



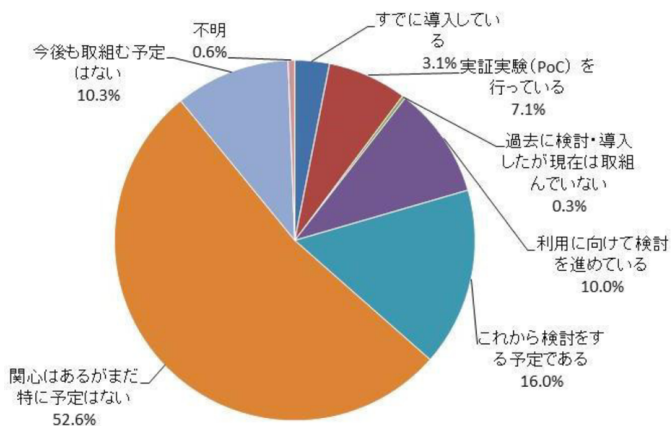
大規模AIクラウド計算システム「ABCI」の構築と 産業界への利用普及に向けた取り組み

小川 宏高

国立研究開発法人 産業技術総合研究所
人工知能研究センター 人工知能クラウド研究チーム長／
実社会ビッグデータ活用オープンイノベーションラボラトリ 副ラボ長

課題：AI技術開発・導入におけるにわとりたまご問題

AIへの関心と導入状況：「既に導入している」はまだ3.1%。
「実証試験を行っている」7.1%を合わせて1割程度。



(出典) 情報処理推進機構「AI社会実装推進調査報告書」(2018年6月19日)より作成

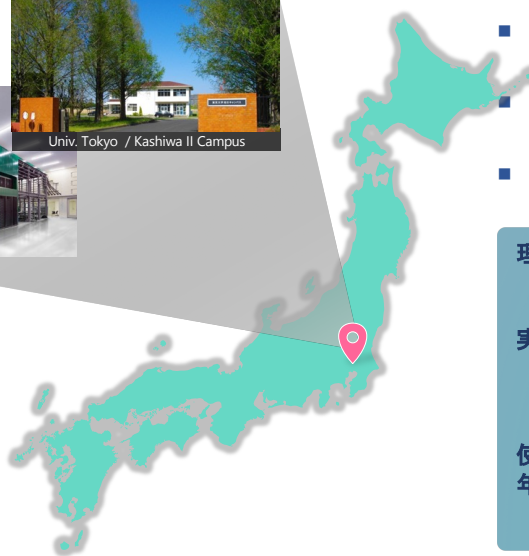


↓ 様々なレベルで問題

- ・ (目的) 基礎研究：
計算機パワーや大量データに支えられて先行してきた米中の技術開発に対抗できない
- ・ (橋渡し) 応用研究：
培ってきた要素技術はあるが、「試す場」がなく、実社会規模の問題にスケールできない
- ・ より一般には参入障壁が高く、そもそも試したことがない

世界最大級・超省電力・オープン A I インフラストラクチャ

東京大学 AIST



ABC I 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A I 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、A I 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

19.88 PFLOPS (TOP500世界5位, 国内1位)

14.423 GFLOPS/W (GREEN500世界4位, 国内2位)

508.85 TFLOPS (HPCG世界5位, 国内2位)

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

A I インフラストラクチャ for everyone

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 **最重要課題への挑戦**にABC Iの全システムを最大24時間提供
- ・ ABCIデータチャレンジ：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

Advanced & Intermediate



- ・ 最大512ノード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
- ・ 高い計算能力を活用した **A I 技術の研究開発・実証を加速、社会実装を推進**

Beginner

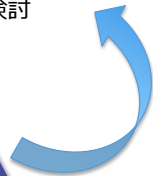


- ・ ディープラーニングの初学者にも使いやすい WebUIベースの開発環境を開発・提供
- ・ B2B2Cの実証プラットフォームとしても利用可



データセンタ事業者

- ・ ABCIをリファレンスアーキテクチャとして事業検討



「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



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

世界トップクラスの実運用される省電クラウド型計算システム

Top500 (Nov 2018) Linpack性能

Rank	System	Rmax
1	Summit (米)	143.5PF
2	Sierra (米)	94.6PF
3	Sunway TaihuLight (中)	93.0PF
4	Tianhe-2A (中)	61.4PF
5	Piz Daint (瑞)	21.2PF
6	Trinity (米)	20.2PF
7	ABCI (日)	19.9PF
8	SuperMUC-NG (独)	19.5PF
9	Titan (米)	17.6PF
10	Sequoia (米)	17.2PF

HPCG (Nov 2018) 共役勾配法性能

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

Green500 (Nov 2018) 省エネ性能

Rank	System	GF/W
1	Shoubu system B (日)	17.6
2	DGX SaturnV Volta (米)	15.1
3	Summit (米)	14.7
4	ABCI (日)	14.4
5	TSUBAME3.0 (日)	13.7
6	Sierra (米)	12.7
7	AIST AI Cloud (日)	12.7
8	MareNostrum P9 (西)	11.9
9	PreE (中)	11.4
10	Taiwania 2 (台)	11.3

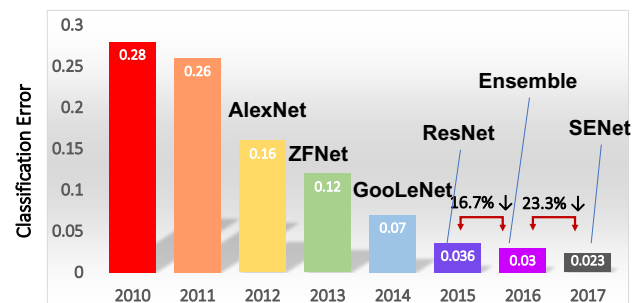


3指標でトップ10に入るのは現在ABCI、米国DOEのSummit、Sierraのみ
実性能と省エネ性能を世界トップレベルで両立
公的かつオープンで、AI技術開発に特化した計算インフラとして世界初の先進的システム

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

- ImageNetデータセット
 - WordNetで定義されたクラスごとにデータを大量に収集、アノテーション
 - 学習データ120万枚、テストデータ10万枚
- ILSVRCではクラス分類(CLS)、位置推定(LOC)、物体認識(DET)名等のタスクでアルゴリズムの優劣を競う
 - 特にクラス分類(CLS)の精度がディープラーニングの進化の指標といわれる
- 2012年以降、ディープラーニングベースの手法が主流となる
 - 2015年のMSRAのResNetは人間の認識能力 (Top-5 errorで5.1%) を超える「ブレイクスルー」
- 以降はHW/SW一体のHPC技術、最適化技術を駆使した学習速度競争の時代に突入

CLS (Top-5 error)

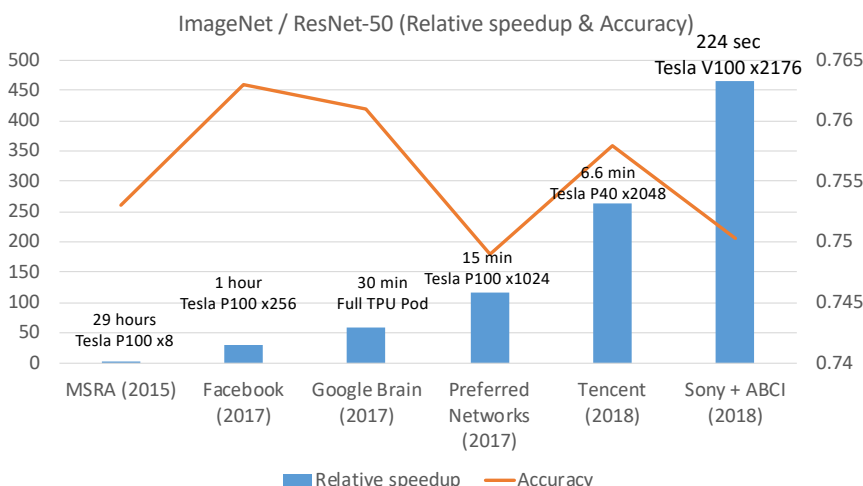


http://imagenet.org/challenges/talks_2017/ILSVRC2017_overview.pdf

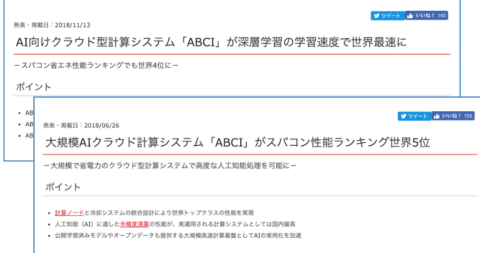


ディープラーニングの学習速度の世界記録を大幅に更新

特筆すべき成果



【プレスリリース】×2 (+ソニー1、富士通1)



【報道等（広報DBによる）】

産総研 スパコン、深層学習で世界最速, 日本経済新聞(WEB), 2018.11.15 をはじめ97件、うちH30年度は91件。

【H30年度の研究成果】

国際会議における招待講演2件、口頭発表7件、その他国際会議論文を複数準備中

その後Googleによってさらに記録を塗り替えられたものの、ABCIにより、Google、Tencentなど先行する技術開発への対抗が可能に

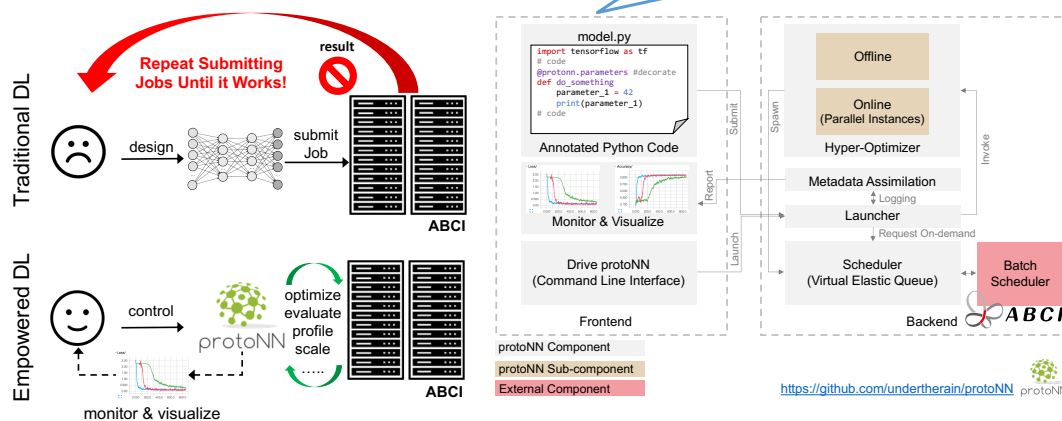


protoNN: Scalable Optimization-driven Prototyping of DDNs for HPC system

特筆すべき成果

- Time for inference = O(seconds)
- Time for training = O(hours or days)
- Time for designing = O(weeks or months)

膨大な計算パワーによりニューラルネットワークのプロトタイピングを支援



<https://github.com/undertherain/protoNN>

ABCIシステムの概要

ABCIの設計・構築

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



大容量ストレージシステム

22 PB GPFS



計算ノード (GPU搭載) x 1088台

- GPU** NVIDIA Tesla V100 SXM2 x 4
- CPU** Intel Xeon Gold 6148 (2.4GHz/20cores) x 2
- 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

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

プロトコルノード等 x 8



2018年8月運用開始

計算ネットワーク (Infiniband EDR)

- Mellanox CS7500 x 2
- Mellanox SB7890 x 229

ファイアウォール等

- Nexus 3232C x2
- FortiGate 1500D x2
- FortiAnalyzer 400E x1

サービスネットワーク (10GbE)

100Gbps

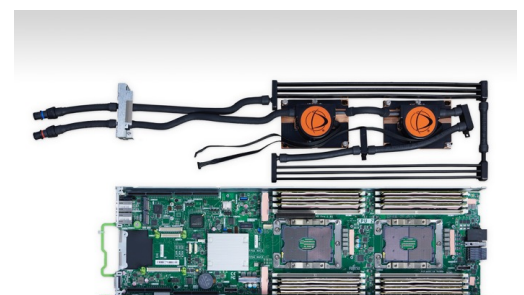
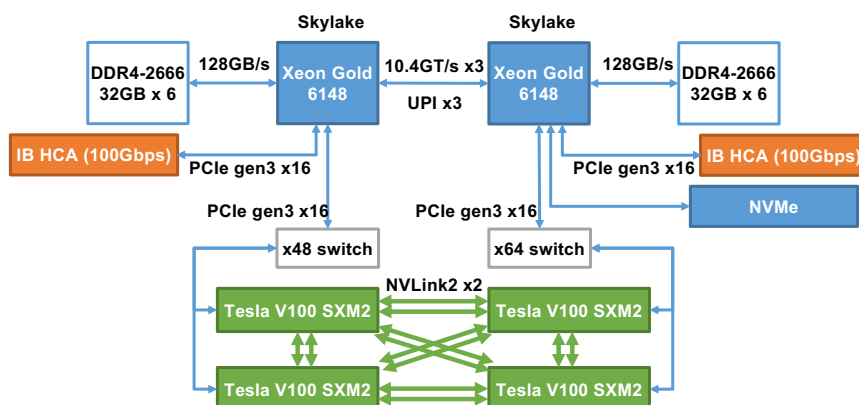
SINET5



ABCI計算ノード

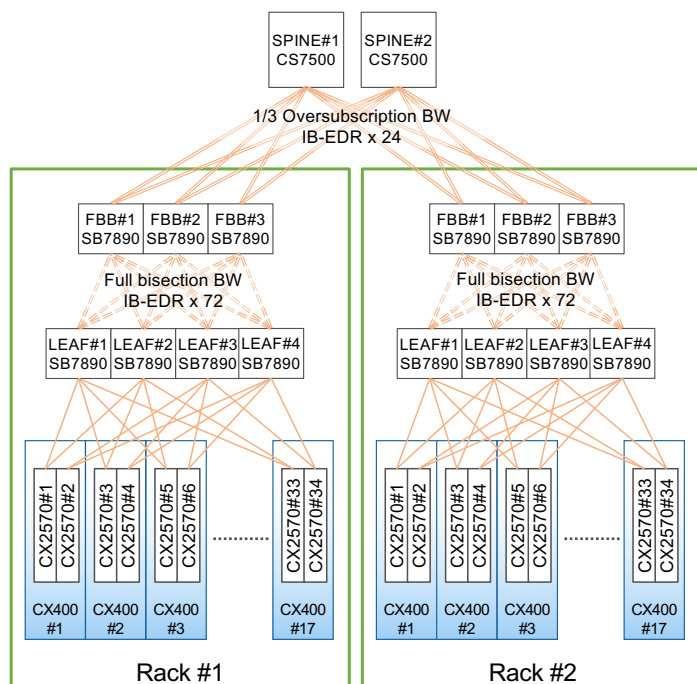
FUJITSU PRIMERGY Server (2 servers in 2U)

CPU	Xeon Gold 6148 (27.5M Cache, 2.40 GHz, 20 Core) x2
GPU	NVIDIA Tesla V100 (SXM2) x4
Memory	384GiB DDR4 2666MHz RDIMM
Local Storage	1.6TB NVMe SSD (Intel SSD DC P4600 u.2) x1
Interconnect	InfiniBand EDR x2





計算ラック・インターコネクト



■ ラックに34ノード、Tesla V100 136基を集約

- ラックあたりの理論ピーク性能：
1.16 PFLOPS (FP64)、17 PFLOPS (FP16)
c.f. Google TPU 2.0 Pod (11.2PF), 3.0 Pod (>100PF)
- ラックあたりの消費電力：67.33 kW

■ ラック内・外インターコネクト

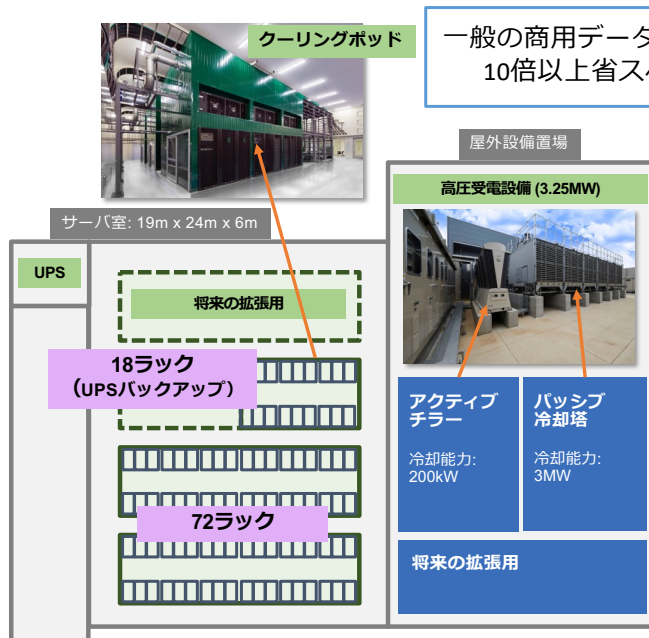
- ラック内：インジェクションバンド幅7200Gbps相当のフルバイセクション構成
- ラック外：1/3オーバーサブスクリプション構成

— InfiniBand EDR x4
- - - InfiniBand EDR x6
— InfiniBand EDR x1



AIデータセンター棟

“1ラック70kW級スパコンのコモディティ化を可能に”



■ 低コストで軽量な建物

■ 耐荷重2トン/m²のコンクリートスラブ上に クーリングポッドとラックを直接設置

■ ラック数

- 初期: 90 (ABCIは41台使用)
- 最大: 144

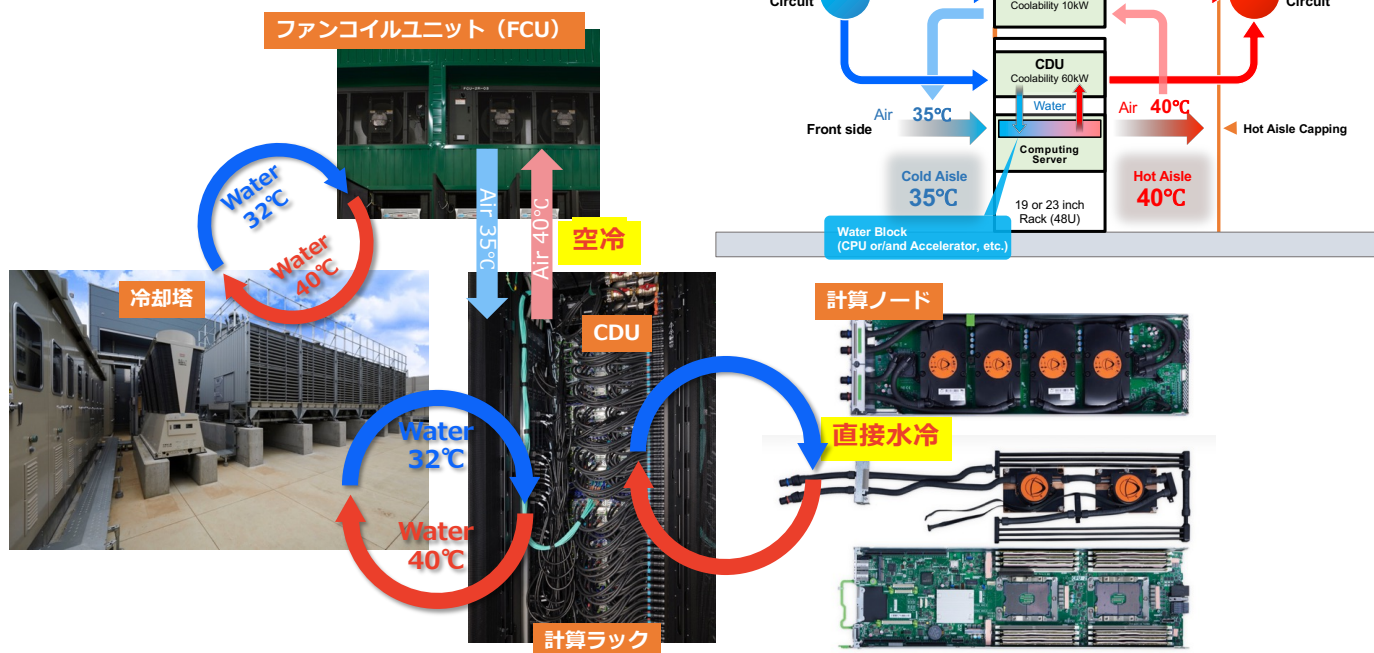
■ 電力容量

- 3.25 MW (ABCIは最大2.3MW使用)

■ 冷却能力

- 3.2 MW (夏季の最小値)
- ラックあたり水冷60kW + 空冷10kW

ハイブリッド冷却システム



ABCI Software Stack (1)

ソフトウェア

※ABCIでは、HPC分野で定評のある最先端ミドルウェア、各種並列化コンパイラ、最新のGPU向け開発環境やライブラリ、各種ディープラーニングフレームワークを提供。

Operating System	CentOS RedHat Enterprise Linux
Job Scheduler	Univa Grid Engine
Container Engine	Docker Singularity
MPI	OpenMPI MVAPICH2-GDR MVAPICH2 Intel MPI
Development Tools	Intel Parallel Studio XE Cluster Edition PGI Professional Edition NVIDIA CUDA SDK GCC, Python, Ruby, R, Java, Scala, Lua, Perl
Deep Learning (available)	Caffe, Caffe2, TensorFlow, Theano, Torch, PyTorch, CNTK, MXnet, Chainer, Keras NVIDIA GPU Cloud (NGC)
Big Data Processing	Hadoop, Spark



ABCI Software Stack (1)

```
[user@es1 ~] $ qsh -g group_name -l rt_F=4 -l h_rt=1:00:00
```

リソースを指定して
インタラクティブ実行

```
[user@g0001 ~] $ module load python/3.6/3.6.5  
[user@g0001 ~] $ module load cuda/9.0/9.0.176.4  
[user@g0001 ~] $ module load cudnn/7.4/7.4.2  
[user@g0001 ~] $ module load openmpi/2.1.5
```

ABCIが提供する
ライブラリをロード

```
[user@g0001 ~] $ python3 -m venv horovod  
[user@g0001 ~] $ source horovod/bin/activate  
(horovod) [user@g0001 ~] $ pip3 install tensorflow-gpu keras torch  
(horovod) [user@g0001 ~] $ pip3 install horovod
```

Python仮想環境を作成

必要なソフトをインストール

```
(horovod) [user@g0001 ~] $ mpirun -n 16 -map-by ppr:4:node python3 keras_mnist.py
```

MPI実行

※ qsubでバッチ実行してもよい（当然）



ABCI Software Stack (2)

コンテナサポート



docker



■ 2つのコンテナエンジンをABCIでサポート

- Docker: 最もユーザベースが大きい
- Singularity: ローレンスバークレー国立研究所で開発されHPC分野での利用が広まりつつある

■ コンテナをベースとしたソフトウェアエコシステムの活用

- グローバルコミュニティで開発された最新の成果をすぐにABCI上で試せる
- ABCIで開発したコンテナをグローバルコミュニティに還元することも容易

■ ABCIに最適化されたコンテナの開発・提供

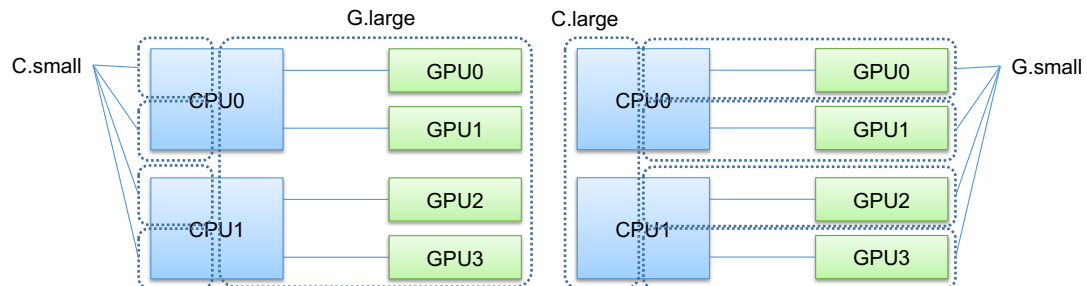
- ABCIに最適化されたChainerMN、Nnablaをはじめとする分散ディープラーニングフレームワークをコンテナ化することで、誰でも大規模データを使ったディープラーニング応用に取り組むことが可能に。

※コンテナとは軽量な仮想化環境の一種。DockerやSingularityは、コンテナをベースとしてソフトウェアの簡便な配布・インストール・実行・再利用を可能にするミドルウェア。

※ABCIでは、ユーザのリソース要求に応じて5つの資源タイプからマッチするものを選択して利用可能。

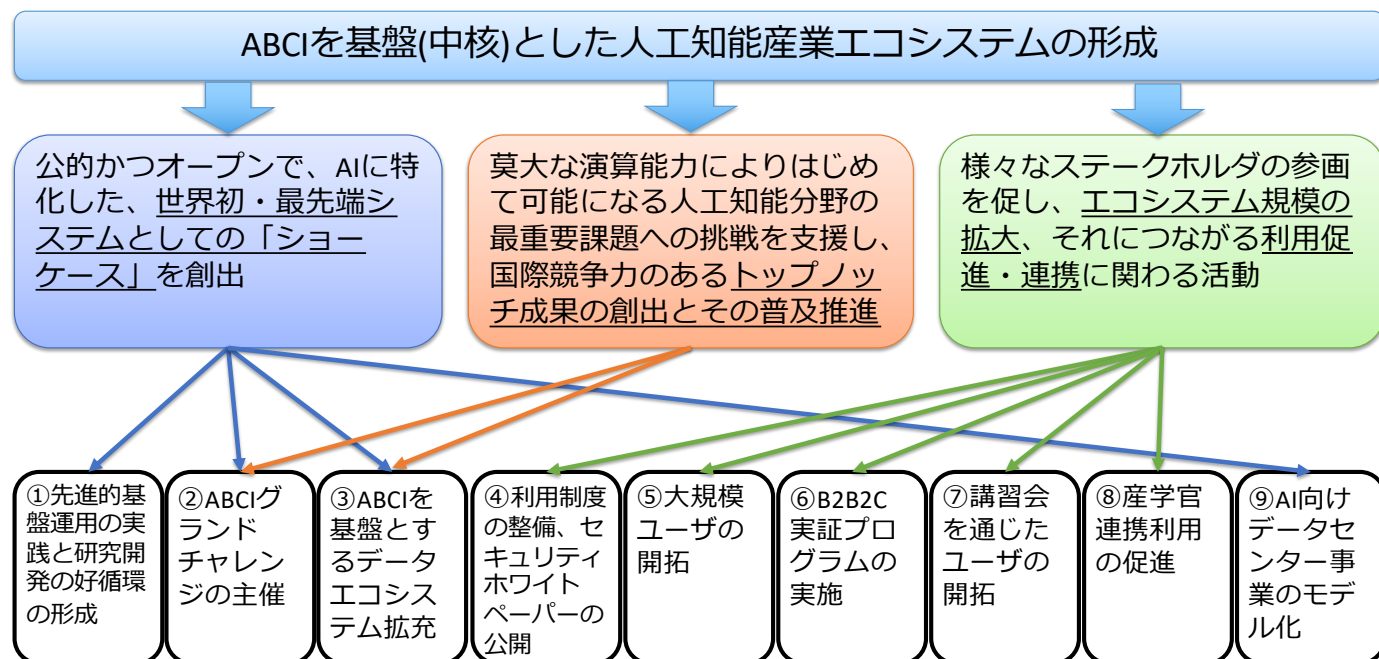
タイプ名	CPUコア 割当量 / 総量	GPU数 割当量 / 総量	メモリ (GB) 割当量 / 総量	ストレージ 割当量 / 総量
F (ノード専有)	40 / 40	4 / 4	360 / 384	1.4 / 1.6
G.large	20 / 40	4 / 4	240 / 384	0.7 / 1.6
G.small	5 / 40	1 / 4	60 / 384	0.175 / 1.6
C.large	20 / 40	0 / 4	120 / 384	0.7 / 1.6
C.small	5 / 40	0 / 4	30 / 384	0.175 / 1.6

※GPUサーバーをまるごと専有することも、論理的に分割した部分のみを利用することも可能。



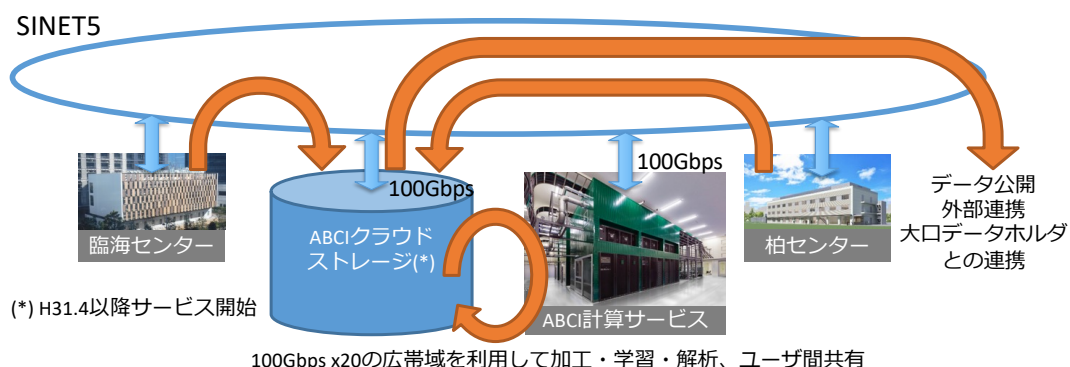
※ABCIでは、処理スループットの最大化を可能とするバッチ型・対話型の実行サービスに加え、自由度の高い予約サービス、IDE、各種ストレージサービスを提供。

サービス名	説明	割当ノード数 (最小/最大)
Spot	バッチ実行型ジョブサービス 利用した時間分のみ秒単位で課金（不使用時間は非課金）	1 / 512
On-demand	インタラクティブ実行型ジョブサービス 利用した時間分のみ秒単位で課金（不使用時間は非課金）	1 / 32
Reserved	日単位での資源予約サービス 予約時間に応じて課金	1 / 32
IDE (開発中)	WebUIベースの統合開発環境	未定
Group Storage	グループ内で共有できるストレージサービス 無償の200GBのホーム領域と、上限を指定して購入できるグループ領域が利用可能	N/A
Secure Object Storage (開発中)	暗号化に対応したオブジェクトストレージサービス Amazon S3に類似したクラウドストレージで、利用量に応じた課金	N/A



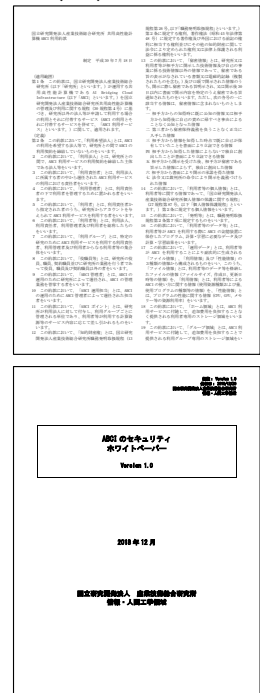
③ ABCIを基盤とするデータエコシステム拡充

■ 「データの価値化」を「実践」するためのデータプラットフォームを構築



④ 利用制度・セキュリティホワイトペーパー

- 国内外の民間企業、大学、国研などによる利用制度を整備
 - － 利用約款（産業利用でも、研究開発、機械学習フェーズ、人材育成であれば利用可）
 - － 利用料徴収
 - － 輸出管理手続き（国外、非居住者向け）
- ABCIのセキュリティ基準の明確化
 - － ABCIはいわゆるクラウドサービス、インターネットに接続する端末からならどこからでも利用可能
 - － こうしたクラウドサービス利用にあたって民間企業が求めるセキュリティ要件との整合性を確認しやすくすることが重要



⑦ 利用者講習会（初級編）

- 講習内容
 - － ABCIシステムの基本概念の理解
 - － 受講者のPCを使ったハンズオン
- 講習会は定期的実施
- ポインタ
 - － https://abci.ai/ja/how_to_use/data/ABCI_User_Seminar_20190207.pdf
 - － https://abci.ai/ja/how_to_use/data/ABCI_hands-on_MAC_20190207.pdf

⑦ ABCI並列プログラミング講習会（上級編）

- 講習内容
 - ABCIを活用した分散学習の実践、ハッカソン
 - 産総研職員、Chainer, TensorFlowのスペシャリスト、NVIDIAのGPUエンジニアがチューターとして参加
- 一人またはチーム（最大3名）での参加
 - 参加にあたり審査があります
 - ハッカソンに必要なABCIの利用権は産総研が無償で提供します
 - 産業界の課題・ユーザを積極的に支援します
- ポインタ
 - https://abci.ai/event/2019/02/27/ja_event.html

2 1 3 ABCI Grand Challenge



2019年度も3回実施予定
(6月/9月/12月)

第1回のRFPは4月予定！

- 莫大な演算能力によりはじめて可能になる人工知能分野の最重要課題への挑戦を支援するため、AI橋渡しクラウドで「ABCIグランドチャレンジ」プログラムを実施します。
- ABCIグランドチャレンジは、ABCIがもつ最大計算ノード数である1,088ノード（4,352GPU）を最大24時間、1研究グループでの占有利用ができる公募型チャレンジプログラムです。
- 画期的な成果の達成が見込まれる、挑戦的な研究課題のご提案をお待ちしております。
- 利用料金はリハーサルを含め、産総研が負担します。

■ 重要日程

	応募締め切り	採択通知	実行予定日
#1	2018年4月30日	2018年5月30日	2018年7月最終週
#2	2018年8月31日	2018年9月30日	2018年10月最終週
#3	2018年11月30日	2018年12月22日	2019年1月最終週

※ 2018年度は、のべ8グループの応募課題が採択、実施されました。

■ HPCI (High Performance Computing Infrastructure)

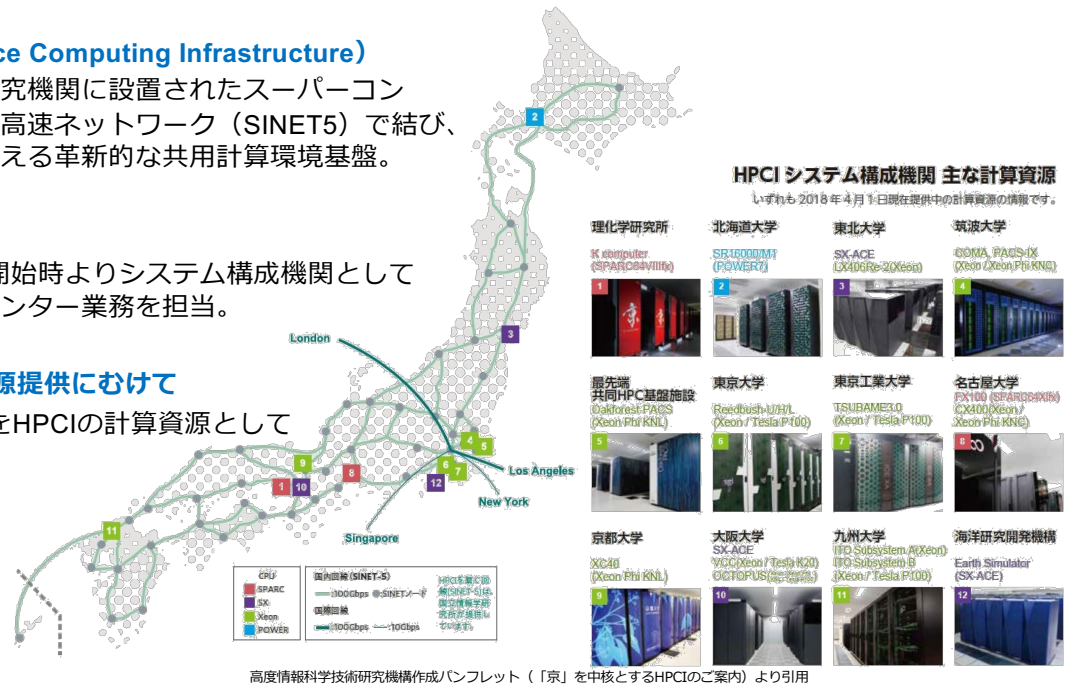
- 「京」と全国の大学や研究機関に設置されたスーパーコンピュータやストレージを高速ネットワーク（SINET5）で結び、多様なユーザーニーズに応える革新的な共用計算環境基盤。

■ 産総研とHPCIの関わり

- 産総研は、HPCIの運用開始時よりシステム構成機関として参画し、プライマリーセンター業務を担当。

■ ABCIによるHPCIへの資源提供にむけて

- 平成31年度から、ABCIをHPCIの計算資源として提供すべく手続き中。



- ABCIはインターネットに接続する端末ならどこからでもアクセス可能なクラウド型サービス
- 利用に必要な情報はすべてインターネットに公開
 - システム概要：https://abci.ai/ja/about_abci/computing_resource.html
 - サービス概要：https://abci.ai/ja/about_abci/cloud_service.html
 - 利用申請手順：https://abci.ai/ja/how_to_use/custom.html
 - 利用料金：https://abci.ai/ja/how_to_use/tariffs.html
 - 利用者ポータルマニュアル：<https://portal.abci.ai/docs/portal/ja/>
 - 利用者マニュアル：<https://portal.abci.ai/docs/ja/>
- その他すべての情報は <https://abci.ai/ja/> から