大規模オープン AI インフラストラクチャ ABCI

Large Scale Open AI Infrastructure ABCI

中田秀基 *

1. はじめに

本稿では，産業技術総合研究所（以下，産総研）が保有，運用している，どなたでも使用できる大規模 AI 向けクラウドシステム ABCI（AI Bridging Cloud Infrastructure: AI 橋渡しクラウド）[1-3] について紹介する．

図 1　ABCI 棟全景

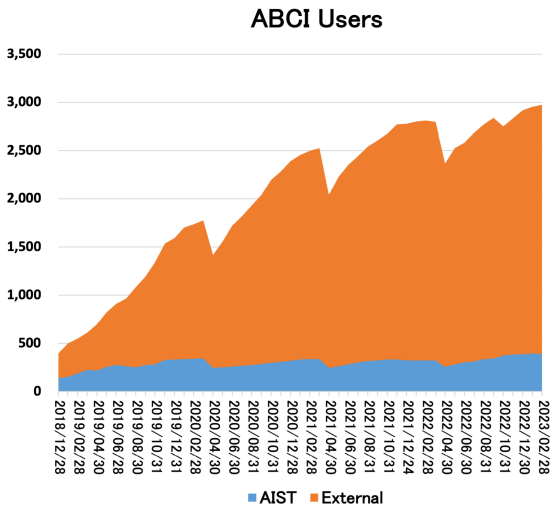

図 2　ユーザー数の遷移

2. ABCI の目的と概要

2000 年代なかばに始まるディープラーニングによる AI の興隆は，インターネットによる大規模な機械可読データの蓄積と，計算機技術の発達による計算性能の大幅な向上によるものであった．これは，AI 技術を応用するためには大規模なデータと計算機を利用する環境が必要だということを意味する．したがって，AI 技術応用を多くの産業に普及するには，大規模計算環境を誰もが容易に利用できる必要がある．

ABCI は AI 技術の産業応用を促進するため，経済産業省からの補助を受けて産総研が整備したオープンな AI インフラストラクチャである．ここでいう「オープン」とは，特定のユーザに提供されるものではなく，広く一般の企業や研究機関等が利用可能であることを意味する．

ABCI は柏の葉キャパス駅間近の東京大学柏 II キャンパス内に設けた「AI データセンター棟」に設置されている．図 1 にデータセンター棟の全景を示す．この建物は ABCI のみが収容されており，小さめの体育館程度の建屋に高密度にノードを配置しており，5000 を超える GPU を提供している．

2018 年 8 月から運用を開始し，2021 年にテクノロジリフレッシュを行っている．

3. ABCI の利用状況

ABCI は，広く一般の企業や研究機関に計算資源を提供するという理念のもと，利用申請を募っている．

ABCI のユーザ数を図 2 に示す．下部の灰色の部分が産総研内部のユーザを示し，上部の黒い部分が

* 産業技術総合研究所 上級主任研究員

外部ユーザを示している．2022 年度末で，3000 人弱のユーザがいることがわかる．また，産総研内部のユーザ数は 300 名程度で伸びていないが，外部のユーザは拡大を続けている事がわかる．

3 月から 4 月にかけて一度減るのは，ABCI が年度単位で運用されており，年度をまたいだ際に一度すべてのユーザの登録を解除するためである．

利用組織はおよそ 300 で，大企業，AI 系スタートアップ企業から，大学，公的研究機関まで幅広い．用途としては，初期のころは画像の解析がドミナントであったが，現在では学習手法の解析，AI 品質評価，材料設計・分析，言語処理，バイオ，医療，生成 AI など，多岐に広がっている．

https://abci.ai/ja/link/use_case.html に，食肉加工への応用，動画解析に流体特性の把握，社会インフラへの展開など，興味深いユーザの利用例を掲載しているので是非ご覧になっていただきたい．

4. ABCI の構成と特徴

4.1 ABCI の構成

ABCI の構成を図 3 に示す. ABCI は, 主に計算ノード，ネットワーク，ファイルシステムから構成される．計算ノードには 2018 年に導入された V ノードと，2021 年度に導入された A ノードがある．

V ノードは 1088 台で，NVIDIA V100 をそれぞれ 4 機ずつ搭載する．A ノードは 120 台と少ないが，NVIDIA A100 をそれぞれ 8 機ずつ搭載する．

V ノード群は Infiniband EDR で，A ノード群は Infiniband HDR でそれぞれ相互に接続している．V ノード群と A ノード群は Infiniband EDR で接続されている．

ファイルシステムとしては大規模な共有ファイルシステムを持つ．このファイルシステムは Infiniband を経由して A ノード，V ノードからマウントされる．ファイルシステムとしては Lustre を用いている．

また，通常の共有ファイルシステム以外にオブジェクト・ストレージを備える．このストレージは，Amazon S3 互換のインターフェイスを持ち，外部から HTTP で直接アクセスすることが可能となっている．このストレージを用いることで外部との大規模なデータの受け渡しをスムーズに行うことができる．

共有ストレージ以外に，個々のノードに高速な SSD を搭載している．これは，学習データを予めステージしておくためや，学習の途中経過をチェックポイントとして記録しておくために用いる．

外部へのネットワークとしては，SINET6[4] に接続している．SINET6 は，国立情報学研究所が管理運用する学術情報基盤ネットワークで，全国の大学や研究機関を高速なネットワークで接続している．SINET6 に参加している機関であれば，10G のネットワーク帯域で ABCI にアクセスすることができる（SINET6 と AI データセンター棟は 400G で接続されているが，ABCI のサービスネットワークの帯域が 10G であるため）．

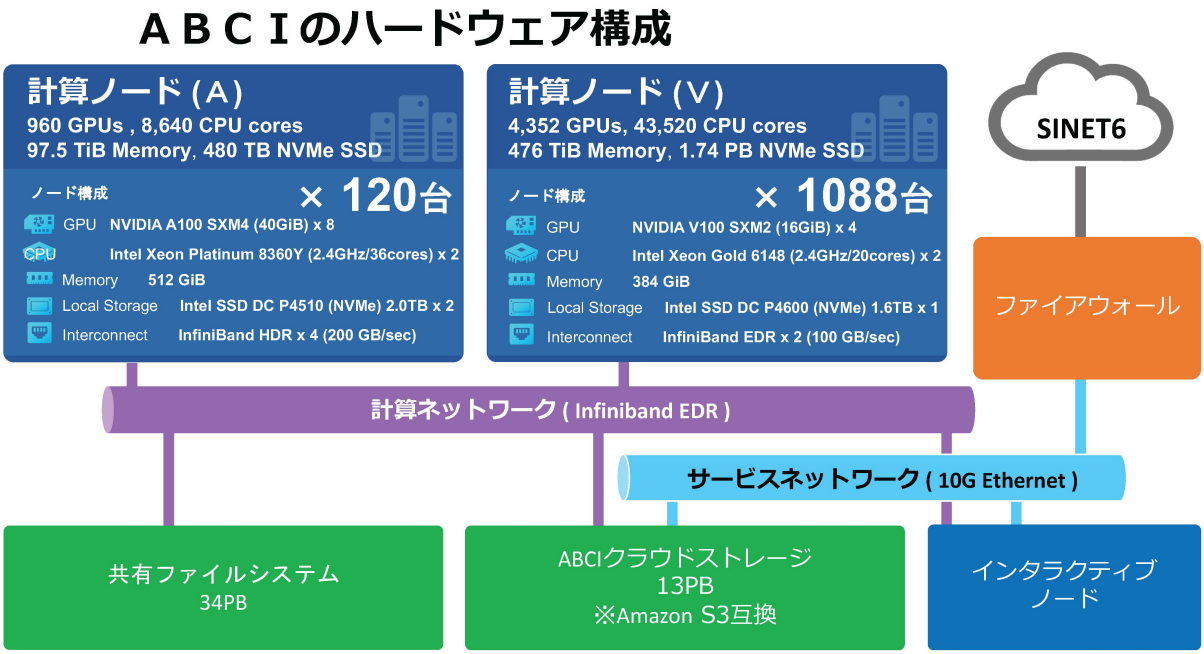

図 3　ABCI の構成

4.2　データセンター

AI データセンター棟は ABCI のために建てられた平屋の建造物で，そのほとんどがサーバ室（19m x 24m 程度）で占められている．データセンターでは，床下に配線などの空間を設けたフリーアクセス構造をとることが多いが，この AI データセンター棟ではコンクリートスラブの上に直接サーバを設置する構造としている．これは昨今の高密度サーバにも耐えられるように，床面の耐荷重を大きくとるためである．

4.3　ABCI の冷却

4.3.1　概要

非常に効率の良い冷却が ABCI の特徴の一つである．昨今の GPU は非常に多くの電力を消費し，それを熱として排出する．これを何らかの方法で排出しなければ，GPU は溶けてしまう．しかし一般には熱の排出，すなわち冷却そのものにもエネルギーが必要であり，この冷却の効率が，データセンターの効率を大きく左右する．

データセンターの冷却効率の指標として PUE（Power Usage Effectiveness）がある．PUE はデータセンターの消費電力全体を，IT 機器の消費電力で割ったもので，この値が小さければ小さいほどエネルギー効率がいいことになる．PUE が 1.0 であれば，全く冷却にエネルギーを使わないということになる．一般的なデータセンターは PUE1.7 程度であると言われている．

これに対して ABCI では，基本を水冷とすること，その際の水の温度を下げすぎないことによって，非常に高い冷却効率を実現しており，年間を通じた PUE は 1.1 程度である．

前述した通り，ABCI では基本的に冷却水を用いて冷却を行う．このためには，空気を媒体とするよりも水を媒体としたほうが効率が良いからである．この際に，水の温度を低く保とうとすると効率が低下する．ABCI では可能な限り水の温度を高く保つことで，冷却効率を保つ．

4.3.2　冷却ポッド

ABCI の冷却の要となるのが冷却ポッドである（図 4）．冷却ポッドは ABCI の冷却の単位で，外部から冷却水の供給を受け内部に設置した機器の冷却を行う．実態としての冷却ポッドは，データセンター内に構築した輸送コンテナのような構造物である（図 5）．冷却ポッドは水冷と空冷の双方を司る．水冷機能としては Coolant Distribution Unit（CDU）を持ち，各サーバに対して 1 次冷却水を提供する．サーバから戻ってきた高温の冷却水を，2 次冷却水で冷却しサーバに戻す．

空冷機能としては，冷却ポッドはホットアイルとコールドアイルを分離し，ホットアイルの空気を冷却してコールドアイルに戻す役割を果たす．冷却ポッド外部がコールドアイル，内部がホットアイルとなる．各サーバは前面がポッドの外に向き，背面がポッドの内側を向く．前方から吸い込んだ比較的低い温度の空気でコンポーネントを冷却し，温度が上がった空気をポッドの中に排出する．

ポッドの上部には Fun Coil Unit（FCU）と呼ばれるラジエータのようなものがある．この FCU がホットアイルの空気と冷却水の間で熱交換を行う．

図 5 に冷却ポッドの外観を示す．緑色の構造物がポッドである．下の段に並んでいるのがサーバで，その上の窓のように見える部分が FCU の出力部分である．左側から伸びている銀色の管が冷却水を循環させている配管である．この図の左側の壁の向こうに図 6 に示す冷却塔がある．

4.3.3　冷却塔

冷却ポッドに供給する冷却水を冷やすのが，冷却塔だ．図 6 に水を冷やすために用いる冷却塔の外観

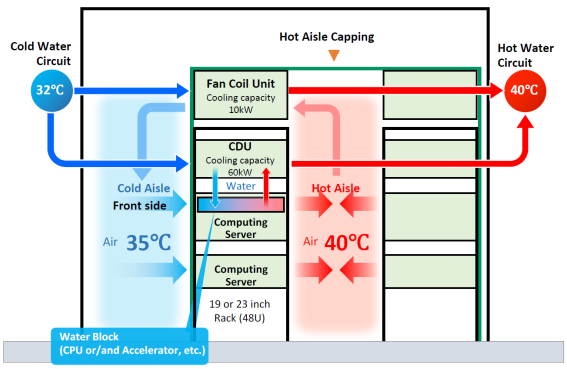

図 4　冷却ポッドの概念図

図 5　冷却ポッドの外観

を示す．左側にあるのが気化熱で水を冷却する冷却塔で，右側の建物がデータセンター棟である．建物との間に循環のためのポンプ設備があるが，この写真では見えていない．

4.3.4　各ノードの冷却

図 7 に A ノードの前面を示す．上下に 2 つノードが並んでおり，それぞれ 5 ペア 10 本ずつ冷却水循環の黒いパイプが接続されている事がわかる．ノード内部では，GPU や CPU などの主要な熱発生源に対して冷却ヘッドが密着する形で設置されており，冷却ヘッドを水冷することで冷却を行う．

ただし，すべての熱を水冷で冷却する事はできない．水冷ヘッドは特に大量の熱を発生するコンポーネントにしか設置していないためだ．例えばメモリなどは水冷できないので，空冷を組み合わせる必要がある．このため冷却ポッドに空冷の機能が必要となるのだ．

5　利用方法

本節では ABCI の具体的な利用方法について述べる．

5.1　利用対象者

安全保障輸出管理上，利用者は国内居住者に限定している[*1]．また，契約者としては国内の法人を前提としている．安全保障輸出管理上要請される確認コストが大きいため，個人としての利用申請は受け付けていない．

図 6　冷却塔

また利用目的は研究や実証実験に限っている．いわゆる実運用サービスのバックエンドとしての利用は想定していない．もっとも ABCI の QoS は比較的低いため，そもそも実運用サービスとしての使用に耐えるものではない．

5.2　ポイント

ABCI では使用時にポイントを事前に購入いただく形となっている．これは，運用者としての産総研にとっての代金徴収のコストを低減し，金銭的なリスクを回避するためである．1000 ポイント（2023 年度は 22 万円）を事前に購入いただき，このポイントを消費する形でご利用いただくことになる．各計算資源に対するポイント消費量は https://abci.ai/ja/how_to_use/tariffs.html に掲載されている．一例を挙げると，NVIDIA V100 を 4 機搭載した V ノードで 1 ポイント 1 時間である．すなわち，1000 ポイントで V100 を 4 機，およそ 42 日間利用できることになる．

また，ポイントは年度末に失効する．これは産総研が単年度会計をとっているため，産総研にとっての債務にあたるポイントを年度をまたいで持ち越すことができないためである．また，科研費を始めとする多くのアカデミアの外部資金も年度をまたいで債権を保有することを認めていないため，年度末に執行したほうが都合が良いという側面もある．

5.3　ジョブの起動

ABCI は多数のノードを管理するために，バッチキューイングシステムと呼ばれる機構を用いている．一般には，計算機を使用する際にはその計算機に直接ログインして計算を実行する事が多い．しか

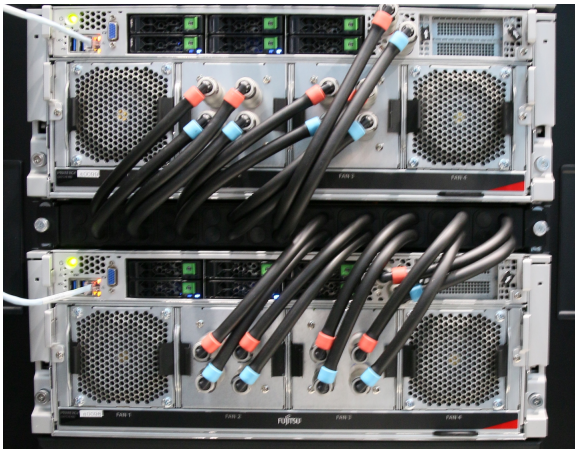

図 7　A ノードの前面

*1　厳密にはさらに詳細な規定があり，居住者でも利用いただけない場合もある．詳細は https://abci.ai/ を参照

```
#!/bin/bash

#$-l rt_F=1
#$-j y
#$-cwd

source /etc/profile.d/modules.sh
module load cuda/10.2/10.2.89

./a.out
```

図 8　ジョブスクリプトの例

し ABCI のように多数の計算ノードを持つ計算機では，どのノードをだれが利用したらいいのかわからない．ノードをユーザに固定的に割り当てればいいと思うかもしれないが，それでは流石にノード数が足りないし，使用率が極端に低下してしまう．

このような問題を解決するために用いられるのがバッチキューイングシステムである．ユーザは計算ノードにログインするのではなく，「ジョブ」を作成し，バッチキューイングシステムに「サブミットする」．バッチキューイングシステムは，サブミットされたジョブを「キュー」と呼ばれる，最初に入れたものが最初に取り出されるデータ構造で管理する．キューイングシステムという名前はこのデータ構造に由来する．

バッチキューイングシステムは，利用可能なノードが存在すると，キューの先頭からジョブを取り出してそのジョブにノードを割り当てる．利用可能なノードがなければ，ジョブはキューの中に滞留することになる．つまり，ABCI が空いていれば，投入されたジョブは即座に実行されるが，混雑しているとかなり待たされることになる．

ユーザの立場からすると使いにくいと感じるだろうが，限りある資源を有効に活用するためにはバッチキューイングシステムが最善であると思われる．

5.4　ジョブスクリプト

前節で「ジョブ」を作成すると書いたが，ジョブの実態は実行するコマンドを羅列したシェルスクリプトである．ただし，スクリプトの冒頭部分に，バッチキューイングシステムへのさまざまなリクエストをコメントとして付加することができる．

例として図8にジョブスクリプトの例を示す．3,4,5 行目のコメント文(冒頭に # がついている)がバッチキューイングシステムへの指示となっている．その後 7 行目，8 行目で環境を設定し，最後の 10 行目で実際のコマンドである ./a.out を実行している．

ジョブの投入は qsub コマンドで行う．その際，ABCI では利用する課金グループをオプションで指

```
qsub -g G0000000 run.sh
```

図 9　ジョブサブミットの例

定しなければならない．例を図 9 に示す．ここで，G0000000 は架空の課金グループ名で，run.sh はジョブスクリプトのファイル名である．

ジョブの実行状態は qstat コマンドで見ることができる．また，qdel コマンドで実行中もしくは実行待機中のジョブを消去することもできる．

通常ジョブの実行結果は，ファイルとして書き出すようにプログラムを書く．また，ジョブプロセスの標準出力と標準エラーはそれぞれ自動的にファイルに書き出される．

5.5　バッチジョブの活用

このようにプログラムの実行をジョブとしてサブミットしなければならないのは，実行までの手間がひとつ増えるため不便ではある．しかし，バッチジョブを用いることで，便利になる点もある．その一つが，複数のジョブの同時実行だ．機械学習では条件を変えて訓練を何度も試行する場合がある．このような場合には条件ごとにジョブスクリプトを作成して，同時に qsub すれば，すべてのジョブが（ノードが空いていれば）並列で実行される．これによって，実験のターンアラウンドタイムを大幅に短縮することができる．

5.6　qrsh の利用

このようにバッチジョブは便利なものではあるが，実行ノードでインタラクティブにプログラムを実行したいという要望もある．このような用途のために qrsh というコマンドが用意されている．qrsh は qsub と同じロジックで，計算ノードを確保し，そのノード上でシェルを実行する．

このコマンドを使用すると通常の SSH でリモート計算機を使用するのと同様の感覚で，ABCI の計算ノードを使用する事ができる．ただし，ノード割当のロジックが qsub と同じであるため，常にすぐノード割当を受けることができるとは限らない点に注意が必要である．場合によっては数時間単位で待たされる可能性もある．

6　ABCI の今後

6.1　利用障壁の低減

現在，ABCI を利用するには SSH でログインし，コマンドラインインターフェイスを用いてジョブを投入するしかない．SSH トンネルを利用することで Jupyter Notebook や JupyterLab[5] などの Web

ベース UI の利用も可能ではあるが，設定は容易ではなく，初心者ユーザの利用障壁となっている．

この問題を解決するために，Open onDemand[6] の導入を検討している．Open onDemand はオハイオスーパーコンピュータセンターが開発した，高性能計算環境のための Web フロントエンドで，多くの計算機センターで導入がすすんでいる．日本でも富岳が導入している [7]．

Open onDemand を用いると，ジョブの投入やファイルのアップロードやダウンロードなどの基本的な操作を Web ベースの UI から行うことができるようになるだけでなく，WebUI をもつアプリケーションを直接使用することも可能になる．例えば JupyterLab などを Open onDemand 経由で直接利用することができる．図 10 に現在テスト中のジョブ投入インターフェイスの様子を示す．

一方で，一般に利便性とセキュリティは常にトレードオフの関係にあり，Open onDemand のようにユーザにとって利便性の高いツールを導入することは，潜在的なセキュリティ低下要因となりうる．セキュリティの観点を重視しつつ慎重に導入を進める予定である．

6.2　今後の更新計画

現在，次期システムの導入を計画している．次期システムは，利用目的として人工知能技術に加えて量子関連技術も視野にいれている．この運用開始は 2025 年春を予定している．全体としては現在の ABCI を大きく上回る演算性能を持つことが期待される．

7　おわりに

本稿では，産総研が管理運用する大規模オープン AI インフラストラクチャ ABCI について解説した．今後あらゆる分野において，AI の応用が広がり続けることが予想される．AI を利用する機会があれば，ぜひ ABCI の利用を検討いただきたい．

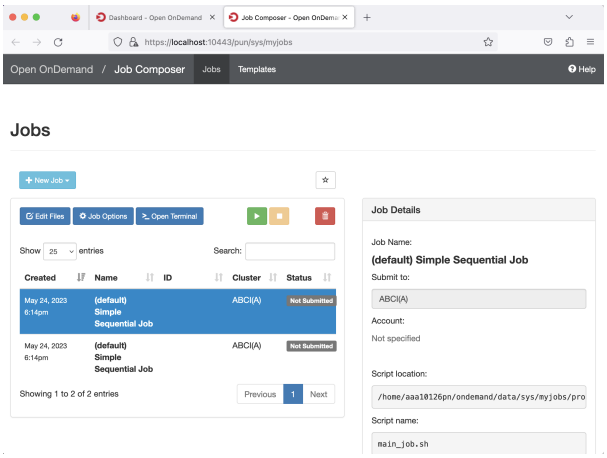

図 10　Open onDemand

謝辞

ABCI の構築，管理，運用に携わるすべての皆さまに感謝します．

参考文献

1）　ABCI AI Bridge Infrastructure: https://abci.ai/. Accessed: 2023-02-01.
2）　小川宏高 , 松岡聡 , 佐藤仁 , 高野了成 , 滝澤真一朗 , 谷村勇輔 , 三浦真一 , and 関口智嗣 . AI 橋渡しクラウド― AI Bridging Cloud Infrastructure（ABCI）― の構想 . In 情報処理学会研究報告 2017-HPC-160, 2017.
3）　Shinichiro Takizawa, Yusuke Tanimura, Hidemoto Nakada, Ryousei Takano, and Hirotaka Ogawa. ABCI 2.0: Advances in Open AI Computing Infrastructure at AIST. In 情報処理学会研究報告 2021-HPC-180, 2021.
4）　学術情報ネットワーク sinet6. https://www.sinet.ad.jp/.
5）　Jupyter. https://jupyter.org/.
6）　Open OnDemand. https://openondemand.org/.
7）　「富岳」Open OnDemand の提供を開始～ Web ブラウザで「富岳」の操作が可能に～ . https://www.r-ccs.riken.jp/outreach/topics/20230530-1/, 2023.

著者略歴

中田 秀基（なかだ ひでもと）
1995 年東京大学情報工学専攻博士課程修了，博士（工学）．
同年 工業技術院電子技術総合研究所入所．
2001 年 産業技術総合研究所に改組．
2001 年から 2006 年まで，東京工業大学客員准教授．
2011 年より筑波大学連携大学院教授．
並列・分散計算，機械学習に興味を持つ．
「プログラミング Rust」「Python ではじめる機械学習」など，翻訳書多数．