HPE

PCCC25（設立25年記念PCクラスタシンポジウム）

AIの価値を引き出すHPEのAI factory ソリューションのご紹介

日本ヒューレット・パッカード合同会社
HPC&AI事業統括本部　第3プリセールス部　長竹 茂紀

2025年12月8日

1. HPE の HPC&AI 製品ポートフォリオと水冷ソリューション

HPE の HPC&AI 製品ポートフォリオ

空 空冷
水 水冷

リーダーシップクラスのスーパーコンピューティング

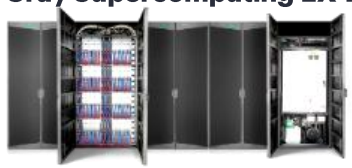
100% ファンレス 直接水冷 (DLC)

HPC と AI の複合的なワークロードに対応して再設計されたスーパーコンピューター

HPE Cray Supercomputing GX5000

水

HPE Cray Supercomputing EX4000



水

HPE Cray Supercomputing EX2500

水

HPE Slingshot はEthernetをベースに独自開発した高性能・高コスト効率のネットワーク。

空 水

アクセラレーテッド AI

空冷と直接水冷 (DLC) のハイブリッド方式、または空冷と liquid to air (L2A) 冷却の組み合わせ

AIモデルの学習、チューニング、推論に最適な 8GPU 搭載 AI サーバー

HPE Cray XD670

空 水

HPE ProLiant Compute XD685

空 水

HPE Compute XD690

空

企業向けAI アプリケーションを加速する HPE ProLiant Compute サーバー

HPE ProLiant Compute DL380a Gen12

空 水

HPE ProLiant Compute DL384 Gen12

空

パラメータ数が 1 兆以上の AI モデル向けラックスケールソリューション

NVIDIA GB200 NVL 72 by HPE

水

NVIDIA GB300 NVL 72 by HPE

水

メインストリームHPC

HPC 向け高密度型スケールアウトサーバー

HPE ProLiant Compute XD230

空 水

HPE Cray XD2000

空 水

Converged HPC/AI

高密度ラックスケールソリューション

NVIDIA GB200 NVL4 by HPE

水

高速ストレージ

HPC、AI、複合ワークロードに対応し、価格性能比に優れたLustreあるいはDAOSストレージ

HPE Cray Supercomputing Storage Systems K3000

DAOS

空

HPE Cray Supercomputing Storage Systems E2000

Lustre

空

HPE Cray Storage Systems C500

Lustre

空

統合された HPC / AI 用ソフトウェアのポートフォリオにより、システム管理とワークロード管理、オーケストレーション、計算環境の拡張、ソフトウェア開発を実現可能

世界中に在籍する HPE Services のエキスパートがお客様のHPC / AI 戦略を強力に支援

大規模 AI モデルのトレーニング、チューニング、推論に最適化された 8GPU搭載サーバー

HPE Cray XD670

- NVIDIA H200 GPU x 8
- 第 5 世代 Intel Xeon プロセッサー x 2
- 空冷または水冷
- 5U 筐体
- MLPerf Inference: Datacenter v5.1 ベンチマークの 6 つのシナリオで 1 位を獲得 [1]
- HPE Performance Cluster Manager (HPCM)

HPE ProLiant Compute XD685

- GPU は以下から選択可能[2]
  - NVIDIA H200 GPU x 8 / NVIDIA B200 GPU x 8 / AMD Instinct MI355X x 8
- 第 5 世代 AMD EPYCプロセッサ
- 水冷 (5U) または空冷 (6U)
  - H200 は水冷・空冷可 / B200 と MI355X は水冷のみ
- HPE iLO, HPE Performance Cluster Manager (HPCM)

HPE Compute XD690

- NVIDIA B300 GPU x 8
- P-cores 搭載 Intel Xeon 6 プロセッサー x 2
- 空冷
- 10U 筐体
- HPE Performance Cluster Manager (HPCM)

[1] HPE delivers several world records in latest MLPerf Inference benchmarks, September 2025
[2] GPU は順次リリース予定

HPE の水冷ソリューション

HPE Cray XDおよび HPE ProLiant Compute XD

Liquid-to-air 冷却

70% 直接水冷 (DLC)

HPE Cray Supercomputing EX/GX

100% ファンレス直接水冷 (DLC)

右に行くほど冷却効率と容量 (kW/rack) が向上

2. HPE の AI factory ソリューション

HPE の AI factory ポートフォリオ

Turnkey AI factory (HPE Private Cloud AI)
一般の企業や組織

AI factory at scale
モデル開発者、サービスプロバイダ

Sovereign AI factory
政府、公共機関

共通のコントロールプレーン: HPE Morpheus, HPE OpsRamp

- 迅速な投資対効果 (ROI) への要求
- NVIDIA Softwareの活用
- 推論、チューニング
- 空冷
- 最大64GPU: RTX PRO 6000, H100/200

- お客様の規模や状況に合わせて調整
- サービス統合型ソフトウェアスタック
- モデル開発、トレーニング、推論
- 直接水冷 (DLC) または空冷
- 100～10,000GPU: H200, B200

- 他システムからの分離
- 厳格なデータ主権
- モデル開発、トレーニング、推論
- 直接水冷 (DLC) または空冷
- 100～10,000GPU: H200, B200

ターンキー型の設計済みシステム

カスタマイズ可能な検証済みソリューション

Turnkey AI factory

Turnkey AI factory | AI factory at scale | Sovereign AI factory

HPE Private Cloud AI (PCAI)

AI モデル	コミュニティ + パートナー + カスタム NVIDIA NIM	
AI ソフトウェア	NVIDIA AI Enterprise	HPE AI Essentials HPE OpsRamp HPE Data Fabric
AI インフラストラクチャ	NVIDIA GPU NVIDIA Spectrum-X	HPE AI サーバー HPE ファイルストレージ
AI サービス	HPE Services インストレーション・サポート	検証済み 3rdパーティー ユースケース

セットアップ、インベントリー、ITおよびアセットライフサイクル管理

ワークロードプロビジョニング、オーケストレーションおよび管理

VM ランタイム
Kubernetesクラスター管理

監視、可観測性および
アプリおよびインフラ向けE2E セキュリティ

HPE Private Cloud コントロールプレーン

Developer System	Small	Medium	Large	Expansion Rack



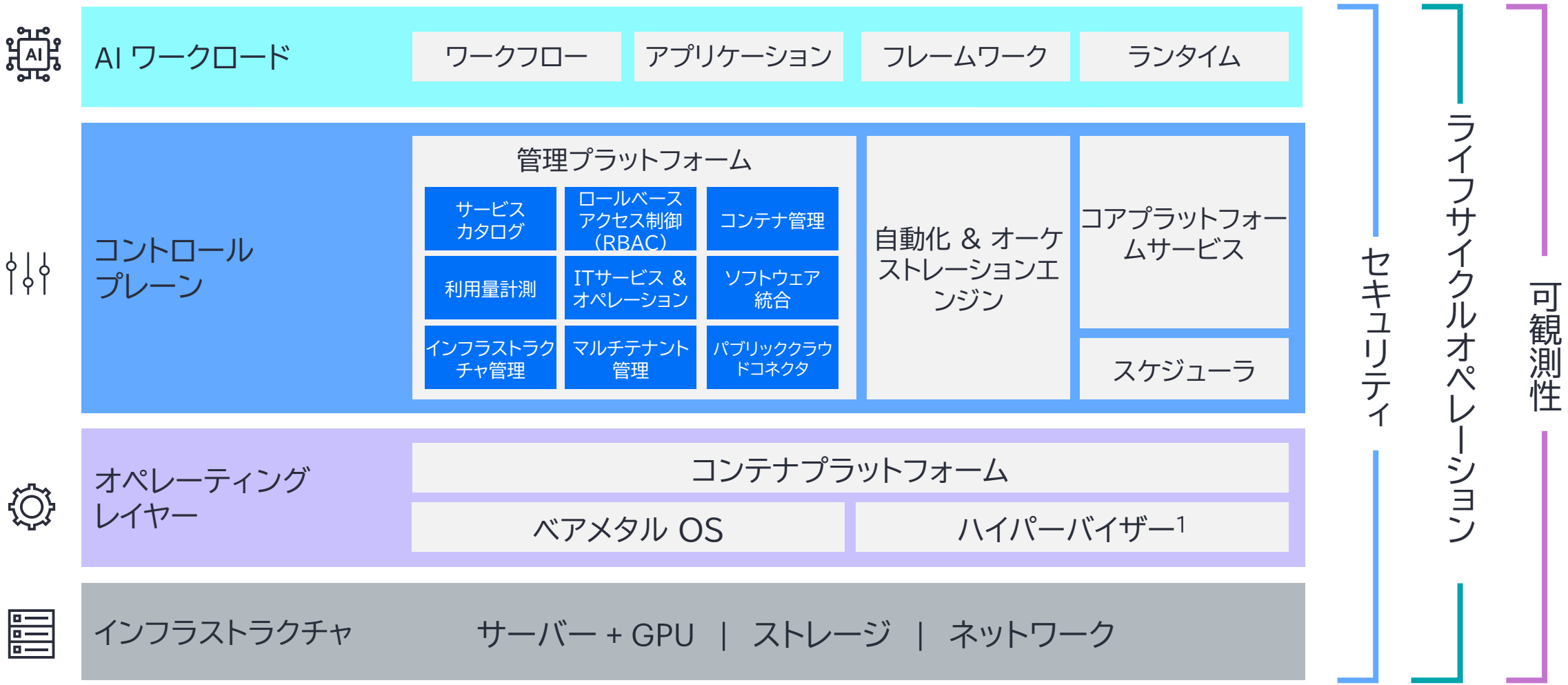


企業における AI ワークロードに合わせた構成

HPE GreenLake cloud

AI factory at scale / Sovereign AI factory の構成要素

Turnkey AI factory | AI factory at scale | Sovereign AI factory



[1] 2026年Q1 以降に対応予定

AI factory at scale / Sovereign AI factory の構成要素

Turnkey AI factory | AI factory at scale | Sovereign AI factory

インフラストラクチャ

AI ワークロード

	GPU サーバー	ネットワーク	HPE AI データストレージ

コントロールプレーン	• HPE Cray XD670
	  • NVIDIA H200 GPU x 8
	• HPE ProLiant Compute XD685
	  • NVIDIA B200 GPU x 8
	2026年対応予定
		• NVIDIA Quantum-2 QM9700 スイッチ
		  • GPU 間通信 (East-West)
		• NVIDIA Spectrum-4 SN5600 スイッチ
		  • In-Band/Storage 用 (North-South)
		• Juniperスイッチ 2026年対応予定
			• Weka ソリューション
			  • HPE ProLiant DL325 Gen11
			  • HPE Alletra Storage Server 4110
			• VAST ソリューション 2026年対応予定
			• HPE Alletra Storage MP X10000

オペレーティングレイヤー	 		

インフラストラクチャ　サーバー + GPU | ストレージ | ネットワーク

AI factory at scale / Sovereign AI factory の構成要素

Turnkey AI factory | AI factory at scale | Sovereign AI factory

オペレーティングレイヤー

AI ワークロード

コントロールプレーン

オペレーティングシステム	ハイパーバイザー*	コンテナプラットフォーム
• Canonical Ubuntu 24.x • Red Hat Enterprise Linux 9.x	• オープンソース Linux KVM • VMware vSphere • HPE Morpheus VM Essentials • RedHat Enterprise Virtualization (RHEV)	• オープンソース Kubernetes • Red Hat OpenShift Container Platform • SUSE Rancher Kubernetes Engine (RKE2)

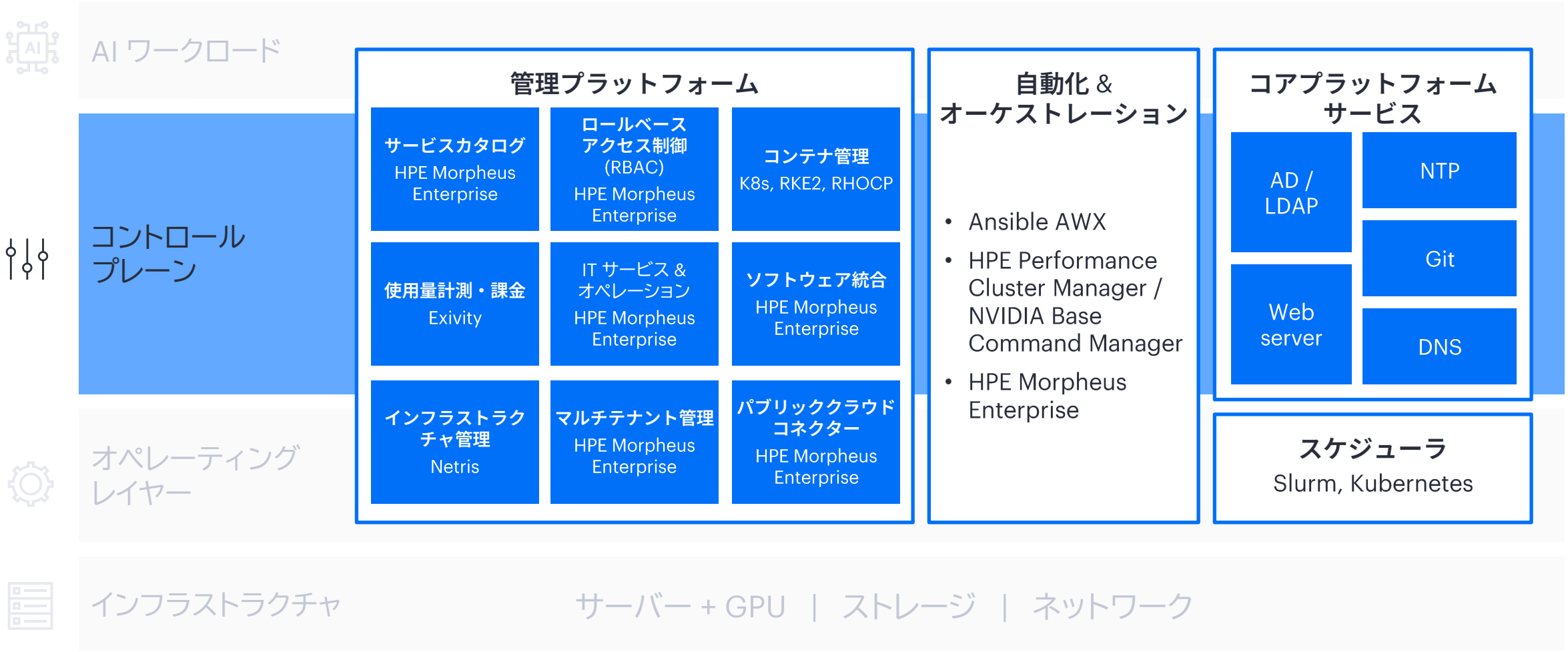
オペレーティングレイヤー

インフラストラクチャ　　サーバー + GPU | ストレージ | ネットワーク

AI factory at scale / Sovereign AI factory の構成要素

Turnkey AI factory | AI factory at scale | Sovereign AI factory

コントロールプレーン



AI factory at scale / Sovereign AI factory の構成要素

Turnkey AI factory | AI factory at scale | Sovereign AI factory

オペレーティングレイヤー

AI ワークロード

ワークフロー	アプリケーション	フレームワーク	ランタイム
チャットボット 自然言語処理 AI エージェント コンテンツ生成	Jupyter LangChain NVIDIA blueprints	PyTorch TensorFlow LlamaIndex	NVIDIA Run:ai ONNX Ray

コントロール
プレーン

オペレーティング
レイヤー

インフラストラクチャ　　サーバー + GPU ｜ ストレージ ｜ ネットワーク

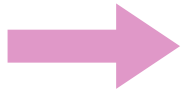
AI factory at scale / Sovereign AI factory におけるマルチテナント

課題

- 複数の「テナント」に対し、リソースを分割して互いにアクセスできない / 影響を与えない形で提供したい

現在の実装

- サーバー：物理ノード単位で分割
- ネットワーク：
  - InfiniBand は Partition Key (PKey) を使用
  - Ethernet は VXLAN を使用
  - Netris[1] ソフトウェアにより管理
- ・ストレージ：
  - CPU、メモリ、ドライブ単位の分割と VLAN を使用



今後のバージョンにおいて、分割単位の選択肢が増える予定

[1] Netris は、InfiniBand の管理を NVIDIA Unified Fabric Manager (UFM) を通じて行い、Ethernet の管理を NVIDIA NetQ を通じて行う。
https://www.netris.io/docs/en/latest/netris-architecture.html

3. まとめ

まとめ

- HPEは、サーバー、ネットワーク、ストレージの幅広いラインナップを展開
  - 比較的小規模なPCクラスタから、TOP500ランキングの上位に位置するスーパーコンピュータまで対応
  - 各種の水冷ソリューションも合わせてご提供

- AIの推論・学習向けに、各種のリファレンスアーキテクチャに準じた構成と、実績のあるソフトウェアの組み合わせをご提案
  - すぐに使える「ターンキー型」のPCAI（Private Cloud AI）ソリューション
  - カスタマイズ可能な大規模向け AI factory at Scale および Sovereign AI factory
  - マルチテナントに対応（分割単位は今後、更新予定）

Thank you