

# Society 5.0 実現に向けた (計算+データ+学習) 融合

中島 研吾<sup>1)</sup>, 埴 敏博<sup>1)</sup>, 伊田 明弘<sup>1)</sup>, 下川辺 隆史<sup>1)</sup>, 三木 洋平<sup>1)</sup>, 星野 哲也<sup>1)</sup>,  
有間 英志<sup>1)</sup>, 田浦 健次朗<sup>1)</sup>, 工藤 知宏<sup>1)</sup>, 関谷 勇司<sup>1)</sup>, 中村 遼<sup>1)</sup>

1) 東京大学情報基盤センター

nakajima@cc.u-tokyo.ac.jp

## Integration of (Simulation + Data + Learning) towards Society 5.0

Kengo Nakajima<sup>1)</sup>, Toshihiro Hanawa<sup>1)</sup>, Akihiro Ida<sup>1)</sup>, Takashi Shimokawabe<sup>1)</sup>,  
Yohei Miki<sup>1)</sup>, Tetsuya Hoshino<sup>1)</sup>, Eishi Arima<sup>1)</sup>, Kenjiro Taura<sup>1)</sup>, Tomohiro Kudo<sup>1)</sup>,  
Yuji Sekiya<sup>1)</sup>, Ryo Nakamura<sup>1)</sup>

1) Information Technology Center, The University of Tokyo

### 概要

スーパーコンピューティングは従来の計算科学シミュレーション中心から、データ科学、機械学習との融合へと移行しつつある。(計算+データ+学習) 融合とそれによる Society 5.0 の実現を目指す東京大学情報基盤センターの取り組みについて紹介する。

## 1 背景：東大センターの現状

東京大学情報基盤センター [1] (以下「当センター」) は 1965 年に東京大学大型計算機センターとして設立されて以来 50 年余り、全国共同利用施設、学際大規模情報基盤共同利用・共同研究拠点の中核拠点、革新的ハイパフォーマンス・コンピューティング・インフラ (HPCI) [2] の構成機関として、国内外の産学官の各機関で実施されているスーパーコンピュータを使用した大規模シミュレーションによる計算科学・計算工学の研究の発展に貢献をしてきた。

2019 年 9 月現在、当センターでは 5 式 (Reedbush-U, Reedbush-H, Reedbush-L, Oakforest-PACS, Oakbridge-CX) のシステム [1] を運用しており、総利用者数は学内外を合計して約 2,000 名である。各システムは、高い計算性能、ユーザーフレンドリなプログラム開発環境、安定した運用が利用者に高く評価されている。

計算科学が「第三の科学 (The Third Pillar of Science)」と呼ばれるようになって久しいが、近年は様々なデータを活用することによって更に新しい科学を開拓する試みが始まっている。

当センターのシステムの利用分野についても、2018 年 3 月まで運用されていた Oakleaf-FX ((旧) 大規模超並列スーパーコンピューターシステム) 及び Oakbridge-FX (長時間ジョブ実行用並列スー

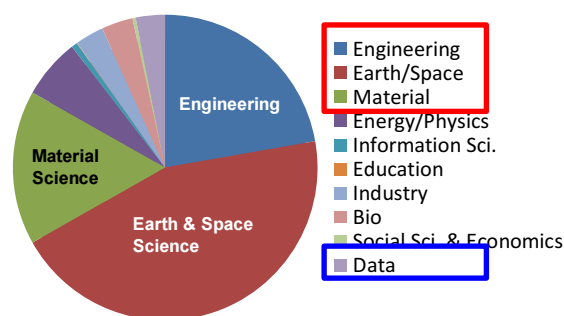


図 1 実行ジョブノード時間の分野別比率 (2017 年度) : (旧) 大規模超並列スーパーコンピューターシステム (Oakleaf-FX) ・長時間ジョブ実行用並列スーパーコンピューターシステム (Oakbridge-FX)

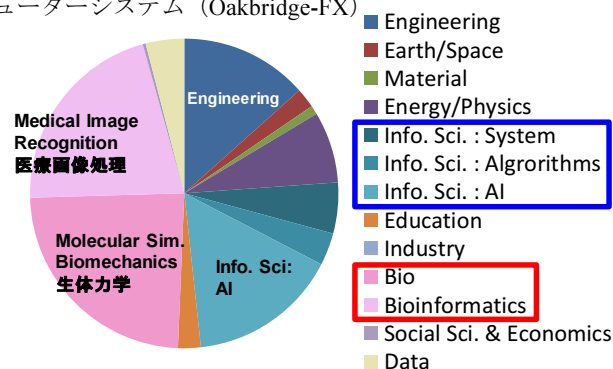


図 2 実行ジョブノード時間の分野別比率 (2018 年度) : データ解析・シミュレーション融合スーパーコンピューターシステム GPU 搭載ノード群 (Reedbush-H)

パーコンピューターシステム) では、ものづくり、地球・宇宙科学、物性科学などの計算科学・計算工学分野で広く利用されてきたが (図 1), 当センター初の GPU 搭載システムとして 2017 年 4 月に運用を開始した Reedbush-H (データ解析・シミュ

レーション融合スーパーコンピュータシステム GPU 搭載ノード群) は人工知能, 医療画像処理などより多様な分野で使用されている (図 2)。

## 2 BDEC システム

スーパーコンピューティングは, 従来の計算科学・計算工学シミュレーションにデータ科学, 機械学習等の知見を融合した多面的な視点に基づく手法の適用によって, 解をより効率的に得ることが可能となり, Society 5.0 が目指す超スマート社会の実現に大きく貢献すると期待される。

海外に目を向けてもアメリカエネルギー省のエクサスケールシステム計画の一つである Aurora/A21 システムのホームページ [3] では「シミュレーション (Simulation) + データ (Data) + 学習 (Learning) (S+D+L)」の融合が謳われている。

当センターでは 2015 年頃からこのような状況を想定し, 「S+D+L」を実現するプラットフォームとして『「計算・データ・学習」融合スーパーコンピュータシステム』(通称「BDEC (Big Data & Extreme Computing) システム」) 構築を目指して, 様々な研究開発を進めてきた。現在当センターで運用中の Reedbush (データ解析・シミュレーション融合スーパーコンピュータ, 2016 年 7 月運用開始), Oakbridge-CX (大規模超並列スーパーコンピュータシステム, 同 2019 年 7 月) はいずれも「BDEC システム」設計のためのプロトタイプ, 実証システムとしても位置づけられている。

BDEC システムは「S+D+L」を実現するために, 「内部ノード (Internal Nodes (INN), 従来のスパコン)」, 外部ネットワークに直接結合された

「外部ノード (External Nodes (EXN))」, 両者の連携・統合を司る「融合ノード (Integration Nodes (ITN))」を有し, 「S+D+L」の様々なワークロードが協調して実行されるこれまでにないタイプのシステムである (図 3, 図 4)。

「外部ノード」の各ノードは SINET 等の外部ネットワークを介して, サーバー, ストレージ, センサーネットワークを含む様々な外部リソースに直接アクセス可能である。

「融合ノード」は「内部ノード」と同じアーキテクチャを有し, 一体として運用される。また「外部ノード」と「融合ノード」は高速ファイルシステム (Fast File System, FFS) を共有している。また共有ファイルシステム (Shared File System, SFS) は, 当センター内の他のシステムから利用可

能とし, 更に BDEC 後継システムからも利用可能となることを想定して導入の予定である。

BDEC システムは東京大学柏 II キャンパスに建設中の「学術高速大容量ネットワーク拠点」に設置され, 2021 年 4 月以降に運用を開始する予定である。同拠点には, 後述の「データ活用社会創成プラットフォーム (データプラットフォーム)」[4] も設置され, BDEC システムとデータプラットフォームは, 連携して (S+D+L) 融合を実現することにより, Society 5.0 が目指す超スマート社会の実現に資するものである。

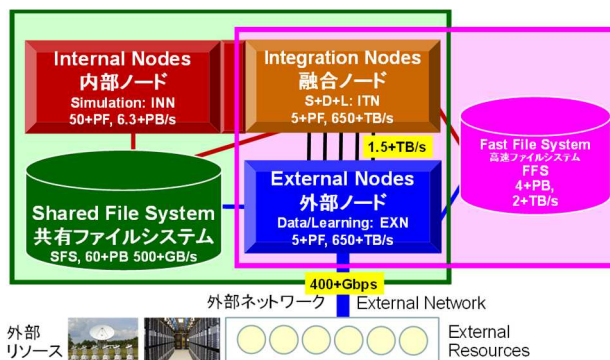


図 3 「計算・データ・学習」融合スーパーコンピュータシステム (BDEC, Big Data & Extreme Computing)

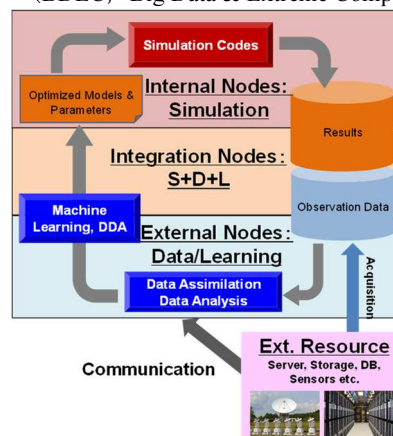


図 4 BDEC システムを使用した (計算+データ+学習 (S+D+L)) 融合

## 3 h3-Open-BDEC

(計算+データ+学習) 融合のためには, BDEC のようなこれまでにない革新的なハードウェアが必要であるが, 様々なアプリケーション, ワークロードを BDEC 上で開発, 実行していくためのソフトウェア群も重要である。当センターでは, (計算+データ+学習) 融合のためのプラットフォームである BDEC システムの能力を最大限引き出し, 最小の計算量・消費電力での計算実行を実現するために:

- ① 変動精度演算・精度保証・自動チューニングによる新計算原理に基づく革新的高性能・高信頼性・省電力数値解法
- ② 機械学習による革新的手法である階層型データ駆動アプローチ

ソフトウェアと共用データの提供

データプラットフォームは、BDEC と同じく「学術高速大容量ネットワーク拠点」内の同じ部屋に設置され、2021年3月までに稼働を開始する予定である。

の2項目を中心とした研究を推進し、その成果を革新的ソフトウェア基盤「h3-Open-BDEC」として整備する(図5)。「h3」はBDECシステムがHierarchical, Hybrid, Heterogeneous (h3)な特徴を持っていることに由来している。また、h3-Open-BDECは、その名称の示す通り、オープンソースとして公開し、BDECシステムのみならず、様々なスパコン上で利用可能とする。

## 5 まとめ

スーパーコンピューティングは従来の計算科学シミュレーション中心から、データ科学、機械学習との融合による新しいスタイルへと移行しつつある。本稿では、BDECシステム、h3-Open-BDEC、データプラットフォームなど、(計算+データ+学習)融合とそれによるSociety 5.0の実現を目指す東京大学情報基盤センターの取り組みについて紹介した。当日の発表では、各計画の進捗状況も併せて紹介する予定である。

## 4 データ活用社会創成プラットフォーム(データプラットフォーム)

東京大学及び当センターでは、Society 5.0におけるビッグデータ等の新たな技術をあらゆる産業や社会生活に取り入れてイノベーションを創出するための基盤として、以下のような機能を持つデータ活用社会創成プラットフォーム構想の検討を、学内各部署の他、全国の大学や諸機関と協力して進めている。

## 参考文献

- ① データ収集機能：IoT データや大規模リアルタイムデータを円滑に扱うことのできるセキュアな大容量通信回線の提供
- ② データ解析機能：高度・高速な解析を実現する高性能計算環境、ストレージの整備
- ③ 応用開発基盤：多様な応用を実現する基盤ソ

- [1] 東京大学情報基盤センター(スーパーコンピューティング研究部門・スーパーコンピューティングチーム)：  
<https://www.cc.u-tokyo.ac.jp/>
- [2] 革新的ハイパフォーマンス・コンピューティング・インフラ(HPCI)；  
<http://www.hpci-office.jp/>
- [3] Aurora/A21 System, Argonne National Laboratory: <https://www.anl.gov/topic/aurora>
- [4] 田浦健次郎, データ活用社会創成プラットフォーム計画について, 東京大学情報基盤センター年報第20号(2018年度), 54-59, 2019
- [5] h3-Open-BDEC: <http://nkl.cc.u-tokyo.ac.jp/h3-Open-BDEC>

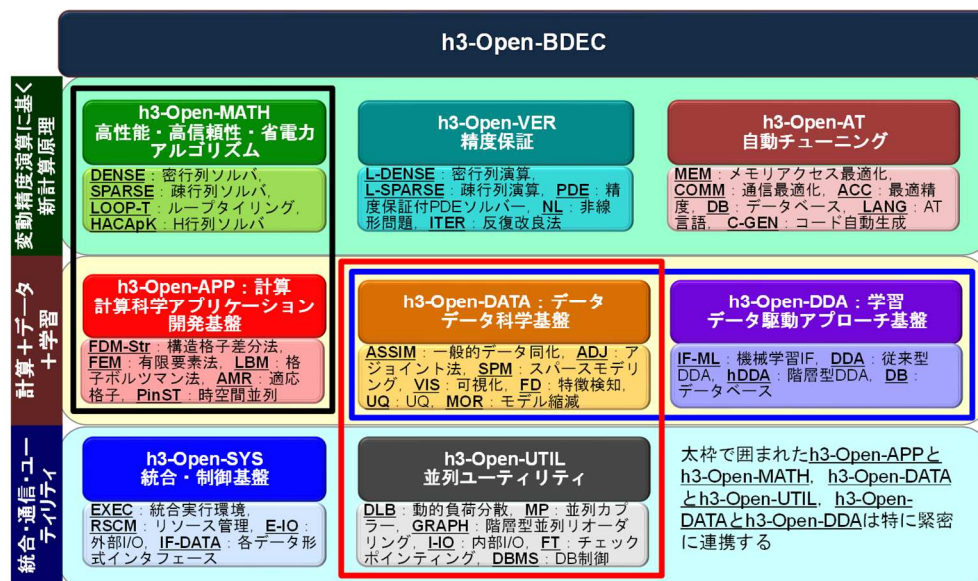


図5 h3-Open-BDECの概要と機能