

国立遺伝学研究所が、 生命科学・ゲノム医療研究を加速させる 「第5世代遺伝研スパコン」を構築

HPEの多様なHPCプラットフォームを全面採用し
従来の約2.2倍となる総理論演算性能、約1.1PFLOPSを実現

“膨大なゲノム断片をつなぎ
合わせて配列を復元するア
センブリ処理では、10TBを
超えるような大規模メモリ
空間が非常に有効です”

—大学共同利用機関法人
情報・システム研究機構
国立遺伝学研究所
DDBJセンター
システム管理部門長
特任准教授
博士(理学) 小笠原 理 氏



大学共同利用機関法人 情報・システム研究機構
国立遺伝学研究所

目的

ゲノム解析およびゲノムデータベース構築のためのスーパーコンピューティングシステムの最新化。増大する解析要求に応えるリソースを確保するとともに、多様な要求に適材適所のリソースで応えること。

アプローチ

メモリ共有型、クラスター型、GPU搭載型など性格の異なるプラットフォームを整備し、多様な解析ニーズに対して最適なリソースを提供する。さらに、事前準備などの研究者の負荷を軽減する環境を整備する。

ITの効果

- 大規模メモリ共有型システム、大規模クラスター型システム、GPUシステムそれぞれにHPEプラットフォームを適材適所で採用し「Right Mix」のゲノム解析基盤を実現
- 12TB メモリを単一OSから利用できるHPE Superdome Flexが、ゲノム配列を復元するアセンブリ処理に大きな威力を発揮
- クラスター管理ツール「Bright Cluster Manager」により、OS、ジョブスケジューラー、AI関連のフレームワークやライブラリなどの管理と展開が容易に

ビジネスの効果

- ゲノム分野ならではの多様な解析要求に最適な計算リソースを割り当て、スーパーコンピュータシステムのコストパフォーマンスを最大化
- 研究者による解析事前準備にかかる作業を解消し、研究業務への注力を可能に
- AIを活用した解析手法の確立とバイオインフォマティクス研究の前進に期待



大学共同利用機関法人
情報・システム研究機構
国立遺伝学研究所
DDBJセンター
システム管理部門長
特任准教授
博士(理学) 小笠原 理氏

2019年3月、国立遺伝学研究所がスーパーコンピューターシステムを刷新し、従来比およそ2.2倍となる総理論演算性能、約1.1PFLOPSを実現した。この「第5世代遺伝研スパコン」は、世界中の研究者が塩基配列データを登録する「国際塩基配列データベース (INSD)」をはじめ、ゲノム関連の多様な解析処理を担う環境として活用され、我が国における生命科学・ゲノム医療研究の中核と位置づけられるものだ。大規模なメモリ共有型システム、クラスター型システム、GPUシステム、統合的な運用管理環境まで、HPEのHPCソリューションが全面的に採用された。

チャレンジ

日本のゲノム医療研究の中核拠点が スパコンシステムを最新化

国立遺伝学研究所(以下、遺伝研)は、我が国における生命科学・ゲノム医療研究の中核拠点であり、遺伝学の先端的研究、遺伝資源の保存と利用、遺伝情報データベースの整備と利用、遺伝学を基盤とした高度な教育と人材育成を担っている。また、大学共同利用機関法人 情報・システム研究機構の一員として、スーパーコンピューターシステムの共同利用や共同研究の場を広く研究者に提供している。遺伝研 特任准教授 博士(理学)であり、DDBJセンターのシステム管理部門長を務める小笠原理氏は次のように話す。

「DDBJ (DNA Data Bank of Japan) センターでは、1996年よりスーパーコンピューターシステム(遺伝研スパコン)を運用しており、そのミッションは大きく2つあります。まず、米国NCBIおよび欧州EBIとの協力による『国際塩基配列データベース (INSD)』の構築。もうひとつは、生命科学・ゲノム医療の研究者への強力な計算リソースの提供です」

世界中のゲノム研究者による研究成果は、DDBJ・NCBI・EBIいずれかのデータベースに登録することが義務づけられている。この日米欧3極のデータベースが相互にデータを交換することで、「世界共通の塩基配列データベース」としての運用が可能になっているのだ。INSDに登録されたデータは科学資料として永久に保存され、公開されたデータは誰でも制限なしで利用できる。いわば人類共有の財産である。遺伝研スパコンは、INSDの構築に加え、個人ゲノムからメタゲノムまで、あるいは分子レベルから生物集団レベルまで、多様かつ膨大な解析処理を長年にわたり担ってきた。

「ヒトゲノムの解析と臨床応用に向けた次世代シーケンサーの利用は、爆発的なデータの増大をもたらしました。私たちは、2012年に遺伝研スパコンのアーキテクチャーを大きく見直し、オープンソースソフトウェアへの移行でコストを抑えながら従来比15倍の性能を獲得しました。本システムで初めて採用したのが、HPE(当時はHP)製のプラットフォームです」(小笠原氏)

以来、コストパフォーマンスに優れたHPEサーバー製品が遺伝研スパコンを支えてきた。そして2019年3月、最新の大規模メモリ共有型システムと大規模クラスター型システムから構成される「第5世代遺伝研スパコン」の運用を開始した。

ソリューション

ゲノム解析の多様な要求に 最適な計算リソースを提供

「第5世代遺伝研スパコン」は、従来比2.2倍となる総理論演算性能、約1.1PFLOPSを発揮する。研究者の要求に合わせて最適な解析リソースの提供が可能で、コンテナ仮想化技術を利用して研究者による解析環境のセットアップを容易にしたこと、仕様の異なるクラスター環境の管理を容易にしたことが大きな特長だ。

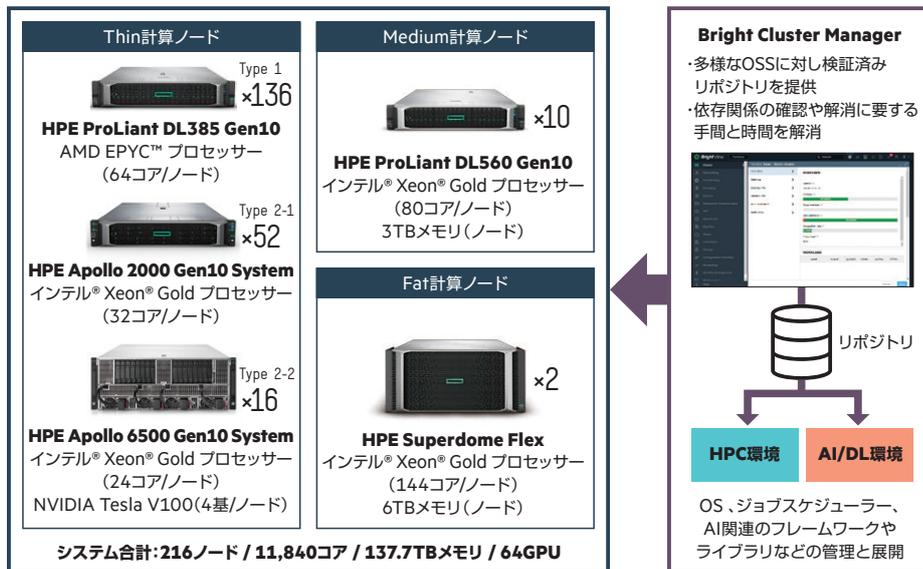
「基本的なアーキテクチャーを継承しながら、解析パフォーマンスの強化とストレージ容量の拡充を図りました。解析リソースは、12TBの大容量メモリを利用できる『Fat計算ノード』、3TBメモリを備える『Medium計算ノード』、計204台のサーバーでクラスターを構成する『Thin計算ノード』から提供されます。AIによるゲノム解析を推進するために最新のGPUマシンも導入しました」(小笠原氏)

Fat計算ノードに採用されたのは「HPE Superdome Flex」。最大48TBの単一メモリ空間を1つのOSから利用できる業界最大級のSMPシステムだ。クロスバーチップをメッシュ状に構成する独自のクロスバーファブリックによりノード間接続のレイテンシを最小化し、HPC領域での「大規模インメモリ処理」のメリットを最大化できる。

「膨大なゲノム断片をつなぎ合わせて配列を復元するアセンブリ処理では、10TBを超えるような大規模メモリ空間が非常に有効です。一般の研究機関では導入の難しい大規模解析システムを、大学共同利用機関法人である遺伝研が提供できることに大きな意義があります」と小笠原氏は話す。



第5世代 遺伝研スーパーコンピューターシステム



Thin計算ノードの中で最大となる136ノードで構成される「Type 1」で目を引くのは、32コアのAMD EPYC™ プロセッサーを搭載する「HPE ProLiant DL385 Gen10サーバー」を採用していることだ。

「参照ゲノム配列に対するマッピング処理など、整数演算性能の高さが有利になるアプリケーションが数多くあります。Thin計算ノードType 1では、AMD EPYC™ プロセッサーのメニーコアと広いメモリ帯域を利用できるメリットが大きいと考えました」(小笠原氏)

「HPE Apollo 6500 Gen10 System」はThin計算ノードType 2-2に分類されている。小笠原氏は、最新のNVIDIA Tesla V100を最大8基搭載可能なGPUサーバーを採用した狙いを、「ひとつは、GPUならではの多数のコアを活かした並列処理への適用、もうひとつは、AIを活用したゲノム解析を推進するためのディープラーニングモデル構築」と説明する。

「第5世代遺伝研スパコン」に登録されているゲノム解析ソフトウェアは2,300種類以上。多様な解析要求に最適な計算リソースで応えるために、本システムは異なる特徴を備えたHPEプラットフォーム群による「Right Mix (適材適所)」でのリソース活用が実現されている。

複雑なクラスター環境の管理にBright Cluster Managerを活用

「第5世代遺伝研スパコン」には、大規模かつ複雑な環境を統合的に管理するとともに、研究者の要求に対して迅速かつ効率的に対応するための工夫が施されている。その重要な役割を担っているのが「Bright Cluster Manager」である。Bright Cluster Managerは、直感的なGUIを備え、プロビジョニング、イメージ管理、クラスター監視など包括的な機能を提供するクラスターマネジメントツール。AI/ディープラーニング環境、HPCシステム双方で豊富な実績がある。

「従来からのゲノム解析要求と、AI/ディープラーニングを利用した新しい解析アプローチに同時に応えるために、かねてから注目していたBright Cluster Managerを採用しました。OS、ジョブスケジューラー、AI関連のフレームワークやライブラリなどの管理と展開が容易になります。Singularityコンテナを扱えることも大きなメリットです」(小笠原氏)

遺伝研では、研究者によるゲノム解析ソフトウェアのインストールを容易するとともに、解析環境の再現性を担保するためにSingularityコンテナを利用している。これにより、アプリケーション、ライブラリ、データをコンテナイメージとしてパッケージ化し、再現性、可搬性、可用性に優れた解析環境を実現している。

「コンテナを管理するKubernetesやMesosなどのミドルウェアを試すにあたって、複雑なインストール作業や正しく動作する組み合わせの確認が大きなストレスになっていたのです。これらを自動化するツールを探す過程でBright Cluster Managerに出会いました。常に新しいテクノロジーを検討・検証しなければならない私たちにとって、理想的なソリューションとなりました」と小笠原氏は振り返る。

Bright Cluster Managerで提供されるリポジトリは、AI/ディープラーニングで利用する多様なOSSが正しく動作する組み合わせで提供される。依存関係の確認や解消に要する手間と時間を解消できるメリットが評価されて、国内でも導入が進んでいる。日本ヒューレット・パカードは、正規代理店としてBright Cluster Managerの日本語サポートに国内で唯一対応。HPE Pointnextによる導入サービスも充実させている。

Case study

業界

大学共同利用機関法人
情報・システム研究機構
国立遺伝学研究所

研究機関

ソリューション概略

導入ハードウェア

- HPE Superdome Flex
- HPE ProLiant DL560 Gen10
- HPE Apollo 6500 Gen10 System
- HPE Apollo 2000 Gen10 System
- HPE ProLiant DL385 Gen10

導入ソフトウェア

- Bright Cluster Manager

“第5世代遺伝研スパコンは、国際塩基配列データベース (INSD) の構築と共有、強力な解析リソースの提供を通じて、生命科学・ゲノム医療の研究者の活動に寄与するために構築されました。計算リソースを強化しただけでなく、事前準備の軽減など研究者の使いやすさを大きく改善させています。ここから、新たなバイオインフォマティクス研究の成果が次々と生み出されることでしょう”

大学共同利用機関法人 情報・システム研究機構 国立遺伝学研究所
DDBJセンター システム管理部門長 特任准教授 博士(理学) 小笠原 理 氏

ベネフィット

多層防御の一環として

Silicon Root of Trustを活用

「第5世代遺伝研スパコン」では、個人ゲノム解析のための区画が新設された。その狙いについて小笠原氏は次のように説明する。

「個人ゲノムの解析要求は年率2倍のペースで増大しており、他の領域の伸びを大きく上回っています。他と大きく異なる要件は厳格なセキュリティが求められることです。世界中の研究者の間でも、個人ゲノムデータを安全に共有するための手法の確立は最重要の課題と認識されています」

個人ゲノム解析区画には、厳格な認証とアクセス制御を適用するとともに、不正アクセスや不審な挙動を検知するシステムが導入された。さらに、小笠原氏は「安心・安全」のためにもうひとつ手を打っている。

「個人ゲノム解析区画で採用しているHPE ProLiant DL385 Gen10サーバーは、ハードウェア主導でセキュリティを強化しており、ファームウェア改ざんなどの不正を検知して修正することができます。これを多層的な防御のひとつとして役立てています。コモディティ化しているx86サーバーの中にあって、HPEの「iLO 5」は際立った特徴ではないでしょうか」

HPE ProLiant DL385やHPE Apollo 6500 SystemなどのHPE Gen10サーバープラットフォームは、独自開発のASIC「HPE Integrated Lights-Out 5 (iLO 5)」を標準で装備。Silicon Root of Trust (シリコンレベルの信頼性) を実装し、CNSA (Commercial National Security Algorithm Suite) による商用で最強クラスのデータ暗号化に対応するとともに、NIST (米国国立標準技術研究所) が発行する「Special Publication 800-53」にも準拠している。

プロジェクトを振り返って、小笠原氏が次のように語って締めくくった。

「第5世代遺伝研スパコンは、国際塩基配列データベース (INSD) の構築と共有、強力な解析リソースの提供を通じて、生命科学・ゲノム医療の研究者の活動に寄与するために構築されました。計算リソースを強化しただけでなく、事前準備の軽減など研究者の使いやすさを大きく改善させています。ここから、新たなバイオインフォマティクス研究の成果が次々と生み出されることでしょう。日本ヒューレット・パッカードには、最先端のHPCソリューションからコンサルティングサービスまでフルポートフォリオでの支援を期待しています」

詳しい情報

HPE ハイパフォーマンスコンピューティングソリューションについてはこちら

hpe.com/jp/hpc

お問い合わせはこちら

カスタマー・インフォメーションセンター

0120-268-186 (または03-5749-8279)

月曜日～金曜日 9:00～19:00

(土曜日、日曜日、祝日、年末年始、および5月1日お休み)

日本ヒューレット・パッカード株式会社
〒136-8711 東京都江東区大島 2-2-1

**Hewlett Packard
Enterprise**

© Copyright 2019 Hewlett Packard Enterprise Development LP

本書の内容は、将来予告なく変更されることがあります。日本ヒューレット・パッカード製品およびサービスに対する保証については、当該製品およびサービスの保証規定に記載されています。本書のいかなる内容も、新たな保証を追加するものではありません。日本ヒューレット・パッカードは、本書中の技術的あるいは校正上の誤り、脱字に対して、責任を負いかねますのでご了承ください。記載されている会社名および商品名は、各社の商標または登録商標です。

CPC14029-01 記載事項は個別に明記された場合を除き2019年7月現在のものです。