

ABCI 2.0: AI橋渡しクラウドのアップグレード

産業技術総合研究所

デジタルアーキテクチャ研究センター

滝澤 真一郎

世界最大級・超省電力・オープンA I インフラストラクチャ



- 我が国における産学官によるA I 研究開発を加速するオープンイノベーションプラットフォーム
- 高い計算能力を活用したA I 技術の研究開発・実証、社会実装の推進、A I 分野の最重要課題への挑戦が目的



Peak Performance:

550 PFLOPS (FP16)

37 PFLOPS (FP64)

Effective Performance: (as of Nov 2020)

19.88 PFLOPS (#14 in TOP500)

14.423 GFLOPS/W (#13 in GREEN500)

508.85 TFLOPS (#11 in HPCG)

Power Usage: < 2.6 MW

Average PUE: < 1.15



2018年8月1日運用開始

<https://abci.ai/ja/>

A I インフラストラクチャ for everyone

Expert



ABCIグランドチャレンジ：
画期的な成果が見込まれる最重要課題への挑戦に
ABCIの全システムを最大24時間、無償提供

Advanced & Intermediate



最大2048GPUまで誰でも利用可能
すぐ使えるソフトウェア、データセット、
学習モデル等を提供

Beginner



初学者にも使いやすい統合開発環境を実現

データセンタ事業者等

企業がクラウドで個人情報等を
扱える水準のセキュリティ機構

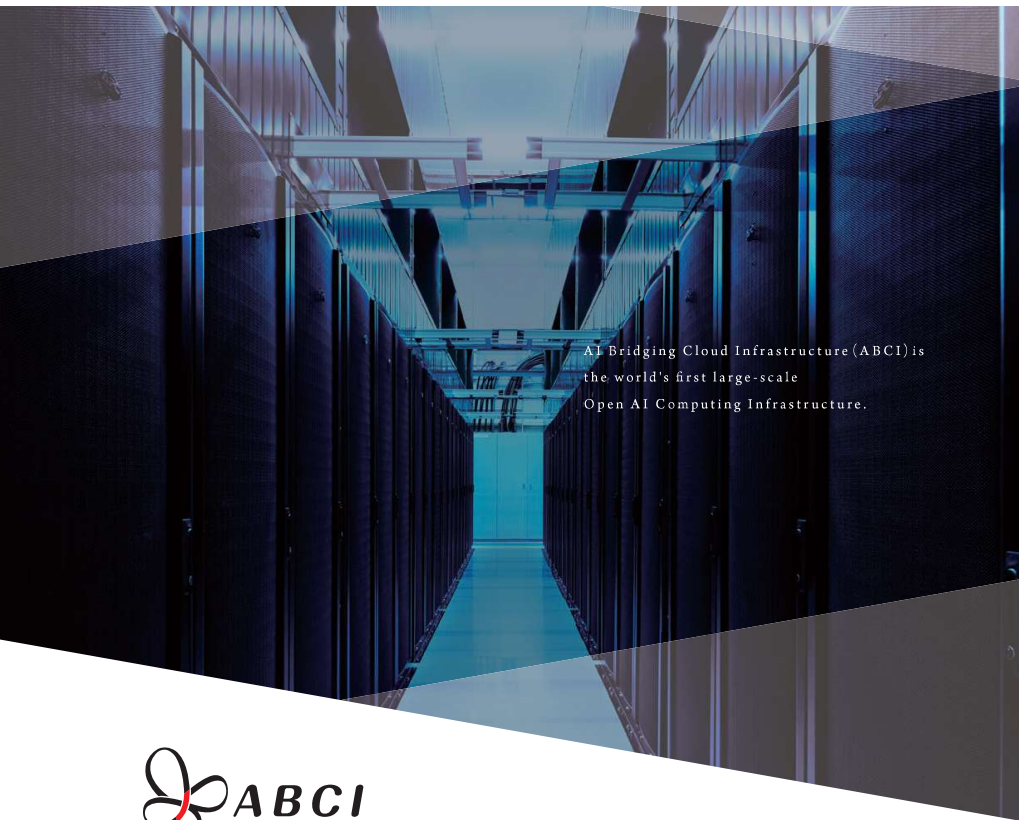
「AIを試す場」
人工知能産業のための
オーブンプラットフォーム形成
最先端のAI研究から
誰でも試して使えるAIまで



数百の研究機関・大学・企業による利用・協業、
数千の研究者・エンジニアによる利用を促進

ABCI の主なユーザ企業

https://abci.ai/ja/link/use_case.html



誰もが利用できるAIクラウド計算システム「ABCI」

利用者インタビュー

事例1 ギリア株式会社 様：長年の積み重ねが開花、強力なコンピュータで実用を迎えたAI

事例2 株式会社 富士通研究所 様：スーパーコンピューティングという「商品」の開発

事例4 株式会社 高電社 様：機械翻訳のスペシャリスト企業として30年、ABCIで実現するAI翻訳

事例5 パナソニック 株式会社 様：家電からソリューションへ、競争力ある技術をABCIで実現

事例6 アイリス 株式会社 様：インフルエンザを防ぐ、AIによる「匠の目」

事例7 Linne(リンネ) 株式会社 様：スマホからはじまる生物多様性保全の環(わ)

事例8 株式会社 トリプルアイズ様：囲碁AI研究は自動車メーカーのF1と同じ。めざせ世界の頂点!

事例9 LeapMind 株式会社様：量子化ニューラルネットワークが実現する、「どこにでもAIがある」エッジAIの世界

事例10 株式会社 リクルートテクノロジーズ様：先進技術から紡ぐ、リクルートの「未来のソリューション」

事例11 株式会社 パスコ様：光と電波で地表をキャッチ、AI解析で災害情報も迅速に

事例12 株式会社 アタリ様：人と見まがうバーチャルヒューマンが人に寄添う未来へ

事例13 オムロンサイニックスエックス 株式会社様：オープンイノベーションで挑むオムロンの非連続な技術進化

事例14 株式会社 IABC様：シミュレーションとAIが切り開く地震工学の新たな地平線

事例15 株式会社 コトバデザイン様：あらゆるモノと自然な“おしゃべり”ができる世界を目指して

事例16 株式会社 日立製作所 様：協創と研究でグローバルな社会課題の解決に挑む

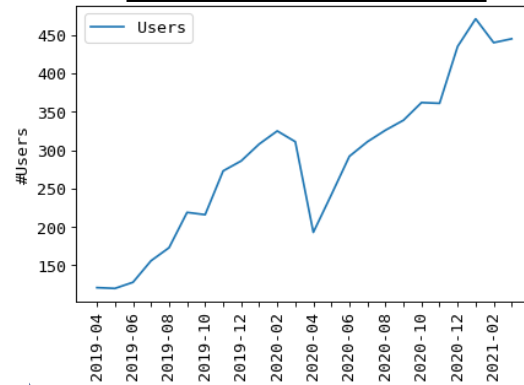
事例17株式会社 YAMATO 様：「ヒト指向型・マルチモーダル」で極める金融AI

運用状況 (1/2)

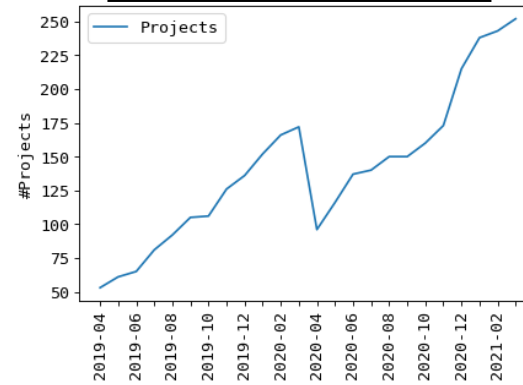
- 利用者数の増加に伴い、利用率が増加
 - 2021年10月時点の利用者
 - **2600のアカウント** (2300は産総研外の利用者)
 - **330の研究プロジェクト** (220は産総研外の利用者によるもの)
 - 2019年度～2020年度の2年間に実行されたジョブの傾向



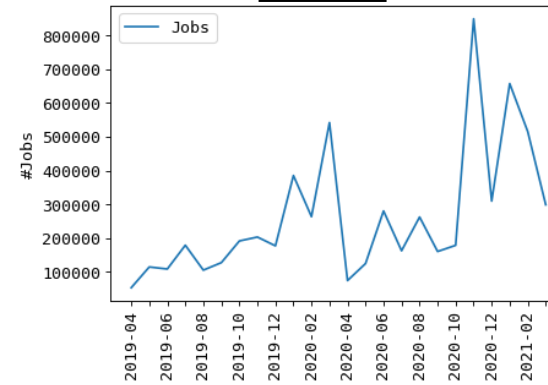
ジョブ投入アカウント数



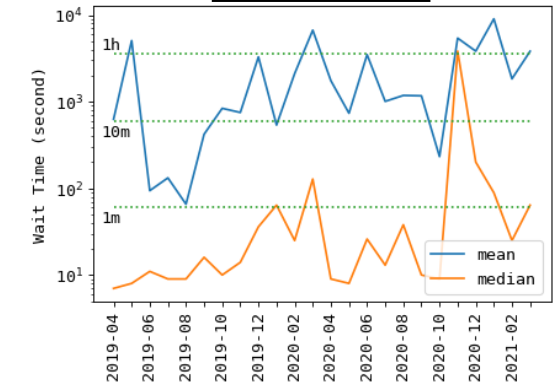
ジョブ投入プロジェクト数



ジョブ数



実行待ち時間



➡ 繁忙期を中心に利用率が高く、実行待ちが多く発生

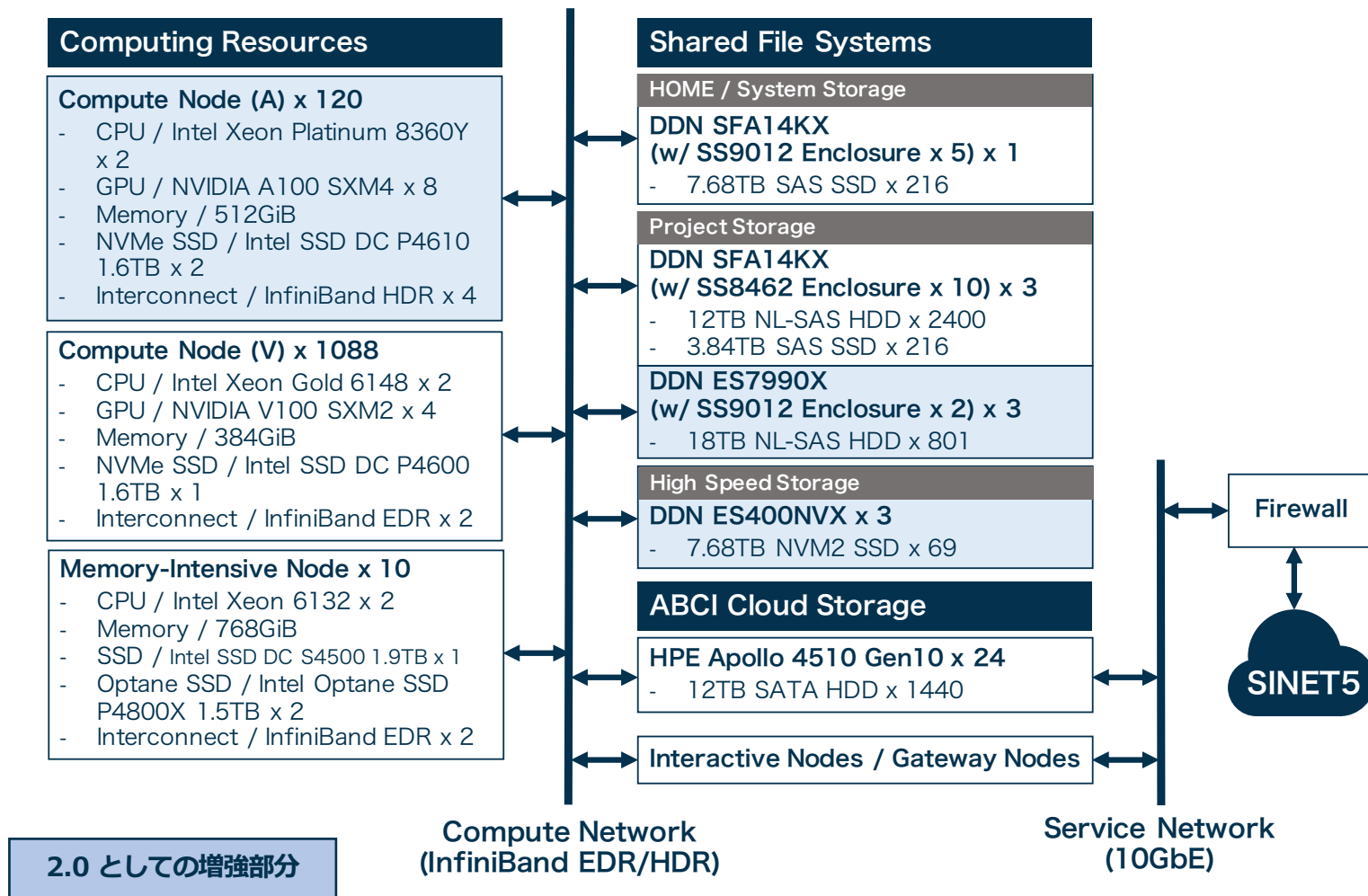
運用状況 (2/2)

- 利用ストレージ容量の増加、I/O性能の干渉
 - 2021年5月時点で共有ストレージの**60%弱を使用**
 - PB級の衛星データ等、大容量データを年度内に格納予定
 - I/O性能共有を意識しない利用者による大量のI/Oが他利用者に影響
 - 多量のデータの並列ダウンロード
 - 多量のデータを共有ストレージから読み込む学習プログラムを実行

➡ ストレージ容量枯渇の心配

➡ 利用者からのストレージ性能の苦情

ABCI 2.0としてシステム増強



- 計算ノード(A) 120台を追加
 - NVIDIA A100を8基搭載
 - 従来ノードは「計算ノード(V)」に改名
 - 資源管理システムを共通化し、容易な使い分けが可能
- 実効9.5 PBの研究プロジェクト用共有ストレージ
- 実効360TBのAll SSDの高速ストレージ

2021年5月10日より運用開始

ABCI 2.0: Computing Resource

ABCI 2.0 (May 2021)

ABCI 1.0 (Aug 2018)

550 PF (FP16), 37.2 PF (FP64)
476 TiB Memory, 1.74 PB NVMe SSD



ABCI Expansion (May 2021)

300 PF (FP16), 19.3 PF (FP64)
97.5 TiB Memory, 384 TB NVMe SSD



Compute Nodes (V) x 1088

19.88 PF

-  GPU NVIDIA V100 x 4
-  CPU Intel Xeon Gold 6148 (2.4GHz/20cores) x 2
-  Memory 384 GiB
-  Local Storage Intel SSD DC P4600 (NVMe) 1.6TB x 1
-  Interconnect InfiniBand EDR x 2 (25 GB/sec)

Compute Nodes (A) x 120

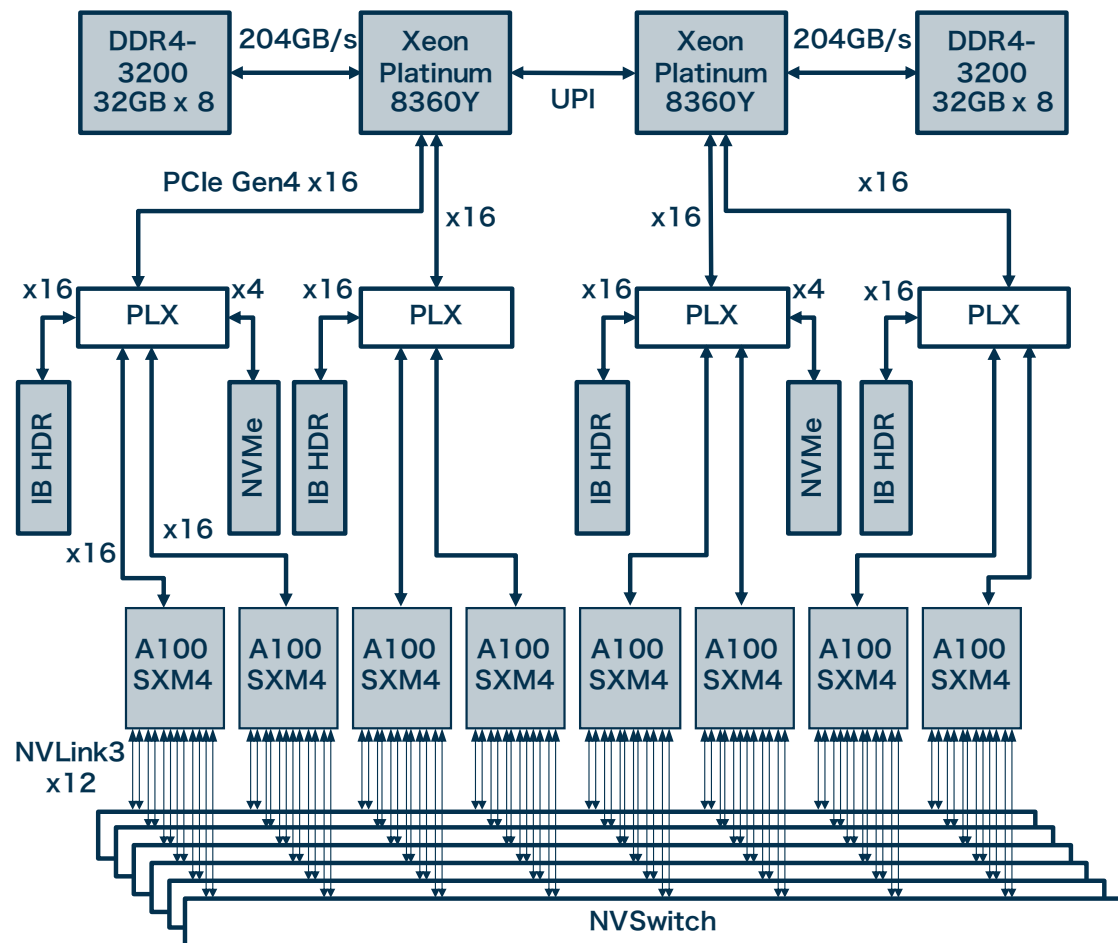
11.48 PF

-  GPU NVIDIA A100 x 8
-  CPU Intel Xeon Platinum 8360Y (2.4GHz/36cores) x 2
-  Memory 512 GiB
-  Local Storage Intel SSD DC P4610 (NVMe) 1.6TB x 2
-  Interconnect InfiniBand HDR x 4 (100 GB/sec)

x 1.11

	ABCI 1.0	ABCI 2.0	Performance Scale-up
HPC (FP64)	37.2 PF 19.88 PF	56.5 PF 22.20 PF	x 1.52
DL Training (FP32/TF32)	75 PF	225 PF	#12 in Top500 June 2021 (#16 in Nov. 2021)
DL Training (FP16/BF16)	550 PF	850 PF	
FP16 w/ Sparsity	550 PF	1.15 EF	x 2.09

ABCI 2.0 で追加した計算ノード – 計算ノード(A)



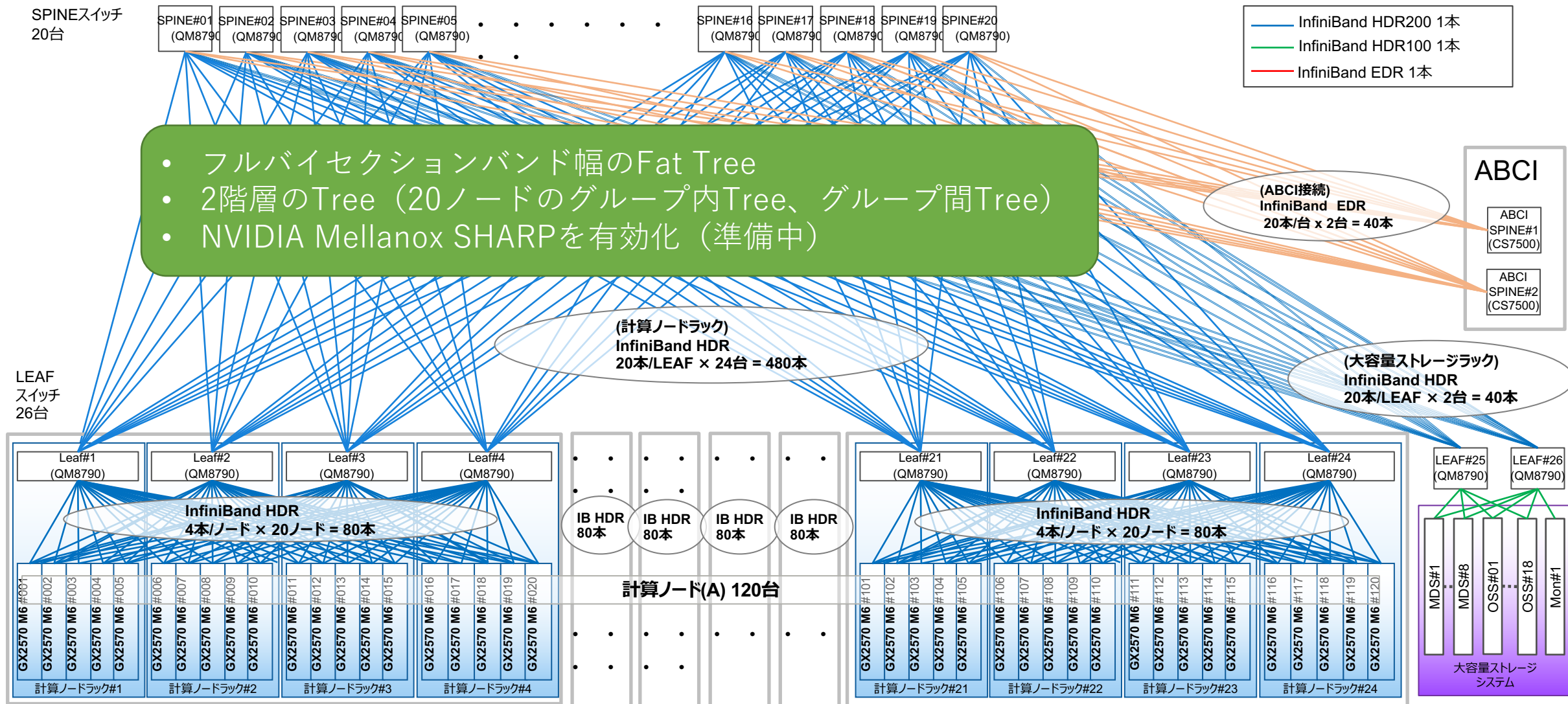
出典: 富士通社製品紹介Webページ

FUJITSU PRIMERGY GX2570 (4U) x 120 nodes

CPU	Intel Xeon Platinum 8360Y (36 cores) x2
GPU	NVIDIA A100 SXM (40GiB) x8
Memory	512GiB DDR4 3200MHz RDIMM
NVMe SSD	Intel SSD DC P4510 2.0TB x2
Interconnect	InfiniBand HDR (200Gbps) x4

- ・ 調達・導入時期で入手可能な最新のHW
- ・ ABCIの旧ノード(計算ノード(V))に比較して高密度実装
 - ・ ノード内並列できるGPU数が増加
 - ・ 1 GPUジョブ向けには、ソフトウェアレベルで資源分割

計算ノード (A) のネットワーク



ABCI 2.0が提供するストレージサービス

ノードローカル
ストレージ



Node (V): 1.6TB
Node (A): 2.0TB x2

- 計算実行中の一時作業領域
- 複数ノード利用時、複数ノードのSSDを集約した一時的な並列ファイルシステムを構築（BeeGFS）

共有ストレージ



DDN SFA 14KXなど

Home Storage

利用者のホーム領域を提供するストレージ
200GB/Account

Project Storage

研究プロジェクトごとのディレクトリを提供するストレージ
1TB～250TB

High Speed Storage (現在は用途限定公開)

一時的に研究プロジェクトに提供する高速ストレージ

オブジェクト
ストレージ



HPE Apollo 4510 Gen10 x24

出典: HPE社製品紹介Webページ

- 17PB (raw) のAmazon S3 API互換オブジェクトストレージ
- ABCI内外からのデータ転送をサポート
- 暗号化通信・サーバサイド暗号化、アクセス制御を提供

共有ストレージ – 種別と構成

ABCI 2.0で追加したストレージ

Type	System	Drive	Filesystem	Size
Project Storage	DDN ES7990X x3 DDN SS9012 Enclosure x6	18TB HDD x801	Lustre	9.5PB
High Speed Storage	DDN ES400NVX x3	7.68TB SSD x69	Lustre	360TB

従来からあるストレージ

Type	System	Drive	Filesystem	Size
HOME Storage	DDN SFA 14KX x1 DDN SS9012 Enclosure x5	7.68TB SSD x185	Lustre	1PB
Project Storage - Old #1, #2, #3	DDN SFA14KX x1 DDN SS8462 Enclosure x30	3.84TB SSD x216 12TB HDD x2400	GPFS	6.6PB for each

現在、単一Lustreストレージとすべく、データ移行実施中

共有ストレージ – Lustreの構成

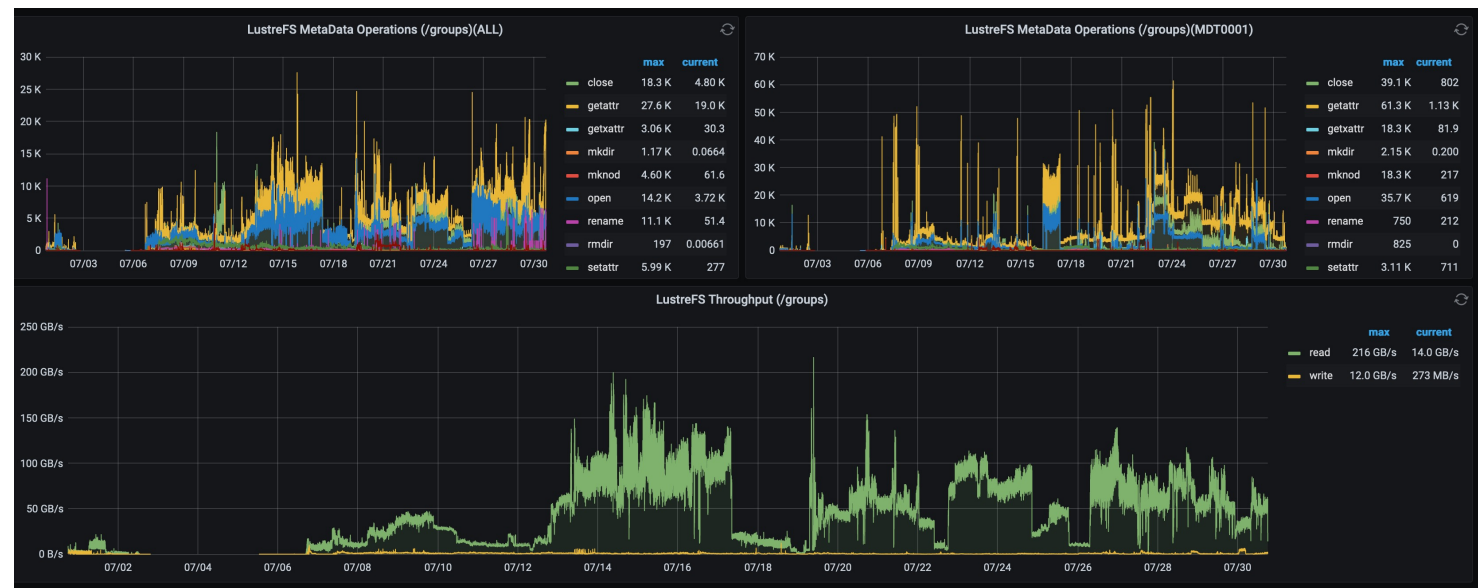
	Project Storage	High Speed Storage	Home Storage
MDS	2 servers	2 servers	2 servers
OSS	6 VMs (3 servers)	12 VMs (3 servers)	4 servers
理論性能	2.4M IOPS 150 GB/s	9M IOPS 300 GB/s	6M IOPS 100 GB/s
容量	9.5 PB	360 TB	1 PB
inode数	18.7 billion	6.2 billion	3.0 billion

- Lustre Version: 2.12.5
- High Speed Storageでのみ、Data On MDT (DOM)の利用が可能
 - inode数減少を懸念して、他では明示的にDisable

共有ストレージのモニタリング

- 2021.4 以前は Spectrum Scale GUI により、GPFS領域をモニタリング
 - Read/Write の転送 Bytes、OPS 等をクライアントやストレージサーバのノード単位で確認可能
- ABCI 2.0 へのアップグレードにおいて DDN Monitoring Tool (LustrePerfMon) を導入
 - ファイルシステム単位でユーザ毎、ノード毎、ジョブ毎の I/O 情報を1 分間隔で取得
 - I/O 情報：open, close, stat 等のメタデータ操作、read, write 操作
 - Grafana による時系列での可視化
 - 現在は運用目的での利用に限定、利用者には未公開

運用チームによる異常検知は、
他の手段と組み合わせて実施



ABCI 2.0へのアップグレードの効果

- 新計算ノードを活用した成果創出が進行中
 - グランドチャレンジ成果報告会を1月に開催予定
- 繁忙期を中心に利用率が高く、実行待ちが多く発生
 - ➡ 利用量は増えたが、キャパも増えたから緩和されていると推測（詳細は後日分析）
 - ポイント購入のペースは昨年度以上、12月末には販売終了予定。一方で、9月、10月の平均利用率は昨年度同水準
- ストレージ容量枯渇の心配
 - ➡ 容量増加により緩和
- 利用者からのストレージ性能の苦情
 - ➡ 引き続き発生しているが、強化した監視システムを活用して迅速な解決を目指す
 - 100万、1000万規模の小さい画像ファイルを用いた学習を実行する利用者が存在
 - モニタリングシステムを活用しつつ、そのような利用者・ジョブを特定し、ローカルストレージの利用や、プログラムの修正を促す

最後に

- ABCI: AI Bridging Cloud Infrastructure
 - AI開発のためのオープンなプラットフォーム
 - NVIDIA V100 GPUを4基搭載するサーバ **1088台**から構成
 - **21PB**の高性能・大容量共有ストレージ、**17PB**のオブジェクトストレージを提供
- ABCI 2.0へのアップグレード
 - NVIDIA A100 GPUを8基搭載するサーバ **120台**を追加
 - **計10PB**のLustre共有ストレージを追加
 - 2021年5月10日より運用開始予定
- Lustre共有ストレージの今後の活用
 - High Speed StorageでのDOMの活用により、多数の小さいファイルを入力とする学習プログラムの性能向上
 - GPUDirect StorageによるLustreからのI/O改善
 - モニタリングツールを活用した、ストレージ過負荷への迅速な対応