

理研計算科学研究センター(R-CCS) におけるSociety5.0への取り組み



理化学研究所 計算科学研究センター
センター長 松岡 聡

HPCIコンソシウム 説明資料
20200127

超大規模ITインフラである富岳のミッション

「京」の主なミッションであるシミュレーション・基礎科学のサポート

CSTIによるSociety5.0早期実現を中心とした国民の関心事の高い成果創出

「京」から踏襲した
研究・開発・運用の
更なる高度化



次世代のIoTの中心イ
ンフラとしての新たな
研究・開発・運用

計算機
システム

「計算の科学」

各種アプリケーションの特性を加味したシステム設計、および、それに適したアプリケーションの最適化

Co-design

富岳

「計算による科学」

各種アプリケーションのさまざまな計算上の特性を一般化・特徴化

ターゲット
アプリ
ケーション

世界トップ性能と世界一の広がり両立

大規模アプリケーションの加速能力が高い / アプリケーションとユーザ層の幅が広い

CSTI中間評価:Society5.0におけるポスト京を含むスパコン技術の位置づけ

- 全ての人とモノがつながり、今までにない新たな価値を生み出す超スマート社会の実現を目指す
Society5.0において、シミュレーションによる社会的課題の解決や人工知能（AI）開発及び情報の流通・処理に関する技術開発を加速するためには、スーパーコンピュータ等の情報基盤技術が必要不可欠

・第5期科学技術基本計画（平成28年度～平成32年度）の柱である「超スマート社会」（Society5.0）を世界に先駆けて実現するためには、新たな価値創造の基盤としてのスーパーコンピュータが必要不可欠

（文部科学省 中間評価）

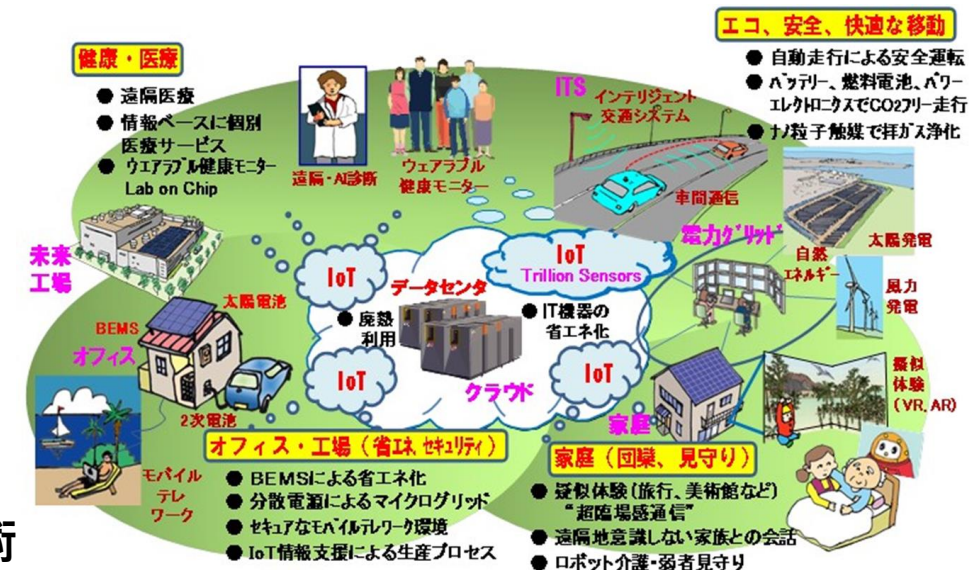
・〔1〕データ駆動型社会の共通インフラの整備

③研究生産性の向上

産学官連携を支え、生産性の飛躍的向上の基盤となる
高速電子計算機施設等の先端的な研究施設・設備の
整備・共用やポスト「京」の開発を進める

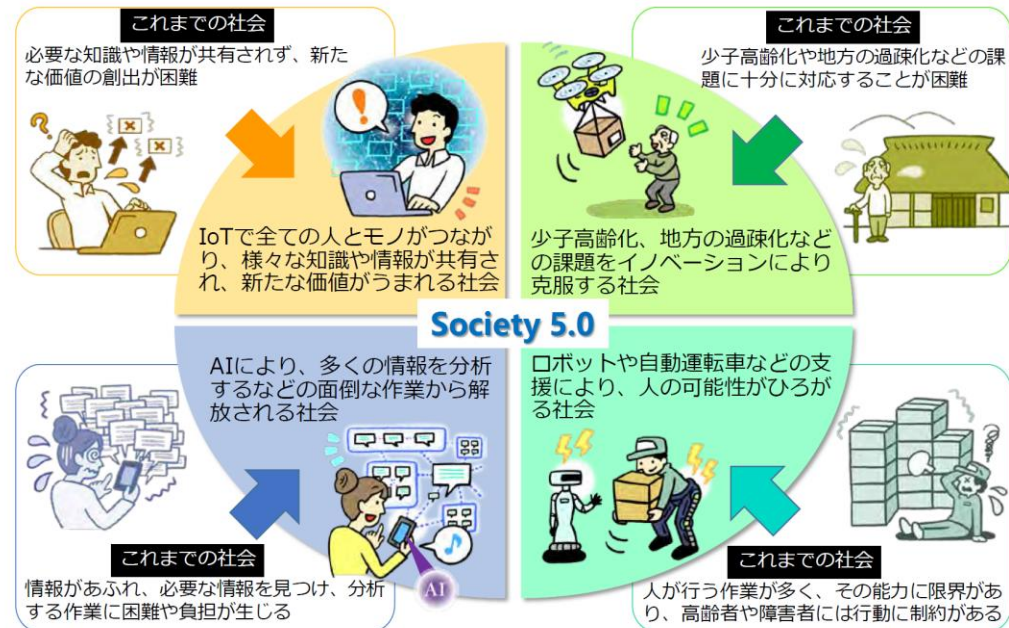
（未来投資戦略2018（平成30年6月15日閣議決定））

このためには富岳におけるA64fxやソフトウェア・アプリのHPC技術が、富岳に留まらず、クラウドにも大幅に波及し、エッジと接続してシミュレーション・AIを駆動することが重要



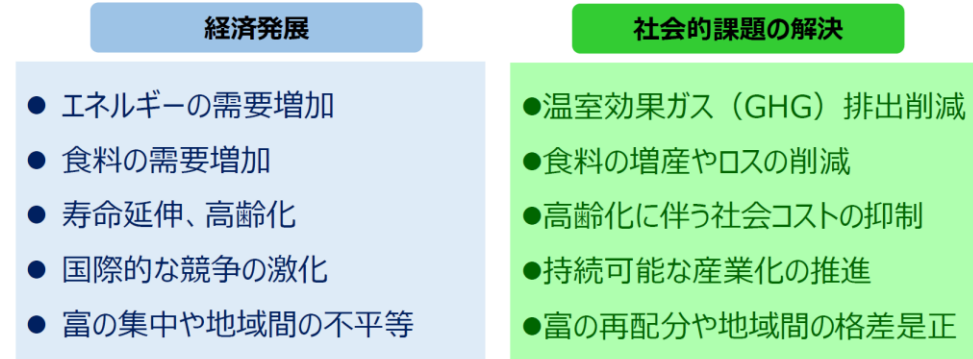
<超スマート社会における人々の生活>

（出典：JST/CRDS 曾根純一上席フェロー講演資料）



経済発展と社会的課題の解決の両立

イノベーションで創出される**新たな価値**により、格差なくニーズに対応したモノやサービスを提供することで、**経済発展**と**社会的課題の解決**を両立



IoT、ロボット、人工知能（AI）、ビッグデータ等の先端技術をあらゆる産業や社会生活に取り入れ、格差なく、多様なニーズにきめ細かく対応したモノやサービスを提供

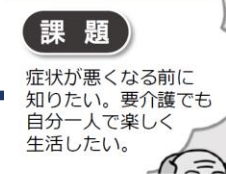
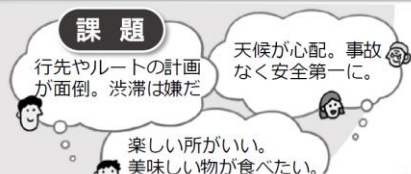
経済発展と社会的課題の解決を両立

サイバー空間とフィジカル空間の高度な融合

フィジカル（現実）空間から**センサー**と**IoT**を通じてあらゆる情報が集積（**ビッグデータ**）**人工知能（AI）**がビッグデータを解析し、高付加価値を**現実空間にフィードバック**

これまでの情報社会(4.0)





富岳によるSociety5.0実現の加速



・富岳上にSociety5.0の医療・環境・サプライチェーン・交通などの仮想空間を作成

- ・ シミュレーションによるデータ生成→少ないデータでのAIの学習
 - ・ 「正しい」サンプリング、正確なラベリング、レアイベントの生成
 - ・ HPC・数値計算・シミュレーションでのトップ技術が最も重要
- ・ 社会実装への実証実験：大規模社会エージェントシミュレーション+ストリーミングデータ処理+AI
 - ・ 実データ+シミュレートデータ

・新たな「データ中心」から「計算中心」のインターネット

- ・ IoT時代の必然、CDNからリアルタイム「処理」へ

・富岳技術の発展と転用によるIoTインフラの構築

- ・ HPCの高速化技術～AIの高速化技術

・GAFAに対抗しうる中心的存在

- ・ シミュレーションとAI、IoTの融合インフラが鍵

富岳



競争力強化・災害対応



人手不足解消・多様なニーズ対応



GHG排出削減・人手不足解消



顧客満足度向上

安全な避難

個人のスマホに避難情報が提示され、安全に避難所まで移動

迅速な救助

アシストスーツや救助ロボットにより被災した建物から救助

物資の最適配送

避難所にドローンや自動配送車により救援物資が配送

- 「富岳Arm」 IoT主流のArmエコシステムの頂点としての開発
 - A64fx CPU：世界最速の汎用プロセッサ(x86, Arm含)
 - HPC・クラウド・AI・IoT全て包括しうるソフトウェアスタック
 - VM, コンテナ、Spack Package Managerなどの
- 「富岳AI」の開発
 - PyTorch, TensorFlow等の高速実装をDNNL for A64fx, Eigen等をベースに開発
 - 富士通・理研・Arm社のMoUによる開発体制
 - その他各種HPCとAIの融合
- 「富岳クラウド」における複数のクラウドプロバイダとの連携
 - 2020年度クラウドサービスの実験を公募、複数クラウドプロバイダ
 - 2021年度本格運用へ
- 「富岳ライブストリーム」で多数のIoTストリームデータのアプリケーションへの提供
 - 種々の研究機関・企業と連携し、一定期間以上保存される観測データの格納・分析・学習・推論機構を提供

「富岳」：世界最高のIT技術とその活用によるイノベーション

- **世界トップレベルの圧倒的性能：米国半導体産業に圧勝**
Green500 世界一位を獲得（2019年11月）で消費電力性能を実証
米国トップのチップ (Intel Xeon) の3倍の性能・効率を達成
- **世界最大規模のAI研究開発基盤：IoT時代の中心的インフラ**
日本のAI劣勢を跳ね返すのに必要不可欠な高性能AIインフラ
- **量子コンピュータ開発にも必須なインフラ**
素子開発、アルゴリズム開発、検証に必要
- **経済・産業界、社会、IoTイノベーションへ活用**
防災や創薬、産業イノベーション



最高の計算能力と省電力性能を両立、Armアーキテクチャ*採用によるIoTとの融合



*...世界のスマホ90%以上で採用が
示す世界標準の広大なエコシステム

Society5.0 実現の重要なITイノベーションおよび基盤

富岳をスマートフォンやサーバと比較すると…？

	スマートフォン		サーバ (クラウド含)		富岳		参考：京
台数	2,000万台 (国内の年間出荷台数の約2/3)	=	30万台 (国内の年間出荷台数の約2/3)	=	1台		30~100台
消費電力	10W×2,000万台= 200MW	=	600-700W×30万台= 200MW (冷却含)	> >	30MW		15MW
CPU種別 基本ソフト	Arm iOS/ Android Linux		x86/Arm Linux (Red Hat 等)/Win		Arm Linux (Red Hat等)		Sparc Linux(独自) 汎用性低い
AI 対応	アプリに応じた 独自回路 推論のみ		GPUなどの汎用 加速チップ、CPU の加速命令		汎用CPU SVE 加速機能		なし

米国の最先端CPUを3倍凌駕した半導体設計(性能・効率)

Green500 世界1位を獲得 (2019年11月、プロトタイプ)

- TOP500内のHPCで電力効率の良い高性能計算の実現を評価するランキング (他の性質の違うアプリでも同様の結果を確認)
- Green500は、これまでGPU搭載や特殊マシンが上位を独占していたが、**富岳は汎用CPU搭載で世界1位獲得**
- 汎用CPUとして、同条件の**Intel機と比べて3倍強**の省電力性能
- チップあたりの絶対性能でも2~4倍の性能⇒AIでは更なる性能差

米HPE/Cray社は、「富岳」に実装されているArmA64FXプロセッサを自社スパコンに採用 (米国製プロセッサ以外では史上初)

2020年度以降、世界市場で販売予定



<納入予定先>

- ・ 米国ロスアラモス国立研究所
- ・ 米国オークリッジ国立研究所
- ・ 米国ストーニーブルック大学
- ・ 英国ブリストル大学 等

日本の半導体産業の復興へ

覇権維持のため開発を継続することが重要



何故汎用CPUで「3倍速い」が実現できたのか

- 企業では困難なターゲットを設定し、**圧倒的な世界トップ**の能力達成
- ⇒イノベーション達成のためリスクテイクできる国家プロジェクトでのみ実現可能
- ⇒国家プロジェクトゆえ、一企業だけでなく、**HPCIのコミュニティを含む日本の叡智が結集**しOne Teamで課題に取り組んだ
 - **世界最高峰Intel製品を凌駕**
- ハードとアプリの双方の開発陣によるCo-design（コデザイン）により、「高いアプリケーション性能」、「省電力」、「使い勝手の良さ」を達成
- 国家プロジェクトであるべき姿を実現できた



Spack enables Software distribution for HPC

- Spack automates the build and installation of scientific software
- Packages are *templated*, so that users can easily tune for the host environment

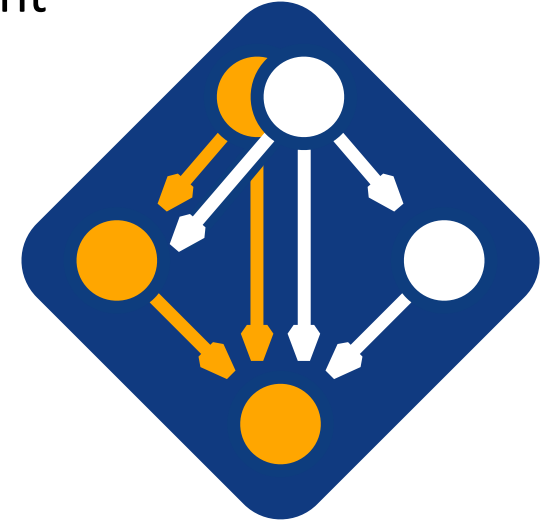
No installation required: clone and go

```
$ git clone https://github.com/spack/spack
$ spack install hdf5
```

Simple syntax enables complex installs

```
$ spack install hdf5@1.10.5
$ spack install hdf5@1.10.5 %clang@6.0
$ spack install hdf5@1.10.5 +threadssafe
$ spack install hdf5@1.10.5 cppflags="-O3 -g3"
$ spack install hdf5@1.10.5 target=haswell
$ spack install hdf5@1.10.5 +mpi ^mpich@3.2
```

- Ease of use of mainstream tools, with flexibility needed for HPC tuning
- Major victories:
 - ARES porting time on a new platform was reduced from **2 weeks to 3 hours**
 - Deployment time for 1,300-package stack on Summit supercomputer reduced from **2 weeks to a 12-hour overnight build**
 - Used by teams across ECP to **accelerate development**



github.com/spack/spack

Spack now understands specific target microarchitectures

- We have developed a cross-platform library to detect and compare microarchitecture metadata
 - Detects based on /proc/cpuinfo (Linux), sysctl (Mac)
 - Allows comparisons for compatibility, e.g.:

```
skylake > broadwell
zen2 > x86_64
```

- Key features:
 - Know which compilers support which chips with which flags
 - Determine compatibility
 - Enable creation and reuse of optimized binary packages
 - Easily query available architecture features for portable build recipes
- We will be extracting this as a standalone library for other tools & languages
 - Hope to make this standard!



```
$ spack arch --known-targets
Generic architectures (families)
  aarch64 ppc64 ppc64le x86 x86_64

IBM - ppc64
  power7 power8 power9

IBM - ppc64le
  power8le power9le

AuthenticAMD - x86_64
  barcelona bulldozer piledriver steamroller excavator zen zen2

GenuineIntel - x86_64
  nocona westmere haswell mic_knl cascadelake
  core2 sandybridge broadwell skylake_avx512 icelake
  nehalem ivybridge skylake cannonlake

GenuineIntel - x86
  i686 pentium2 pentium3 pentium4 prescott
```

```
class OpenBlas(Package):
    def configure_args(self, spec):
        args = []
        if 'avx512' in spec.target:
            args.append('--with-avx512')
        ...
        return args
```

Simple feature query

```
$ spack install lbann target=cascadelake
$ spack install petsc target=zen2
```

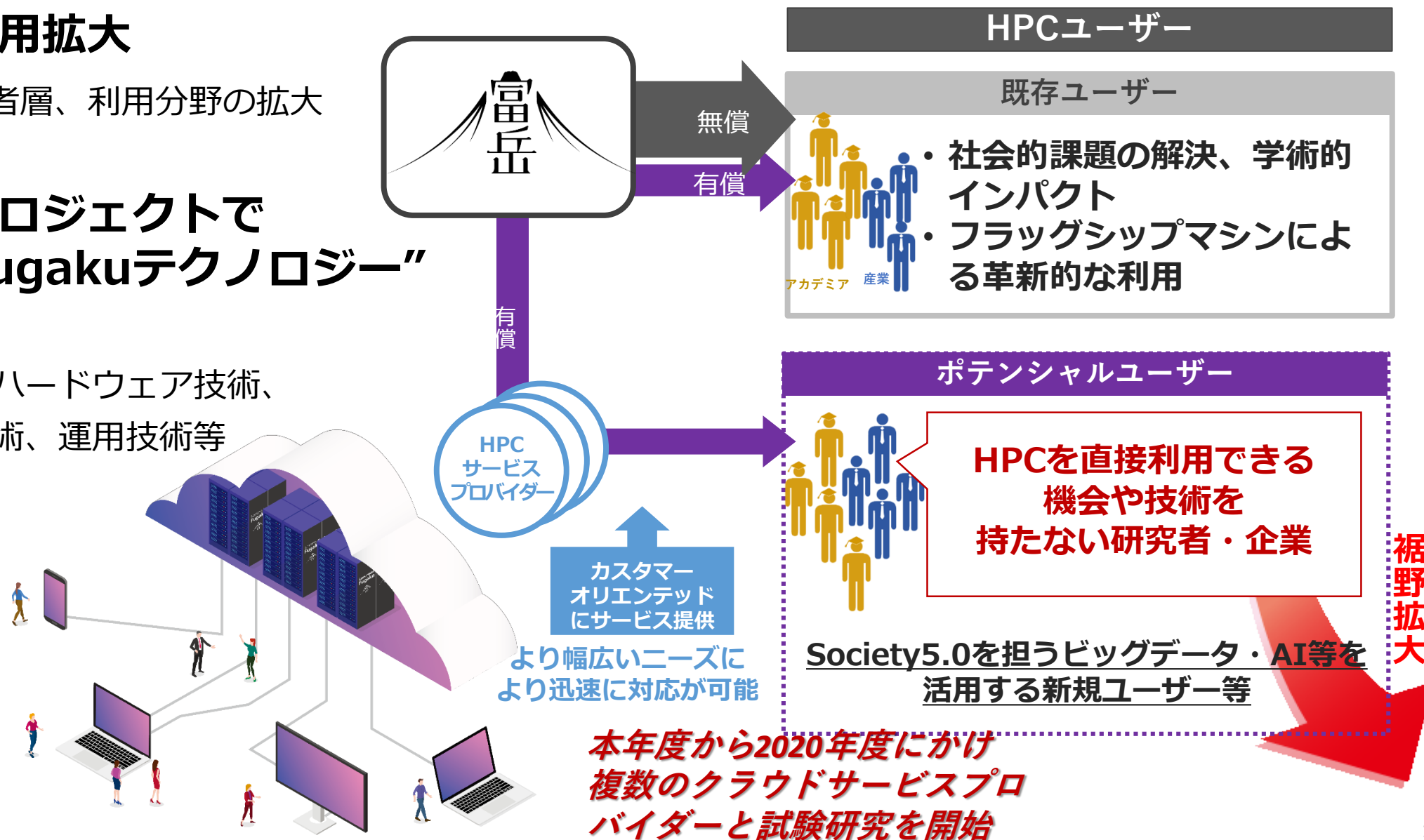
Specialized installations

■「富岳」の利用拡大

利用者数、利用者層、利用分野の拡大

■「富岳」のプロジェクトで 培われた“Fugakuテクノロジー” の普及

HPCサービス、ハードウェア技術、
ソフトウェア技術、運用技術等



富岳：世界最大規模のAI基盤

日本がAIで劣勢を跳ね返すには

- 莫大な学習データの収集
 - ✓ 今後の日本の政策により、統計的に有意なサンプル数は収集可能
- 高性能AIインフラ
 - ✓ Deep learning世界トップレベルのABCIの15～20倍以上の性能
- 高度人材
 - ✓ 上記条件が揃えば世界から人材が集まる



GAFAに対抗し、追い越すことも可能
(ABCIでは既に富士通・ソニーが一部)

- ◆ 「富岳試作機評価環境の利用に関する覚書」を富士通と締結し、「富岳」上にAIフレームワークを創り上げ「富岳」を中心とした世界トップクラスのAI学習・推論・利活用の計算機環境基盤を構築する。
- ◆ 大規模AIベンチマーク「HPL-AI」(2020年6月創設) に向け研究開発を推進中、世界一位獲得可能性大



超高速なAIソフトウェアを「富岳」、商用機、クラウド等に展開しSociety5.0の中心的なインフラとする。

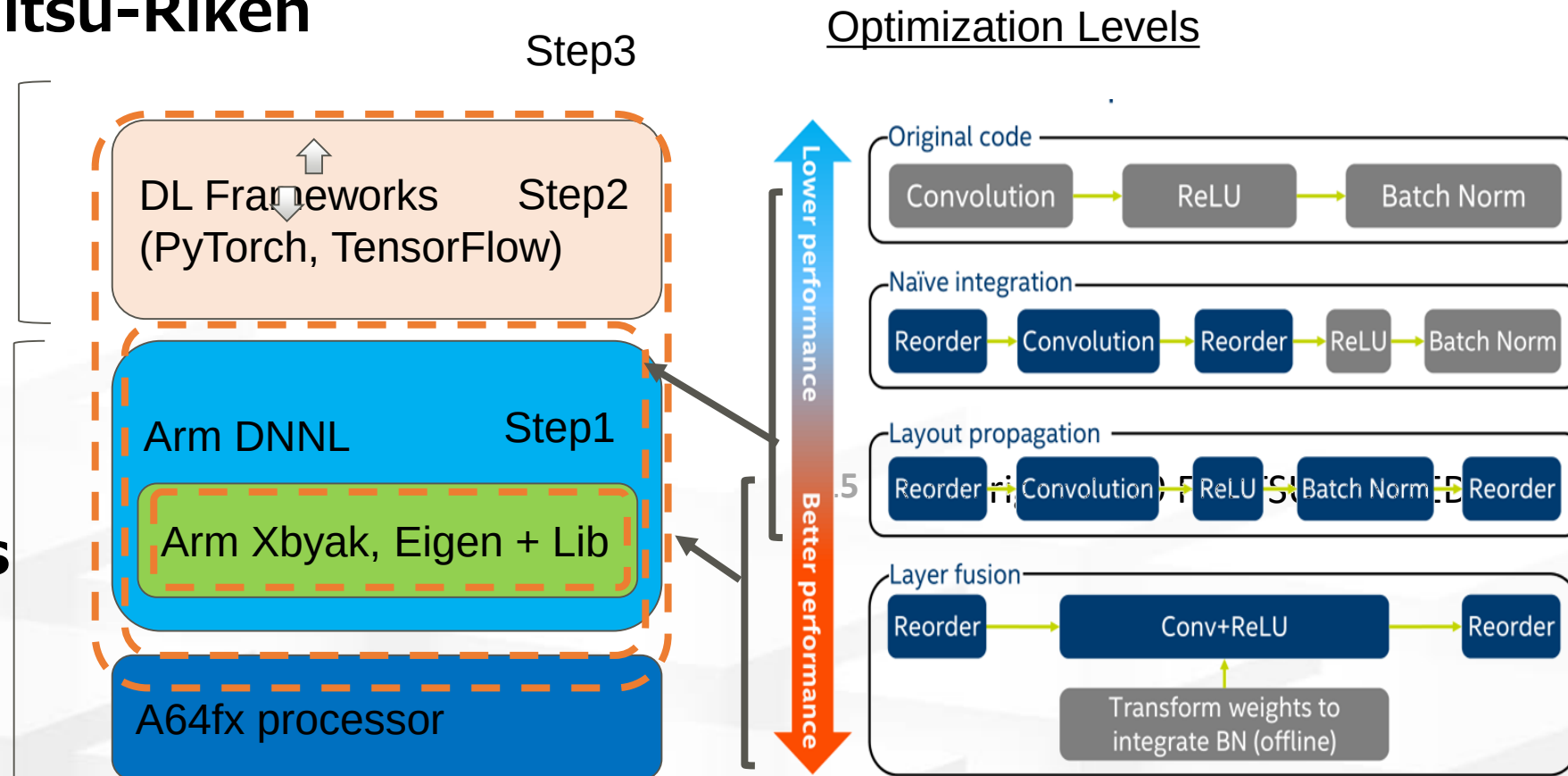


Fujitsu-Riken-Arm joint effort on AI framework development on SVE/A64FX

- **MOU Signed Fujitsu-Riken Nov. 25, 2019**



- **w/Arm in works**



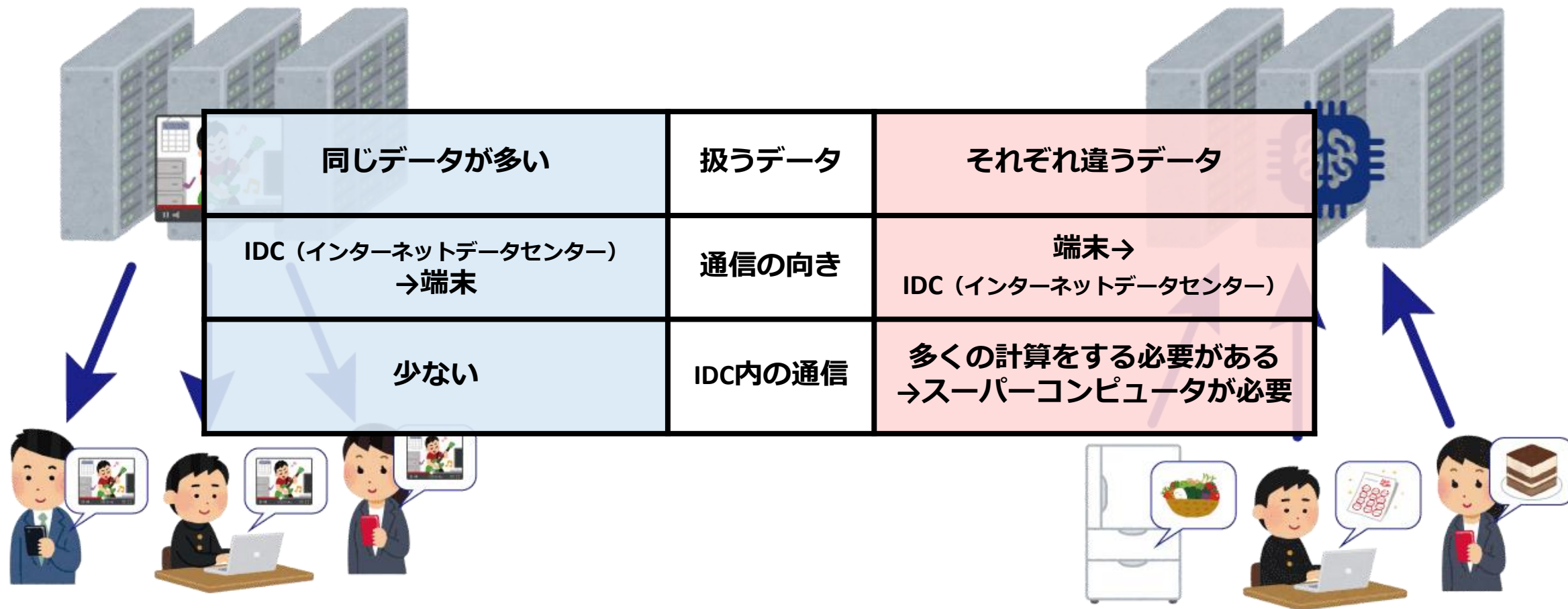
Exaops of sim, data, and AI on Fugaku and Cloud

我が国の公的大規模AI計算インフラ

	Deployed	Purpose	AI Processor	Inference Peak Perf.	Training Peak Perf.	Top500 Perf/Rank	Green500 Perf/Rank
Tokyo Tech. TSUBAME3	July 2017	HPC + AI Public	NVIDIA P100 x 2160	45.8 PF (FP16)	22.9 PF / 45.8PF (FP32/FP16)	8.125 PF #22	13.704 GF/W #5
U-Tokyo Reedbush-H/L	Apr. 2018 (update)	HPC + AI Public	NVIDIA P100 x 496	10.71 PF (FP16)	5.36 PF / 10.71PF (FP32/FP16)	(Unranked)	(Unranked)
U-Kyushu ITO-B	Oct. 2017	HPC + AI Public	NVIDIA P100 x 512	11.1 PF (FP16)	5.53 PF/11.1 PF (FP32/FP16)	(Unranked)	(Unranked)
AIST-AIRC AICC	Oct. 2017	AI Lab Only	NVIDIA P100 x 400	8.64 PF (FP16)	4.32 PF / 8.64PF (FP32/FP16)	0.961 PF #446	12.681 GF/W #7
Riken-AIP Raiden	Apr. 2018 (update)	AI Lab Only	NVIDIA V100 x 432	54.0 PF (FP16)	6.40 PF/54.0 PF (FP32/FP16)	1.213 PF #280	11.363 GF/W #10
AIST-AIRC ABCI	Aug. 2018	AI Public	NVIDIA V100 x 4352	544.0 PF (FP16)	65.3 PF/544.0 PF (FP32/FP16)	19.88 PF #7	14.423 GF/W #4
NICT (unnamed)	Summer 2019	AI Lab Only	NVIDIA V100 x 1700程度	~210 PF (FP16)	~26 PF/~210 PF (FP32/FP16)	????	????
C.f. US ORNL Summit	Summer 2018	HPC + AI Public	NVIDIA V100 x 27,000	3,375 PF (FP16)	405 PF/3,375 PF (FP32/FP16)	143.5 PF #1	14.668 GF/W #3
Riken R-CCS Fugaku	2020 ~2021	HPC + AI Public	Fujitsu A64fx > x 150,000	> 4000 PO (Int8)	>1000PF/>2000PF (FP32/FP16)	> 400PF #1 (2020?)	>16 GF/W #1
ABCI++ (speculative)	2020 ~2021	AI Public	Future GPU 1000 ~ 2000	>100 PO (INT8)	>200PF (16 bit)	~50PF (w/ABCI)	???

現在のインターネット

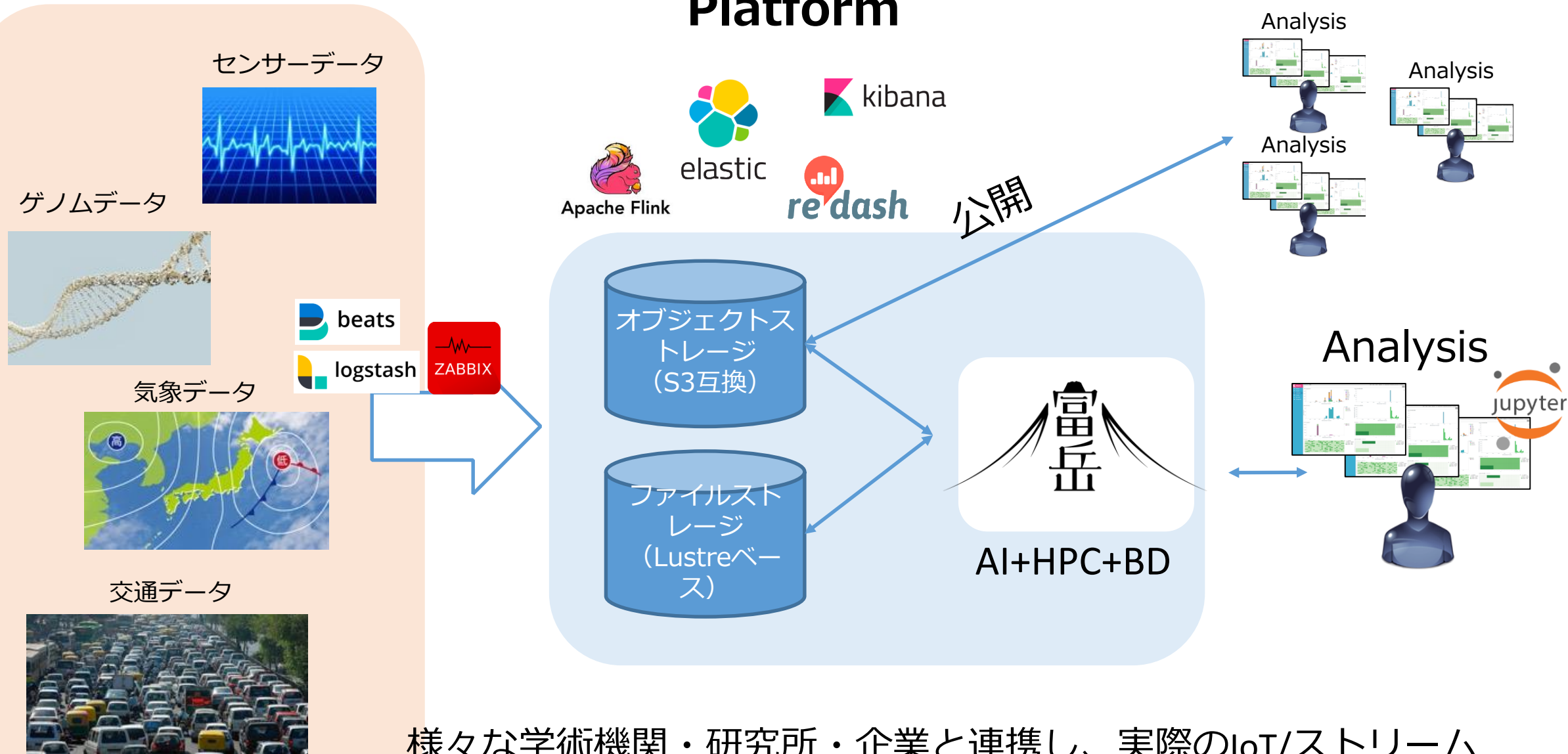
Society5.0時代のインターネット



ビデオ配信が通信の90%

データ中心から計算中心の時代へ

Fugaku LiveStream: Society 5.0 Data Storage & Analysis Platform

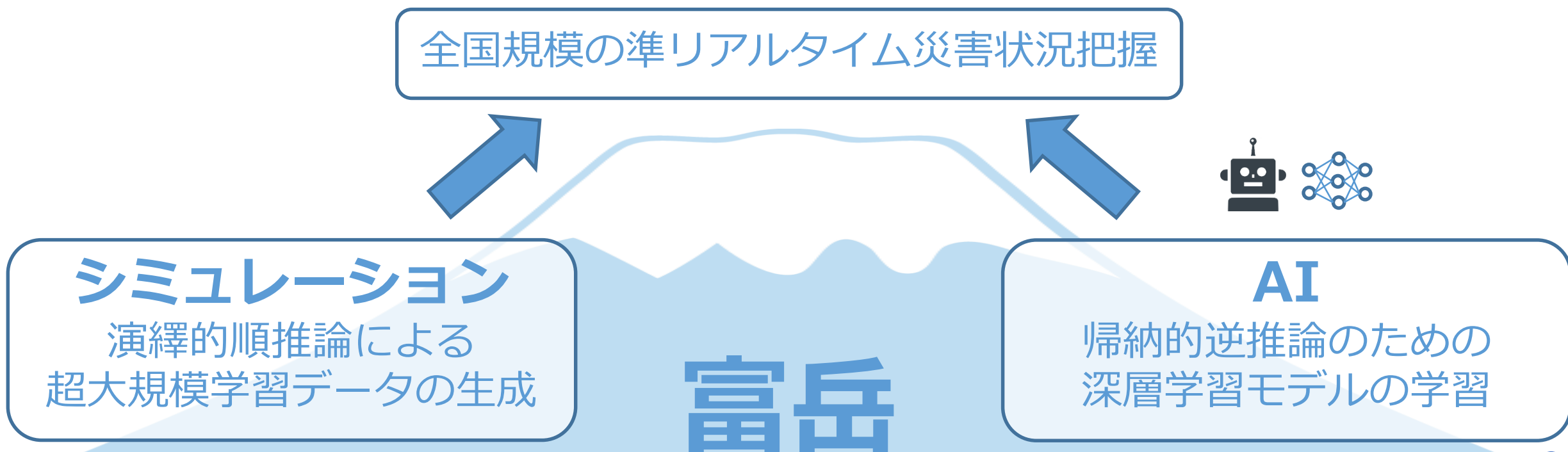


様々な学術機関・研究所・企業と連携し、実際のIoT/ストリームデータを多数一定期間格納し、Society5.0アプリの開発、及び社会実装へ

シミュレーション x AI @ 富岳

⇒ 全国規模の準リアルタイム災害状況把握

- シミュレーションによる大規模学習データの生成と大規模深層学習を可能とするためには世界最高水準の計算性能が必須
- 富岳で初めて全国規模の準リアルタイム災害状況把握が実現



宇宙 x AI x 国際協力 ⇨ 迅速な世界の災害状況把握

革新的災害状況把握基盤技術の創出

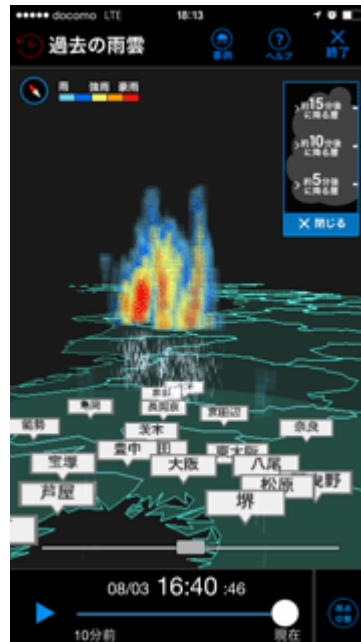


今後の展開

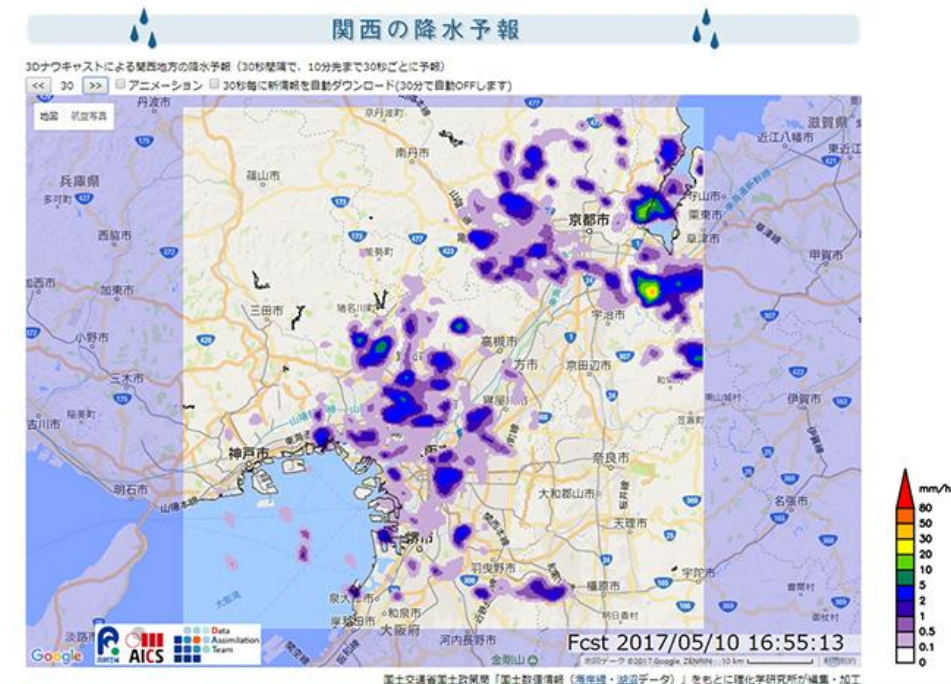
- 富岳を用いた大規模シミュレーションとAIによる超高速・高精細災害把握
- AIP内異分野連携による説明可能な空間情報AI

R-CCS三好ら：AI、ビッグデータとシミュレーションが融合した、リアルタイム気象予報実験：東京オリンピックでの社会実装実験

- 大阪・神戸のフェーズドレイ気象レーダを用いた降水短時間予報について、2017年に予報業務許可を取得し、現在までリアルタイムで予報の公開を行っている
- 気象予報士が常駐し、予報を随時確認した上でWebサイト上で配信(平日9時~5時のみ)
- 共同研究契約を結んでいる株式会社エムティーアイが別に予報業務許可を取得し、スマートフォンアプリ「3D雨雲ウォッチ」(2019/9時点でダウンロード数247,000以上)により予報を配信
- 富岳は間に合わないが、東大のスパコンの1/4を用いて東京オリンピックのリアルタイム気象予報を行う予定



<https://www.mti.co.jp/?p=21823>



<政策的位置づけ> 総合科学技術・イノベーション会議から文部科学省への指示（2018年11月22日）

- ・「富岳」を利用したSociety5.0実現のためには、ビッグデータの活用において国全体で進めていくことが重要。
- ・関係府省との連携を取りながら、「富岳」利用の仕組みをつくること。

理研の取り組み ～内閣府、文科省の協力等をいただきつつ～

本格的な産学官協働体制

個別研究者ネットワークに基づく産学官協働

自動車用次世代 CAEコンソーシアム
(2017年設立)

トヨタ自動車株式会社をはじめとする産業界（14法人）と大学研究機関（9法人）が参画。

Society5.0への貢献に必須な経団連レベルでの協働構築に向けて準備中

（経団連、内閣府、文科省、理研にて相談開始）

経団連・COCNプロジェクトに「富岳」の利用が本格的に位置付けられること等を目指す

戦略的国際連携

- **米DOE**と、ポストムーア、AI、量子コンピュータに関する連携に向けて準備中。
- **欧州**と、富岳テクノロジーの汎用性を活かした連携等を強化中。
- **ASEAN**のHPCネットワーク構想と日本の富岳中核HPCIとの連携に向け対応中。
(2月末のシンガポールでのスパコン分野の国際会議（SCアジア）にて、CSTI委員が講演予定。)

世界的拠点化を目指す

人材育成など

- **Society5.0を支える人材育成**
 - ・データ関連の創造的・高度人材の産業界を含めた育成に貢献
 - ・兵庫県、神戸市、地元機関等と連携
- **社会発信の強化**
 - ・2020年3月4日には、「富岳フォーラム（仮称）」@東京を開催予定。Society5.0との関係を中心に内閣府審議官にご講演頂く予定。
 - ・先般のGreen500での世界トップ獲得の際など、機会あるごとに分かりやすい情報提供を実施

イノベーションの基盤たる人材育成と、「富岳」時代の社会発信を目指す