

巻頭言

中島 研吾

東京大学情報基盤センター

新しい年、2020 年が幕を開けた。今年は東京でオリンピック、パラリンピックが開催される。2020 年 1 月現在、東京大学情報基盤センター（以下「当センター」）では 5 式（Reedbush-U, Reedbush-H, Reedbush-L, Oakforest-PACS, Oakbridge-CX）のシステムを運用しており、総利用者は学内外を合計して約 3,000 名強と過去最大となっている。

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。当センターのスパコンはものづくり、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。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、機械学習による医療画像処理など、更に多様な分野で使用されている。昨今増加している、機械学習、人工知能分野の利用者は、よりインタラクティブな利用環境を求める傾向にあり、多様なニーズに応えることも当センターの重要な課題である。

スーパーコンピューティングそのものも変革を遂げつつあり、従来の計算科学・計算工学シミュレーションとデータ科学、機械学習の融合が、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ものと期待されている。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 システムでも「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に注目し、「S+D+L」融合を実現するプラットフォーム構築を目指して、ハードウェア、ソフトウェア、アプリケーション、アルゴリズム等の様々な観点から研究開発を継続して実施してきた。

当センターでは、2021 年春の運用開始を目指して、「（計算・データ・学習）融合スーパーコンピュータシステム（BDEC, Big Data & Extreme Computing）」、「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, mdx）」の 2 つのシステムの導入を進めている。BDEC は「S+D+L」の「S」に着目し、伝統的なスパコンの延長線上で、データ科学・機械学習の活用により、計算科学・計算工学シミュレーションをより効率的に実施することを目指している。mdx は「D+L」に力点を置いたクラウド的なシステムであり、より柔軟でインタラクティブな利用環境を提供する。両システムを合わせて倍精度浮動小数点演算性能換算で 60PLOPS 以上の理論演算性能を有する超大規模なシステムである。両システムは現在本学柏 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」の同じマシンルームに設置される予定である。また、ハードウェアだけでなく、ソフトウェア、アプリケーション開発環境の整備も同時に進めている。特に BDEC 向けに開発が進められている h3-Open-BDEC（h3: Hierarchical, Hybrid, Heterogeneous）は「S+D+L」融合を、計算科学・計算工学の専門家がデータ科学・機械学習の専門家の力を借りることなく実現できるアプリケーション開発環境の実現を目指している。

BDEC と mdx によって「S+D+L」融合が実現するまで、あと 1 年あまり、2020 年は新しいスーパーコンピューティング実現に向けた重要な年となる。