

最先端AI技術の研究開発・社会実装を
加速するABCI 3.0



2025/02/03

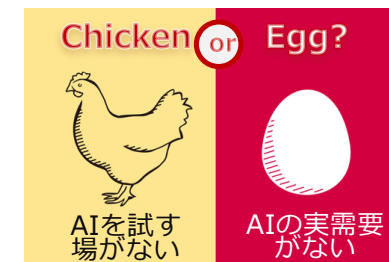
株式会社 AIST Solutions

小川宏高

1. ABCIとは

ABCI導入の背景と設計コンセプト

- 2016年当時、AIの革新的な成果が海外で先行（Googleなど）。計算資源を集約し、**AIモデルのさらなる大規模化**が予想されていた
- 一方、AIについては産業界の関心の高さの割に、国内で導入が進んでいないことが課題となっていた → **大規模AIを簡単に試せる場がない！**
- 産総研は2000年代初頭からスパコン研究開発を継続しており、**最先端技術の導入による大規模な並列分散システムの構築・運用に強み**



ABCI (AI Bridging Cloud Infrastructure) の設計コンセプト

- 産総研が培った**スパコン構築・運用技術**を大規模AIスパコンに導入
- コモディティ(汎用製品)を活用し、**民間クラウド事業者が模倣しやすく費用対効果に優れた設計、運用コストの圧縮、省電力性**を実現
- プロトタイプシステムでの検証を通して、**産総研が先駆的に実現**

2018.8.1 1.0運用開始
2021.5.10 2.0運用開始



Expert



LLM構築支援プログラム・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の計算資源を有償・無償で大規模に提供

Advanced & Intermediate



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
大規模言語モデルハッカソン・ハンズオン：
LLM学習の実践的ノウハウを参加者と共有

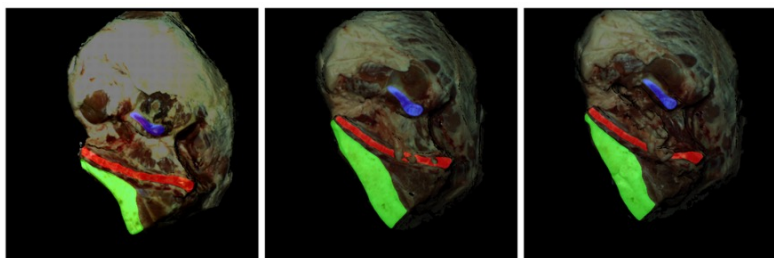
「AIを試す場」
人工知能産業のための
オープンプラットフォーム形成
最先端のAI研究から
誰でも試して使えるAIまで



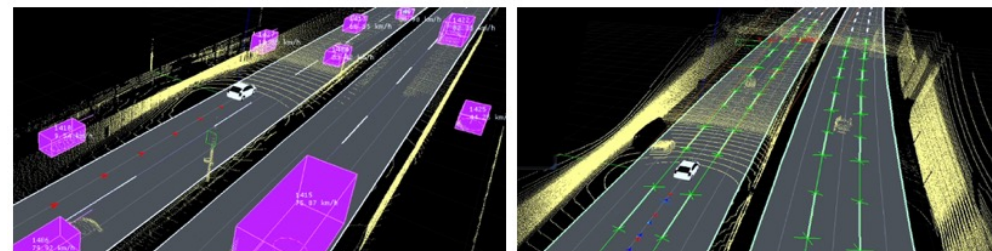
数百の研究機関・大学・企業による利用・協業、
数千の研究者・エンジニアによる利用を促進

ABCIの利用者数・主な利用分野

- 2018年8月の運用開始以来、利用者数は右肩上がり増加
- 2024年3月現在の利用者数は約2700人(うち外部利用が約87%)
- AIスタートアップから総合電機メーカーまで利用が拡大し、世界に伍する研究が可能になるとともに、我が国のAI研究全体を支える重要基盤に成長
- 利用事例は以下でご紹介
 - https://abci.ai/ja/use_case/



画像認識で食肉から骨を見つけ出す、食肉加工機械の進化（株式会社前川製作所）



AI・自動運転技術で新しい物流インフラを構築する（株式会社T2）

本ページの写真を利用する場合は、出典を記載した上でご活用ください。

生成AI時代のABCI 2.0

大規模言語モデル構築支援プログラム(2023.8-)

- 生成AI・LLM開発の提案を公募し、有識者による審査の結果採択された提案に、計算資源を優先的に割り当て
- Preferred NetworksのPLaMoやELYZAの日本語LLM、東工大-産総研のSwallowなどの成果を創出
- 今年度は産総研、NICT、理研、ELYZAの計6提案を採択



LLM構築支援プログラム採択者 第1回



第2回



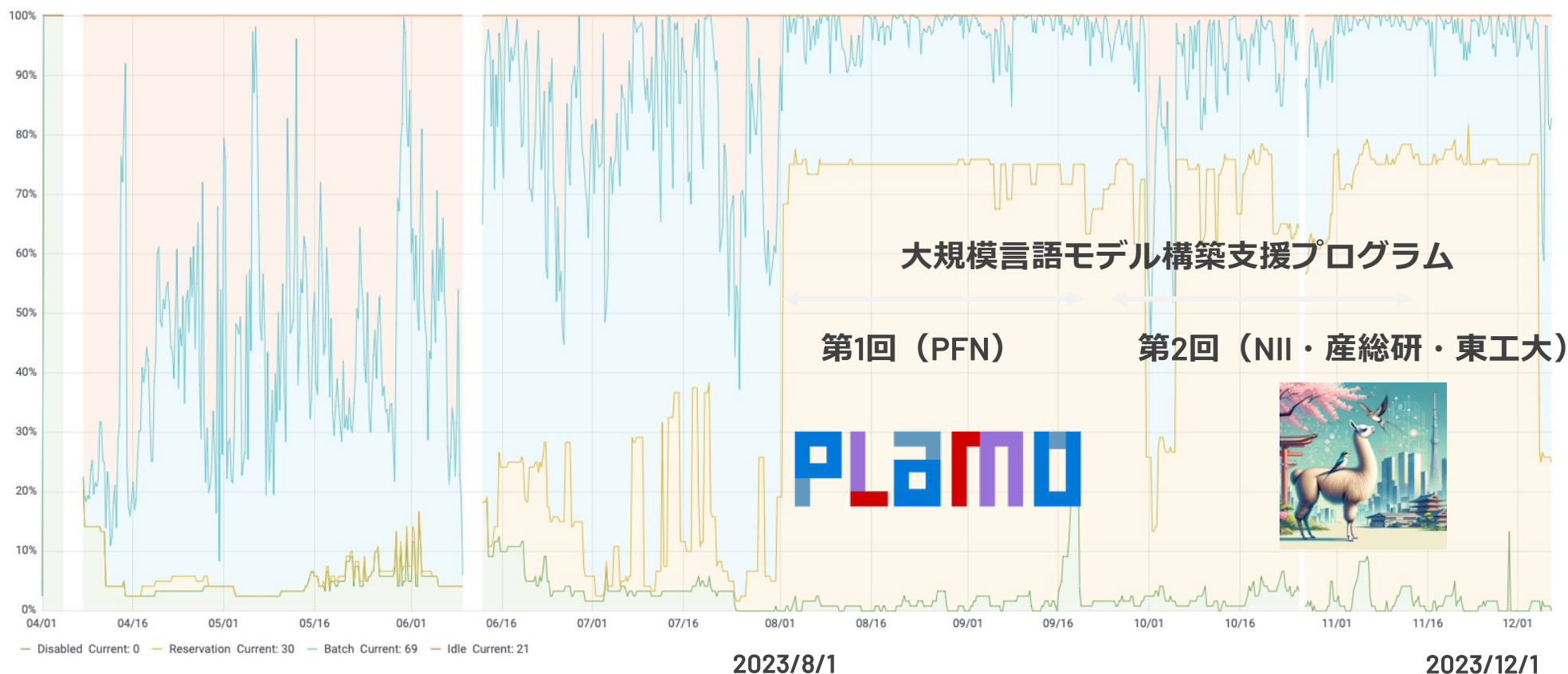
LLMハッカソン: 国内生成AI人材交流・育成の取り組み (2023.7)

- 産総研研究者やNVIDIA技術者等がメンター・チューターとなり、LLM学習やファインチューニングに関する実践的ノウハウを共有
- 20チーム（民間16、大学3、国研1）が参加



生成AI時代のABCI 2.0

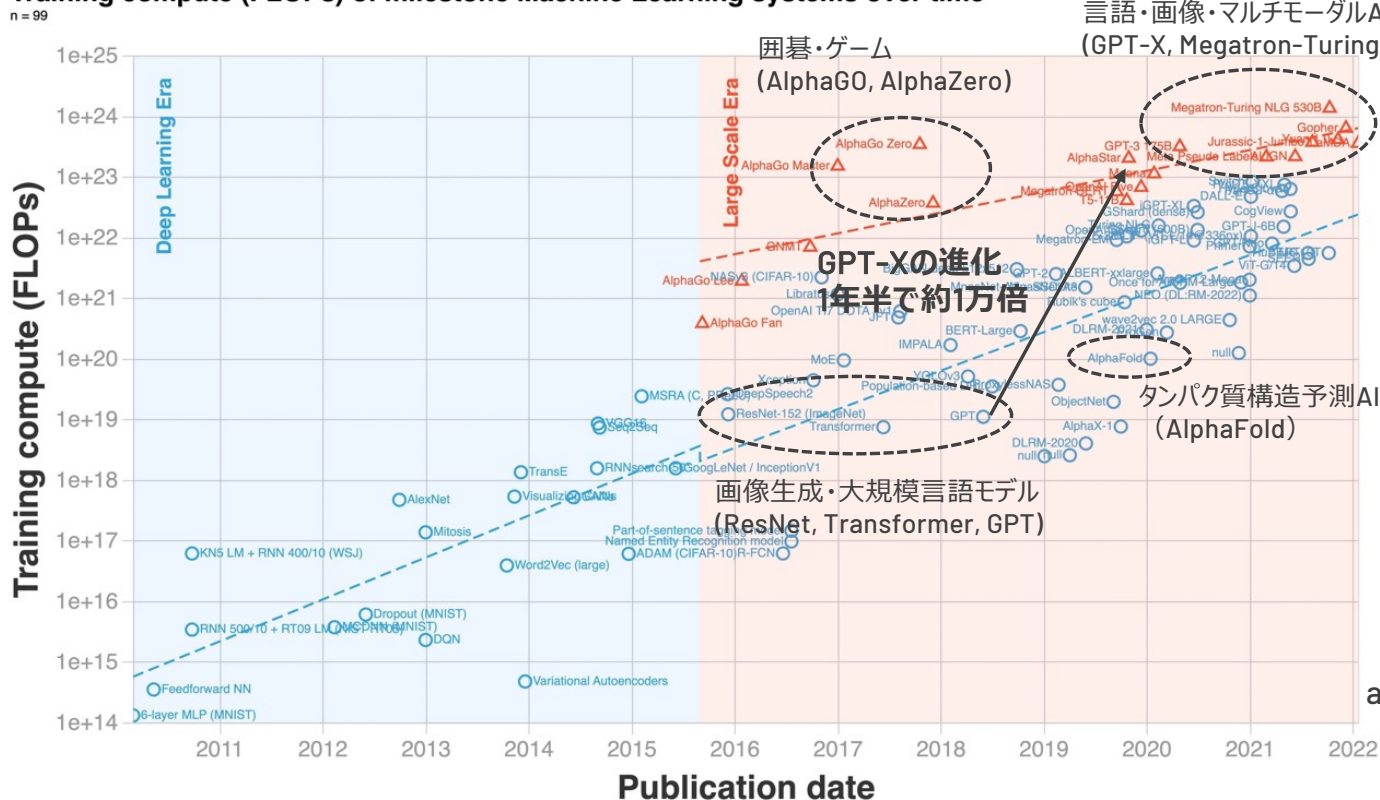
- 生成AI開発への機運の高まりを踏まえ、2023年8月より大規模言語モデル構築支援プログラムを開始
- その結果利用率100%が恒常化し、当該プログラム以外の開発利用の妨げに



生成AI開発に必要な計算量とABCIの性能のギャップの拡大

- 世界的に競争力のある生成AIの開発・活用には計算資源の抜本的な拡充が必要
- 世界最先端の基盤モデル開発には、ABCI 2.0の10倍以上計算能力が利用されたとされる

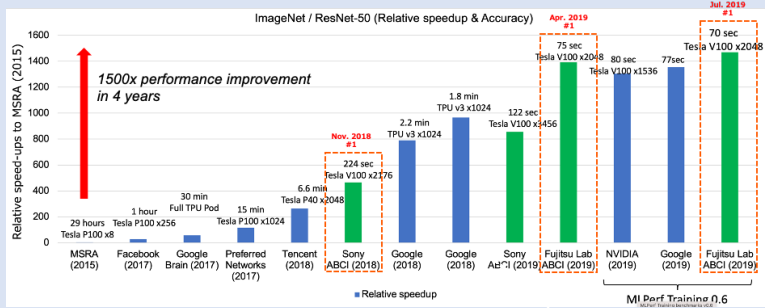
Training compute (FLOPs) of milestone Machine Learning systems over time



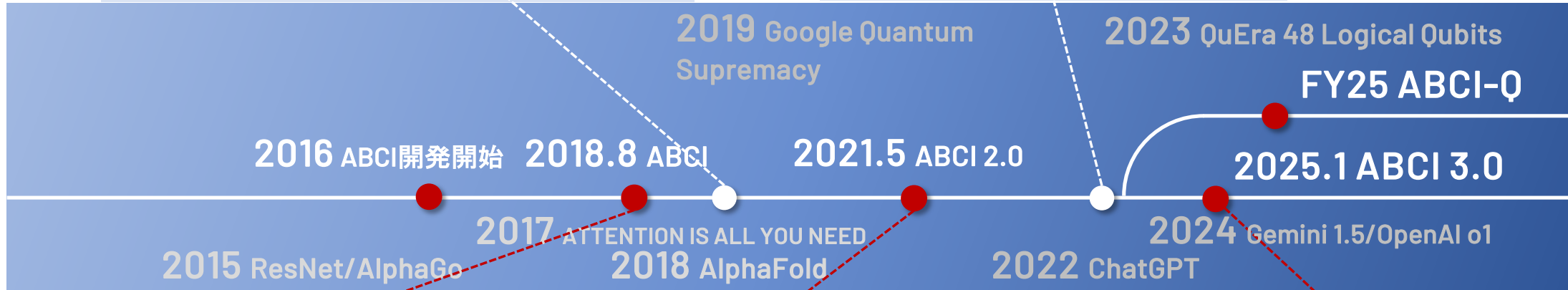
(出典) Sevilla et al., Compute trends across three eras of machine learning, 2022

これまでのABCIと今後

“MLPerf training” benchmarkの世界記録達成



大規模言語モデル構築支援プログラム (FY23-)



ABCI 1.0: 0.55 Exa FLOPS
NVIDIA V100 x4352



ABCI 2.0: 0.85 Exa FLOPS
+NVIDIA A100 (40GB) x960



ABCI 3.0: 6.22 Exa FLOPS



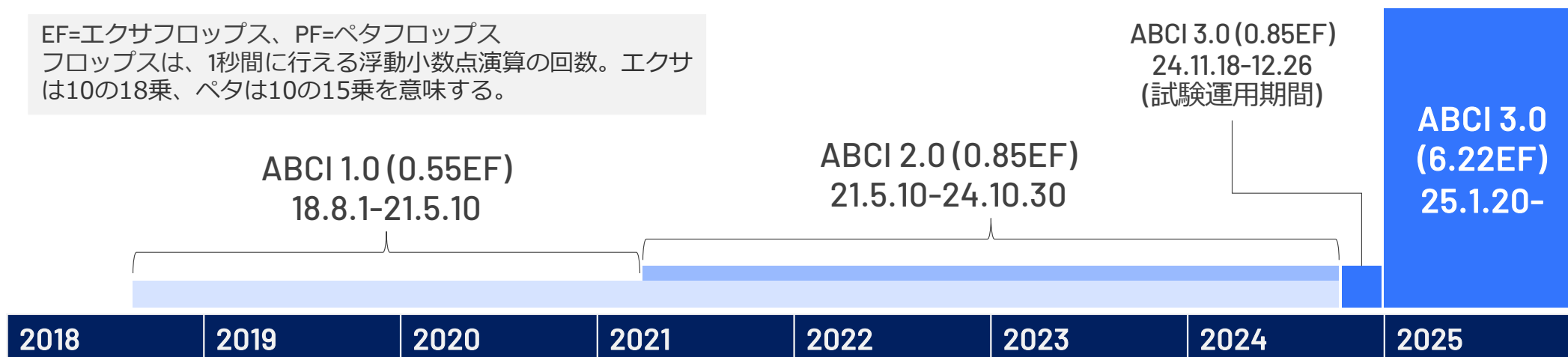
2. ABCI 3.0



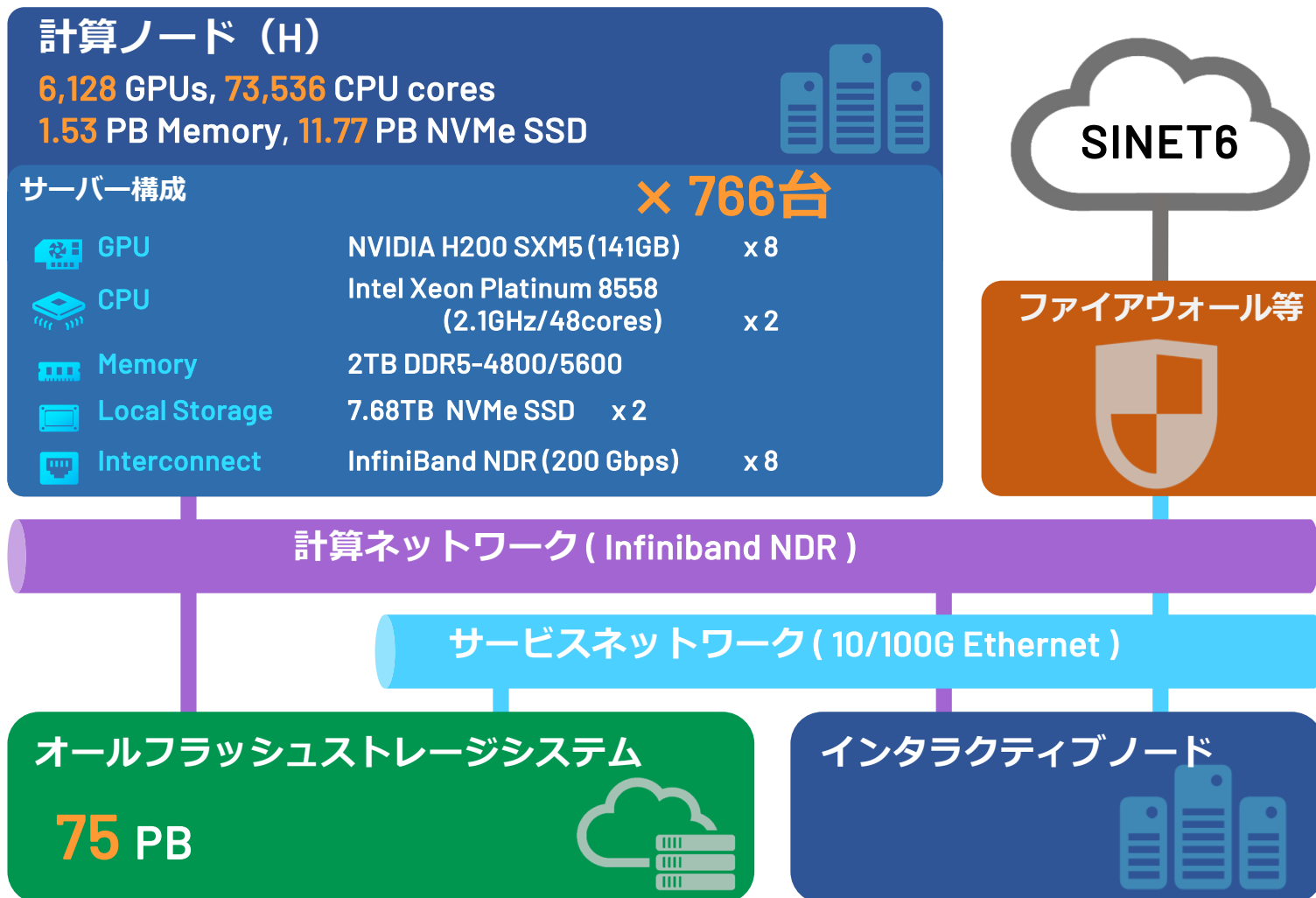
ABCI 3.0 -最先端AI技術の研究開発・社会実装を加速-

- **生成AIモデルをはじめとした最先端AI技術の研究開発能力の強化**を目的として、公的計算基盤として国内最大規模の計算能力（半精度6.22エクサフロップス、従来比約7倍）を整備、**国内の産学官、特に国研・大学、スタートアップ等による公的な研究開発利用を中心に提供**
 - 経済産業省「生成AIの基盤的な開発力強化に資する計算資源の整備」（令和5年度補正）の一環
- 従来不可能だった規模の大規模言語モデル開発、実世界データを用いた基盤モデル開発、それらを組み合わせたマルチモーダルAI開発等、**世界に伍する最先端AI技術開発が可能に**
- **既設のAIデータセンター棟設備を活用（一部拡張）することで、ごく短期間で一般提供開始**
 - 24.10末に2.0運用終了、24.11より一部システムの試験運用を開始、**25.1.20正午より一般提供を開始**

EF=エクサフロップス、PF=ペタフロップス
フロップスは、1秒間に行える浮動小数点演算の回数。エクサは10の18乗、ペタは10の15乗を意味する。



ABCI 3.0システム概要



技術的な特長（1）

- **最新の高性能・省電力GPUサーバー 766台**
 - 高性能で省電力の最新GPU「NVIDIA H200 SXM5」をサーバー 1台あたり8基、766台で計6128基搭載。
- **ABCI 3.0のピーク性能は7~13倍に拡大**
 - 半精度: 6.2エクサフロップス（従来比約7倍）
 - 単精度: 3.0エクサフロップス（従来比約13倍）
 - 倍精度: 415ペタフロップス（従来比約7倍）

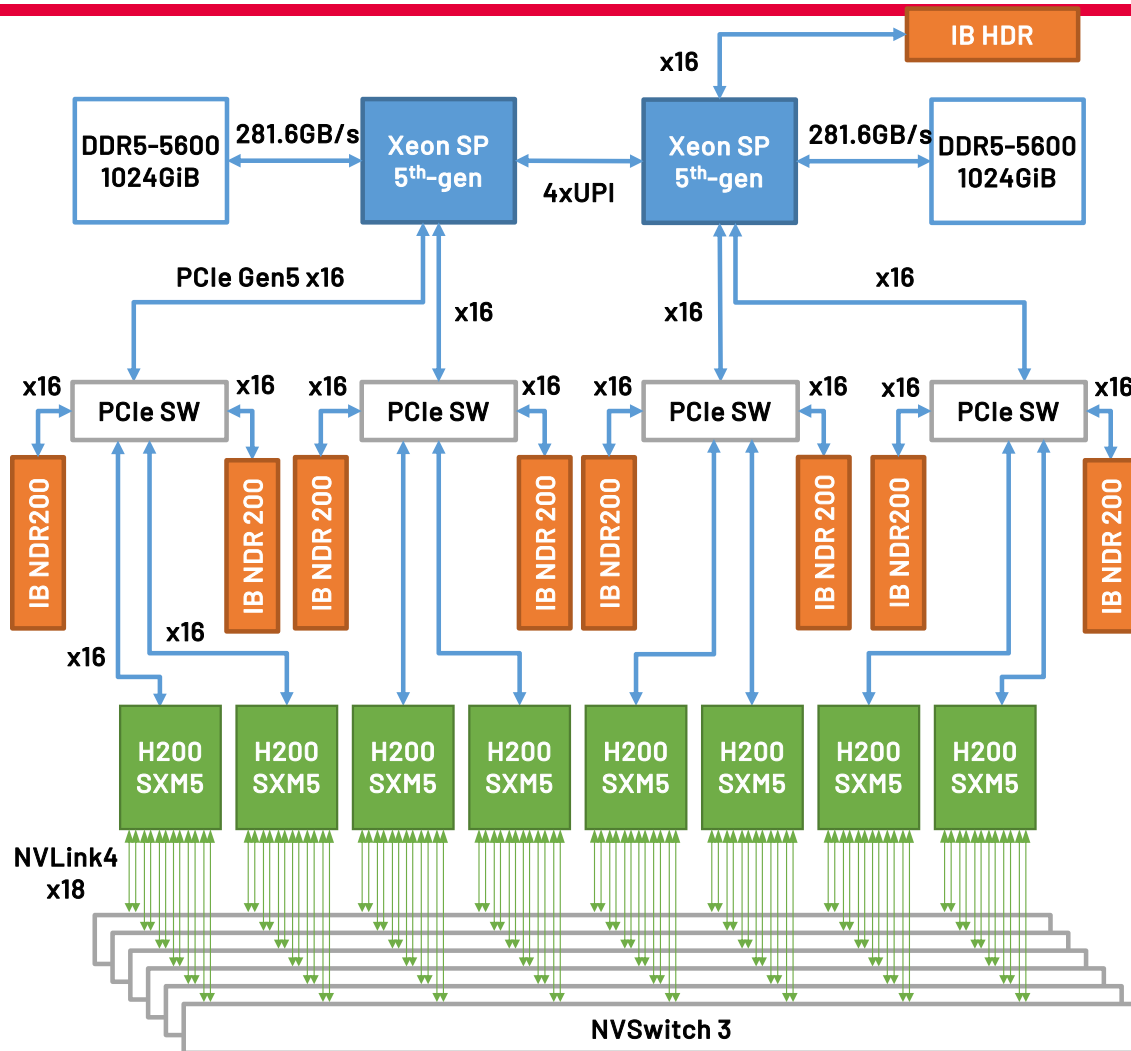
	ABCI 1.0	ABCI 2.0	ABCI 3.0
半精度 (FP16/BF16)	0.55 EF	0.85 EF	6.2 EF
単精度 (FP32/TF32)	75 PF	225 PF	3.0 EF
倍精度 (FP64)	37.2 PF	56.5 PF	415 PF

EF=エクサフロップス、PF=ペタフロップス
フロップスは、1秒間に行える浮動小数点演算の回数。エクサは10の18乗、ペタは10の15乗を意味する。

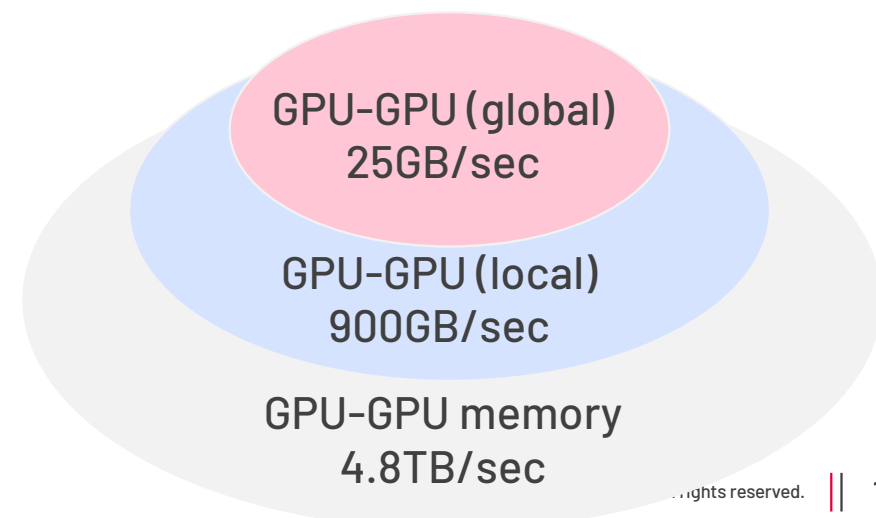


HPE Cray XD670

技術的な特長 (1)



HPE Cray XD670 Server (5U) x 766 nodes	
CPU	5th-gen Intel Xeon-SP (Emerald Rapids, 48core) x2
GPU	NVIDIA H200 SXM5 141GB, 4.8TB/s x8
Memory	2048GiB DDR5-4800/5600 DIMM
NVMe SSD	U.3 NVMe SSD 7.68TB x2
Interconnect	InfiniBand NDR200 (200Gbps) x8
Storage Interconnect	InfiniBand HDR (200Gbps) x1



技術的な特長（2）

- **大容量・高速オールフラッシュストレージシステム**

- QLC（クアドレベルセル）の高密度フラッシュストレージを採用、物理容量75PBの大容量・高速ストレージシステムを実現。
- 従来システムに比べてストレージ容量、理論読み書き性能ともに2倍以上、設置スペースはEIA標準ラックのユニット数換算で70%に。
- 従来読み書き性能がボトルネックとなっていた大規模なデータ処理の性能向上と、限られたデータセンター面積の有効活用が期待できる。

DDN ES400NVX2 ×17台で構成
うち16台を共有ファイルシステム、
1台をオブジェクトストレージに利用



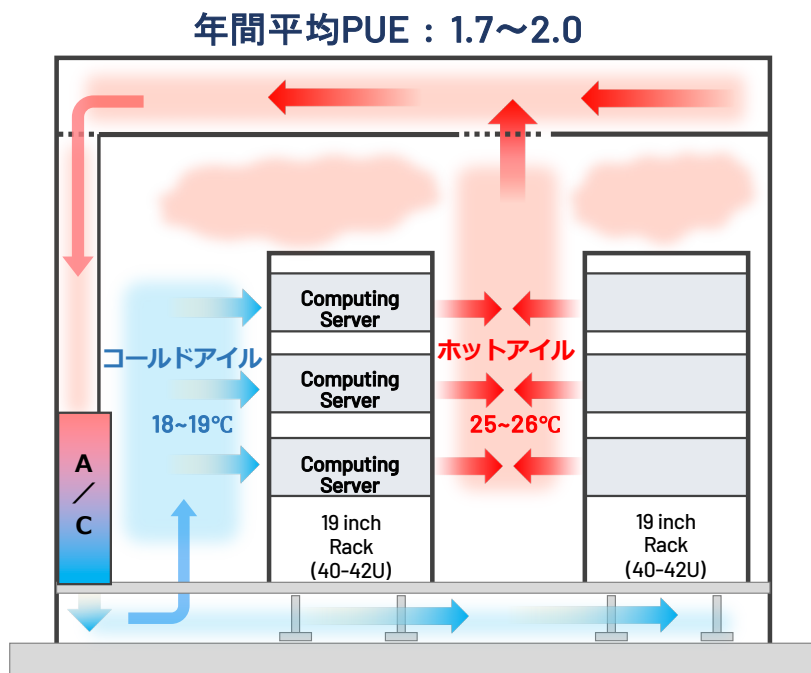
技術的な特長（3）産総研が独自設計した超省電力・超高密度コンテナの活用

産総研グループ

多湿な日本で温水冷却・高温排気の有効性をABCI 1.0で初めて実証（一部拡張により3.0に対応）

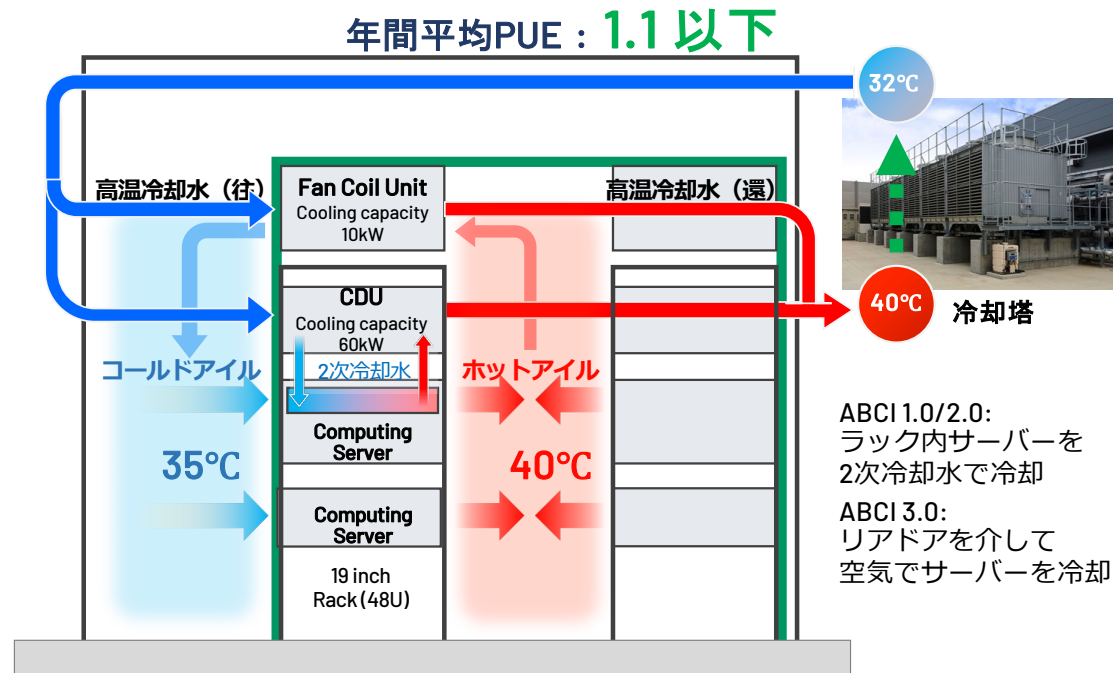
従来型データセンターの冷却システム

「床吹き出し方式」による全体空調のモデル



ABCIの冷却システム

「水冷+空冷のハイブリッド冷却方式」
 冷却塔において気化熱を利用して「32°C以下の冷却水」を作り、
 その水でラック内サーバーと、ホットアイル内の空気を冷却



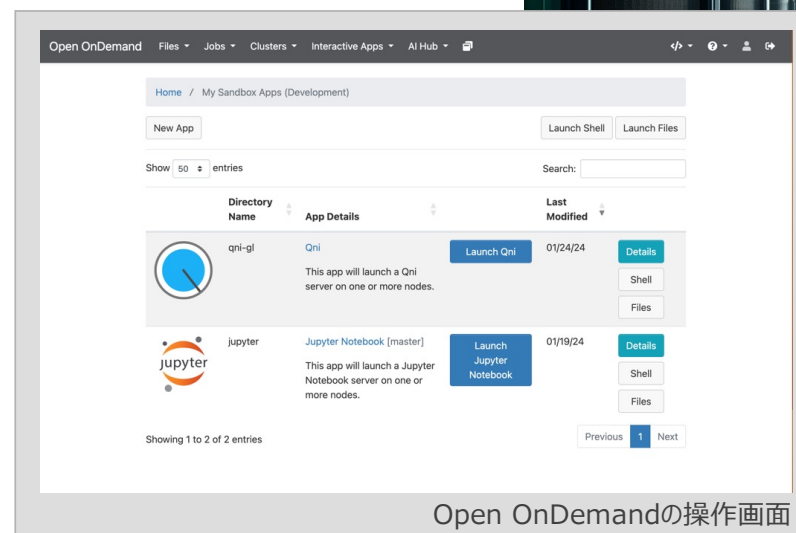
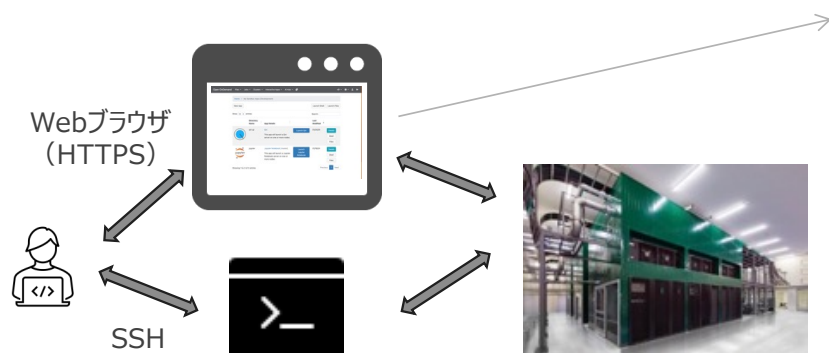
$$PUE = \frac{\text{データセンター全体の消費電力}}{\text{IT機器による消費電力}}$$

※PUE (Power Usage Effectiveness)
 データセンターにおける電力効率を示す指標で、1.0に近いほど高効率。

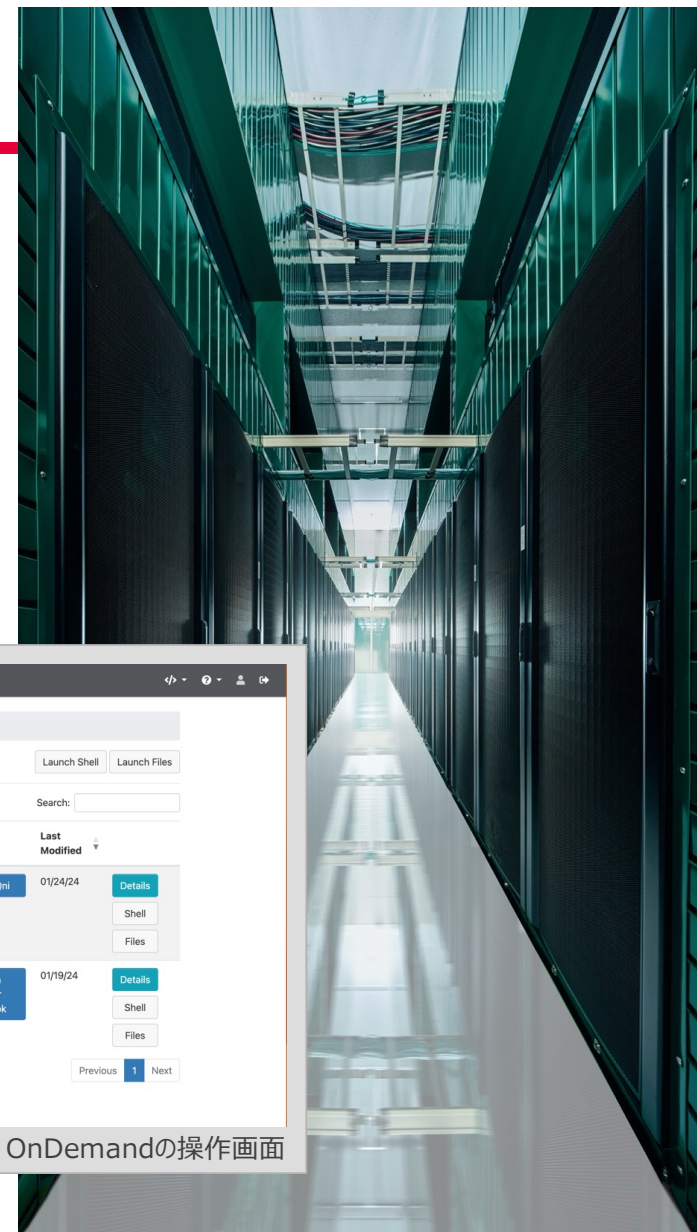
技術的な特長（４）

• 従来ABCI同様の使いやすい利用サービス

- これまで産総研グループが培ってきたABCIの技術資産を生かして、利用者にとって使いやすい利用サービスを提供。
- AIの開発がすぐに始められるソフトウェアスタックを備えるほか、学習済みモデルの再利用などAI開発を容易にする「AI Hub」サービスを提供。ウェブポータル「Open OnDemand」の導入により、ウェブブラウザからGUIベースで簡単に操作できる環境も提供。



Open OnDemandの操作画面



ABCI 3.0の運用上の特長

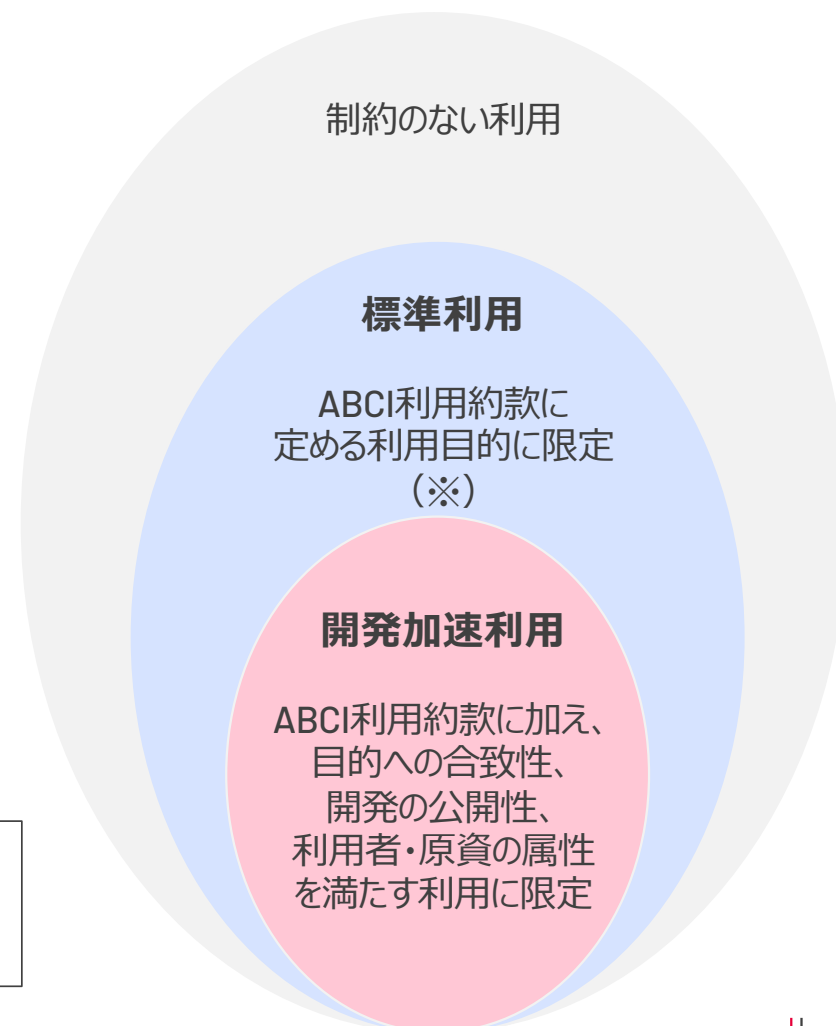
- **最先端AI技術の研究開発・評価・人材育成（開発加速利用）に優先的に提供**
- NEDOポスト5G事業の一環として実施している**生成AI開発加速プロジェクトGENIACと連携**し、開発成果の迅速な社会実装に向けた取り組みを推進
- **非東京圏でのAI開発を支援するため、地方優先枠を導入**
 - 東京圏（東京都、神奈川県、千葉県、埼玉県）外に所在する利用法人による利用を一部優遇
- **生成AIハッカソンを通じた国内生成AI人材交流・育成の取り組み**
- など



ABCI 3.0の利用制度

- 基盤モデル、生成AI、マルチモーダルAI等の最先端AI技術の研究開発・評価・人材育成を目的とした公的な利用に対して重点的に提供
- 「標準利用」と「開発加速利用」の2クラス料金制を導入
 - 標準利用における単価は市場価格を参考に定め、開発加速利用は標準利用からディスカウント
 - 開発加速利用は、申請に基づき、目的への合致性、開発の公開性、利用者・利用原資の属性等を審査の上、要件を満たす利用のみを認定
 - その上で標準利用の割合を一定以下に抑制
- ただし2025年3月までは経過措置として一律「開発加速利用」の料金を適用し、FY25から2クラス料金制を適用

(※) ABCI利用約款第6条3の二
利用目的が**研究、開発及び評価、並びにそれらに通じた人材の育成**を目的としたものであること。



GENIAC連携：共創型データ実証支援プログラム

- GENIAC (Generative AI Accelerator Challenge) https://www.meti.go.jp/policy/mono_info_service/geniac/
 - 経産省・NEDOが推進している生成AI開発支援プログラム（2023年度～）
 - 生成AI開発企業への計算資源の支援や、データ・生成AIの利活用に係る先進事例の調査、生成AI開発加速に向けた新たなデータセットの構築に関する調査を実施
- 共創型データ実証支援プログラム
 - GENIACと連携し、ABCI 3.0へのデータ持ち込みとそれを用いた利活用実証を行う産学官を支援するプログラムを実施
 - データ作成企業、AI開発企業等の共同（複数社による体制を必須とする）によるデータ利活用実証課題を公募



ABCI生成AIハッカソン

- 対象：ABCI利用者及びABCI利用を検討中の方またはチーム（1チーム3人程度）
- 内容：ABCIを活用した生成AIモデルの構築（ファインチューニング）や推論の最適化
- 参加者のタスク：
 - データセットを持ち込む
 - 既存の言語モデルをマルチノードで訓練（ファインチューニング）
 - 単一GPU、複数GPUを用いた推論の最適化
- 開催日：2/4～2/13
 - 最終日には実施内容やハッカソンで得た学びを共有する発表会を開催
- メンターによるサポート：
 - ABCIスタッフ（産総研職員）
 - 深層学習や自然言語処理のスペシャリスト
 - GPUエンジニア



画像は前回2023.7のものです

https://abci.ai/event/2024/12/23/ja_abci_3.0_genai_hackathon.html

- 生成AIモデルをはじめとした最先端AI技術の研究開発能力の強化を目的として整備された、ABC1 3.0が2025年1月20日正午に一般提供を開始しました。
- ABC1 3.0は、AI開発用の公的計算基盤として国内最大規模の計算能力を有し、国内の産学官、特に国研・大学、スタートアップ等による公的なAI研究開発・評価・人材育成に優先的に提供されます。
- 計算能力のみならず、高度でより使いやすいAI開発環境の提供、生成AI開発に資するデータ整備・実証の支援等もあわせて推進することで、日本国内の生成AI開発能力の強化に寄与します。
- 今後は、産総研における実世界の画像・音響・3次元点群等のデータによるフィジカル領域の基盤モデル開発、それらを組み合わせたマルチモーダル生成AIの構築と応用といった先進的なAI研究開発、国研等によるさまざまな生成AI技術の研究開発に活用される計画です。
- この他、AI開発環境および計算インフラ構築ノウハウを国内クラウド事業者とも共有することで、計算インフラから生成AIのサービス提供に至る、幅広い産業の競争力強化に貢献してまいります。

