AIST

ABCI

# 大規模AIクラウド計算システム「ABCI」

2020年9月9日

国立研究開発法人 産業技術総合研究所
情報・人間工学領域
ディジタルアーキテクチャ推進センター

# 世界最大級・超省電力・オープンＡＩインフラストラクチャ

東京大学 THE UNIVERSITY OF TOKYO　AIST

Univ. Tokyo / Kashiwa II Campus

ABCI　AI Bridging Cloud Infrastructure

- 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共同による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人工知能技術の研究開発・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

**理論ピーク性能**
- 550 PFLOPS　(半精度)
- 37 PFLOPS　(倍精度)

**実効性能**
- TOP500 (HPL):　19.88 PFLOPS
- HPCG　:　508.85 TFLOPS
- GREEN500　:　14.423 GFLOPS/W

**使用電力 (最大)**　:　2.3 MW

**年間平均PUE**　:　1.1以下 (推定値)

**2018年8月1日運用開始**

# 世界トップクラスの実運用される省電力クラウド型計算システム

Top500 (June 2019) Linpack性能

| Rank | System | Rmax |
|---|---|---|
| 1 | Summit | 148.6PF |
| 2 | Sierra | 94.6PF |
| 3 | Sunway TaihuLight | 93.0PF |
| 4 | Tianhe-2A | 61.4PF |
| 5 | Frontera | 23.5PF |
| 6 | Piz Daint | 21.2PF |
| 7 | Trinity | 20.2PF |
| 8 | ABCI | 19.9PF |
| 9 | SuperMUC-NG | 19.5PF |
| 10 | Lassen | 18.2PF |

HPCG (June 2019) 共役勾配法性能

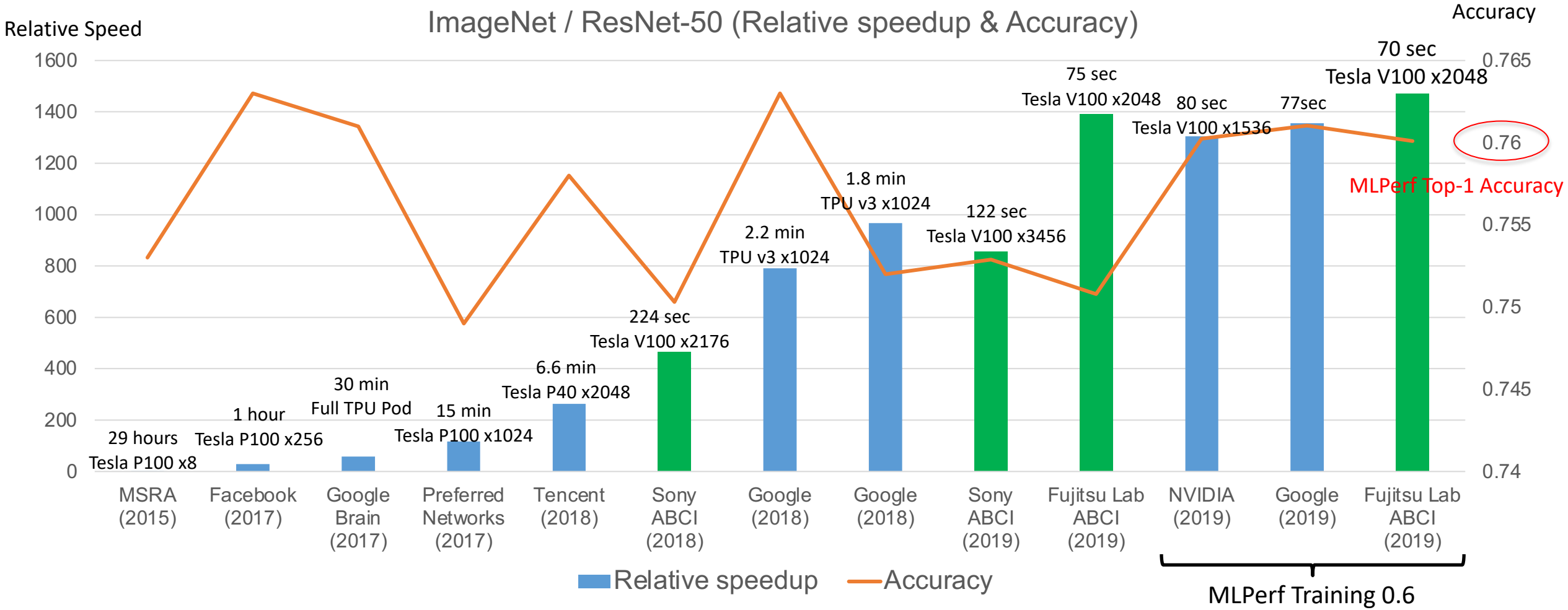
| Rank | System | HPCG |
|---|---|---|
| 1 | Summit | 2.9PF |
| 2 | Sierra | 1.8PF |
| 3 | K computer | 603TF |
| 4 | Trinity | 546TF |
| 5 | ABCI | 509TF |
| 6 | Piz Daint | 497TF |
| 7 | Sunway TaihuLight | 481TF |
| 8 | Nurion | 391TF |
| 9 | Oakforest-PACS | 385TF |
| 10 | Cori | 355TF |

Green500 (June 2019) 省エネ性能

| Rank | System | GF/W |
|---|---|---|
| 1 | DGX SaturnV Volta | 15.1 |
| 2 | Summit | 14.7 |
| 3 | ABCI | 14.4 |
| 4 | MareNostrum P9 CTE | 14.1 |
| 5 | TSUBAME3.0 | 13.7 |
| 6 | PANGEA III | 13.1 |
| 7 | Sierra | 12.7 |
| 8 | PreE | 11.4 |
| 9 | Taiwania 2 | 11.3 |
| 10 | Huawei G5500 | 10.8 |

**３指標でトップ１０に入るのは現在ABCI、米国DOEのSummit、Sierraのみ**
**実性能と省エネ性能を世界トップレベルで両立**
**公的かつオープンで、AI技術開発に特化した計算インフラとして世界初の先進的システム**

# ディープラーニングの学習速度の世界記録を大幅に更新




**今年７月、富士通研究所がABCI上で記録した70秒が現在の世界記録**
**ABCIにより、Google、Tencentなど先行する技術開発への対抗が可能に**

# ＡＢＣＩのハードウェア構成

## 高性能計算システム

550 PFlops (FP16), 37.2 PFlops (FP64)
476 TiB メモリ, 1.74 PB NVMe SSD

### 計算ノード (GPU搭載) x 1,088台

| | |
|---|---|
| GPU | NVIDIA Tesla V100 SXM2 x 4 (計 4,352基) |
| CPU | Intel Xeon Gold 6148 x 2 (計 2,176基) |
| Memory | 384GiB |
| Local Storage | Intel SSD DC P4600 (NVMe) 1.6TB x 1 |
| Interconnect | InfiniBand EDR x 2 |

**マルチプラットフォームノード (GPU非搭載) x 10台**
- Intel Xeon Gold 6132 (2.6GHz/14cores) x 2
- 768GiB Memory, 3.8TB NVMe SSD

インタラクティブノード x 4　　管理サーバ・ゲートウェイノード x 15

## 大容量ストレージシステム

22 PB GPFS + 1 PB フラッシュストレージ +
17 PB クラウドストレージ

**DDN SFA14K (w/ SS8462 Enclosure x 10) x 3**
- 12TB 7.2Krpm NL-SAS HDD x 2400
- 3.84TB SAS SSD x 216
- NSD Server x 12

**DDN S**
- 7.68T
- MDS S

**HPE A**
- 12TB
- 3.2TB

## 計算ネットワーク (Infiniband EDR)

- Mellanox CS7500 x 2
- Mellanox SB7890 x 229

## サービスネットワーク (10GbE)

- FortiAnalyzer 400E x1

SINET5

### ■ポイント1: 世界最大級の最新GPUファーム

全部で4352基。誰でも2048基までの最新GPUを使って大規模・高速なディープラーニングを実行可能。

### ■ポイント2: スパコン由来のネットワーク技術の活用

すべてのGPUとCPUが2.5μ秒以下の低遅延、200Gbpsの広帯域で通信可能。
(商用クラウドでは数10倍の遅延、1/2以下の帯域)

# 従来スパコンにはないABCIの運用上の特徴

## ■ オープンなクラウドサービス

- 産総研のみならず、国内の企業、大学、研究機関に所属している方なら誰でも利用申請して、均一料金で利用可
- インターネットを介してどこからでもアクセス可能
- 利用に必要な情報はすべてインターネットに公開

## ■ セキュアなクラウドサービス

- セキュリティ管理体制、セキュリティ実装、脆弱性管理など、セキュリティ要件を策定
- 商用クラウドベンダーと同様にセキュリティ要件をホワイトペーパーとして公開し、ユーザ企業による導入検討を支援

## ■ 先進的でマネージドなクラウドサービス

- 最新のOS、デバイスドライバ、ミドルウェア、開発環境等の計画的導入により、up-to-dateな利用環境をユーザに提供（公的なスパコンとしては稀）
- コンテナミドルウェアをサポートすることにより、ユーザが社内システムや商用クラウドで開発したソフトウェアをそのままABCIに移行可能

# ＡＩインフラストラクチャ for everyone

ImageNet-1k 学習速度の世界最速記録

## Expert

- ABCIグランドチャレンジ：学術的・産業的に画期的な成果が見込まれる最重要課題へのABCIの全システムを最大24時間提供
- オープンデータを用いたオープンサイエンスを推進、コンテスト開催

ＡＩ囲碁

リアルタイム動画解析(産総研開発)

## Advanced & Intermediate

- 最大512ノード/2048GPUまでの計算リソースを誰でも利用可
- すぐ使えるソフトウェア、データセット、学習済みモデル等を提供
- 先端技術者を招いて上級者向けの講習会・ハッカソンの開催

ABCI × SONY Neural Network Console

## Beginner

- ABCI-WebUI実証プラットフォームを提供し、初学者にも使いやすい統合開発環境の実現を支援
- 初級者向けの利用者講習会の開催

データセンタ事業者

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」**

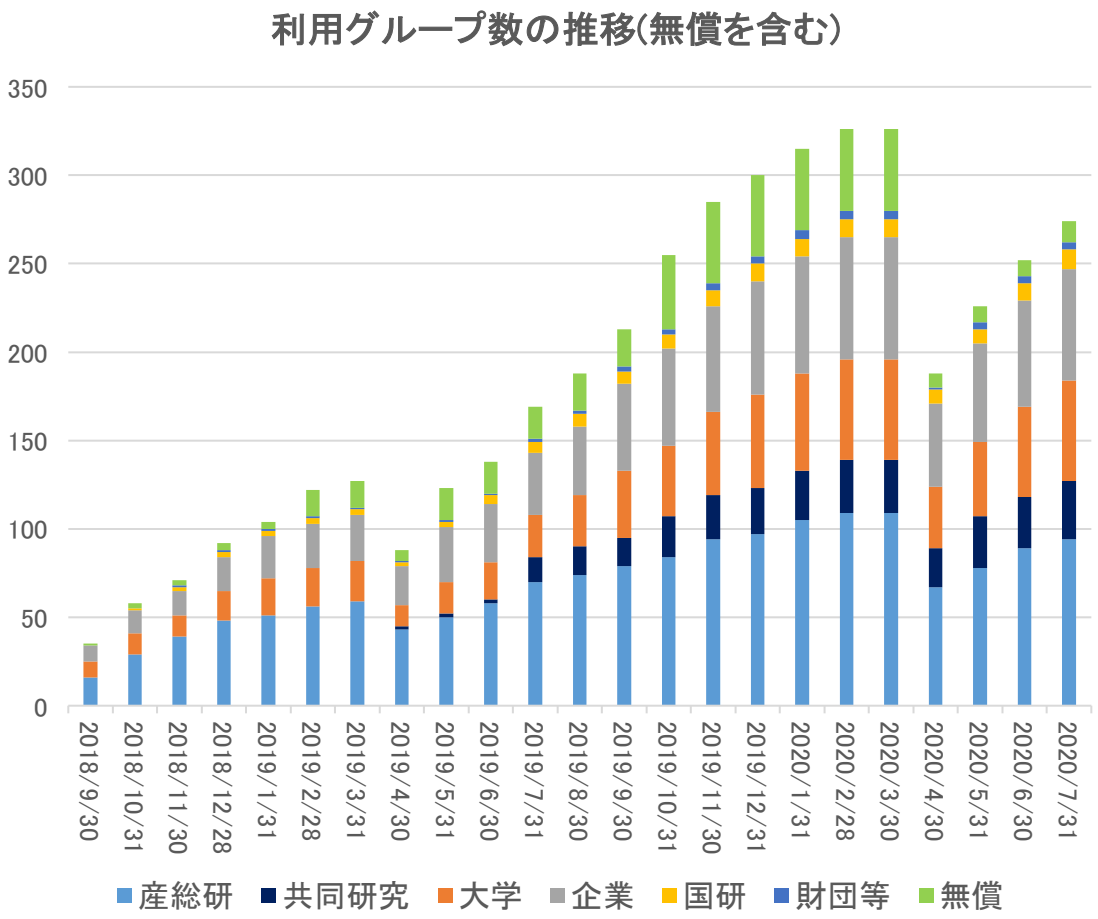
**人工知能産業エコシステムの形成**

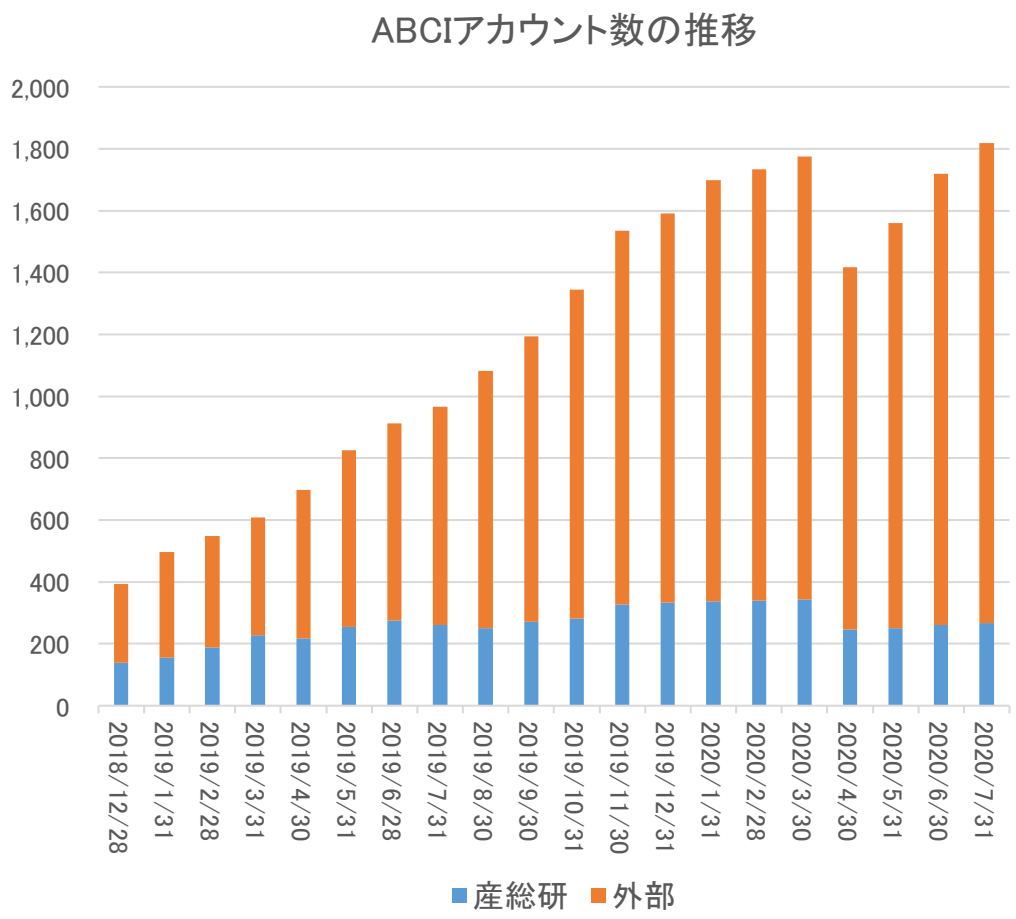
最先端のAI研究から
誰でも試して使えるAIまで

2020年3月時点で326グループ(うち60%以上が外部利用)、1,774ユーザ(うち80%が外部利用)

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

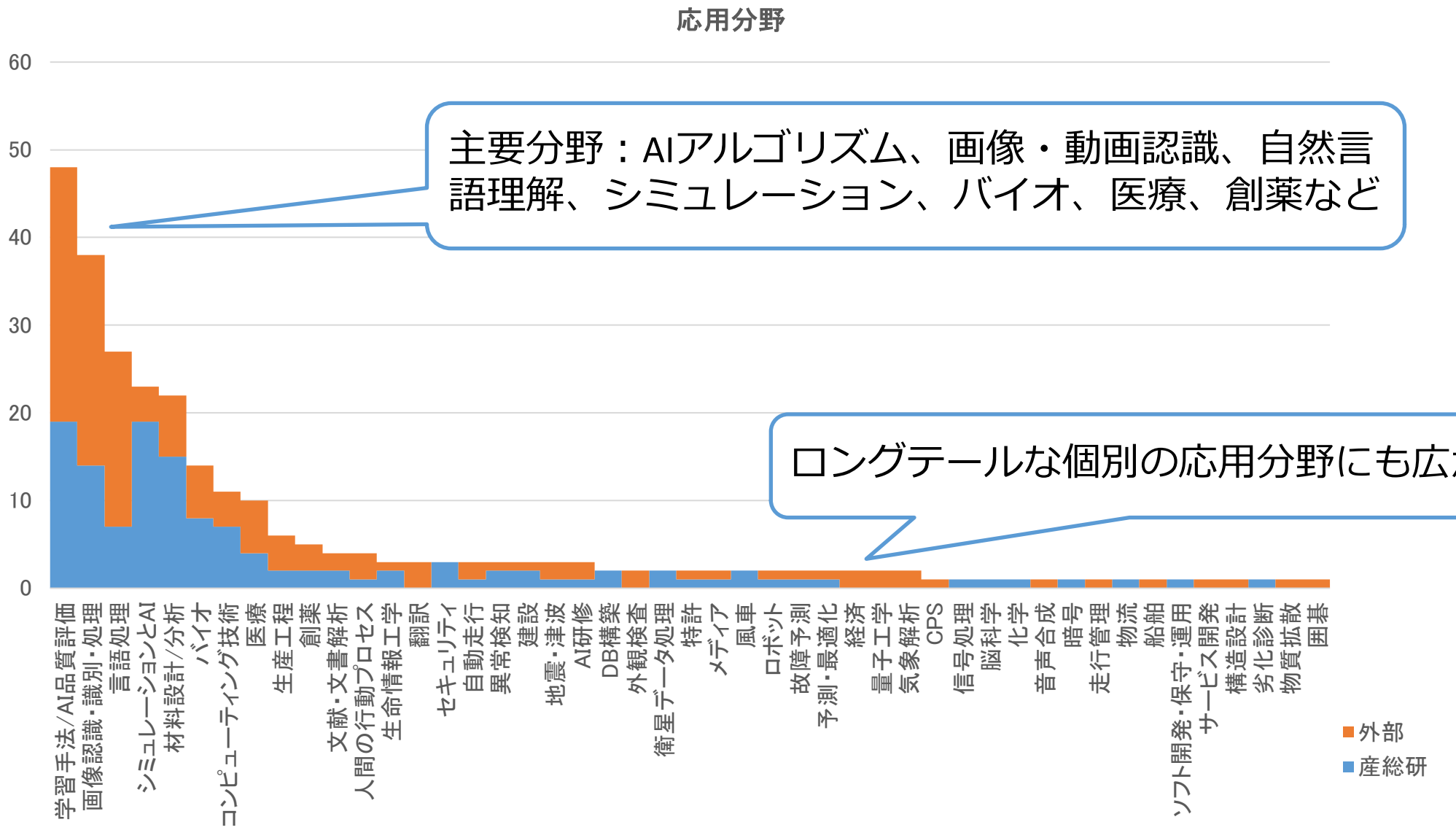
# ABCI利用者統計




「無償」とは、ABCIグランドチャレンジ等での利用をいう。

# 応用分野
## 2020/2現在




主要分野：AIアルゴリズム、画像・動画認識、自然言語理解、シミュレーション、バイオ、医療、創薬など

ロングテールな個別の応用分野にも広がり

# ABCIサービス諸元

AIST

# 1. 計算ノードの利用
## ジョブ割当リソース

| 資源タイプ | 資源タイプ名 | 説明 | 割り当て物理CPUコア数 | 割り当てGPU数 | メモリ(GiB) | ローカルストレージ(GB) | 資源タイプ課金係数 | 割当可能な個数 |
|---|---|---|---|---|---|---|---|---|
| Full | rt_F | ノード占有 | 40 | 4 | 360 | 1440 | 1.00 | 1-512 (1-32)* |
| G.large | rt_G.large | ノード共有 GPU利用 | 20 | 4 | 240 | 720 | 0.90 | 1 |
| G.small | rt_G.small | ノード共有 GPU利用 | 5 | 1 | 60 | 180 | 0.30 | 1 |
| C.large | rt_C.large | ノード共有 CPUのみ利用 | 20 | 0 | 120 | 720 | 0.60 | 1 |
| C.small | rt_C.small | ノード共有 CPUのみ利用 | 5 | 0 | 30 | 180 | 0.20 | 1 |
| M.large | rt_M.large | ノード共有 CPUのみ利用 | 8 | 0 | 800 | 480 | 0.40 | 1 |
| M.small | rt_M.small | ノード共有 CPUのみ利用 | 4 | 0 | 400 | 240 | 0.20 | 1 |

New!

＊バッチジョブは最大512個、インタラクティブジョブは最大32個

資源タイプ名と数量で資源量を指定
例） 資源タイプ(Full)を1個指定： -l rt_F=1

AIST

# 2. 計算リソース
## 制限事項

hour:minute:second
時間：分：秒

| | インタラクティブジョブ（上限/デフォルト） | バッチジョブ（上限/デフォルト） | 予約 |
|---|---|---|---|
| rt_F | 12:00:00/1:00:00 | 72:00:00/1:00:00 | 1日単位<br>最大30日間 |
| rt_G.large<br>rt_C.large | | | NA |
| New! rt_M.large<br>rt_M.small | | | |
| rt_G.small<br>rt_C.small | | 168:00:00/1:00:00 | |
| ノード時間積 | 12 | 2,304 | |

- 72時間（3日間）以上必要な学習は、「予約」を利用する。
- 小さなジョブは、一度に多数投入できる（分散処理向き）。

AIST

# 3. 計算ノードの予約利用

| 項目 | 説明 |
|---|---|
| 最小予約日数 | 1日 |
| 最大予約日数 | 30日 |
| システムあたりの最大同時予約可能ノード数 | 442ノード |
| 1予約あたりの最大予約ノード数 | 32ノード(1ジョブで実行可能な最大値) |
| 1予約あたりの最大予約ノード時間積 | 12,288ノード時間積 |
| 予約受付開始時刻 | 30日前の午前10時 |
| 予約受付締切時刻 | 予約開始前日の午後9時 |
| 予約取消受付期間 | 予約開始前日の午後9時 |
| 予約開始時刻 | 予約開始日の午前10時 |
| 予約終了時刻 | 予約終了日の午前9時30分 |

予約ノードを利用する際は、qsubコマンド又はqrshコマンドにて、「-ar ar_id」オプションを使う。

AIST

# 4. ABCI料金表：　2019/2020年度

| | | インタラクティブジョブ | バッチジョブ | 予約 |
|---|---|---|---|---|
| 計算資源 | rt_F | 1.0 ポイント/時間 | | 36ポイント/日<br>（1.5ポイント/時間） |
| | rt_G.large | 0.9ポイント/時間 | | NA |
| | rt_C.large | 0.6ポイント/時間 | | |
| | rt_G.small | 0.3ポイント/時間 | | |
| | rt_C.small | 0.2ポイント/時間 | | |
| New! | rt_M.large | 0.4ポイント/時間 | | |
| | rt_M.small | 0.2ポイント/時間 | | |
| ストレージ | グループ領域<br>（共有ディスク） | 5 ポイント/TB・月<br>（ABCIアカウント毎に割当られる200GBのホーム領域は無償） | | |
| New! | ABCIクラウドストレージ | 0.0001ポイント/GB・日<br>（従量制） | | |

# 5. ジョブ実行上の特長
## 大規模演算を必要とする場合、有効

- バッチジョブ(Spot)では、**最大５１２ノード**までを利用可能
- インタラクティブジョブ(On-demand)でも、**最大３２ノード**まで利用可能
- ジョブ投入数（ユーザあたり）

  1）アレイジョブあたりの最大投入可能タスク数：　**75,000**

  2）同時投入可能ジョブ数：　**1,000**

  3）同時実行ジョブ数：　**200**

# グランドチャレンジ 採択課題2019年度

| | クラス | 所属機関名 | 代表者名 | 課題名 | 研究の概要 |
|---|---|---|---|---|---|
| 第3回 | Large | 富士通研究所 | 大石裕介 | 想定外を想定する津波即時予測 AI の構築 | 2019-3-L-1 |
| 第3回 | Large | 東京工業大学 | 横田理央 | プラズマの挙動予測のための再帰型ニューラルネットワークの大規模並列深層学習 | 2019-3-L-2 |
| 第3回 | Large | 東京工業大学 | 横田理央 | 巨大な言語モデルの分散並列深層学習 | 2019-3-L-3 |
| 第2回 | Large | 東京工業大学 | 横田理央 | KFACの計算時間をSGDとほぼ同等に抑える事によるImageNet学習のさらなる記録更新 | 2019-2-L-1 |
| 第2回 | Large | ソニー | 影山雄一 | ABCIアーキテクチャに最適な分散深層学習向けデータ同期通信手法、および、大きなモデルの学習におけるパフォーマンス課題の明確化 | 2019-2-L-2 |
| 第2回 | Medium | 東京工業大学 | 秋山泰 | 超高速タンパク質間相互作用予測システムMEGADOCK 5.0による細胞内タンパク質間相互作用の網羅的解明 | 2019-2-M-1 |
| 第1回 | Large | 富士通研究所 | 田原司睦 | 大規模ミニバッチによる分散並列深層学習 | 2019-1-L-1 |
| 第1回 | Large | ソニー | 影山雄一 | ABCIアーキテクチャに最適な分散深層学習向けデータ同期通信手法 | 2019-1-L-2 |
| 第1回 | Large | 産総研 | 陳鵬 | A Scalable Framework for Instant High-resolution Image Reconstruction | 2019-1-L-3 |
| 第1回 | Medium | 東京工業大学 | 横田理央 | KFACのChainerX実装を用いたImageNetの１分以内の学習 | 2019-1-M-1 |

# 成果発表（グランドチャレンジ ）

論文

- Kates-Harbeck, J., Svyatkovskiy, A., Tang, W., “Predicting disruptive instabilities in controlled fusion plasmas through deep learning”, Nature (2019 Apr).

国際会議・学会発表

- Osawa, K., Swaroop, S., Jain, A., Eschenhagen,R., Turner, R. E., Yokota, R., Khan, M. E., “Practical Deep Learning with Bayesian Principles”, The 33rd Conference on Neural Information Processing Systems (NeurIPS 2019).
- M. Yamazaki, “Yet Another Accelerated SGD: ResNet-50 Training in 70.4 sec.”, 8th ADAC Workshop, Oct. 2019.
- T. Narihara, H. Suganuma, “Sony's deep learning development environment for massively large scale training”, 8th ADAC Workshop, Oct. 2019.
- Y. Tsuji, K. Osawa, Y. Ueno, A. Naruse, R. Yokota, and S. Matsuoka, “Performance Optimizations and Analysis of Distributed Deep Learning with Approximated Second-Order Optimization Method”, The 1st Workshop on Parallel and Distributed Machine Learning 2019 (PDML19).
- Y. Ueno and R. Yokota, “Hierarchical Topology-aware Communication for Scaling Deep Learning to Thousands of GPUs”, The 19th Annual IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing (CCGrid 2019).
- K. Osawa, Y. Tsuji, Y. Ueno, A. Naruse, T. Yokota, and S. Matsuoka, “Large-scale Distributed Second-order Optimization Using Kronecker-factored Approximate Curvature for Deep Convolutional Neural Networks”, Conference on Computer Vision and Pattern Recognition (CVPR 2019).
- Yokota, R. “Kronecker Factorization for Second Order Optimization in Deep Learning”, SIAM CSE, Spokane, USA, February 25-March 3 (2019).
- Massively Distributed SGD: ImageNet/ResNet-50 Training in a Flash　(arXiv:1811.05233) [v2] Tue, 5 Mar 2019.
- ImageNet/ResNet-50 Training in 224 Seconds　(arXiv:1811.05233) [v1] Tue, 13 Nov 2018.

# 成果発表（一般）

国際会議・学会発表

- Kazuki Uehara, “MULTI-SCALE EXPLAINABLE FEATURE LEARNING FOR PATHOLOGICAL IMAGE ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS”, 27th IEEE International Conference on Image Processing (ICIP2020), Oct. 2020.
- Dario Bertero, Takeshi Homma, Kenichi Yokote, Makoto Iwayama, Kenji Nagamatsu, “Model Ensembling of ESIM and BERT for Dialogue Response Selection”, The Eighth Dialog System Technology Challenge (DSTC8) Workshop, Aug. 2020.
- S. Takizawa, “ABCI’s Scheduling Design for Accommodating Various AI Jobs", 8th ADAC Workshop, Oct. 2019.
- Kohei Ozaki and Shuhei Yokoo, “1st place in Retrieval Challenge” and "3rd place in Recognition Challenge", CVPR 2019 Second Landmark Recognition Workshop, 16 June 2019.
- Kohei Ozaki and Shuhei Yokoo, "Large-scale Landmark Retrieval/Recognition under a Noisy and Diverse Dataset" (arXiv:1906.04087) [v1] Mon, 10 June 2019.
- Peng Chen, Mohamed Wahib, Shinichiro Takizawa, Ryousei Takano, Satoshi Matsuoka, "iFDK: A Scalable Framework for Instant High-resolution Image Reconstruction", SC19, 17-22 November 2019.
- Peng Chen, Mohamed Wahib, Shinichiro Takizawa, Ryousei Takano, Satoshi Matsuoka, "A Versatile Software Systolic Execution Model for GPU Memory Bound Kernels", SC19, 17-22 November 2019.
- Atsuro Okazawa, Tomoyuki Takahata, Tatsuya Harada, “Simultaneous transparent and non-transparent objects segmentation with multispectral scenes", 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 4-8 November 2019.

AIST

# ABCI

AI Bridging Cloud Infrastructure

https://abci.ai/