



Hewlett Packard
Enterprise

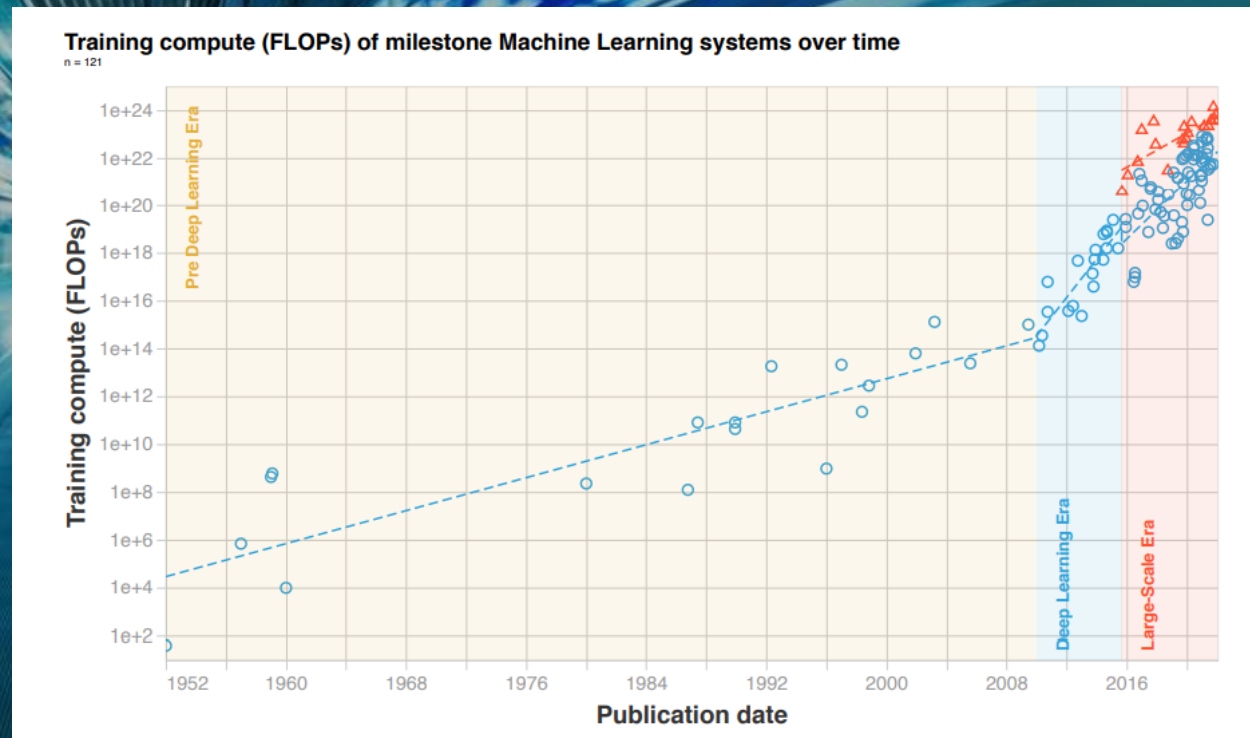
HPEの生成AIへの取り組み

日本ヒューレット・パカード合同会社
HPC&AI事業統括本部
カテゴリーマネージャー
高橋 健

大規模言語モデルにはスーパーコンピュータは不可欠

コンピューティングリソースの増強とモデル精度の向上は比例

大規模 AI モデルの
コンピューティング需要は、2016 年から
2022 年まで 10.7 か
月ごとに 2 倍に



Source: [Compute trends across three eras of machine learning](#). University of Aberdeen, Centre for the Governance of AI, University of St. Andrews, MIT, University of Tübingen, Complutense University of Madrid, March 2022

HPE は政府機関向けに 前例のない大規模な AIシステムを提供

研究者たちが人類の壮大な課題に取り組むのを支援

6,144

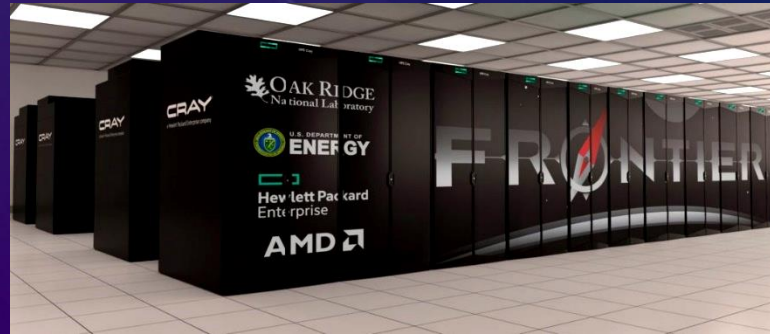
GPUs

37,632

GPUs

63,744

GPUs



100% liquid-cooled HPE Cray EX supercomputer

High performance GPU accelerated blades

HPE Slingshot exascale interconnect

Cray ClusterStor file systems

最も優れたトランスフォーマーエンジンGPUを もっと優れた実装密度で提供

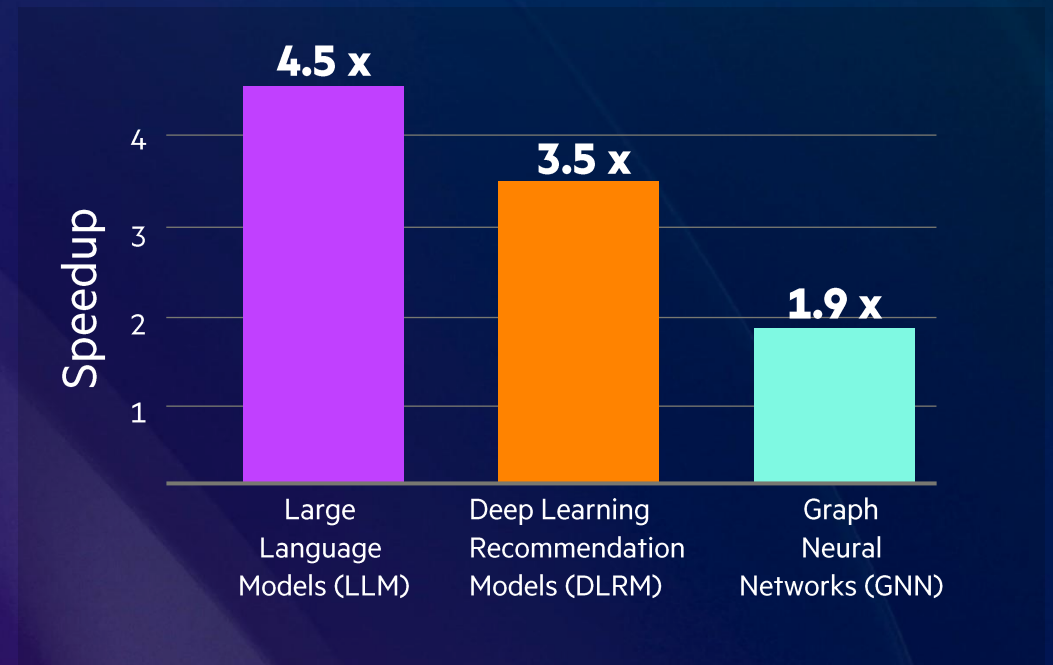
168 NVIDIA GH200 Superchips を21 compute blades/1 rackで

HPE Cray SC EX254n blade for HPE Cray EX supercomputers

- 100% liquid-cooled compute blade with 8 NVIDIA GH200 Superchips
- Two 4-socket nodes with 576 CPU cores and 147,456 GPU cores per blade
- 768 GB HBM3 and 1,024 GB LPDDR5 per blade
- 8 HPE Slingshot injection ports per blade

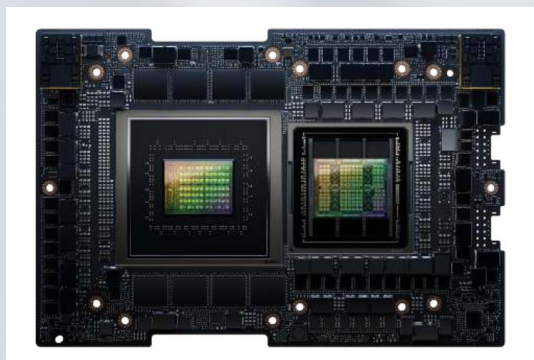
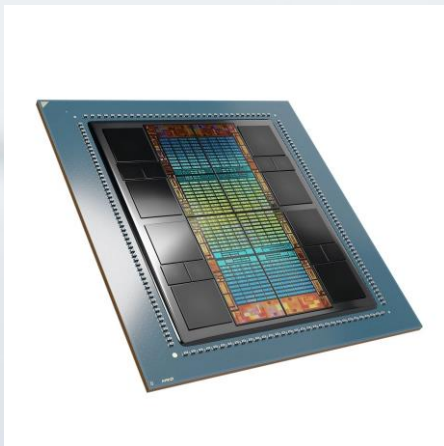


Application performance speed up vs x86 + Hopper¹

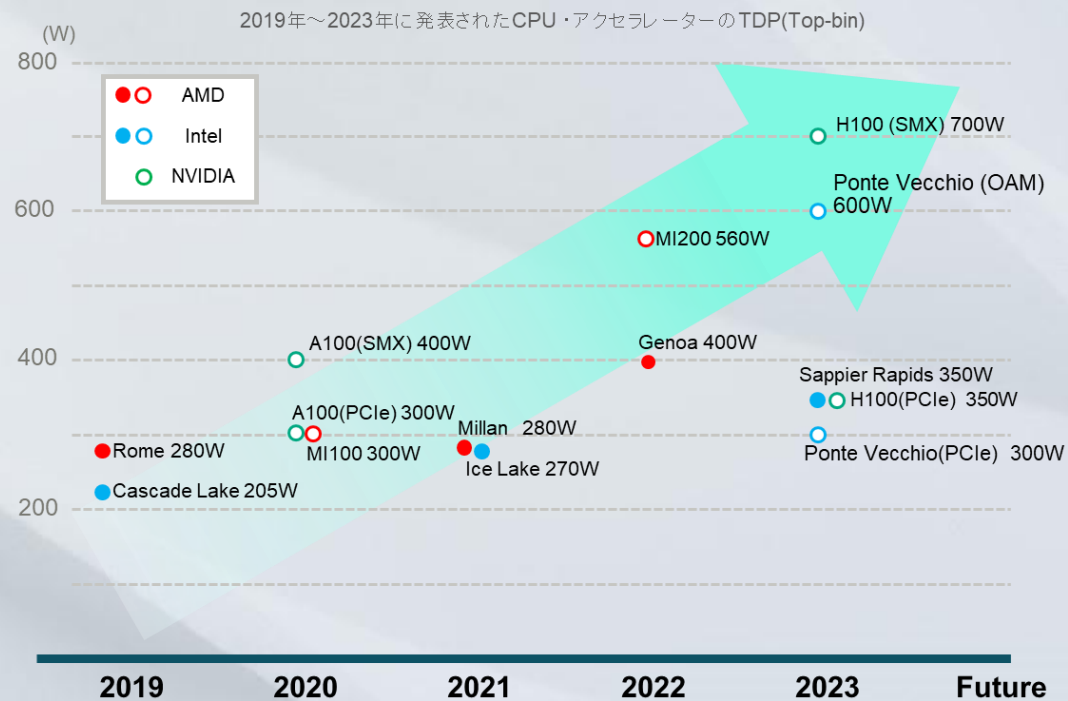


1 NVIDIA Grace Hopper Superchip Architecture Whitepaper

アクセラレータの“過熱”トレンド



各社GPUの競争過熱
CPU/GPU統合チップや、大容量HBM版の準備
1チップ当たりの高価格化も



TDPの過熱トレンド
高い性能/効率性を求めるには、高いTDPに対応したシステム
やファシリティが必要不可欠

Ref:

<https://www.amd.com/ja.html>

<https://www.intel.co.jp/content/www/jp/ja/homepage.html>

<https://www.nvidia.com/ja-jp/>

性能の追求と同時に、効率的な運用が求められる

消費電力を抑え、性能を高める冷却

データフローの整備、
計算リソースを無駄なく活用

Performance

最高性能のCPU/GPUをサポート
ターボモードを持続



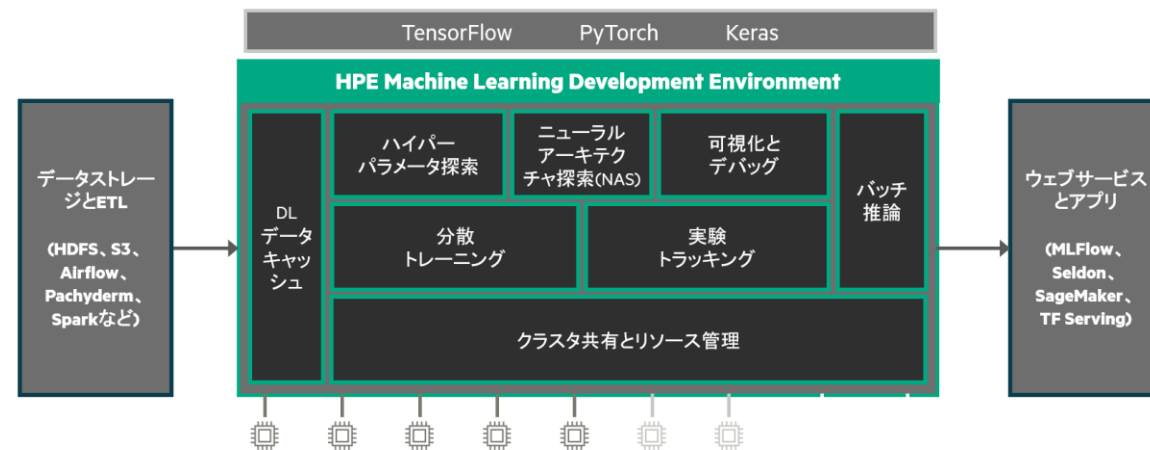
Liquid Cooling Solution

Density

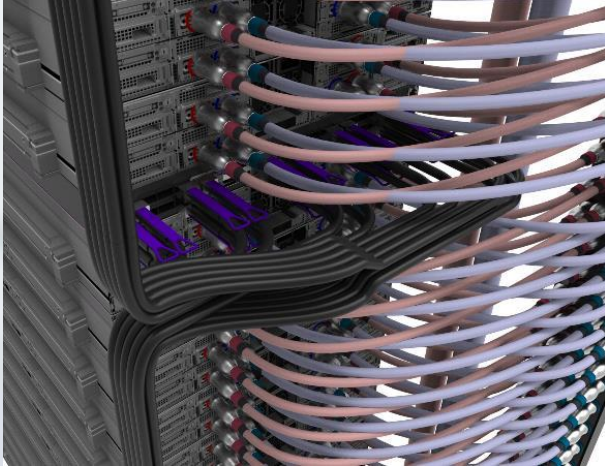
サーバー搭載数を増やし、
必要なラック数を削減

Efficiency

効率的に熱を取り除き、
消費電力を削減



HPE Direct Liquid Cooling Solution



効率性

- 水のほうが空気よりも冷却効率が高い(熱の伝導率20倍以上)
- 水のほうが空気より少ない力で循環させられる
- 高密度実装も比較的容易(スペース効率が高い)
- 故障率の低下

性能

- 高TDPのGPU/CPUの搭載が可能に
- 性能向上も期待可能

特性

従来のIT(デジタル)とは異なる、アナログな知見・経験が重要
- Ex.配管、チラー、水特注のトラブルシューティング



HPE Machine Learning Development Environment : MLDE

最適なモデル開発において、負担となる作業要素にフォーカスしたトレーニング環境

ハイパーパラメータ探索

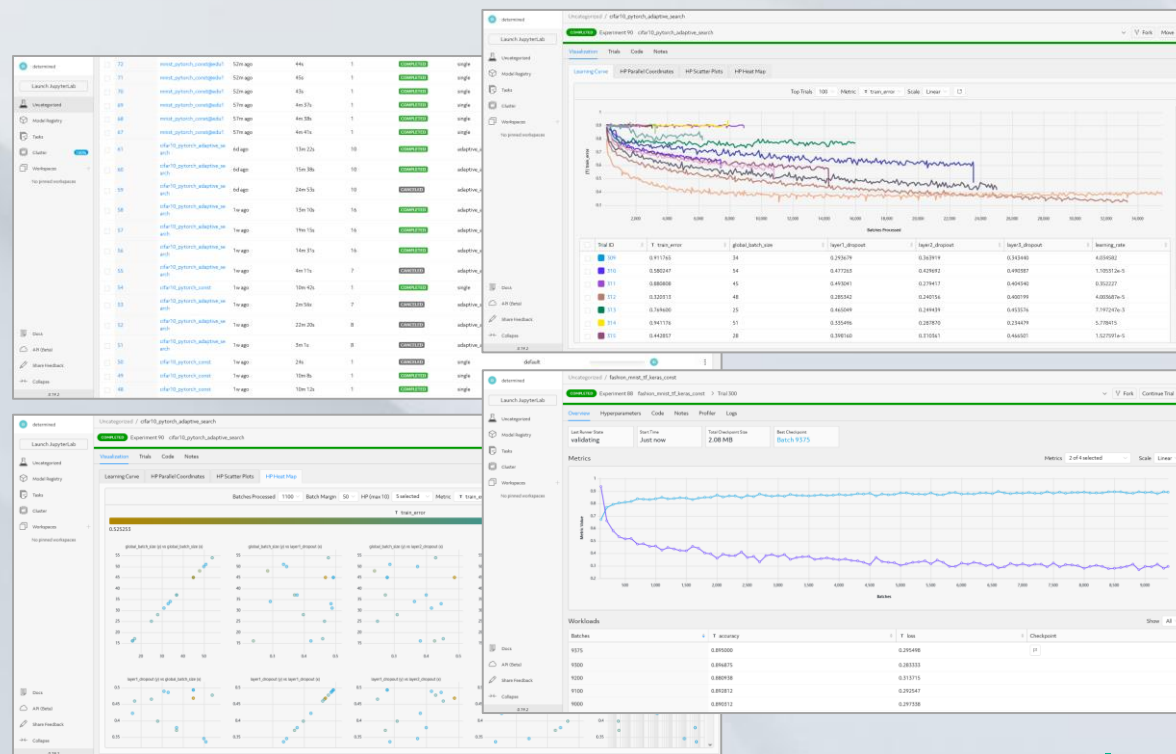
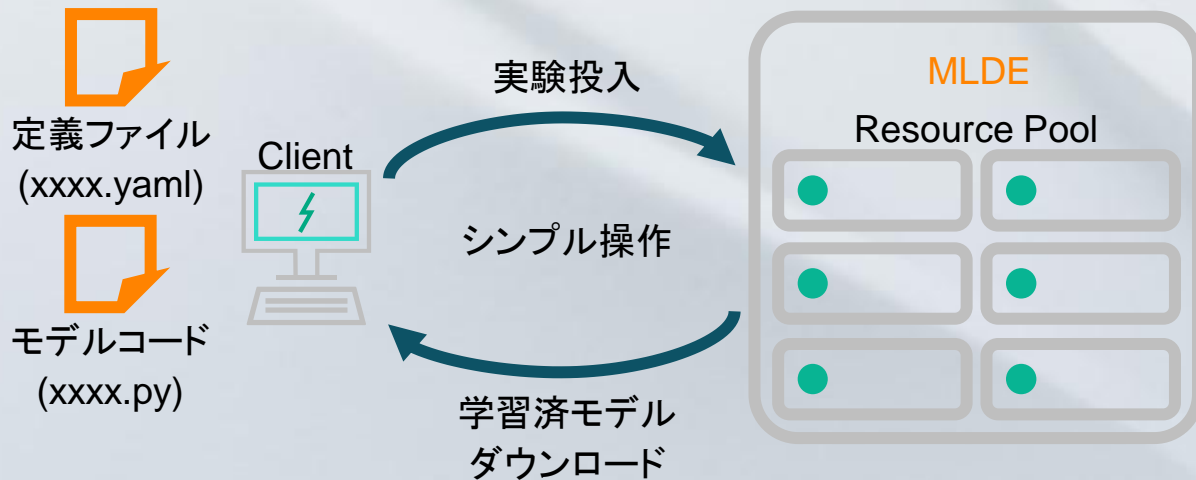
- ・複数の最適化手法を容易に利用
- ・パラメータ定義と試行モデルを自動生成
- ・独自GPUスケジューラとの緊密な連携

分散トレーニング

- ・マルチGPU利用による速度向上
- ・分散化に向けたコード改修負担の軽減
- ・MLエンジニアがDevOpsに悩まされない

チーム利用ツール

- ・インフラでなくモデルに集中できる環境
- ・クラウドインスタンスを約70%節約
- ・モデル追跡や共有による再利用促進



HPE Machine Learning Development Environment
Determined AI 概要デモンストレーション



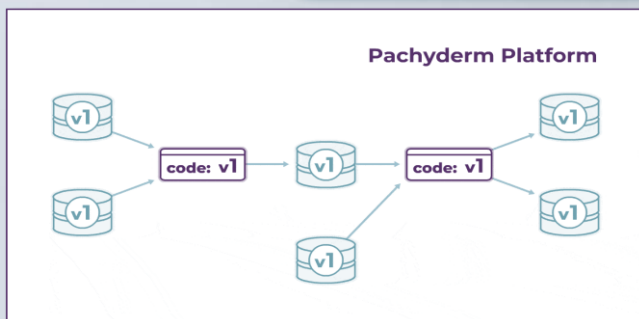
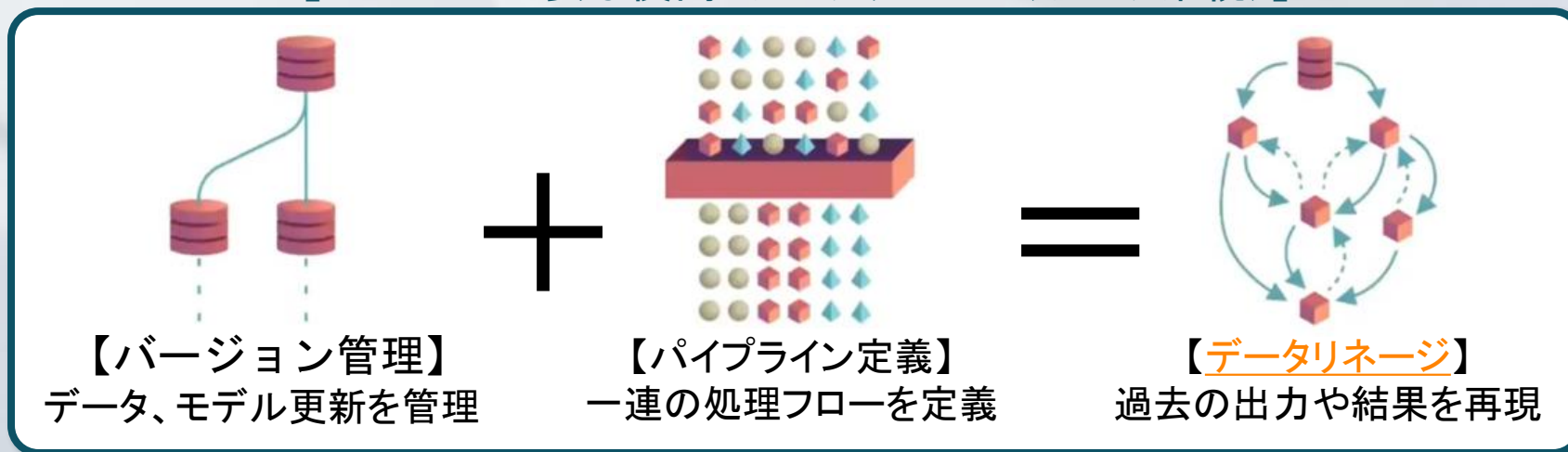
2022/12/02 Determined AI 概要デモンストレーション [11min]

注) 動画のWebUIが最新版と異なります。再生時は、Youtubeの字幕をOffにしてご視聴ください。

HPE Machine Learning Data Management : MLDM

MLDMは、大規模なAIアプリケーションを対象としたAI/MLパイプラインを自動化するためのソフトウェアです。主要機能はバージョン管理とパイプライン定義であり、それを組み合わせることで一連処理の再現性を実現します。

【MLDMの主要な役割: データリネージ(データ系統)】



データや処理コードが更新されるとそれ以降のデータを自動更新して最新に保ちます。

膨大なデータを学習するAIモデルの一連管理に有効

- ・どのデータをどこに格納したのか？
- ・データを誰がどのように移動、更新したか？

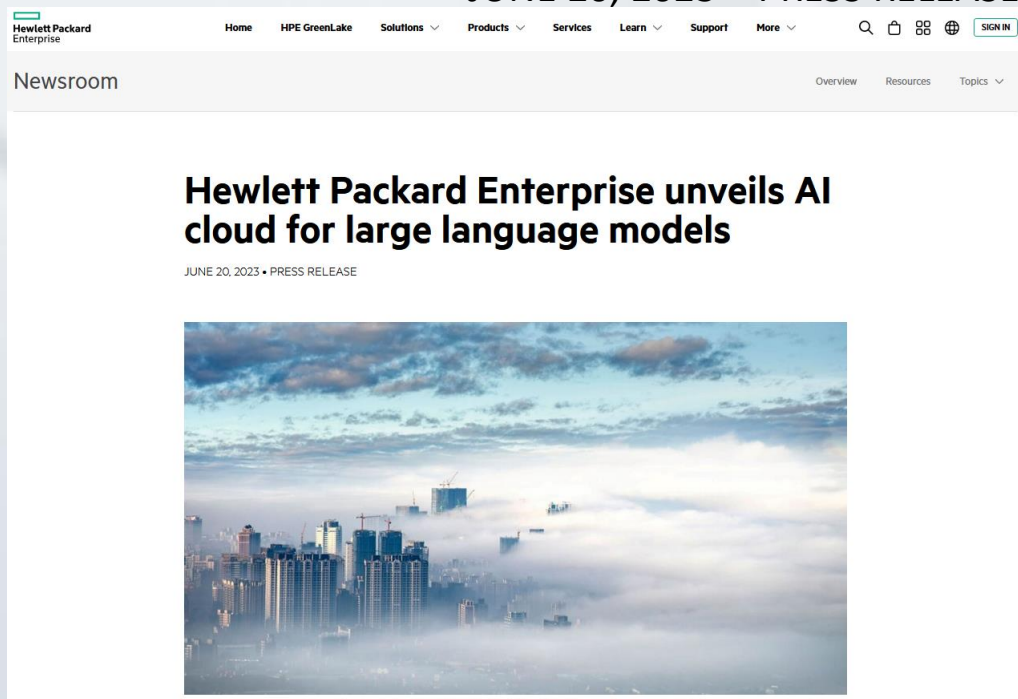
データ品質確保と向上、**トラブルシューティング対応**に効果を発揮します。

データリネージによって明確になる点

- ・データ作成に関する情報
- ・データ変換履歴と変換方法
- ・データ保管場所
- ・データ移動経路
- ・データ利用者、アクセス者

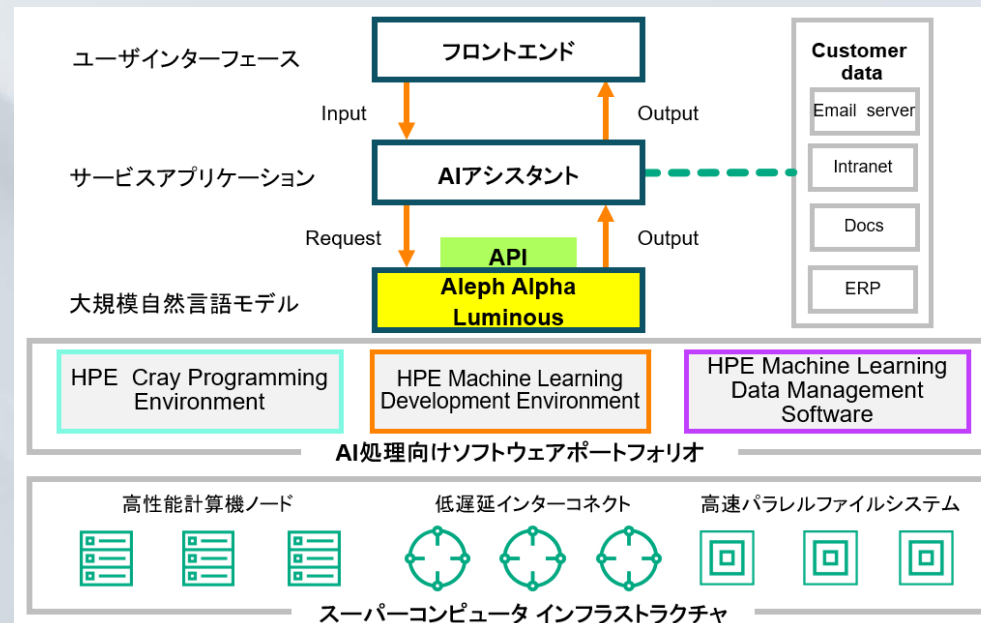
HPE GreenLake for LLMs

JUNE 20, 2023 • PRESS RELEASE



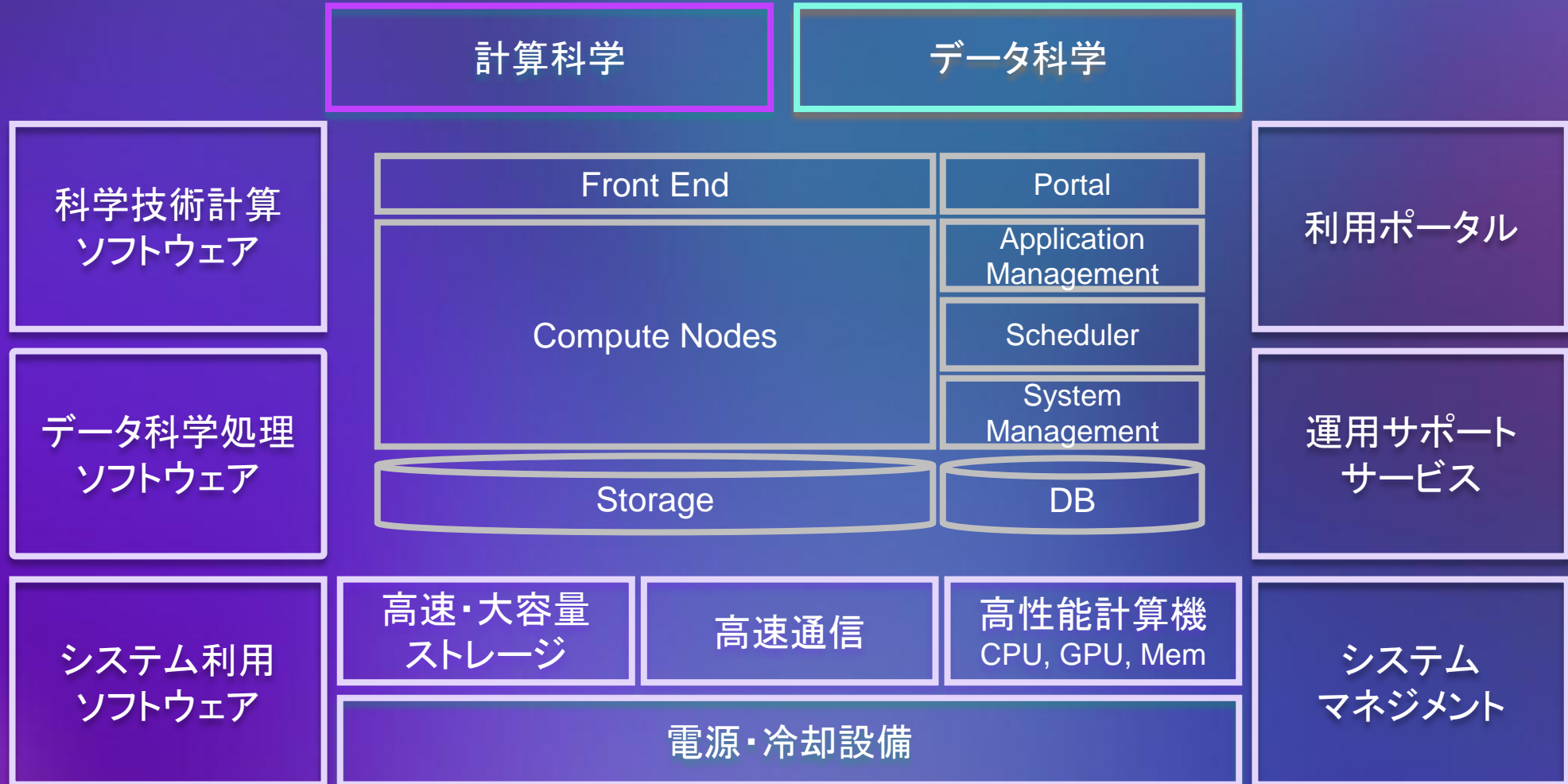
[HPE GreenLake for Large Language Models \(LLM\)](#)は、HPEのAIソフトウェアと市場をリードするスーパーコンピューティングプラットフォームを使用して、大規模なAIをプライベートにトレーニング、調整、展開することができます。HPE GreenLake for LLMは、初の提携パートナーであるAleph Alpha社様にて開発されるLLM(Luminous)を搭載し、テキスト、画像、分析における処理を必要とするユースケースに対応します。

HPE GreenLake for LLMは、世界で最も強力なスーパーコンピューターであるHPE Cray XDスーパーコンピュータを利用でき、お客様が自社で計算機を購入・管理する必要はありません。また、このスーパーコンピューティング・プラットフォームには、大規模モデルを迅速にトレーニングするHPE Machine Learning Development Environmentや、様々なデータを統合、追跡、監査し、データやモデルを正確に維持管理できるHPE Machine Learning Data Managementソフトウェアなど、HPEのAI/MLソフトウェアをサポートしています。



スーパーコンピュータ構築・運用はONE TEAM

スーパーコンピュータは、最新技術とパフォーマンスを引き出すノウハウが結集した未来へ導く脳
プロセッサのみでは完結しない、多様な要素の集合体



まとめ

- 生成AIの要となるアクセラレータの複数の選択肢を持つことが重要
- これからのアクセラレータには必須となる水冷システムには、知見・経験が重要
- AIソフトウェアにより、LLM開発・ファインチューニングを効率化

HPEは、ファシリティからソフトウェア、クラウドまで、
幅広くAI開発を支援いたします



Thank you

ken.takahashi@hpe.com

