ABCIの現状と課題

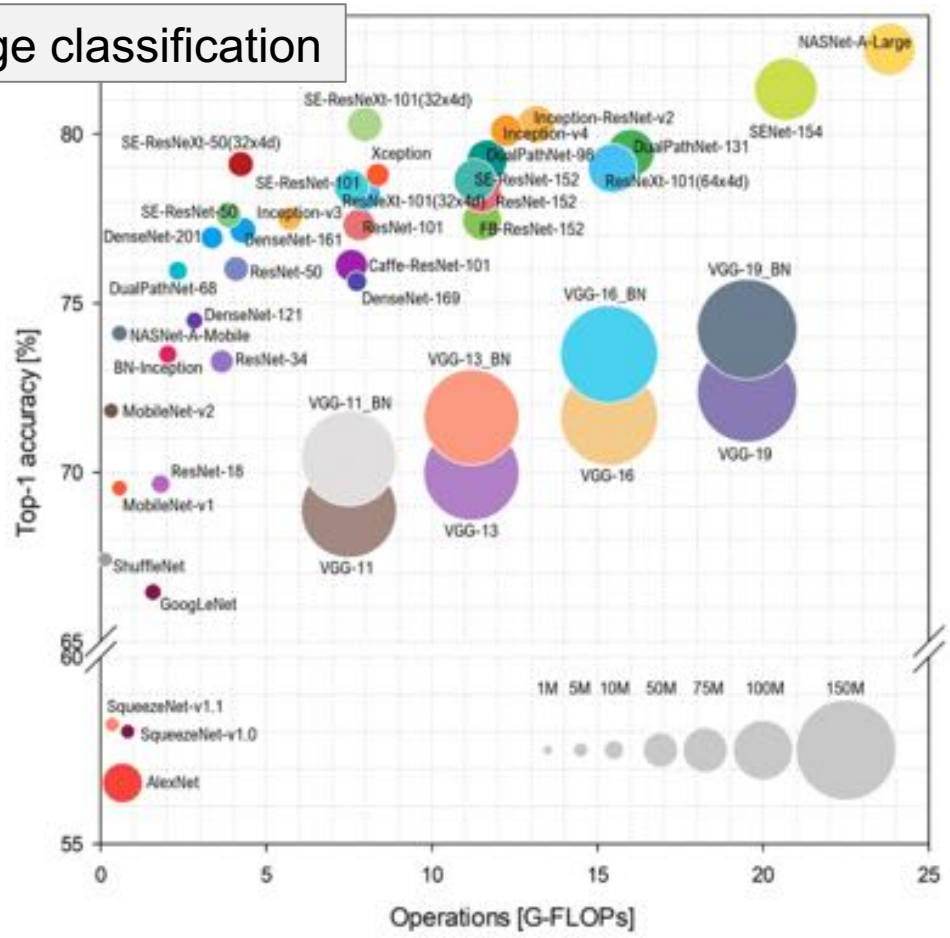
国立研究開発法人 産業技術総合研究所
高野 了成

2022年6月23日　PCクラスタワークショップ in 神戸2022「クラウドとHPC」

AI研究開発における大規模インフラの重要性

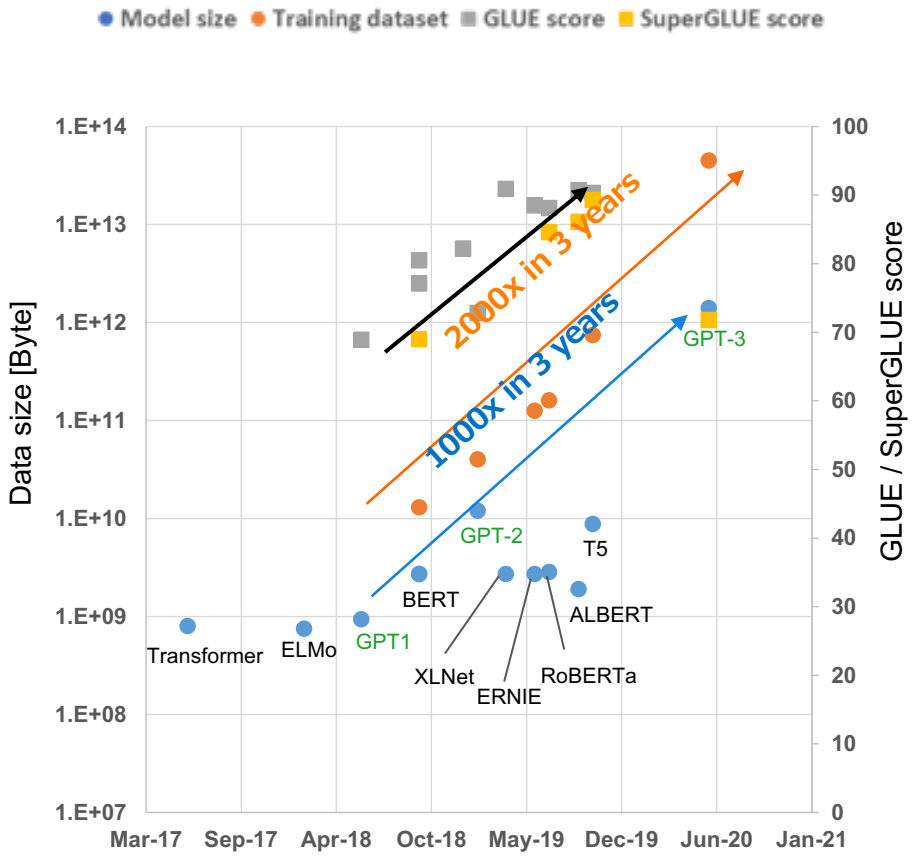
大きな計算機パワー、大きなデータセットを用いるほど、学習済みモデルの精度は向上する

Image classification



S.Bianco et al.: Benchmark Analysis of Representative Deep Neural Network Architectures, IEEE Access, 2018

Natural Language Processing



Courtesy of A. Kasahara, Fujitsu

世界最大級・超省電力・オープンＡＩインフラストラクチャ

ABCI

- 経産省「人工知能に関するグローバル研究拠点整備事業」（H28二次補正）の一環として整備
- 我が国における産学官によるＡＩ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ＡＩ技術の研究開発・実証、社会実装の推進、ＡＩ分野の最重要課題への挑戦が目的
- 新型コロナウイルス感染症対策に無償提供
- 経産省「人工知能に関する橋渡しインフラ拡張」（R1補正）により昨年度末、大幅アップグレードを実現

2018年8月1日　運用開始
2021年5月10日 2.0運用開始

ＡＩ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

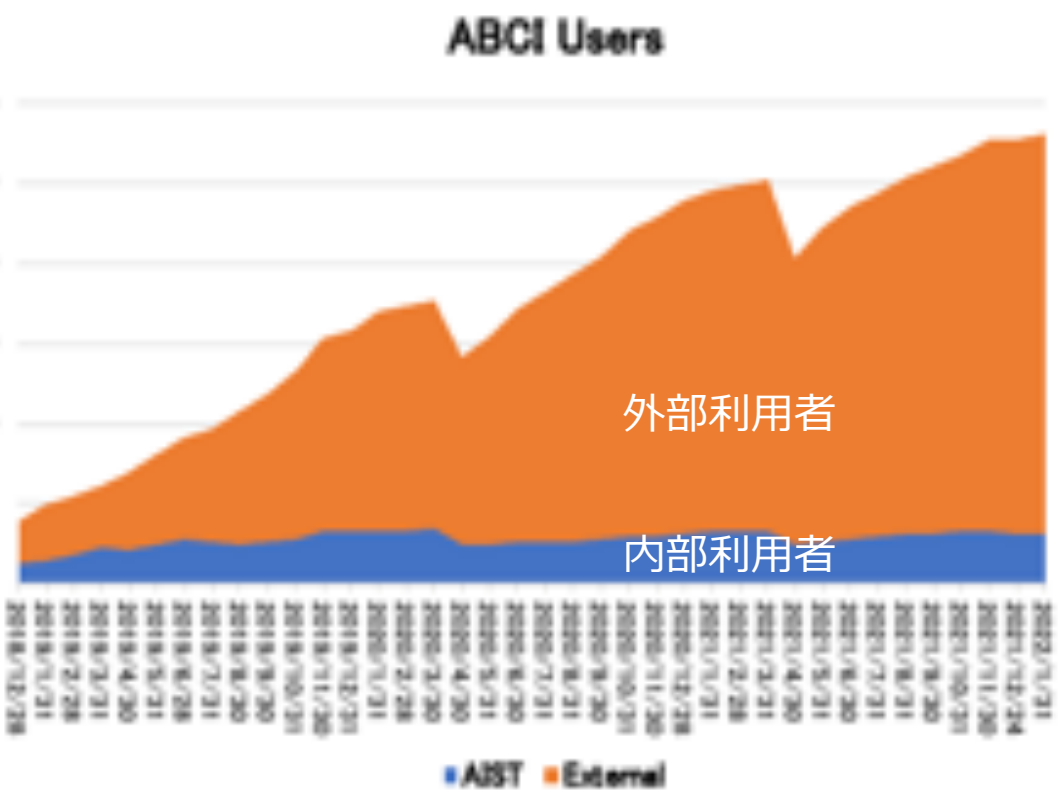
企業がクラウドで個人情報を
扱える水準のセキュリティ機構

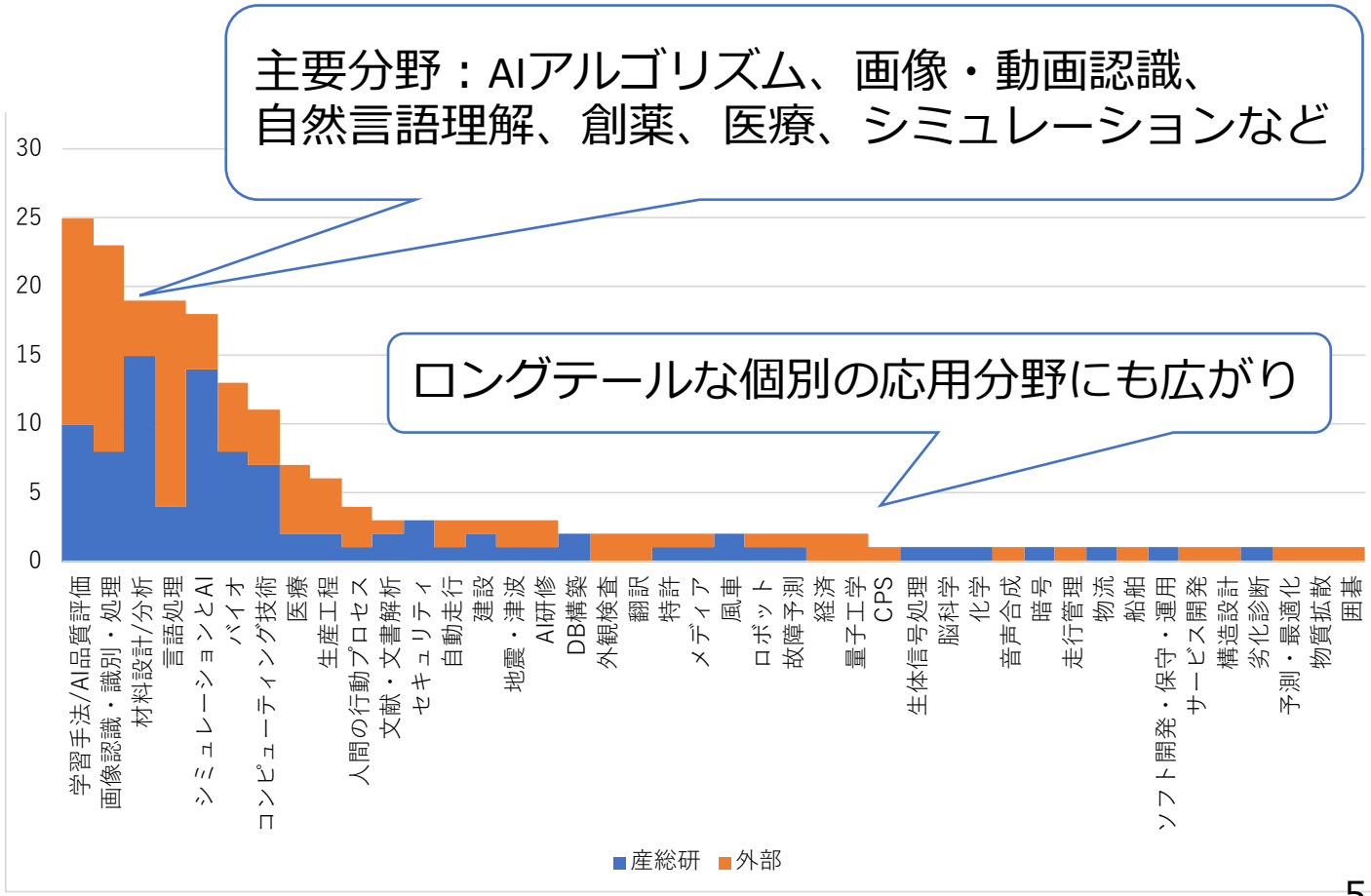
「AIを試す場」
人工知能産業のための
オープンプラットフォーム形成
最先端のAI研究から
誰でも試して使えるAIまで

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で増加
- 2022年1月現在の利用者数は2800以上 (うち外部利用が約88%)
- AIに関する様々な応用に幅広く利用





ABCIの主な利用機関

AIスタートアップ～中小企業
- ギリア株式会社*
- 株式会社高電社*
- アイリス株式会社*
- Linne株式会社*
- 株式会社トリプルアイズ*
- LeapMind株式会社*
- 株式会社アタリ*
- 株式会社IABC*
- 株式会社コトバデザイン*
- 株式会社YAMATO*
- 株式会社 Laboro.AI*

大企業
- 株式会社富士通研究所*
- パナソニック株式会社*
- 株式会社リクルートテクノロジーズ*
- 株式会社パスコ*
- オムロンサイニックエックス株式会社*
- 株式会社日立製作所*
- ソニー株式会社
- 日本電信電話株式会社
- NHK放送技術研究所
- オリンパス株式会社
- JFEスチール株式会社
- トヨタ自動車株式会社
- 株式会社東芝
- ルネサスエレクトロニクス株式会社

大学・国研
- 国立研究開発法人 日本原子力研究開発機構*
- 国立研究開発法人 医薬基盤・健康・栄養研究所*
- 特定国立研究開発法人理化学研究所AIP
- 東京工業大学
- 千葉工業大学
- 東北大学
- 東京大学
- 京都大学

グループ数: 480
- 産総研 158
- 共同研究 39
- 大学 128
- 企業 102
- 国研 24
- 財団等 7
- 無償 22

利用者数: 2,803
- 産総研　324
- 外部 2,479

2022/1末時点

＊ 利用事例公開(2022/1時点)
https://abci.ai/ja/link/use_case.html

ABCIのハードウェア構成

計算ノード (A)

960 GPUs , 8,640 CPU cores
97.5 TiB Memory, 480 TB NVMe SSD

× 120台

ノード構成

GPU	NVIDIA A100 SXM4 (40GiB) x 8
CPU	Intel Xeon Platinum 8360Y (2.4GHz/36cores) x 2
Memory	512 GiB
Local Storage	Intel SSD DC P4510 (NVMe) 2.0TB x 2
Interconnect	InfiniBand HDR (200 Gbps) x 4

計算ノード (V)

4,352 GPUs, 43,520 CPU cores
476 TiB Memory, 1.74 PB NVMe SSD

× 1088台

ノード構成

GPU	NVIDIA V100 SXM2 (16GiB) 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 (100 Gbps) x 2

SINET6

ファイアウォール等

計算ネットワーク (InfiniBand)

サービスネットワーク (10G Ethernet)

共有ファイルシステム
34 PB

ABCIクラウドストレージ
13 PB
※Amazon S3 互換

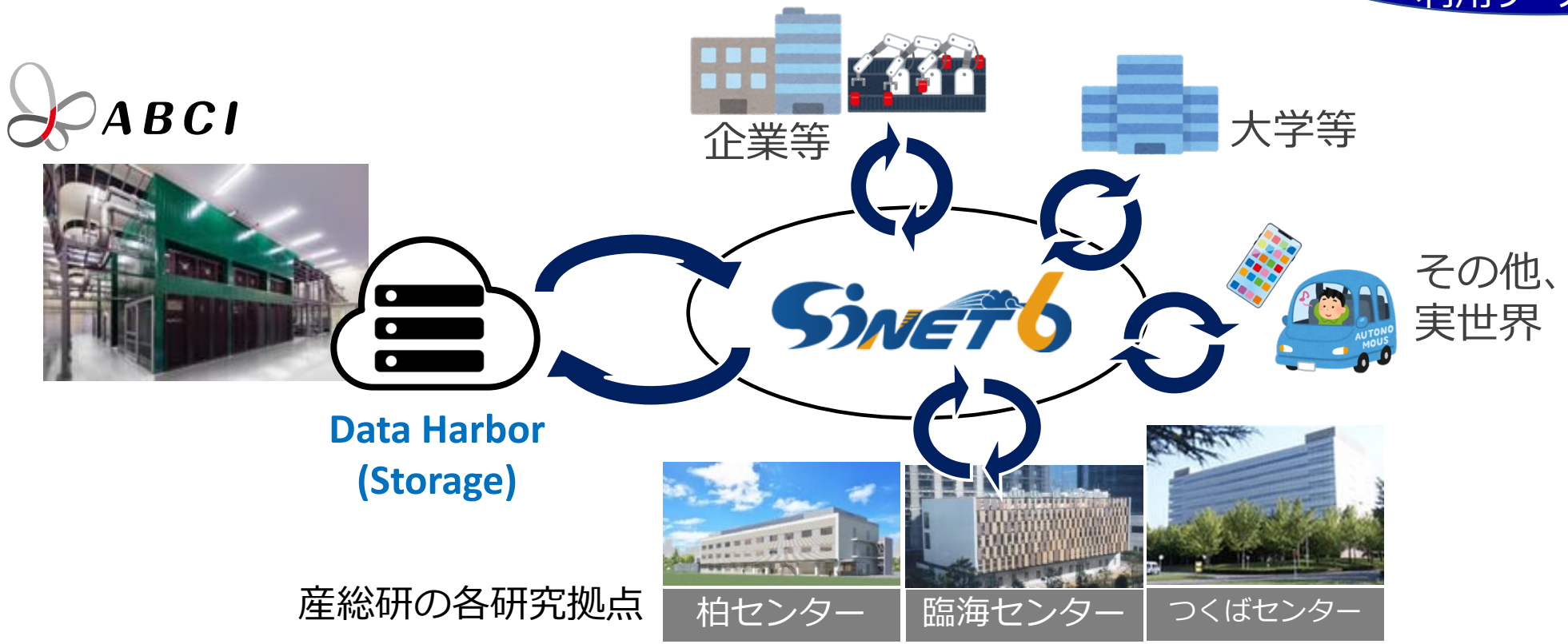
インタラクティブノード

データの価値化を実践する「場」の提供を目指して

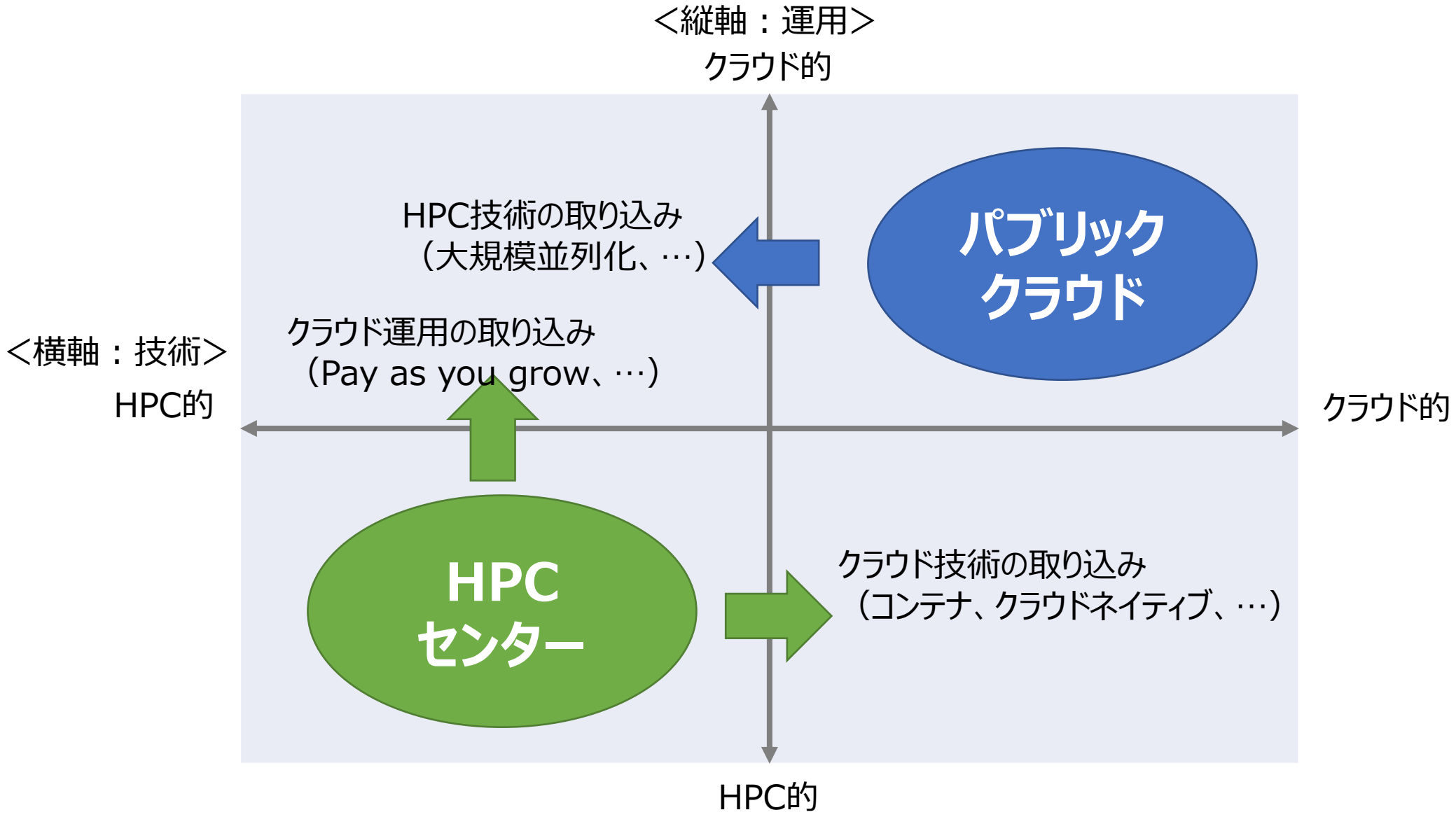
- クラウドストレージは、ABCI のフロントに位置し、SINETに直結した「データハーバー」としての役割を担う
  - SINETに直結した各機関から、高速・安全にデータを収集・蓄積
  - ABCI上で生成された、高性能な汎用学習モデル等の共有・配布

ABCI Datasets サービス

学習済みモデル 利用ワークフロー



クラウドとHPCの融合：技術と運用



（他のHPCセンターと比較した）ABCIの特徴

- AI分野における標準的・オープンな技術を用いており、開発したアプリケーションやアルゴリズムを事業として容易に社会実装に提供できる。
  - アイデアから検証までのTurn-Around-Timeが短い
  - ABCIで検証した技術がそのまま産業移転可能

- クラウド的に運用により、企業利用の参入障壁が低くなるよう制度を長年に亘り整備しており、実績の蓄積がある。
  - 共用促進法*の縛りがなく、約款ベースで即時利用が可能
  - 利用者累計4543名、アクティブ利用者2487名（うち外部利用は88%）

- 主要な部分を汎用品のCPUとGPU等で構成し、短期間でシステムの導入、バージョンアップ、加速演算部等の組込が容易である。

*共用法の制約を受けるのは富岳だけで、他のセンターは特にそれはないです。一方で大学は学術機関であり、単なる商用利用を認めるのは難しいという側面はあります。（塙先生コメント）

企業利用事例（1）

目的と成果について

FUJITSU

- 目的：誰でも容易に大規模演算機環境を利用する事ができ、迅速、且つ、高精度な言語モデルを作成
  - 参入障壁：高速化のためにはシステム構成に依存した処理/知識が必要

システム依存部分

環境構築

↓

ステージング等前処理

ABCIグランドチャレンジ

学習処理(pre-training)

↓

学習処理(fine-tuning)

↓

評価

各システムに適した構築環境を登録1ステップで実行環境構築可能

実際の構築環境で大規模モデルを学習し評価
→CANDAR2021にて発表
Outstanding paper受賞

FUJITSU

ホーム > 研究開発 >

スーパーコンピュータを活用したAIモデルの開発において、世界トップクラスの高速演算性能と認識精度を実現できるソフトウェア技術を開発

スーパーコンピュータを活用したAIモデルの開発において、世界トップクラスの高速演算性能と認識精度を実現できるソフトウェア技術を開発

2021年11月24日

近年、様々なデータを解析し活用するために、AIは進化し大規模な処理を高性能に行えるようになりました。例えば、人が作業したのと同じような品質の高い翻訳や文書生成が実現されてきています。

しかし、一般のコンピュータでは大規模なAIモデルを学習するには、数か月以上かかることがあります。スーパーコンピュータ（以下、スパコン）を使えば瞬く間にできますが、使いこなすには専門の知識が必要です。

そこで、富士通はスパコンでの数値計算のノウハウを活かし、大規模なAI学習を簡単に高速処理できるソフトウェア技術を開発しました。大規模な日本語AIモデルの学習にこの技術を適用し、短時間での学習に成功しました。今回構築した規模のAIモデルの学習処理では通常数カ月以上かかりますが、スパコンを利用し4.7時間で学習できる様になりました。

本技術は、2021年11月23日から開催の国際会議CANDAR 2021で発表します。

富士通笠置氏より提供「2021年度第1回ABCIグランドチャレンジ日本語大規模モデルの分散並列学習」

大規模日本語BERT(6.7Bパラメータ)による BERT-base(110Mパラメータ)への蒸留

中町 礼文, 李 聖哲, 小林 滉河, 佐藤 敏紀 (AIカンパニー NLP開発チーム)

企業利用事例（2）

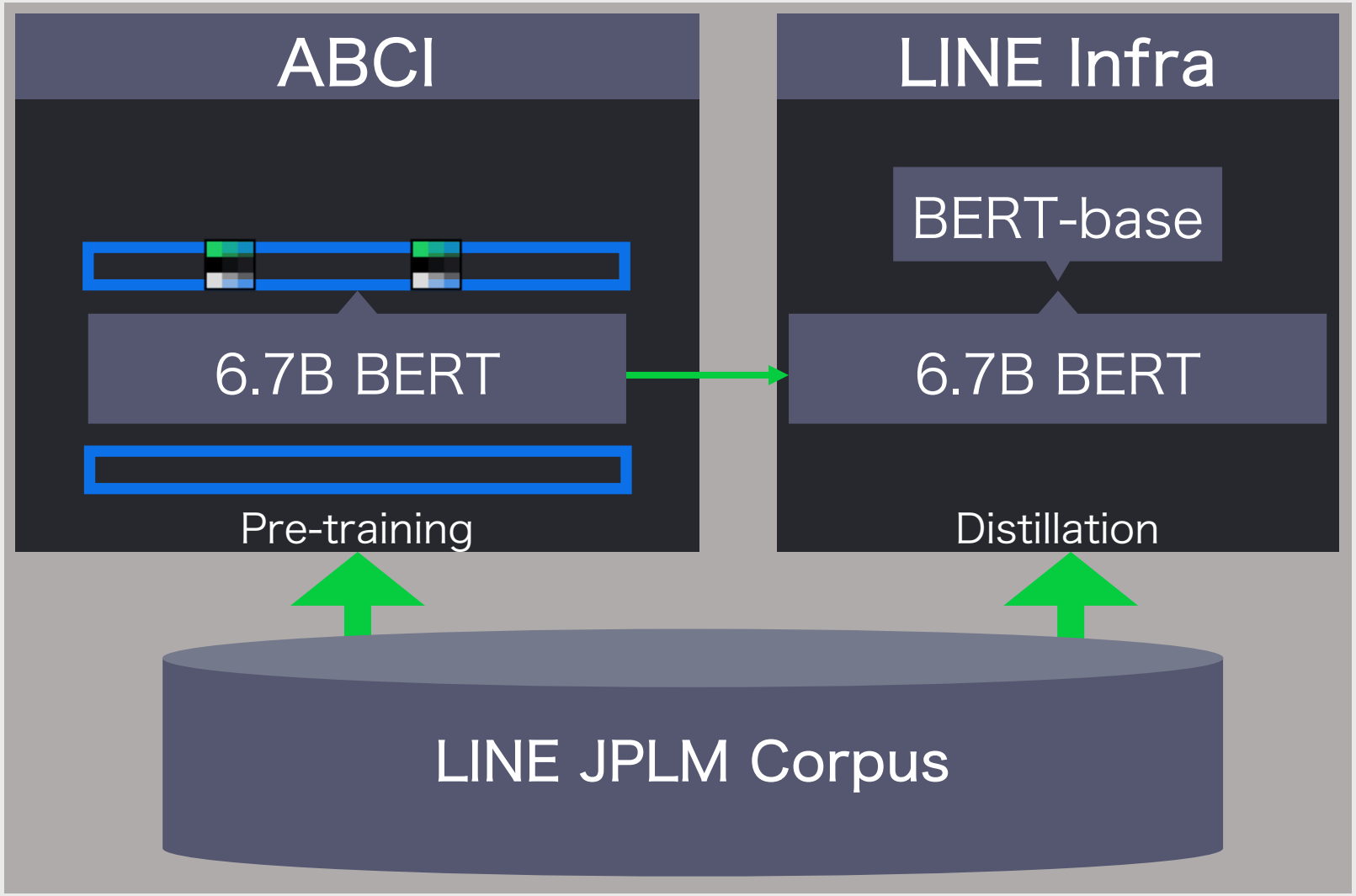
LINE

LINEにおけるABCI活用事例

ABCIチャレンジで提供して頂いた計算機群を活用し, LINEのTB級コーパスによって約6.7Bパラメータの日本語BERTの事前訓練を行った.

現在は, ABCI上で高速に作成できた6.7B BERTからBERT-baseへの蒸留を, LINEのGPU基盤上で行っている.

今後の予定として, モデルの評価を実施し, 成果をOSSとして提供する.



企業利用事例（３）

HITACHI
Inspire the Next

News Release

2020 年 12 月 2 日
株式会社日立製作所

自然言語処理の国際コンペティション
「CoNLL 2020 Shared Task」と「SemEval 2020」の複数部門で1位を獲得
日立が研究開発した AI・自然言語処理の基礎技術により、日本初の同時獲得

株式会社日立製作所(執行役社長兼 CEO:東原 敏昭／以下、日立)は、このたび、自然言語処理の技術を競う世界的なコンペティションである「CoNLL 2020 Shared Task」と「SemEval 2020」において、それぞれの複数部門(タスク)で 1 位を獲得しました。両コンペティションでの 1 位同時獲得は本邦初であり、さらに、「CoNLL Shared Task」での 1 位獲得も日本企業初となります。

CoNLL: Conference on Natural Language Learning
SemEval: Semantic Evaluation

https://www.hitachi.co.jp/New/cnews/month/2020/12/1202.html

技術面での現状と課題

HPC寄り	クラウド寄り
ジョブスケジューラによる資源管理 並列ジョブでは近傍ノードを割当	ノードを分割して資源提供 オーバサブスクリプションはないが、複数利用者でノード共有可能
管理者権限なし	
Environment modules, Spackによるソフトウェアパッケージ管理	Dockerコンテナイメージを用いたLinuxコンテナ利用 （ランタイムはSingularity）
UNIXアカウント・グループベースでの認証	S3互換クラウドストレージ （柔軟な認証認可）
UNIXパーミッションによるアクセス制御	クラウドサービスのバックエンドとして連携動作

今後の課題

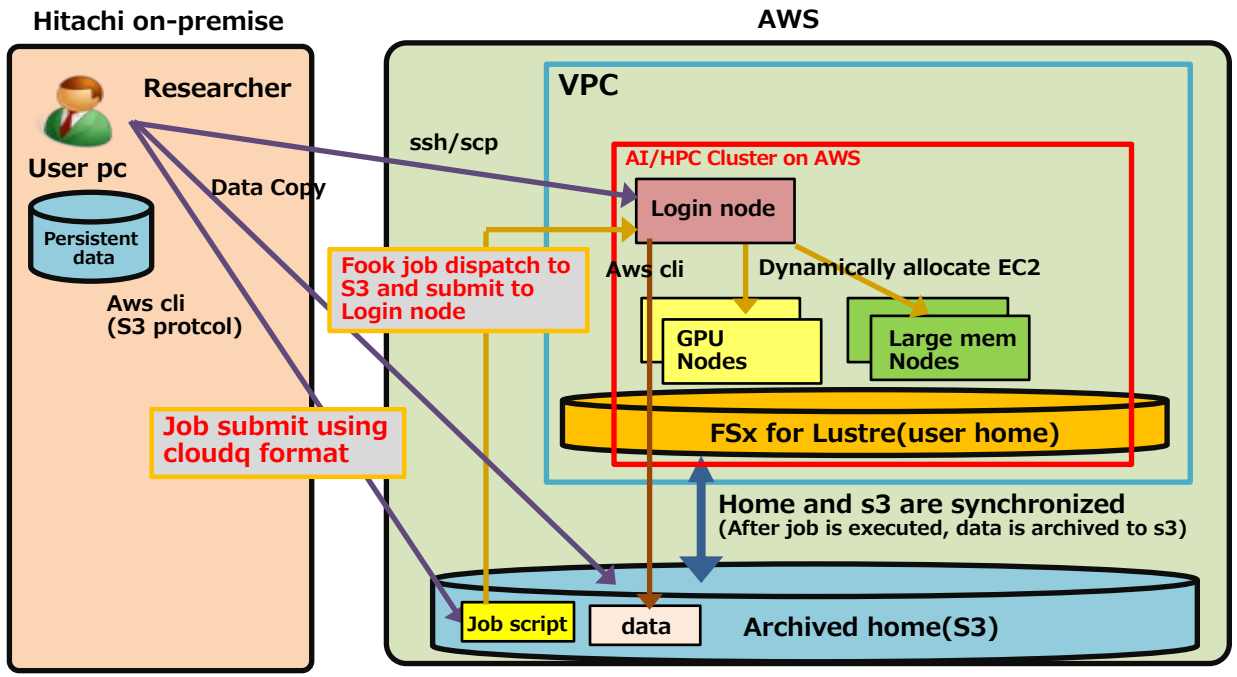
- **（常駐）サービス**のデプロイ
  例えば、MLOps
- HPCとクラウドの最適な使い分け

- エッジ・クラウドとの連携を考えると、**ユーザ管理・データ共有モデルの再考**が必要ではないないか。
  - SSHの権限強すぎ問題
  - いつまで**UNIXベースのセキュリティモデル、ファイルシステム**に縛られるのか？

ABCIとクラウド連携

日立/産総研 CloudQ

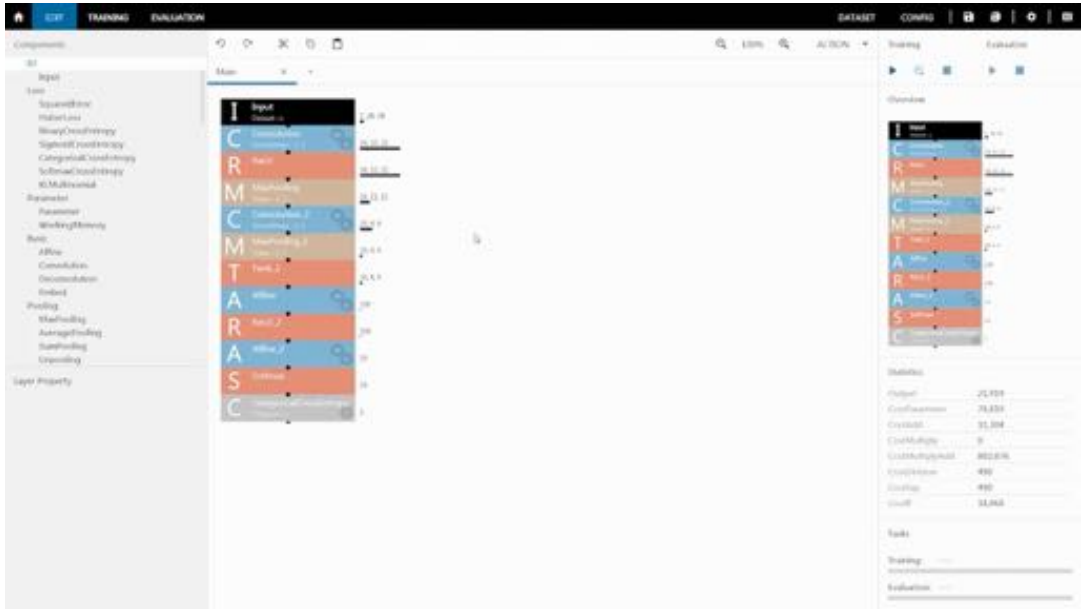
- クラウドストレージをジョブサブミッション機構に使用したクラウドバースティングの実現
- クラウドとしてABCIとAWSに対応



Courtesy of M.Shimizu, Hitachi
※詳細はSWoPP2022にて発表予定

SONY Neural Network Console (NNC)

- GUI操作で深層学習のプログラミングが可能
- バックエンドとしてABCIに対応



https://dl.sony.com/ja/cloud/abci/index.html

運用面の課題

- 原則9:00-17:00のユーザサポート
- 1システム1ロケーションだと可用性に限界（cf.マルチリージョン）
- 前払式ポイント制度
  - 年度繰越できない
  - 資金決済法改正対応
- 非居住者の利用に輸出管理申請が必要
- 計画的なシステム更新
  - 黒字運用して更新費用積立が困難

  →補助金適正化法（収益納付ルール）等の縛り

運用の外部委託化、
例えば、PPP (Public Private Partnership) /PFI (Private Finance Initiative) *等の検討も必要では

そこまでやり遂げて
「橋渡し」の完了か！？

*)公共サービスの提供に際して公共施設が必要な場合に、従来のように公共が直接施設を整備せずに民間資金を利用して民間に施設整備と公共サービスの提供をゆだねる手法

参考：ABCI利用者からの要望（１/３）

1. **ポイントの利用期限**
   - ポイントを年度をまたいで使えると良いですね。先払いしたポイントを年度内に消費しなければいけないのはちょっとプレッシャーです。（F社）
   - ポイントが基本的には年度末で持ち越しできない仕組みになっている。民間会社の場合は適切に会計処理を行えば、特に問題なく繰り越しでも自社の予算に関しては使えるという側面もある。ABCIも制度を改善してポイントを繰り越せるとありがたい。（O社）
   - 毎年3月末でポイントが消失するのはやはりおかしいと感じる、またその〆切が1月末で2ヶ月以上前なのも予算管理上非常に厳しい。思いとして、当社はできる限りABCIをメインで利用したいので改善していただきたい。（H社）
2. **可用性の向上**
   - （イベント等による）ダウンタイムがなくてずっと使えると嬉しいですね。（I社）
   - 年度末になってしまうと、少し圧迫される資源も多くなってくるのでその解消をしてほしい。（P社）
   - (実際に利用している研究者からは)年度末３月にかけて繋がりにくいとか、ジョブが集中するのでキャンセルされるとの話を聞くが改善いただければありがたい。（O社）
   - 小規模ノード実行ジョブが大量にあり大規模ジョブ (256～512ノード) がなかなか入らないのがなんとかなるとうれしいです。

参考：ABCI利用者からの要望（2/3）

3. 利用サービスの拡充

- AIを使うためのツールはいろいろありますが、(ABCIでも)それがもっと充実すると利用のハードルはさらに下がります。（P社）
- 自社のIoTの実験環境が5Gを介してABCIと簡単に繋がるようになると良い。（H社）
- 自然言語処理、AI OCRに関しても、日本人が作ったAIモデルの方が日本語の認識精度が優れる事が多いと感じています。日本の中でも産総研のような公的機関でもデータセットやツールを整備して、公開していただけるとありがたい。（R社）
- 各ノードでの大容量データ処理を実現するために、VPN経由でover 10Gbpsでのリアルタイム入出力機能のサポートをお願いします。

4. 利用者サポートの充実

- 「DockerユーザのためのSingularity講座」のようなセミナーを企画してもらえればもっと普及が進むのかもしれない。（K社）
- 産総研の方々の中でAIに関するセミナーとかいろいろ開催されてるかと思いますが、いろんな企業・大学などと交流できるのは貴重ですので、今後も続けていただきたい。（P社）
- 講習会などに加えてオンライン実習など含めてサポートを強化していただくと、もっと間口が広がっていくと思う。（H社）
- 新型コロナ感染症に対する研究への無償提供と同じように、自然災害の対策に向けた防災・減災の研究に対しても公募などで計算資源のご支援をいただけると大変助かります。（I社）

参考：ABCI利用者からの要望（3 / 3）

5. スタートアップ支援

- スタートアップや、国内の大学研究者などがGAFAと同じような成果を出していくためには、公的研究機関がABCIのような計算基盤を整備して、安価に提供しているのは非常にありがたい。ABCIにも投資を継続して、国内のAI開発を支え続けていただきたい。（L社）
- ベンチャー企業向けには安く提供してもらいたい。（T社）
- スタートアップ向けにはコスト的なところでも利用者が拡張されるような取組があると良い。例えば産総研主導でベンチャー向けのコンペティションを行うなど。（A社）

6. その他

- さらに大きなGPUメモリがあれば良い。（L社）
- データのステージング時間が短縮されるとうれしいです。
- ABCIの活用は、機密データやソースコードの流出リスクが懸念されるため、論文や公開データセットの追試のための利用に留まっています。データ暗号化を組み合わせたABCIの活用方法等の情報がありましたらWebページ等でご紹介いただけると、企業側のユーザとしては安心して利用できる環境になるのでは、と考えております。
- 限られた人しか使えないもののようなイメージがありましたので、もっと広く使えることをアピールしてもらえると良い。（T社）

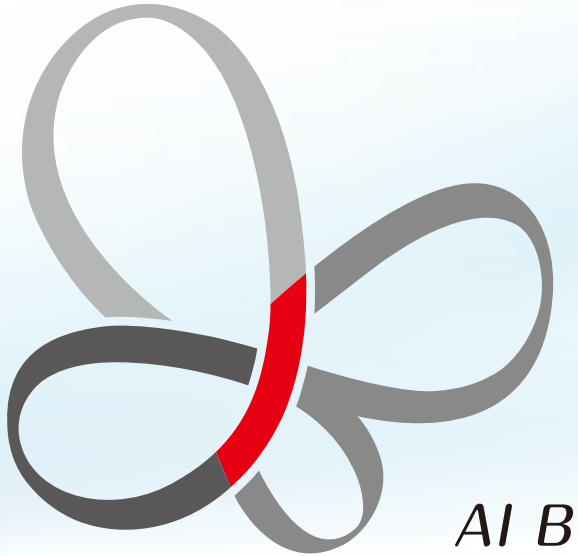
まとめ

- ABCIは2018年の運用開始以来、AIに対する旺盛な研究開発ニーズの後押しもあり、順調にユーザベースを伸ばし、成果を上げてきた
- より使いやすいインフラを提供するために、クラウド技術のさらなる導入が必要。技術的にも運用的にもまだまだ課題が多い
  - 今後のエッジ・クラウドとの連携を考えると、ユーザ管理・データ共有モデルを再考する時期に来ているのではないか。
  - 運用の外部委託化

ご清聴ありがとうございました

明日のパネルに続く（？）

AIST



ABCI

AI Bridging Cloud Infrastructure

https://abci.ai/