


AI時代のデータインテンシブHPCプラットフォームの開発と展開

Software Defined Rack Scale Structure
Hyper Matrix HPC

2024/6/27

コアマイクロシステムズ株式会社 鶴見利章

ミックスワークロードデマンド

- SIMULATION / MODELING
- AI / ML & DL
- BIGDATA ANALYSTIC

ソフトウェア定義型トレンド

- DISAGGREGATED COMPOSABLE
- HYPER CONVERGENCE
- SDS / SDN / SDWAN

クラウド分散 / 広域化トレンド

- HYBRID CLOUD
- MULTICLOUD ROUTING
- MP EDGE/GRID CLOUD

階層ファブリック構成

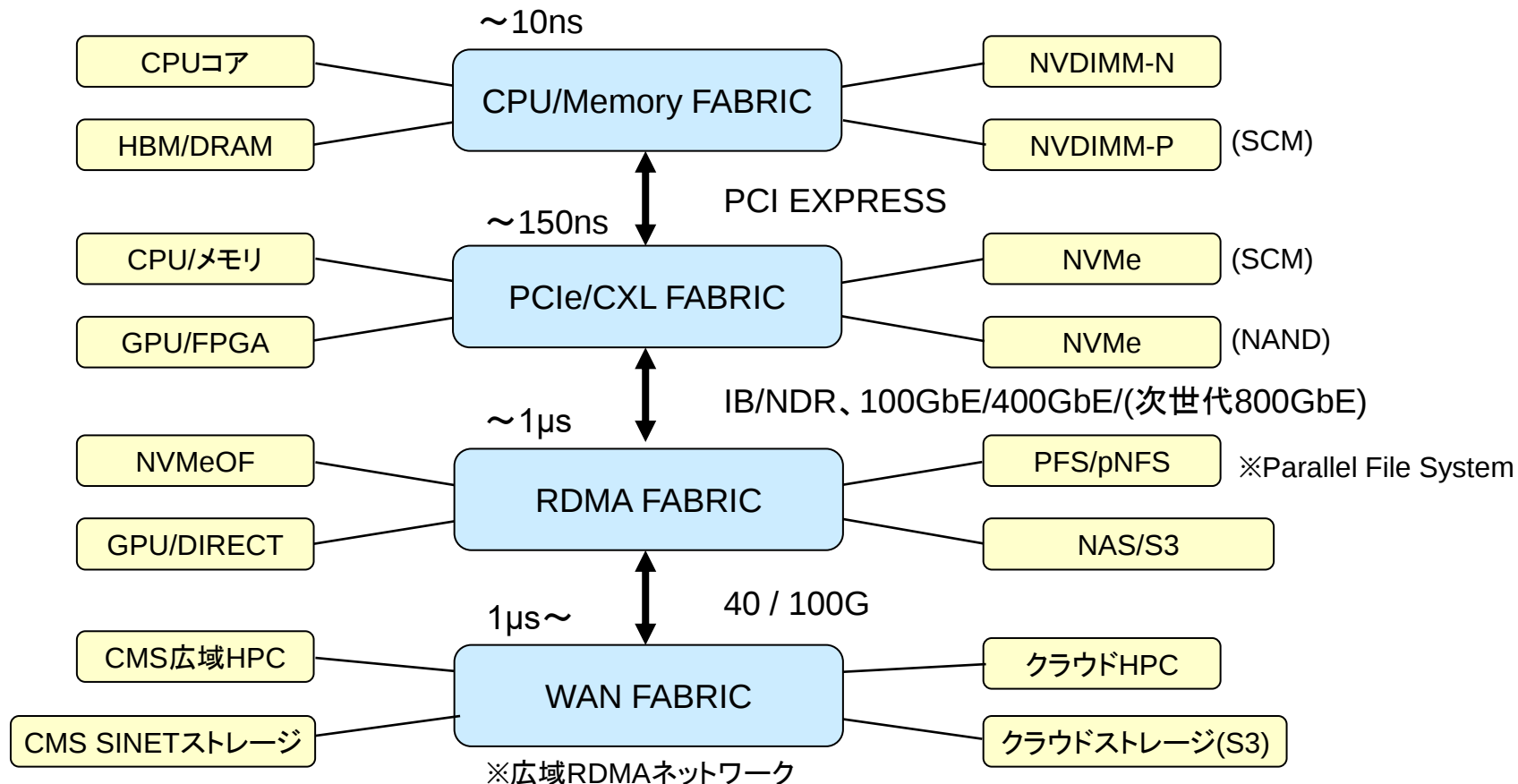
- MEMORY FABRIC
- PCIe/CXL FABRIC
- RDMA FABRIC
- WAN FABRIC

ハイパースケール構成

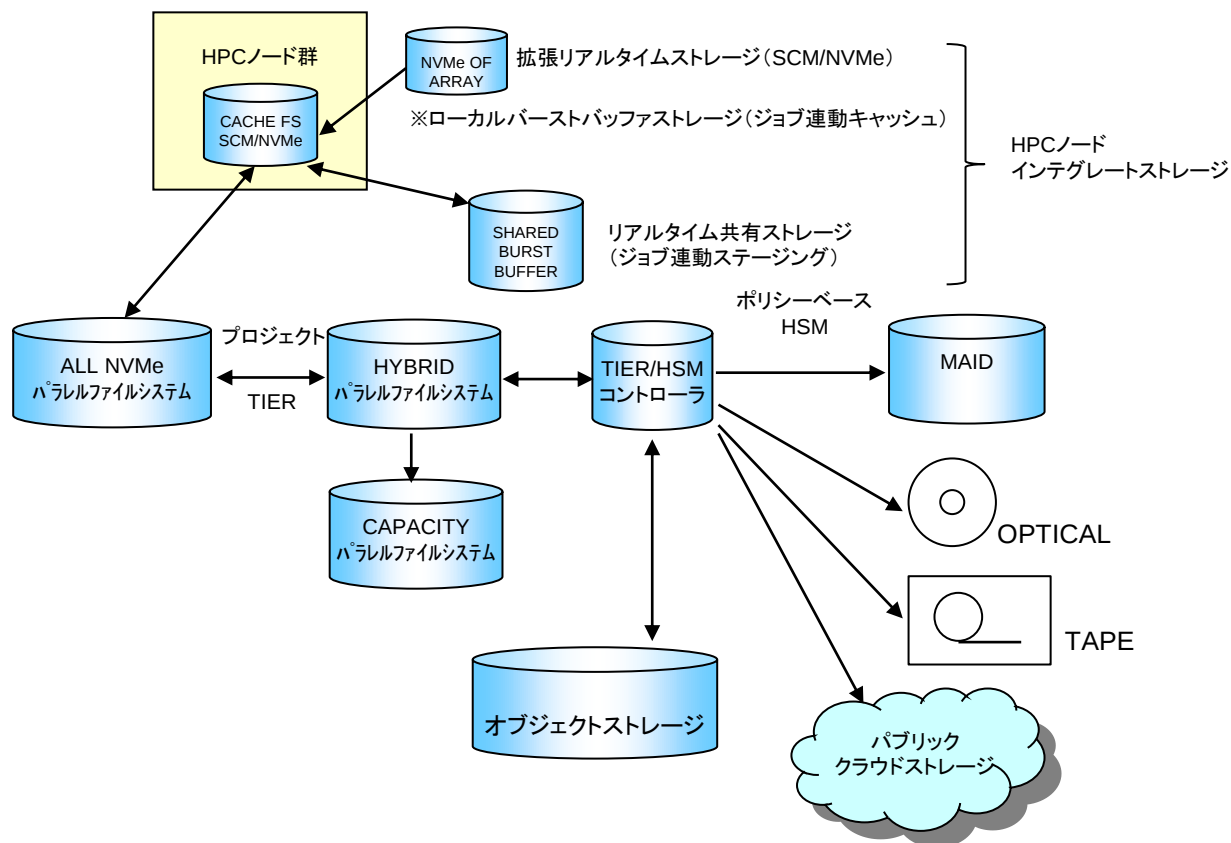
- RACK SCALE FUSION HPC
- HYPER CONVERGED HPC
- DISAGGREGATED COMPOSABLE HPC

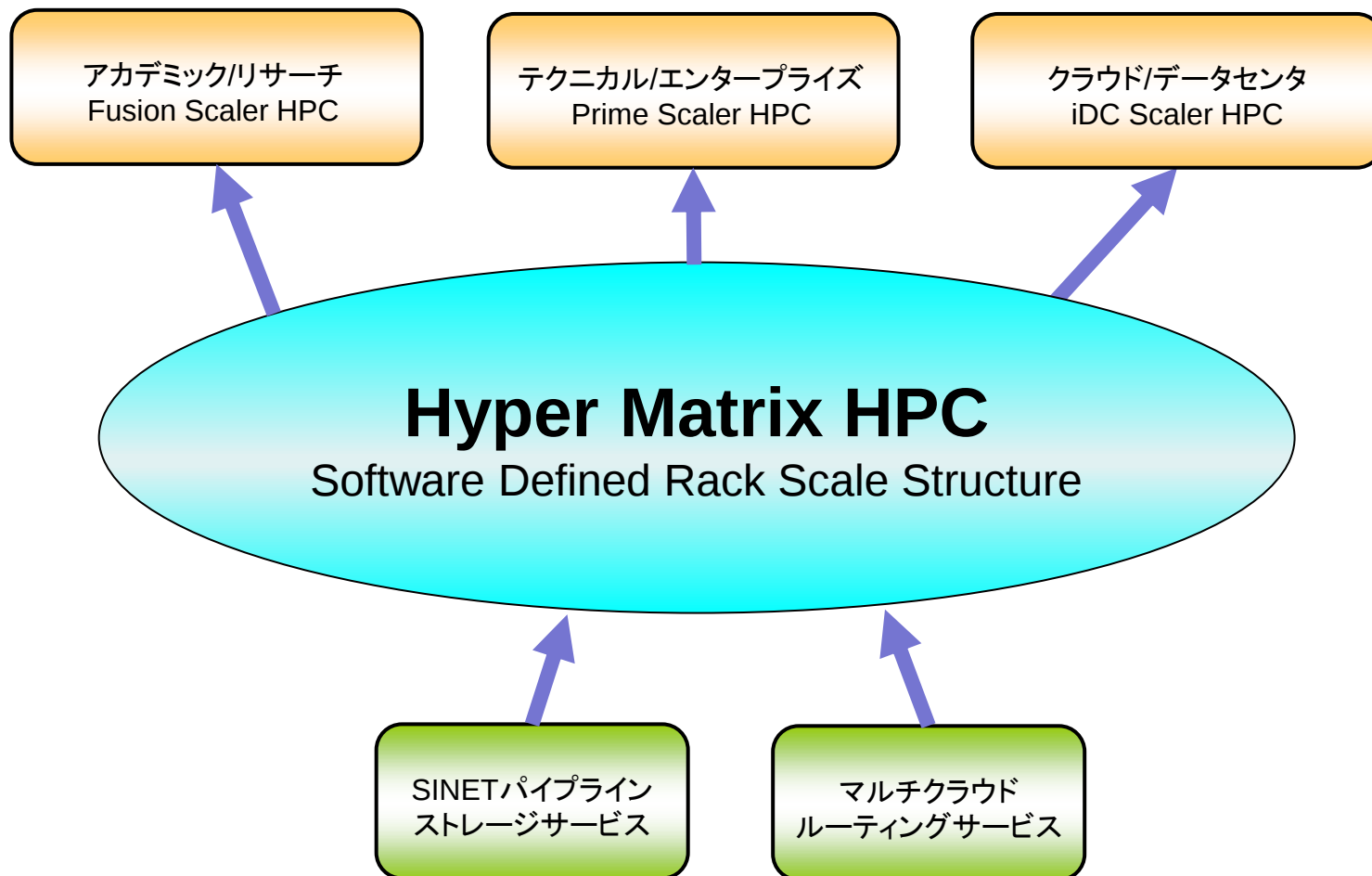
グローバルスケール構成

- MULTI SITE HPC
- HYBRID CLOUD HPC
- MULTI CLOUD HPC

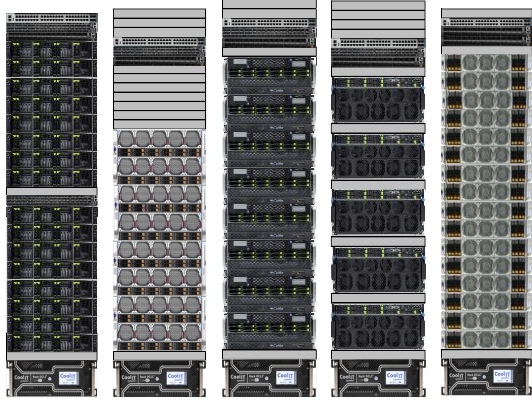


HPCストレージシステムのレイヤ構成イメージ





アカデミックリサーチ向け
Fusion Scaler HPC
Modeling & Simulation
AI/Analytics Hybrid
■ Large Cluster HPC ■ Hybrid Cluster HPC ■ Flex HPC Resource
• Fusion Pod MAX CPU • Fusion Pod BIG MEM • Fusion Pod Flex GPU • Fusion Pod Pro GPU • Fusion Pod Pro APU



テクニカル/エンタープライズ向け
Prime Scaler HPC
AI/CAE/EDA/BIO
BigData Hybrid
■ HCI/CDI Cluster HPC ■ Single Big Machine HPC ■ Hybrid HPC Resource
• Prime Pod HCI • Prime Pod CDI • Prime Pod GPU MAX • Prime Pod MEM MAX



クラウド/データセンタ向け
iDC Scaler HPC
AI/Analytics
HPC Hybrid
■ AI Native BIG HPC ■ Composable AI HPC ■ Large AI Resource
• iDC Pod GenAI • iDC Pod Green • iDC Pod Dynamix
※フルコンポーザブルAIプラットフォーム(開発中)

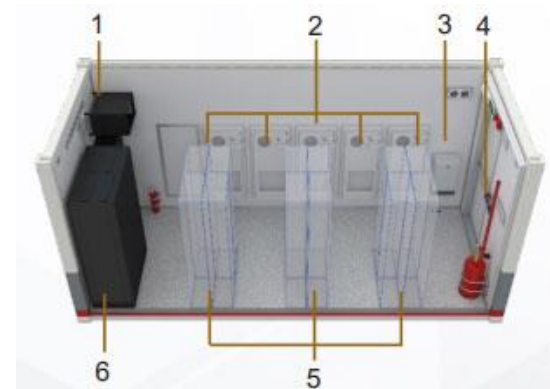


■屋内プレハブ方式

- ・In-Row冷却

■屋外コンテナ方式

- ・In-Row冷却
- ・Rear Door冷却
- ・Immersion冷却
- ・外気連動冷却



■アカデミック/リサーチ向けFusion Scaler HPC

■Fusion POD MAX CPU

■Fusion POD BIG MEM

■Fusion POD Flex GPU

■Fusion POD Pro GPU

■Fusion POD Pro APU

■ Fusion Pod MAX CPU

■ ラック内機器構成

- ・2U/4ノード(8CPU/AMDまたはIntel)サーバ x 16
- ・1U/64ポート IB NDR(400G)スイッチ x 2
- ・1U/32ポート100GbEスイッチ x 1(オプション)
- ・1U/48ポート管理スイッチ x 2

■ 性能指標(FP64)

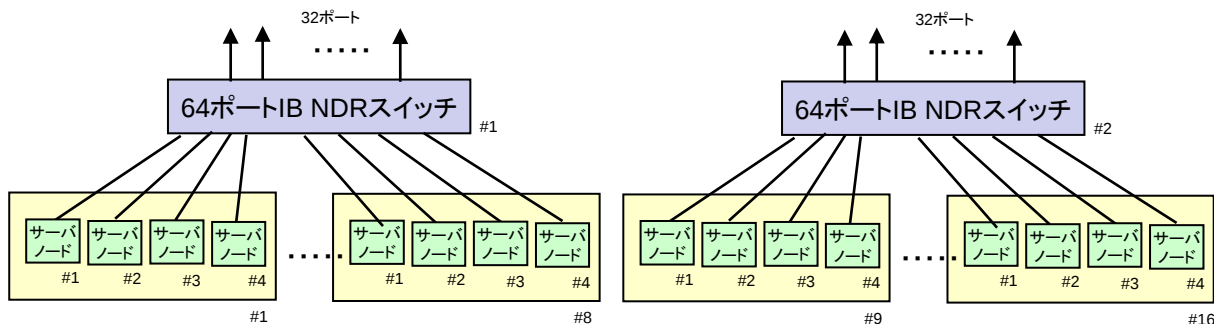
- ・CPU群(AMD GENO構成):最大589TFLOPS
- ・CPU群(Intel XEON構成):最大707TFLOPS

■ 冷却システム

- ・コールドプレート型液体冷却
- ・冷却能力:80KW

■ HPCノード群インタコネクト構成

- ・フルバイセクションFAT-TREE



■ Fusion Pod BIG MEM

■ ラック内機器構成

- ・6U/8CPU (Intel) / 128DIMM (124NVMe) サーバ x 4
- ・1U/64ポートIB NDR (400G) スイッチ x 1
- ・1U/32ポート100GbE x 1 (オプション)
- ・1U/48ポート管理スイッチ x 1

■ 性能指標 (FP64)

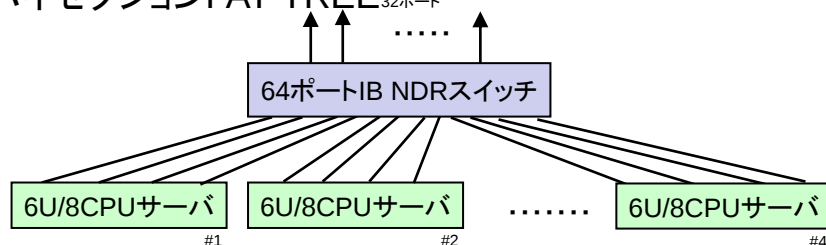
- ・CPU群: 最大120TFLOPS
- ・GPU群: 最大1040TFLOPS

■ 冷却システム

- ・リアドア型水冷熱交換器 (アクティブファン型)
- ・冷却能力: 40KW

■ HPCノード群インタコネクト構成

- ・フルバイセクションFAT-TREE^{32ポート}



■ Fusion Pod Flex GPU

■ ラック内機器構成

- ・4U/8GPU (PCIe型)/2CPU (AMD)/12NVMeサーバ x 8
- ・1U/64ポートIB NDR (400G) スイッチ x 1
- ・1U/32ポート100GbEスイッチ x 1 (オプション)
- ・1U/48ポート管理スイッチ x 1

■ 性能指標 (FP64)

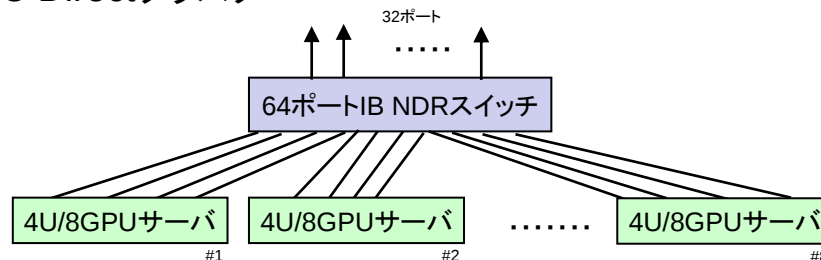
- ・CPU群: 最大73TFLOPS
- ・GPU群: 最大2176TFLOPS

■ 冷却システム

- ・リアドア型水冷熱交換器 (アクティブファン型)
- ・冷却能力: 40KW

■ HPCノード群インタコネクト構成

- ・フルバイセクションFAT-TREE
- ・GPU Directクラスタ



■ Fusion Pod Pro GPU

■ ラック内機器構成

- ・5U/8GPU (SXMまたはOAM)/8NVMeサーバ x 4
- ・1U/64ポートIB NDR (400G) スイッチ x 1
- ・1U/32ポート100GbEスイッチ x 1 (オプション)
- ・1U/48ポート管理スイッチ x 1

■ 性能指標 (FP64)

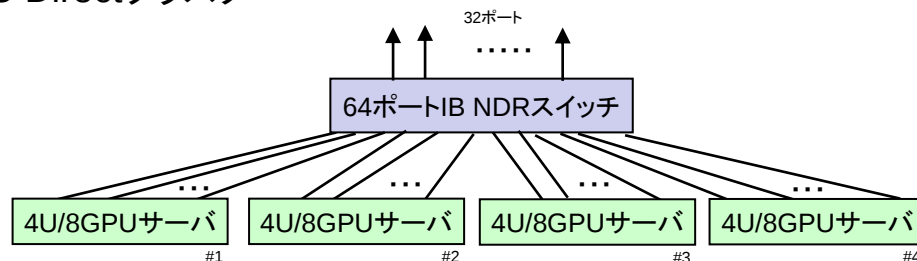
- ・CPU群: 最大32TFLOPS
- ・GPU群: 最大2614TFLOPS

■ 冷却システム

- ・コールドプレート液体冷却
- ・冷却能力 80KW

■ HPCノード群インタコネクト構成

- ・フルバイセクションFAT-TREE
- ・GPU Directクラスタ



ラックスケール型高密度ユニファイドメモリAPU HPCビルディングブロック

■ Fusion Pod Pro APU

■ ラック内機器構成

- ・2U/4APU (AMD Mi300A) / 8 NVMe サーバ x 16
- ・1U/64ポート IB NDR (400G) スイッチ x 2
- ・1U/48ポート 管理スイッチ x 1

■ 性能指標 (FP64)

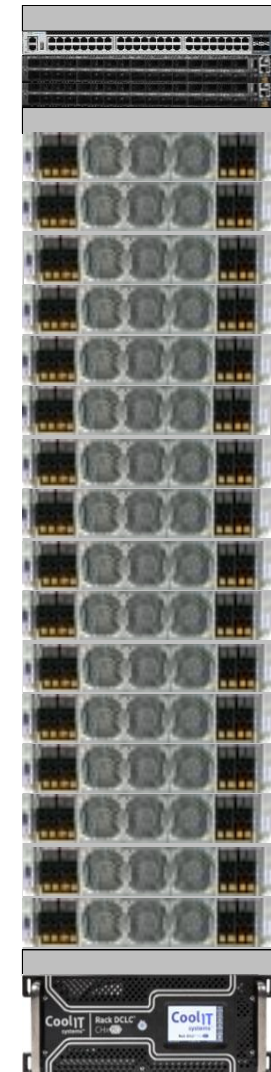
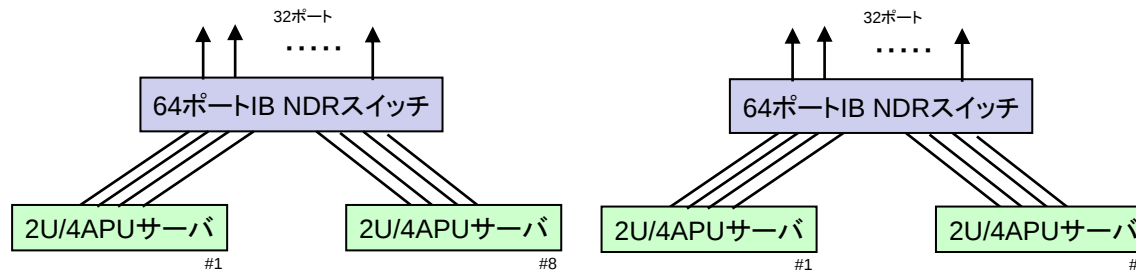
- ・APU群: 最大3.92 PFLOPS

■ 冷却システム

- ・コールドプレート型液体冷却
- ・冷却能力: 80KW

■ HPCノード群インタコネクト構成

- ・フルバイセクションFAT-TREE



- Prime Pod HCI
- Prime Pod DCI
- Prime Pod GPU MAX
- Prime Pod Mem MAX

■Prime Pod HCI

■ラック内機器構成

- ・1U/2CPU(AMD/128コアGENOA)/1GPUサーバ x 32
- ・1U/64ポート IB NDR(400G)スイッチ x 1
- ・1U/32ポート100GbEスイッチ x 1(オプション)
- ・1U/48ポートGbE管理スイッチ x 1

■性能指標(FP64)

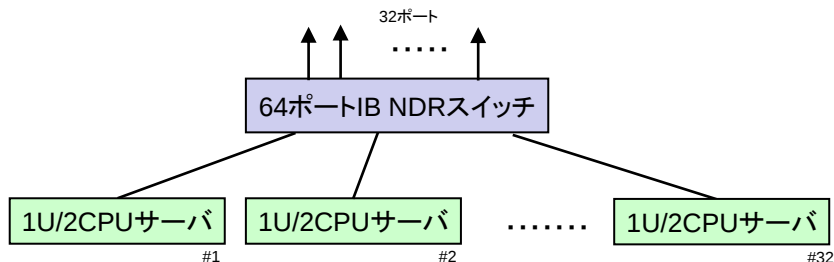
- ・CPU群: 最大294TFLOPS
- ・GPU群: 最大832TFLOPS(オプション)

■冷却システム

- ・リアドア型水冷熱交換器(アクティブファン型)
- ・冷却能力: 40KW

■HPCノード群インタコネクト構成

- ・フルバイセクションFAT-TREE



■ Prime Pod CDI

■ ラック内機器構成

- ・1U/2CPU/1GPU/12NVMe x 16
- ・4U/8GPU/JBOG x 2
- ・2U/32スロット JBOM x 2
- ・2U/24NVMe JBOF x 2
- ・1U PCIe/CXLスイッチ x 1
- ・1U/32ポート100GbEスイッチ x 1(オプション)
- ・1U/48ポート管理スイッチ x 1

■ 性能指標

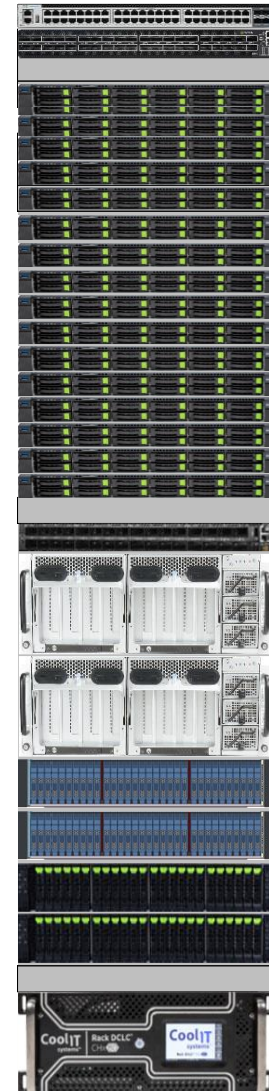
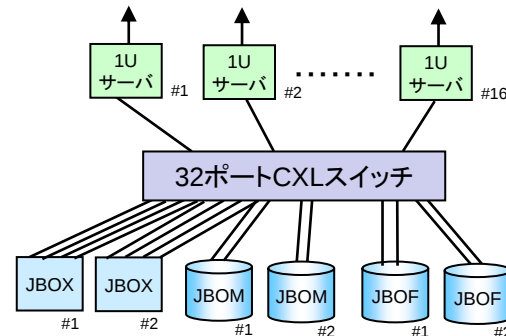
- ・構成依存
- ・単体マシン化/複数マシン化

■ 冷却システム

- ・リアドア型水冷熱交換器(アクティブファン型)
- ・冷却能力:40KW

■ HPCノード群ローカルインタコネクト構成

- ・PCIe/CXLコンポーザブル



ラックスケール型ラージGPUシングルマシン HPCビルディングブロック

■ Prime Pod GPU MAX

■ ラック内機器構成

- ・6U/8CPU(Intel)/8GPU/128DIMM/24NVMeサーバ x 1
- ・4U/8GPU JBOG x 8
- ・1U/48ポート管理スイッチ x 1

■ 性能構成

- ・CPUコア数: 最大448コア
- ・GPU数: 最大64+(8)
- ・メモリ総量: 最大32TB

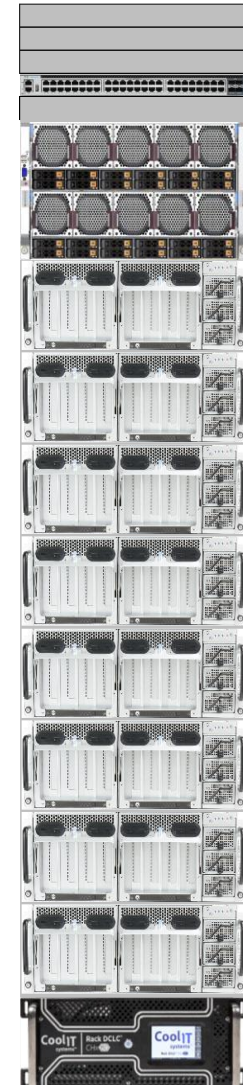
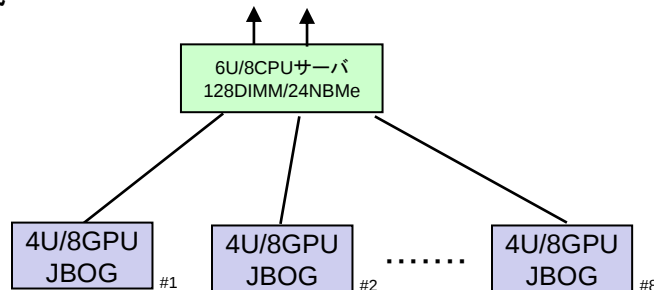
■ 性能指標(FP64)

- ・CPU群: 最大44TFLOPS
- ・GPU群: 最大2176TFLOPS

■ 冷却システム

- ・リアドア型水冷熱交換器(アクティブファン型)
- ・冷却能力: 40KW

■ プラットフォーム構成



ラックスケール型ラージメモリスシングulumashin

HPCビルディングブロック

■Hyper Pod Mem MAX

■ラック内機器構成

- ・6U/8CPU(Intel)/DIMM/24NVMeサーバ x 6
- ・1U/16ポート(x16) CXL2.0スイッチ x 3
- ・1U/32ポート100GbEスイッチ x 1(オプション)
- ・1U/48ポート管理スイッチ x 1

■性能構成

- ・CPUコア数:最大2688コア
- ・メモリ総量:最大196TB
- ・NVMe総量:最大4.32PB

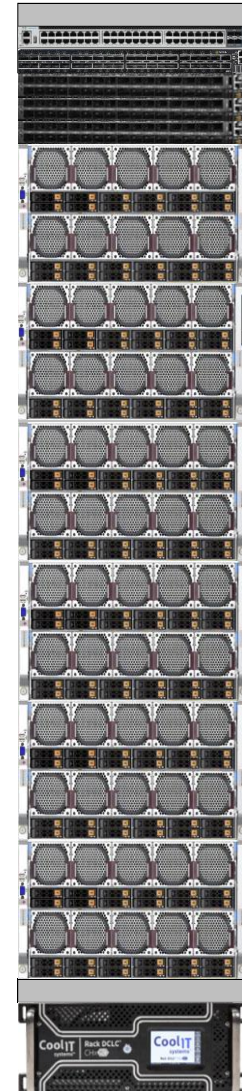
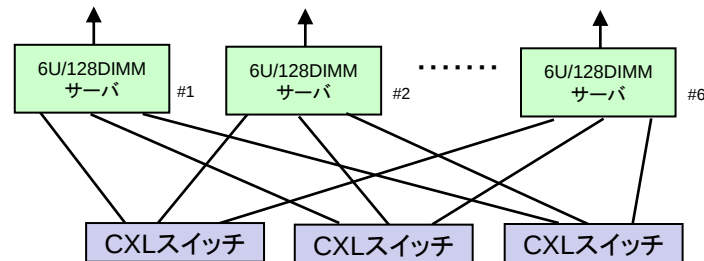
■性能指標(FP64)

- ・CPU群:最大265TFLOPS

■冷却システム

- ・リアドア型水冷熱交換器(アクティブファン型)
- ・冷却能力:40KW

■プラットフォーム構成



■ iDC Pod GenAI

■ iDC Pod Dynamix ※計画中(コンポーザブルリソースラック群/CPU/メモリ/GPU/NVMe)

■ iDC Pod Green

ラックスケール型ラージAPUクラスター HPCビルディングブロック

■ iDC Pod GenAI

■ ラック内機器構成

- ・NVIDIA GB200ブレード x 18
- ・NVLINKスイッチ x 9
- ・1U/64ポート IB NDR(400G)スイッチ x 3
- ・1U/32ポート 100GbEスイッチ x 2
- ・OCP電源ユニット x 4

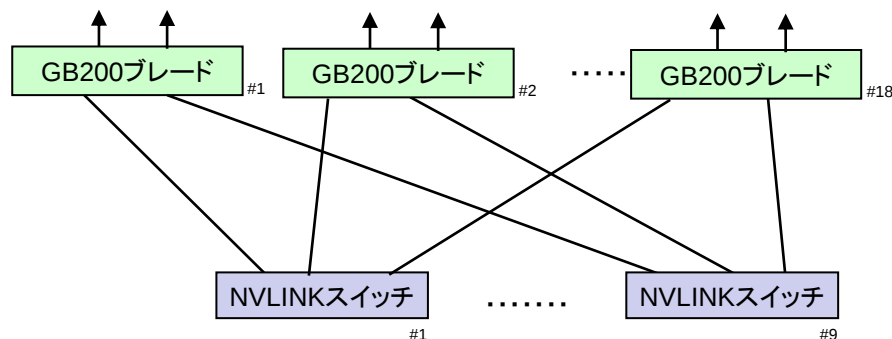
■ 性能指標

- ・FP64: 最大3.24PFLOPS
- ・FP16/AI: 最大360PFLOPS
- ・FP4/AI: 最大1440PFLOPS

■ 冷却システム

- ・コールドプレート型液体冷却
- ・冷却能力: 130KW

■ プラットフォーム構成



ラックスケール型液浸冷却専用GPU HPCビルディングブロック

■ iDC Pod Green

■ ラック内機器構成

- ・1U/4GPU (PCIe) / 2CPU (AMD/Intel) サーバ (OCP3.0) x 32
- ・1U/64ポートIB NDR (400G) スイッチ x 1
- ・1U/32ポート100GbEスイッチ x 1 (オプション)
- ・1U/48ポート管理スイッチ x 1
- ・DPU NICオプション
- ・OCP電源ユニット x 2 (4)

■ 性能指標 (FP64)

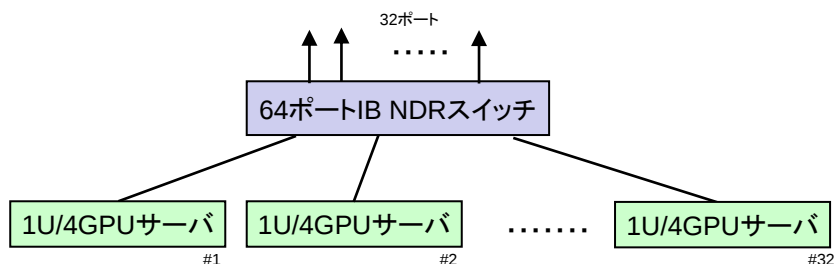
- ・CPU群: 最大294TFLOPS
- ・GPU群: 最大3328TFLOPS

■ 冷却システム

- ・Single Phase液浸冷却
- ・冷却能力: 最大130KW

■ HPCノード群インタコネクト

- ・フルバイセクションFAT-TREE



Flash Scalerシリーズ

- Flash Scaler BB
- Flash Scaler Primery
- Flash Scaler Central

ラックスケール型バーストバッファ HPCストレージビルディングブロック

■Flash Scaler BB

■ラック内機器構成

- ・2U/24NVMe CiB NVMeOFストレージ x 8
- ・1U/12NVMeストレージサーバ(Intel DAOSベース) x 16
- ・1U/64ポートIB NDR(400G)スイッチ x 2
- ・1U/48ポート管理スイッチ x 1

■性能指標

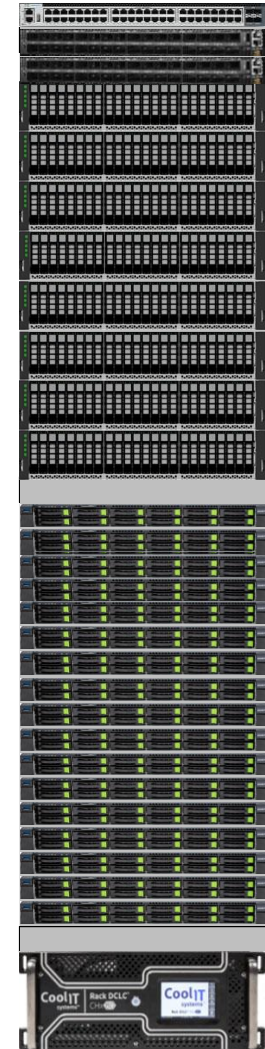
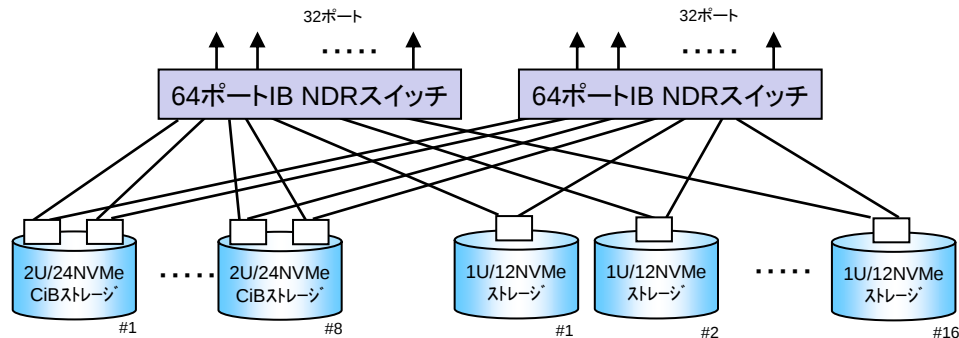
- ・最大スループット: 2.3TB/s + 0.76TB/s(最大1.52TB/s)
- ・最大IOPS: 128M + 128M
- ・最大容量: 5.7PB + 5.7PB

■冷却システム

- ・リアドア型水冷熱交換器(パッシブファン型)
- ・冷却能力: 40KW

■ストレージノード群インタコネクト構成

- ・2重化スイッチ構成
- ・フルメッシュ接続



ラックスケール型プライマリー NVMe HPCストレージビルディングブロック

■ Flash Scaler Primary

■ ラック内機器構成

- ・2U/24NVMe CiB Lustre PFSストレージ x 16
- ・1U/64ポートIB NDR (400G) スイッチ x 2
- ・1U/48ポート管理スイッチ x 1

■ 性能指標

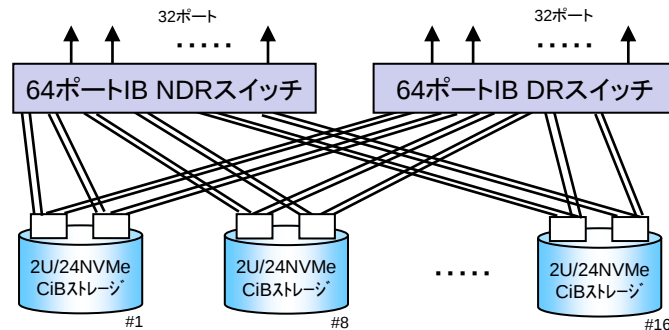
- ・最大スループット: 4.6TB/s
- ・最大IOPS: 256M
- ・最大容量: 11.5PB

■ 冷却システム

- ・リアドア型水冷熱交換器 (パッシブファン型)
- ・冷却能力: 40KW

■ ストレージノード群インタコネクト構成

- ・2重化スイッチ構成
- ・フルメッシュ接続



ラックスケール型セントラル NVMe HPCストレージビルディングブロック

■ Flash Scaler Central

■ ラック内機器構成

- ・8U/160NVMe CiB Lustre PFSストレージ x 4
- ・1U/64ポートIB NDR (400G) スイッチ x 2
- ・1U48ポート管理スイッチ x 1

■ 性能指標

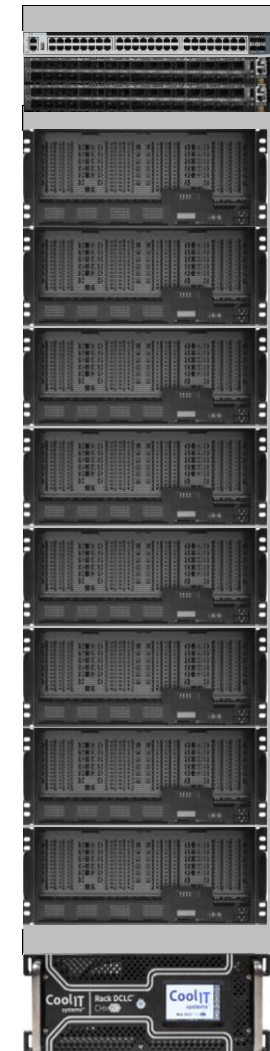
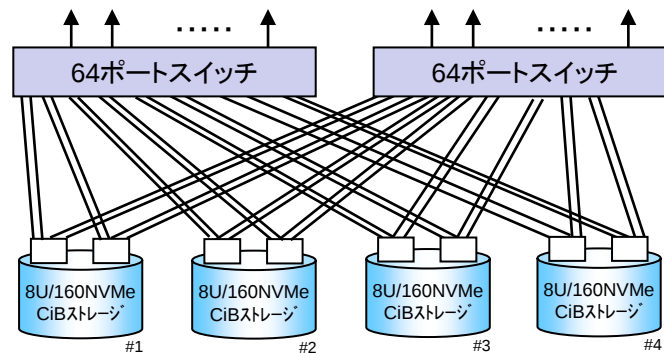
- ・最大スループット: 768GB/s
- ・最大IOPS: 32M
- ・最大容量: 38.4PB (PLC NAND構成時)

■ 冷却システム

- ・リアドア型水冷熱交換器 (パッシブファン型)
- ・冷却能力: 40KW

■ ストレージノード群インタコネクト構成

- ・2重化スイッチ構成
- ・フルメッシュ接続



■オブジェクトストレージ

- ・CMS Object STOR
- ・900HDD(21.6PB/物理容量)/ラック
- ・Cephストレージクラスタ

■MAID省電力ストレージ

- ・CMS MAID STOR
- ・816HDD(最大19.5PB/物理容量)/ラック
- ・最大90%の省電力コントロール

■光ディスクライブラリー

- ・SONY PETA SITE
- ・最大 2.9PB/ラック

■テープライブラリ

- ・Quantum Scaler i6000(LTO9)
- ・最大 46.8PB/ラック

■階層ストレージ管理ソフトウェア

- ・HSM/TIERストレージ統合(NODEUM)
- ・広域ストレージメタデータ統合(iRODS)

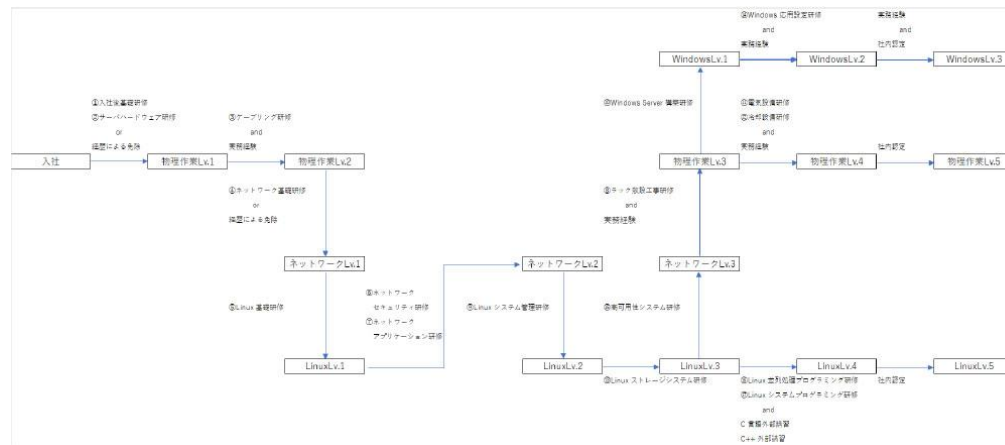
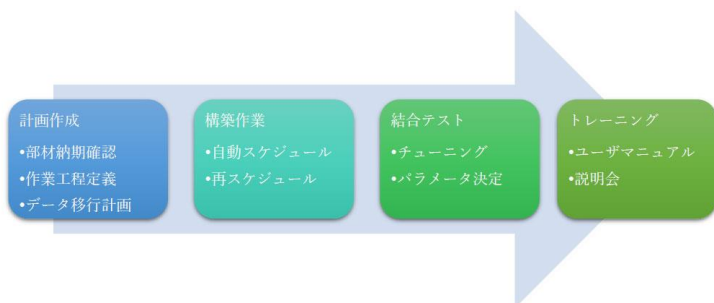


CMS Object

CMS MAID

SONY
PETA SITE

Quantum
Scaler i6000



運用保守監視業務

保守監視業務には、ログ集中管理とAIによる自動異常検知システム (CMS Log Ponder) を適用します。このシステムは、本システムより出力された各種システムログをそのまま保存する「ログ集積機能」、集められたログを種類とエラー内容に従い分類する「ログ分類機能」、ログの統計的な特徴を分析する「ログ統計機能」、ログ内容から異常を検知する「ログ異常検知機能」を有しています。

1. ログ集積機能

本システムからはシステム状態を通知するログやセンサー値が時系列データとして出力されます。これらは CMS Log Ponder の時系列データベースへ集約保存されます。

2. ログ分類機能

ログ分類機能は、ログを出力したホスト、メッセージクラス、特定の文言など、内容に応じてログデータを分類します。

3. ログ統計機能

ログ統計機能は、ログに対する様々な統計情報を提供します。これには、特定期間内での発生件数、センサー値の増減、時間ウィンドウを変化させた場合の、平均値、中央値、標準偏差など基本的な統計量の変化などのデータを、管理者様の要求に応じて作成します。

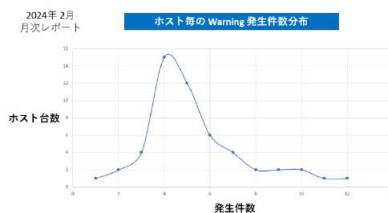


図 4 統計レポート作成例

4 ログ異常検知機能

本システムには多数の計算ノード、ストレージノード、ストレージ装置、ネットワーク機器などが含まれており、それらを単純な閾値ベースの監視のみで運用するのは大変な困難を伴います。弊社は AI モデルを適用し定常状態の時系列ログデータを学習させることにより、AI モデルによる能動的な異常検知を実現することに成功しました。これにより、運用管理者様の運用負荷を軽減するとともに、事態が悪化する前に対応いただくよう通知することが可能となります。

ご清聴どうもありがとうございました。

HPC4U !



コアマイクロシステムズ株式会社

Core Micro Systems, Inc.

URL: <https://www.cmsinc.co.jp/> Mail: sales@cmsinc.co.jp

TEL: 03-6279-8501 FAX 03-6279-8502

〒174-0041 東京都板橋区舟渡1-12-11 ヘリオスIIビル3F