

# PCCC20 パネル討論 次世代先端的 計算基盤のあり方を考える

パネリスト：佐藤賢斗

理化学研究所 計算科学研究センター

高性能ビッグデータ研究チーム

# 自己紹介：佐藤賢斗

## • 経歴

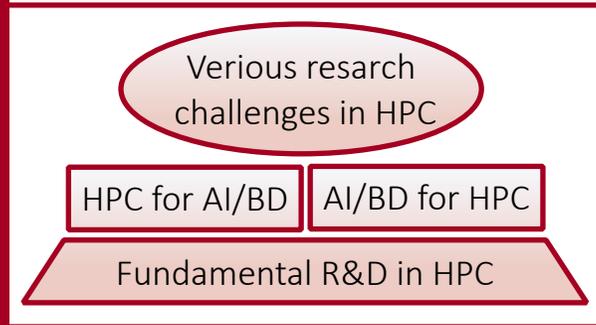
- March, 2014
  - Ph.D. in Science, Department of Mathematical and Computing, Tokyo Tech
- April, 2014 (5 months)
  - Postdoc, Global Scientific Information and Computing Center, Tokyo Institute of Technology
- September, 2014 (3 years)
  - Postdoc, Lawrence Livermore National Laboratory
- September, 2017 (1 year)
  - Computer scientist, Lawrence Livermore National Laboratory
- November, 2018 ~ present
  - Team leader, RIKEN R-CCS
  - Associate Professor at Kobe Univ. (Graduate school) 2019~present

- ▶ Research interest: System software
  - ▶ Debugging tools for MPI applications
    - ▶ Record-and-replay
  - ▶ Fault tolerance
    - ▶ Checkpoint/restart
  - ▶ I/O
    - ▶ Data compression
    - ▶ File system
  - ▶ Cloud computing
    - ▶ Optimization with VM migration
  - ▶ AI/Big data

# Mission of High perf. Big Data Research Team

## Convergence of AI, Big Data and HPC

Addition to fundamental R&D in HPC,  
we study and develop software to accelerate AI and Big data applications (HPC for AI/Big data) while we also use AI techniques to resolve various research challenges in HPC (AI/Big data for HPC)



- **Fundamental R&D in HPC**

- Reproducibility in HPC applications
- Design space exploration for the next-gen supercomputers in the Post-Moore era

- **AI/BD for HPC**

- Big data compression with AI techniques
- System software optimization with AI techniques
- System log analysis and prediction

- **HPC for AI/BD**

- Deep learning framework tuning on Fugaku
  - PyTorch, TensorFlow for Fugaku's Arm processor (A64FX)

# NGACIホワイトペーパー作成での役割

- 分科会
  - 新利用形態の取りまとめ
- WG:
  - システムソフトの取りまとめ

# 今後のHPC開発に重要と思われる部分の簡単なお紹介

不変的に必要とされているソフトウェア・技術は、下位のHWの変更や上位のアプリの要求に応じて継続的に研究開発をする必要がある。

大テーマ	小テーマ
1	基盤ソフトウェア (OS、仮想化、コンテナ)
2	大規模並列/高性能計算 (汎用CPU、アクセラレーター)
	大規模並列/高性能計算 (次世代特化型ハードウェア: 量子コンピュータ、ニューロモーフィック、データフロー型 etc.)
	大規模並列/高性能計算 (メモリ階層、データ移動)
	大規模並列/高性能計算 (通信、ネットワーク、インターコネクト、データ転送)
3	大規模並列/高性能計算 (ストレージ、I/Oシステム、データ管理)
	プログラミング環境 (プログラミングモデル)
	プログラミング環境 (コンパイラ)
	プログラミング環境 (デバッグ、テスト、再現性)
	データ学習&推論フレームワーク (TensorFlow, PyTorchなど)
データ処理&解析フレームワーク (Hadoop, Sparkなど)	
4	性能解析ツール (プロファイラ、トレーサー、性能解析方法論)
	性能解析ツール (性能モデル、シミュレーション)
5	利用高度化ツール (省電力化システム: DVFSなどアプリの消費電力最適化)
	利用高度化ツール (耐障害性システム: 高信頼通信ライブラリ、冗長符号化、チェックポイントリスタートなど)
	利用高度化ツール (ディスアグリゲーション技術)
6	資源管理 (ジョブスケジューラ、システム全体の電力マネジメント)
	資源管理 (リアルタイム処理)
7	外部資源連携 (分散ストレージ、広域ネットワーク、クラウド連携)

# お題

- 1) 10年後にあるべきスパコンの姿として、汎用指向 vs 専用指向のどちらを目指した開発をすべきか？
- 2) 次世代先端的計算基盤としてSociety5.0関連のアプリ（例えばAIやリアルタイムCPSアプリ）の重要性もさらに増すと考えられるが、従来のHPCアプリとの違いは何が考えられるか。またシステム面や運用面での課題は何か？
- 3) 今後最も重点的に取り組むべき研究開発要素は何か？
- 4) （もしあれば）今後のスパコン開発にあたってHPC関連ベンダーには技術面や運用面などで何を期待しているか？