



平成 30 年 6 月 13 日

国立大学法人東北大学東北メディカル・メガバンク機構
国立研究開発法人日本医療研究開発機構

次世代医療の開発を加速！
ゲノム医科学用供用スーパーコンピュータ
AMED/ToMMo のシステムを拡張して全国利用体制を構築へ

【発表のポイント】

- ・国立研究開発法人 日本医療研究開発機構 (AMED) はゲノム医療研究の推進加速のため、東北大学東北メディカル・メガバンク機構 (ToMMo) に運用を委託しているスーパーコンピュータの 2/3 を公開・分譲区画 (外部からアクセス可) となるよう更新し、AMED 事業のみならず全国の研究者が利用可能とした。
- ・データ及び計算・解析機能の共有 (シェアリング) により様々な機関による医療、医学の研究が効率的になる見込み。「Society5.0^{*1}」時代のヘルスケアとして提唱され、東北メディカル・メガバンク計画が目指してきた個別化医療をはじめとする次世代医療の開発が近づく。
- ・AI による解析・ディープラーニング^{*2}に適した GPGPU^{*3}ノードを強化し、ビッグデータ解析に向けて最新の解析環境を整備。ストレージも追加拡張して保存データ量が増大、世界でも有数の複合バイオバンクの基盤整備へ。

【概要】

解析技術の急速な進展等により、ライフサイエンスに関わるデータ量は膨大となり、研究にはビッグデータの解析、つまりスーパーコンピュータの利用が必須となりつつあります。しかし、スーパーコンピュータを所有できる研究組織は限られ、多くの研究者が計算資源の確保に苦慮しています。

そこで、ToMMo は、AMED からの委託のもと、運用してきたスーパーコンピュータの更新に取り組みました。従来の東北メディカル・メガバンク計画推進のためのゲノ

ム・オミックス解析、情報の保管、分譲に加えて、全体の 2/3 を公開・分譲区画(外部からアクセス可)とし、AMED が推進する「疾病克服に向けたゲノム医療実現プロジェクト」における供用のほか、全国の研究者が利用可能なものとなりました。この更新は、昨今重要性が叫ばれるデータ及び計算・解析機能の共有(シェアリング)に対応したものです。

また今回の更新では、①東北大学外部からのアクセスが可能な公開・分譲区画の大幅な拡張、と共に、②GPGPU ノードを増強し人工知能(AI)を活用したディープラーニング(深層学習)を含む多様な計算に対応、③ストレージの拡張によるバイオバンク機能の増強がはかられ、日本のゲノム医療研究を支える基盤としての機能をより一層充実させたものとなっています。

本システムは、全国規模のゲノム医療研究を加速させるための解析データの利活用基盤として用いられ、政府の第5期科学技術基本計画に定められている「Society5.0」時代のヘルスケアとして求められている次世代医療の本格的な実現に向けて大きく寄与すると考えられます。

【詳細】

ToMMo は、大規模なコホート調査 *4とそれに由来する多数の生体試料のゲノム・オミックス解析 *5によって生じる膨大なデータを解析し、セキュリティを保って保管するために、スーパーコンピュータシステムを平成 26 年に導入し、同年 7 月から本格運用させ、これまで利活用を行ってきました(納入企業:株式会社日立製作所)。

一般的にコンピュータ機器は定期的なリプレイス等が必要となります。ToMMo が導入したスーパーコンピュータシステムは、導入当時で東京以北の生命科学系で最大規模のものでしたが、国際的な競争力を維持するためにも、更新が必要とされてきました。

ToMMoとAMEDは、ToMMoが導入したスーパーコンピュータシステムが本格運用から約 4 年が経過したのを機に更新するにあたり、今般我が国の科学研究において大きな課題となっているデータ及び計算・解析機能のシェアリングを推進するよう設計しました。平成 29 年度に、システム構築が完了し、平成 30 年度に供用利用を本格的に開始します。

今回更新されたスーパーコンピュータシステムの特長は以下の通りです。

- 計算ノード数は台数を最適化し、GPU を 16 基から 24 基(1 台あたり 8 基搭載している NVIDIA の DGX-1(Volta)を 3 台導入)とし、理論演算性能が GPU として 10 倍(倍精度で 18,720GFLOPS*6→187,200GFLOPS)となった。Volta は半精度(FP16)計算に適した Tensor コアを有しており、ディープラーニングに適した構成となった。

(計算速度向上の具体例)

同一サンプルの全ゲノム解析を実行して比較した場合、以前のシステムでは 30 時間かか

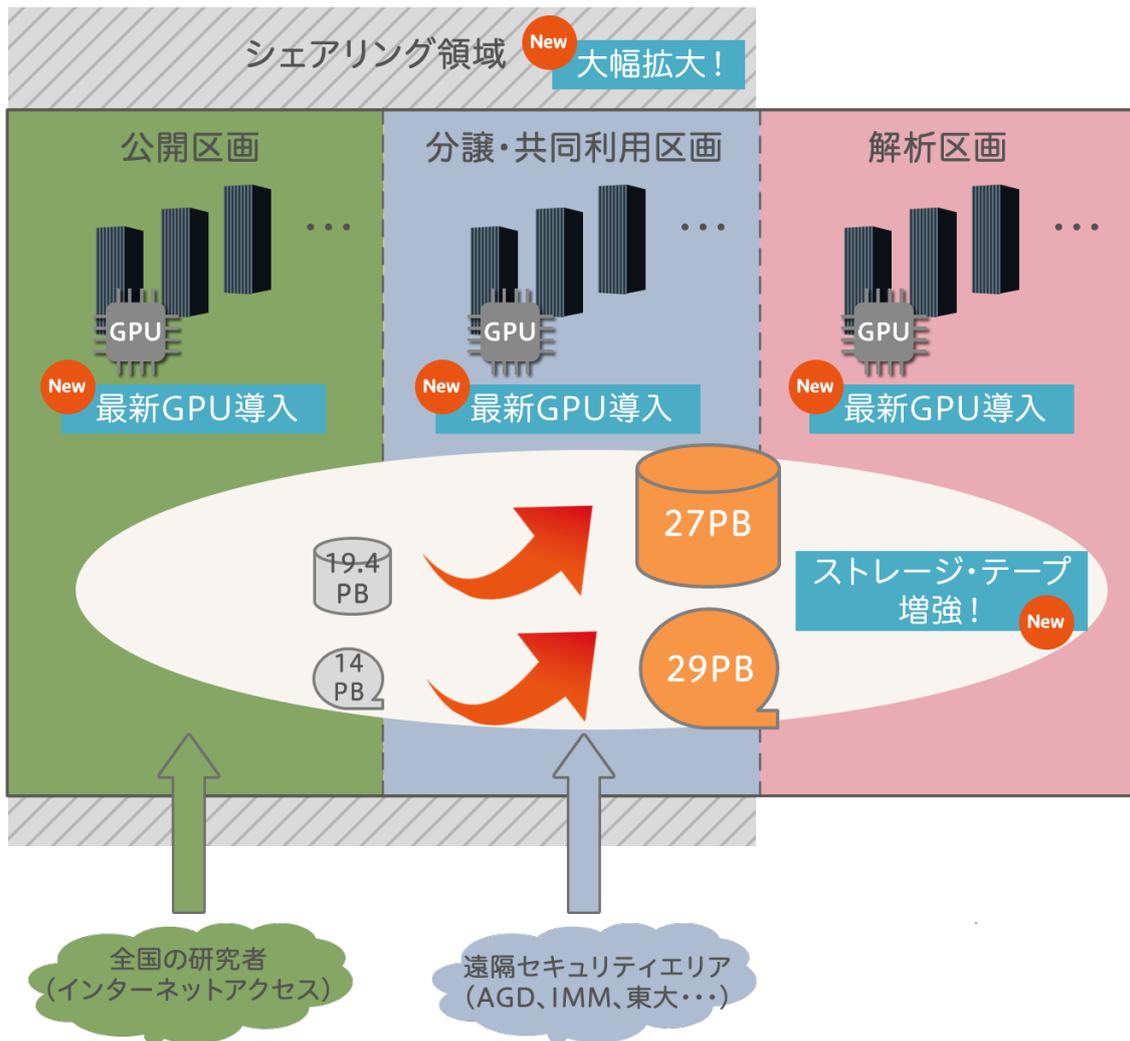
っていたものが、今回の更新により 22 時間で終了するようになった。2～3 割程度の速度短縮が実現 (CPU のみの計算結果)。GPGPU (汎用画像処理装置) を活用することにより、処理内容によっては 40 倍程度にも達する高速化が期待される。

- ストレージは既存システムの 19.4PB を一部利用停止したうえで 21PB 増強し、合計 27PB に到達した。うち、データシェアリングに利用している容量は、4.2PB から 8.5PB と倍増した。
- バックアップ用のテープカートリッジも従来の 14PB に加え、15PB を増強し、合計 29PB に到達した。
- CPU はスーパーコンピュータとしては最新世代の Skylake を導入し、前システムから 3 世代更新された。

また、現在、推進・計画されている全国の研究機関による供用は以下の通りです。

- 遠隔セキュリティエリア *7 を通じた利活用
岩手医科大学いわて東北メディカル・メガバンク機構、東京大学大学院医学系研究科人類遺伝学教室、国立成育医療研究センター、ライフサイエンス統合データベースセンター、日本製薬工業協会など、全国で 15 カ所程度。
- AMED 事業を通じた利活用
AMED が推進する「疾病克服に向けたゲノム医療実現プロジェクト」の下で推進される「ゲノム医療実現推進プラットフォーム事業」や「臨床ゲノム情報統合データベース整備事業」の研究課題での利活用を推進しており、今後、AMED 全体のゲノム解析等を伴う研究開発課題に拡大を予定しています。

ToMMo と AMED は今後も構築したシステムを活用して、多様な機関による利活用を推進し、日本のゲノム医療基盤を強固なものにするべく取り組んでいきます。そして、患者一人ひとりに最適な治療を提案する「個別化医療」を推進し、社会に成果を還元していきます。



【参考】

<東北メディカル・メガバンク計画について>

東北メディカル・メガバンク計画は、東日本大震災からの復興と、個別化予防・医療の実現を目指しています。東北大学東北メディカル・メガバンク機構と岩手医科大学いわて東北メディカル・メガバンク機構を実施機関として、東日本大震災被災地の医療の創造的復興および被災者の健康増進に役立てるために、平成 25 年より合計 15 万人規模の地域住民コホート調査および三世代コホート調査等を実施して、試料・情報を収集したバイオバンクを整備しています。東北メディカル・メガバンク計画は、平成 27 年度より、国立研究開発法人日本医療研究開発機構 (AMED) が本計画の研究支援担当機関の役割を果たしています。

【用語等説明】

* 1. Society 5.0 : 超スマート社会。フィジカル空間(現実空間)とサイバー空間(仮想

空間)を高度に融合させた社会、システムを指す。今後大量に生み出される M2M(Machine to Machine)、IoT 等のビッグデータを AI、人工知能を用いて解析し、フィジカル空間にフィードバックすることで高付加価値情報、提案、機器への指示を可能とする。中でもヘルスケア分野においては健康寿命の延伸に向けて、QOL(Quality of life)/QOS(Quality of society)の向上に寄与し、治療、診断においては未病ケア、予防への寄与、個別化医療の実現が期待される。

- *2. ディープラーニング：機械学習の手法のひとつである深層学習のこと。十分なデータがあれば、人間からの指示がなくとも機械的にデータから特徴を抽出、学習を高精度に行うことができる手法で、ビッグデータの解析手法のひとつとして大きな注目を集めている。
- *3. GPGPU : **General-purpose computing on graphics processing units**; GPU による汎用計算。GPU はリアルタイム画像処理に特化した演算装置ないしプロセッサで、定型的な大量の演算を並列に処理する性能に優れている。GPGPU は GPU を画像処理以外の目的に汎用化させること。
- *4. コホート調査：ある特定の人々の集団を一定期間にわたって追跡し、生活習慣などの環境要因・遺伝的要因などと疾病発症の関係を解明するための調査のこと。
- *5. ゲノム・オミックス解析：生物の遺伝情報の総体であるゲノム情報と、メタボローム(代謝物)、プロテオーム(タンパク質)、トランスクリプトーム(メッセンジャー RNA)などの総称であるオミックス解析を合わせて指す。解析結果は膨大なデータ量となる。
- *6. FLOPS：フロップス、**F**loating-**p**oint **O**perations **P**er **S**econd はコンピュータの性能指標の一つ。一秒間に実行できる浮動小数点演算の回数を示す。本文書内にある GFLOPS、TFLOPS はそれぞれ、ギガ (G)フロップス(10^6 フロップス)、テラ(T)フロップス(10^9 フロップス)である。
- *7. 遠隔セキュリティエリア：東北大学東北メディカル・メガバンク機構に置かれたスーパーコンピュータに対して、高いセキュリティを保持して、外部からアクセスするために設けられた管理エリア。入退室が監視カメラ等で記録されると共に、生体認証等が導入され、シンクライアントを通じた画像転送によってスーパーコンピュータ内の情報を閲覧することができる仕組み。

【お問い合わせ先】

(研究に関すること)

東北大学東北メディカル・メガバンク機構

ゲノムプラットフォーム連携センター

センター長 木下賢吾 (きのした けんご)

電話番号：022-274-5952

Eメール：kengo@ecei.tohoku.ac.jp

(報道に関すること)

東北大学東北メディカル・メガバンク機構

長神 風二 (ながみ ふうじ)

電話番号：022-717-7908

ファクス：022-717-7923

Eメール：f-nagami@med.tohoku.ac.jp

(AMED 事業に関すること)

国立研究開発法人日本医療研究開発機構

基盤研究事業部 バイオバンク課

電話番号：03-6870-2228

Eメール：genome-support@amed.go.jp