

AIST replaces its ABCI AI cloud with ABCI 3.0, built on 6,128 NVIDIA H200 GPUs

Source copy kept by wheresthe.ai for the story "AIST replaces its ABCI AI cloud with ABCI 3.0, built on 6,128 NVIDIA H200 GPUs" (<https://wheresthe.ai/d/jp-aist-abci-3-supercomputer-2024-10/>).

Source: <https://kyodonewsprwire.jp/release/202410097802>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: a6304991da6476073d069b9103019ba303d1d9b9f44d0e6309a131ffc83ba0a

TOP 政府・官公庁 産総研 大規模AIクラウド計算シ...

産総研

2024/10/10 14:00

ポイント

- ・ 2025年1月中旬までに「ABCI 3.0」の一般提供をスタート
- ・ ピーク性能は半精度で6.2エクサフロップス、単精度で3.0エクサフロップスとなり、従来システムの7〜13倍に
- ・ 生成AIをはじめとした最先端AI技術の研究開発・評価・人材育成に優先提供

概要

産総研グループ（国立研究開発法人 産業技術総合研究所（以下「産総研」という）、および株式会社AIST Solutions）は、2025年1月中旬までに大規模AIクラウド計算システム「ABCI 3.0」の一般提供を開始します。

ABCI 3.0は、従来システムの GPU より高性能で省電力の最新GPUが6128基搭載された計算サーバー、およびオールフラッシュストレージシステムへリプレイスされます。このリプレイスにより、ピーク性能は、半精度で6.2エクサフロップス、単精度で3.0エクサフロップスとなり、従来システムに比べ、7〜13倍に増強されます。またストレージ容量、理論読み書き性能ともに2倍以上に拡大します。

ABCI 3.0の一般提供により、生成AIをはじめとした最先端AI技術の研究開発・評価・人材育成の加速が期待されます。

下線部は【用語解説】参照

開発の経緯

従来システムである「AI橋渡しクラウド2.0（AI Bridging Cloud Infrastructure 2.0、以下「ABCI 2.0」という）」は、わが国の人工知能技術開発の加速を目的として、産総研が設計・開発を行った計算システムで、産総研 柏センターのAIデータセンター棟に導入し、2021年5月に運用を開始しました。産総研の成果活用等支援法人であるAISolの設立に伴い、2024年4月からはABCI 2.0の運用業務をAISolに移管し、両機関の共同体制で運営しています。これまでに、ABCI 2.0を活用した国内機関等が、世界トップレベルの生成AIの基盤となる大規模言語モデル構築に成功したことを始め、多くの機関の利用により、顕著な成果を達成しています。

その一方、国内の産学官による生成AIの開発需要の急拡大に伴い、利用を開始できるまでの待ち時間が非常に長い、開発に必要なリソースが確保できないなどといった課題がありました。また、生成AI開発はいまだ黎明期にあり、現在は主に自然言語を対象とした段階で、製造業等あらゆる産業で活用するには、実世界から取得される大量の画像、音響、センシングデータなどを用いて構築する実世界基盤モデルの開発が重要になります。こうした世界に伍する最先端AI技術の研究開発・応用実証には、さらに高い計算能力の確保が急務となっています。

産総研は、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）の一環としてABCI 2.0の後継システムであるABCI 3.0（調達件名：「実世界基盤モデル開発向け大規模クラウド基盤」）を整備中です。ABCI 3.0は、ABCIの従来の技術資産を生かしてABCI 2.0をリプレイスするシステムとして、日本ヒューレット・パッカート合同会社の技術を採用し、産総研グループが設計・開発しました。

ABCI 3.0 の主な特長

- 1) 最新の高性能・省電力GPUサーバー 766台

GPUサーバー「計算ノード(H)」は、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台計6128基搭載します。これにより、ABCI 3.0のピーク性能は、半精度で6.2エクサフロップス（従来比約7倍）、単精度で3.0エクサフロップス（従来比約13倍）、倍精度で415ペタフロップス（従来比約7倍）となります。

2) 大容量・高速オールフラッシュストレージシステム

QLC（クアドレレベルセル）の高密度フラッシュストレージを採用し、物理容量75PBの大容量・高速ストレージシステムを搭載します。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%になります。これにより、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ます。

3) 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します。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」サービスを提供します。また、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します。

提供する計算資源や利用サービスの詳細は、ABCIウェブサイトにて順次公表する予定です。

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

ABCI 3.0 の利用について

ABCI 3.0は市場に準じた提供価格としつつ、「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」を目的とした利用に対して重点的に提供するため、料金をディスカウントします。具体的には、「標準利用」と「開発加速利用」の2クラス料金制を導入し、前者は市場価格を参考に定め、後者はその1/2の価格とします。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するものです。

計算ノード(H) 1台あたりの2024年度利用料金は表1のとおりです。ただし2025年3月までは経過措置として、すべての利用に対して「開発加速利用」クラスの料金を適用し、2025年4月から新しい2クラス料金制を適用する予定です。

料金クラス

利用形態

時間単価

標準利用

バッチ

3300円

予約

4950円

開発加速利用

バッチ

1650円

予約

2475円

表 1. ABCI 3.0の2024年度利用料金（抜粋）

※「バッチ」はベストエフォート型の利用形態で、サービスの混雑状況により利用開始まで待ちが発生する場合があります。継続利用時間にも制限があります。「予約」は利用者が計算資源を占有するため、バッチのような制限がありません。

※GPU 1基のみ・CPUのみの利用、共有ストレージの利用も可能であり、これらの料金は別途定めます。

今後の予定

ABCI 3.0は、2024年10月から段階的に導入され、11月から一部システムの試験運用を開始し、12月までに従来システムからのリプレースを完了する予定です。その後、各種調整を経て2025年1月中旬までに「ABCI 3.0」として一般提供を開始する予定です。

ABCI 3.0は、高度でより使いやすいAI開発環境を構築・提供するとともに、基盤モデル構築時に利用可能なデータ群の整備などを進めて、日本国内の生成AI開発能力の向上に寄与します。加えて、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ます。

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や、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研究開発でもABCI 3.0は活用されるとともに、国内の産学官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。

展示会等出展の予定

ABCI 3.0は、2024年11月から一部システムによる試験運用、2025年1月中旬までに一般提供を開始する予定です。一般提供に先立ち、産総研グループでは複数の展示会・講演でABCI 3.0の紹介を行う予定です。

◆CEATEC 2024（2024年10月15日～18日）

公式サイト：<https://www.ceatec.com/ja/>

●カンファレンス&イベント「企業のトップは成AIをどう見ているのか？」

2024年10月18日 13:30～14:30

幕張メッセ コンベンションホールA

<https://www.ceatec.com/ja/conference/detail.html?id=2563>

●産業技術総合研究所ブース内セミナー（計2回）

2024年10月15日 12:20～12:50

2024年10月18日 14:45～15:15

幕張メッセ ホール7 ブース番号7H410

◆NVIDIA AI Summit（2024年11月12日～13日）

公式サイト：<https://www.nvidia.com/ja-jp/events/ai-summit/>

●[SJP1063] ソブリンAIの日本戦略とABCI3.0

2024年11月13日（水）14:00～14:40

ザ・プリンス パークタワー東京 B2F

◆SC24（2024年11月17日～22日）

公式サイト：<https://sc24.supercomputing.org/>

Georgia World Congress Center（米国ジョージア州アトランタ市）

Hall B, Booth 3701 (AIST Booth)

用語解説

GPU（Graphics Processing Unit）

同時に多くの計算を並列で実行できる、コンピューターグラフィックス専用のプロセッサのこと。グラフィックス処理が複雑化するにつれて性能や汎用（はんよう）性が増し、現在では高性能計算向けの汎用ベクトル・行列演算プロセッサに進化している。深層学習（ディープラーニング）の高速化にも広く用いられている。

精度

数値（実数）のコンピューター内での表現方法。倍精度は8バイト（有効桁数約16桁）、単精度は4バイト（有効

桁数約7桁)、半精度は2バイト(有効桁数約3.3桁)で表現する。最新のGPUなどを用いると、半精度・単精度では倍精度よりも大幅に高速な演算処理が可能になるため、機械学習・AI分野において活用が進んでいる。

フロップス

フロップス(FLOPS, Floating-point Operation Per Second)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。

QLC

(Quad-Level Cell、クアッドレベルセル)

1セルに4ビットの情報が保存可能なフラッシュメモリ。従来のMLC(Multi-Level Cell、マルチレベルセル)、TLC(Triple-Level Cell、トリプルレベルセル)に対して耐久性に劣るものの、安価で大容量という特徴がある。

EIA(米国電子工業会)標準ラック

データセンターなどで使用されるサーバーやネットワーク機器を収納するラックの規格。1ユニットは幅19インチ、高さ1.75インチと定められている。

AI Hub

ABCI上で提供されている大規模な汎用学習済みモデルの再利用等を行うためのツールやサービス群。

Open OnDemand(OOD)

オハイオ州立大学で開発され、オープンソースで提供されている、ウェブブラウザからスーパーコンピューターを利用するためのウェブサービス。ABCI 2.0では2024年8月からOpen OnDemandのベータ版を提供しており、ABCI 3.0でも継続して提供予定。

関連記事

(1) 大規模 AI クラウド計算システム「ABCI」が「ABCI 2.0」にアップグレード

- 産官学共同による AI 研究開発、実証、社会実装を加速 -

(2) 産総研の計算資源 ABCI を用いて世界トップレベルの生成AIの開発を開始

- 産総研・東京工業大学・LLM-jp(国立情報学研究所主宰)が協力 -

プレスリリースURL

https://www.aist.go.jp/aist_j/news/announce/pr20241010.html

本プレスリリースは発表元が入力した原稿をそのまま掲載しております。また、プレスリリースへのお問い合わせは発表元に直接お願いいたします。

プレスリリース添付画像

1

概要図.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI1tn_9z0wW7VW.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI1im_181w4tID.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI1lg_MG469L19.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI1fl_5IYHBJJN.png

2

計算ノード.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI2tn_WMZ1K1qz.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI2im_Olqv2EhJ.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_PI2lg_n5gEfiP5.png

https://cdn.kyodonewsprwire.jp/prwfile/release/M107968/202410097802/_prw_P12fl_5w7BA6sj.png

画像を一括してダウンロード

エリア

東京都

キーワード

研究開発、生成AI、スーパーコンピューター、ABCI

このプレスリリースには、報道機関向けの情報があります。

プレス登録

ログイン

プレス会員登録を行うと、広報担当者の連絡先や、イベント・記者会見の情報など、報道機関だけに公開する情報が閲覧できるようになります。

プレスリリース受信に関するご案内

このプレスリリースを配信した企業・団体

名称 国立研究開発法人産業技術総合研究所

所在地 茨城県

業種 政府・官公庁

URL <https://www.aist.go.jp/>

SNSでも最新のプレスリリース情報をいち早く配信中

x

facebook

過去に配信したプレスリリース

オール電気めっきで超薄型パラジウム水素分離膜を高耐圧化

10/15 14:00

#産総研25周年特別公開 in つくばセンターの開催について

10/14 14:00

排ガスからアンモニア資源を製造、社会実装に向けた活動を加速

10/7 14:00

8th RD20 Conference to Be Held in South Africa, Its First in Africa

10/5 15:00

発光強度の変化でタイヤ添加剤とアルブミンの結合を「見える化」

9/28 14:00

合成音声読み上げ時の音声情報理解力を定量化

9/25 14:00

第8回RD20国際会議開催のご案内

9/18 18:00

AIの予測と人によるスクリーニングが創薬を変える

9/18 14:00

産総研、BioJapan 2026に出展

9/18 11:00

ルーブル美術館がランス別館での特別展でアザラシ型ロボット「パロ」を展示

9/17 17:00

遺伝子治療の可能性を広げる「HSVアンプリコンベクター」の高効率産生技術を開発

9/16 14:00

国内最大級のデジタルイノベーションの総合展CEATEC 2026に出展

9/15 11:00

もっと見る