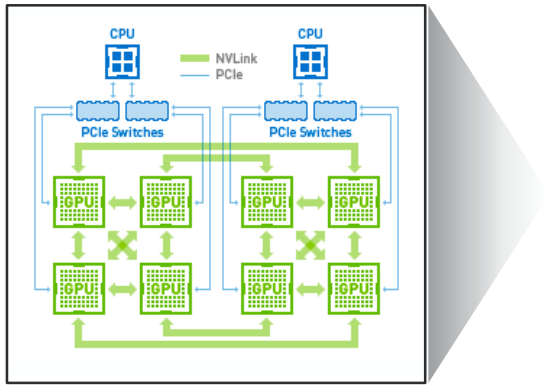
FlashBlade1台で…

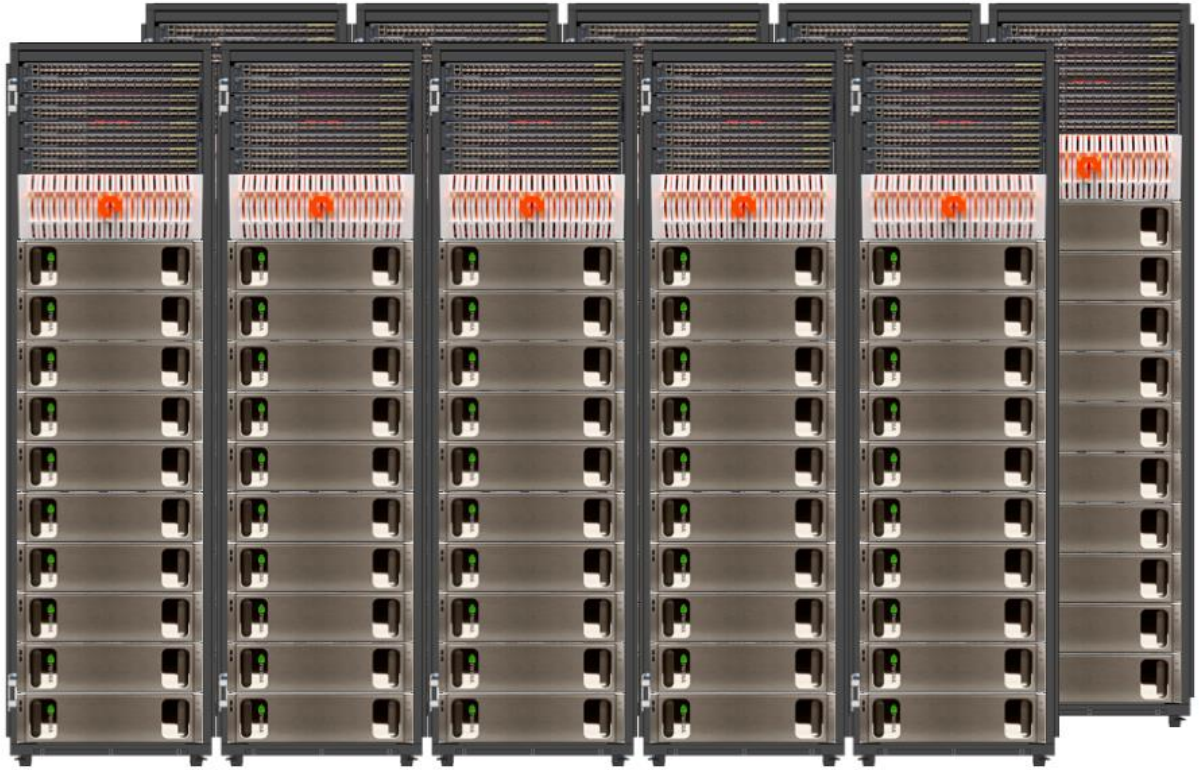
FLASHBLADE

15 GB/s でデータ供給し、
ラック内の10台の DGXs で
絶え間なくトレーニングを実施

DGX-1

13K Images/Sec for each DGX-1
Assume 115KB on average for images





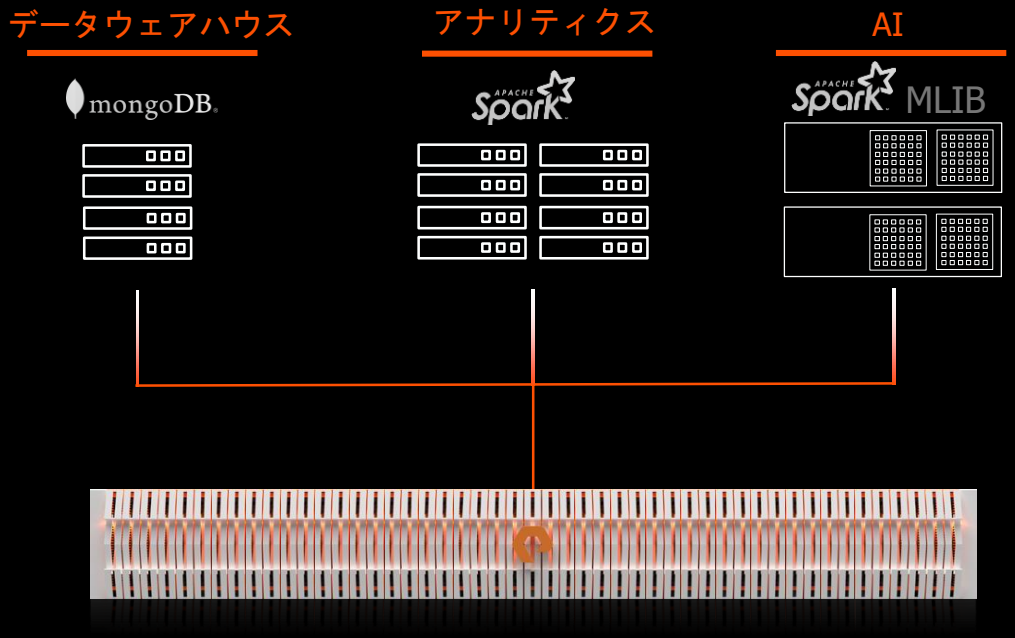
For DGX-1 13K images per second performance: http://files.shareholder.com/downloads/AMDA-1XAJD4/4389242263x0x918093/50C3BC56-468D-4A02-941B-C0599570915A/JHH_SC16_FINAL_PUBLISHED.pdf

PURESTORAGE®

お客様事例

TIME-TO-MARKET の優位性を手に入れる

世界最大のヘッジファンド事例



Man AHL

"Our quants want to test a model, get the results, and then test another one-all day long. So a **10-20X improvement in performance is a game-changer** when it comes to creating a time-to-market advantage for us."

Gary Collier, co-CTO, Man AHL

お客様事例

FLASHBLADEと DGX-1で 構成した AI パイプライン

5台の DGX-1 (8GPU/ノード)

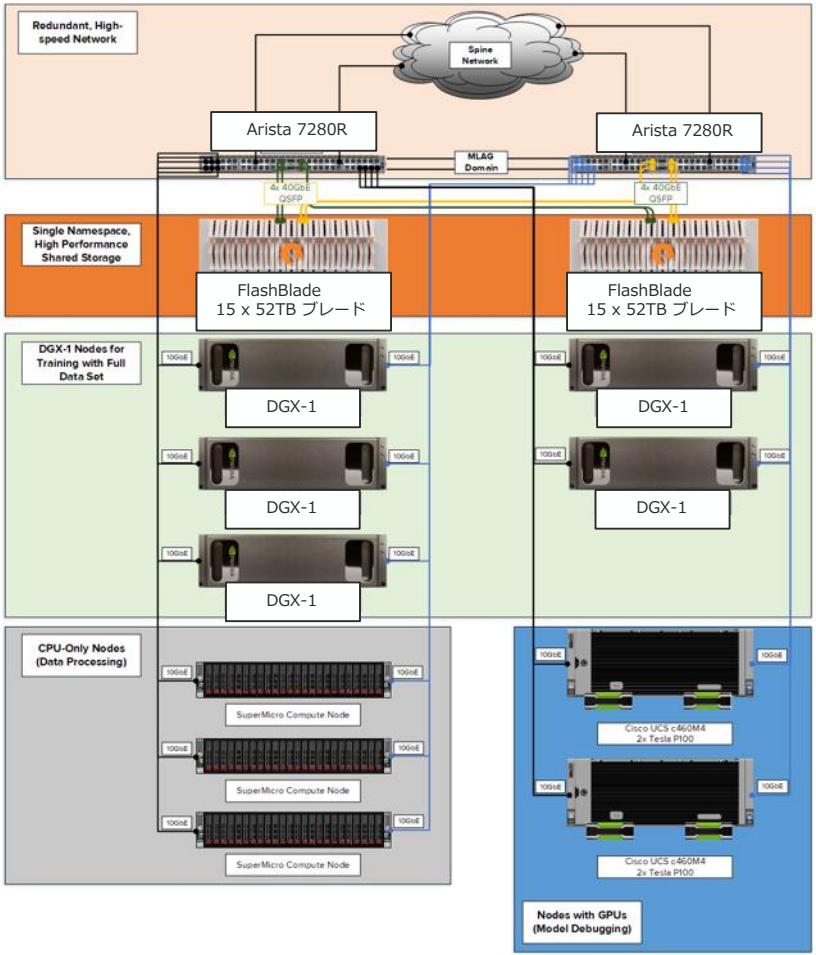
DL Training Cluster

2 セットの FLASHBLADE

1 PBを超えるトレーニングデータと更なる性能の余裕

全ての AI パイプラインを一つのHUBで

プリプロセッシング、エクスプローリング、
トレーニングを FlashBlade 上で実施



AI PIPELINEにおける従来のアプローチ

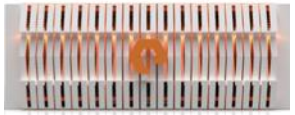
従来のストレージベースの課題とサイロ

アクセスパターン	**シーケンシャル**	**シーケンシャル またはランダム**	**ランダム**	**ランダム**
ReadかWriteか	**write**	**read & write**	**read**	**read**
ファイルサイズ	**メタデータ：小 データ：小～大**	**小～大**	**小～大**	**小～大**
同時アクセス数	**多い**	**多い**	**少ない**	**多い**
使用される ストレージ	専用NAS	専用NAS Hadoopなど	専用NAS	分散ストレージ ex) DDN(GPFS/Lustre)

AI PIPELINEにおけるPUREの回答

SIGNIFICANT CHALLENGE TO LEGACY STORAGE, NOT FOR FLASHBLADE

	INGEST	CLEAN & TRANSFORM	EXPLORE	TRAIN
	センサーやマシンから	CPU サーバ	GPU サーバ	GPU クラスター
アクセスパターン	シーケンシャル	シーケンシャル またはランダム	ランダム	ランダム
ReadかWriteか	write	read & write	read	read
ファイルサイズ	メタデータ：小 データ：小～大	小～大	小～大	小～大
同時アクセス数	多い	多い	少ない	多い



よく聞かれる質問

大規模AIに求められるストレージシステムの要件とは？

大規模AI環境に求められるシステム要件

・システムの電力消費が抑えられること
（限られた電源環境での計算応力の最大化）
・運用管理が簡単なこと
・特殊なNWに依存しないこと
・データのパイプライン処理のためファイル・オブジェクトでの
　アクセスが容易なこと
・必要に応じて性能がスケールさせられること

FlashBladeが選ばれる理由＞他の製品では実現不可能

8KVAの電源供給ラックでDGX-1　２台と、必要スループットを賄えるストレージ

3 KVA
3 KVA

＜8 KVA

＜2 KVA

FlashBladeが選ばれる理由＞運用管理性

- 運用負荷が高い
- 大量の電力消費、冷却コスト
- アップグレード時は買い替え

144ノードのスケールアウトNAS

４Uに集約されたスパコンのストレージ
設置後はクラウド管理で手間いらず

PURESTORAGE®

大規模AI環境に求められるシステム要件

- システムの電力消費が抑えられること
  ＞＞15GB/sのスループットを1850Wattで
- 運用管理が簡単なこと
  ＞＞GUIで設置後は手間いらず、リモート保守サービス付き
- 特殊なNWに依存しないこと
  ＞＞Infinibandではなくイーサネット
- 必要に応じて性能がスケールさせられること
  ＞＞マルチノードのトレーニングがリニアにスケール
- データのパイプライン処理のためファイル・オブジェクトでのアクセスが容易なこと
  ＞＞NFS/S3に対応

Best Innovation in AI Hardware


AIRI™
AI-READY INFRASTRUCTURE
From
PURESTORAGE®
Powered by
NVIDIA
NVIDIA
PREFERRED
PARTNER

AIRI

業界初の

AI完全対応インフラストラクチャ

ハードウェア

NVIDIA® DGX-1™ | 2or4x DGX-1 システム | 2-4 PFLOPSのDL性能
PURE FLASHBLADE™ | 7or15x 17TB ブレード | 1.5M IOPS
ARISTA | 2x 100Gb RDMA対応Ethernetスイッチ

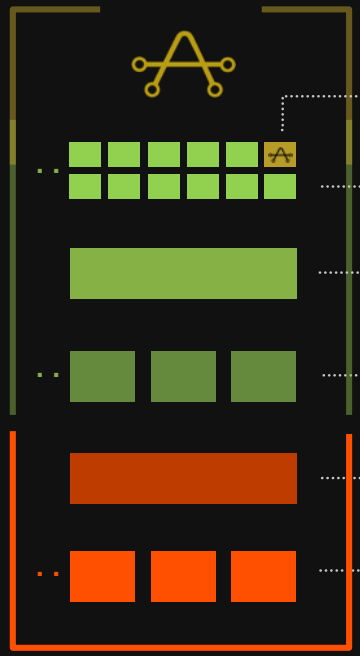
ソフトウェア

NVIDIA GPU CLOUD DEEP LEARNING STACK | 最適化フレームワーク
AIRI SCALING TOOLKIT | 簡素化されたマルチノードトレーニング

AIRI

テクノロジースタック

シンプルになった大規模AI実装



AIRI テクノロジースタック

NVIDIA GPU CLOUD DEEP LEARNING STACKと
AIRI SCALING TOOLKITを含む

マルチノードトレーニング
AIRI Scaling Toolkit

NVIDIAの最適化ディープラーニング フレームワーク
Caffe Caffe2 mxnet TensorFlow Theano torch PYTORCH

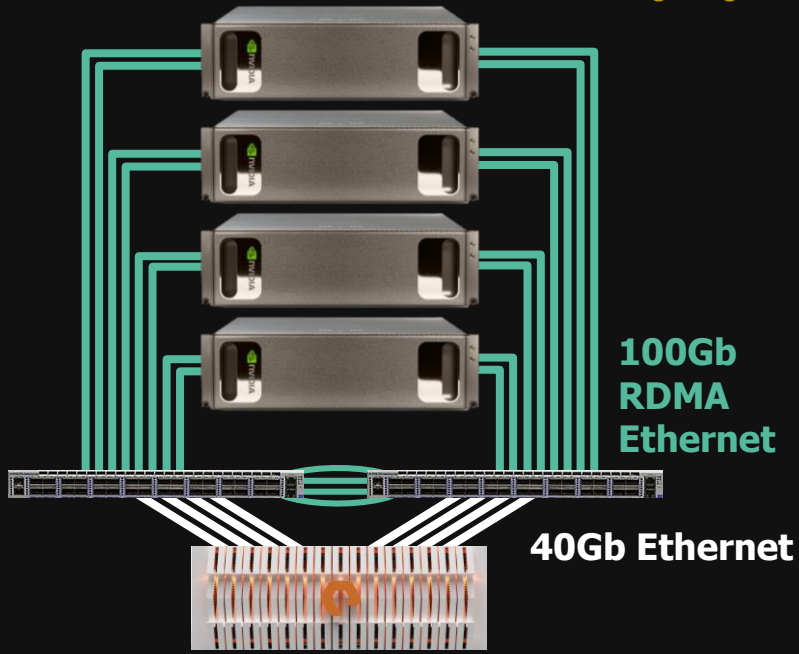
コンテナ化
GPU最適化 Docker

スケールアウト対応GPUコンピューティング
Tesla V100 GPUs搭載のNVIDIA DGX-1

スケールアウト対応ファイル&オブジェクト プロトコル
Pure Storage Purity//FB

スケールアウト対応フラッシュストレージ
Pure Storage FlashBlade

AIRI



4x NVIDIA® DGX™-1 SYSTEMS
4 PetaFlops of DL Performance
32x Tesla™ V100 GPUs

CONVERGED FABRICS
2x 100Gb Ethernet Switches with RDMA
CISCO Nexus9000 or ARISTA 7060X

PURE® FLASHBLADE™
1.5M NFS IOPS
15x 17TB blades

AIRIが生まれた背景

DGX-1
マルチノード

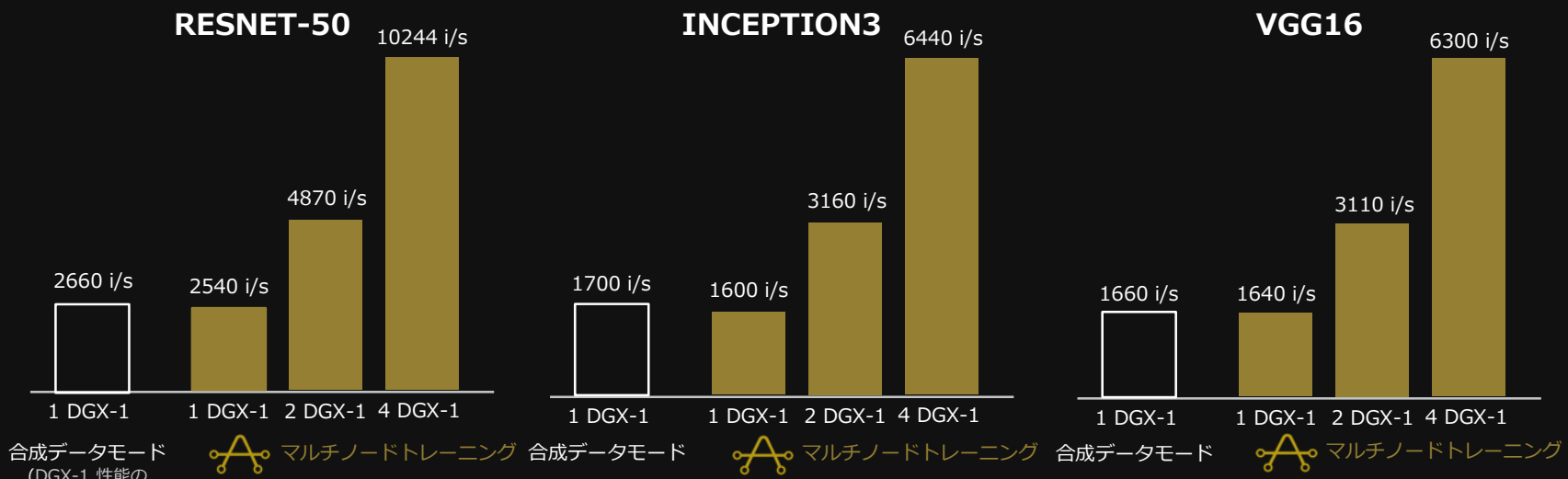
Infiniband
ではなく

AI
パイプライン

消費電力

データボトルネックを完全に排除

AIRI™ がトレーニング時間を 1 /4に削減、データサイエンティストの生産性を向上



TensorFlowおよびRDMA対応 100Gb Ethernetを使用してGPU使用率を向上

すべての価格ポイントでエラスティックなスケールを提供

SCALE-OUT ARCHITECTURE BUILT TO GROW PERFECTLY WITH YOUR AI JOURNEY

AIRIとは

-DGX-１ユーザーが、マルチノード環境に取り組みたい時に、スケールアウト性能を即座に提供する実績のあるソリューション

大規模AI環境に求められるシステム要件

- システムの電力消費が抑えられること
  ＞＞15GB/sのスル　　Wattで
- 運用管理が簡単なこ
  ＞＞GUIで設置　　ート保守サービス付き
- 特殊なNWに依存
  ＞＞Infiniband
- 必要に応じて性能　　こと
  ＞＞マルチノー　　ニアにスケール
- データのパイプライ　　イル・オブジェクトでの
  アクセスが容易なこと
  ＞＞NFS/S3に対応

ELEMENT AI

“AIRIは、企業のAI採用の突破口を開く存在として、インフラストラクチャの複雑さという障壁を解消し、あらゆる組織がAIイニシアチブを迅速に指導させるための道を開きます。

AIRIには、ElementAIが社内および顧客向けに幅広く使用しているのと同じ中核ソリューション、すなわちNVIDIA DGX-1とPureStorage FlashBladeが搭載されています。”

ジェレミー・バーンズ氏
ElementAI社チーフアーキテクト

PAIGE

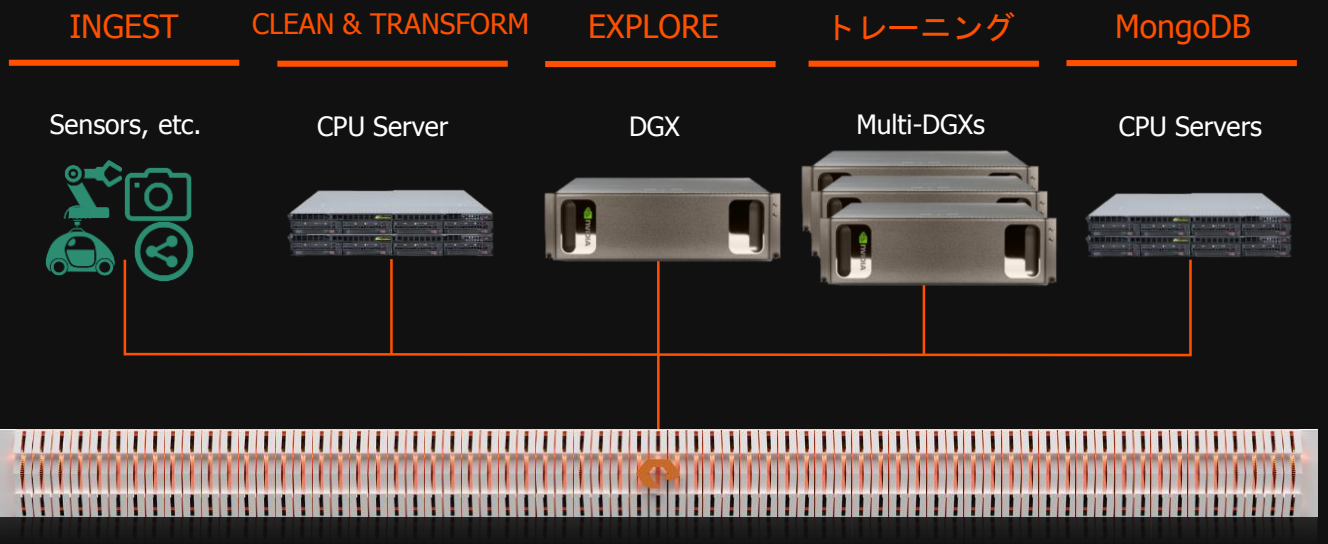
"データはAI革命を推進する燃料です。世界最大規模の腫瘍病理アーカイブを利用するにあたり、私たちは、膨大な量のデータを臨床的に検証されたAIアプリケーションへとすばやく変える最先端のディープラーニングインフラストラクチャを求めていました。

DGX-1とFlashBladeの強力な組み合わせが、AIを使って医療業界を根底から変えるという私たちのミッションに拍車をかけてくれます。 AIRIは、弊社のAIインフラストラクチャと同じコアテクノロジーに基づいて設計されています。AIRIを使用してAIイニシアチブを推進する企業に、どのような可能性がもたらされるのかを楽しみにしています。"

トーマス・フックス博士
ファウンダー兼最高科学責任者
Twitter @ThomasFuchsAI

データパイプラインを支えるアーキテクチャ

DATA STRATEGY GOES BEYOND AI INITIATIVES



FlashBlade は予測可能な性能を多岐にわたるワークロードに対して提供


AIRI™
AI-READY INFRASTRUCTURE
From
PURESTORAGE®
Powered by
NVIDIA
NVIDIA
PREFERRED
PARTNER