

国立遺伝学研究所のスーパーコンピュータシステムを構築

新型 DNA シーケンサーのビッグデータ解析を加速

株式会社日立製作所執行役社長:中西 宏明/以下、日立)は、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(所長:小原 雄治/以下、遺伝研)が国際 DNA データベース構築事業を行うためのスーパーコンピュータシステムを構築し、このたび本格稼働を開始しました。

遺伝研は、本システムを活用して、米国・欧州の研究機関などと連携した国際塩基配列データベースの構築と、世界中の研究者に対する公開、研究者が新型 DNA シーケンサー^(*)の出力情報などのビッグデータ解析を加速するための IT リソースの提供を行います。これにより、遺伝研は世界中の研究者による生活習慣病やがんなどの原因解明をはじめ、さまざまな病気の予防、診断、治療に寄与する医学研究や、生物の遺伝学や分子生物学に関わる基礎および応用研究のさらなる発展をサポートしていきます。

遺伝研は、米国立生物工学情報センター(以下、NCBI)により運営されている GenBank および、欧州バイオインフォマティクス研究所(以下、EBI)により運営される EMBL-Bank との密接な連携のもと、国際塩基配列データベースを構築しています。国際塩基配列データベースは目的や国籍に拘らず閲覧転用できる世界科学の共有財産であり、世界中の研究者は NCBI、EBI、ならびに遺伝研の 3 つの機関を通じて国際塩基配列データベースに自らの研究データを登録することができます。この国際塩基配列データベースが扱うデータ量は近年爆発的に増加しています。このデータ量増加の背景としては、昨今、ゲノム研究分野での新型 DNA シーケンサーの発展が著しく、ヒトゲノムをはじめとした各種生物のゲノム情報の網羅的解析が加速していることがあります。最新の新型 DNA シーケンサーでは、断片配列データと呼ばれる DNA の配列を解析単位に断片化したデータが、一回の計測で約 60 億個(約 1.8TB)も生成されます。この新型 DNA シーケンサーの登場により、世界中の研究機関で生成される断片配列データ量は爆発的に増加しており、2010 年に生成された塩基配列^(*)の断片データ量は PB(ペタバイト)^(*)単位となります。さらに、第 3 世代、第 4 世代の DNA シーケンサーの発展が予想され今後もデータ量の増加が見込まれます。

このため遺伝研では、スーパーコンピュータシステムを刷新するにあたり、刻々と発生するビックデータの解析を迅速に行うことができる DNA 解析計算速度と、ビックデータを迅速に扱えるデータストレージなどのシステム環境を求めていました。また、消費電力の削減も実現したいという希望を持っていました。

日立は、これまでのゲノム分野でのスーパーコンピュータシステムの構築を含めた、さまざまなライフサイエンス関連のシステムでの構築、運用などで培ったノウハウを活用し、DNA シーケンサーから発生するビックデータの解析を可能とするだけでなく、これまでより省電力で効率的なゲノム解析を実現するスーパーコンピュータシステムを構築しました。

なお遺伝研は 2014 年に、今回導入したシステムと同等規模以上のシステム増強を予定しています。

■用語説明

*1 新型 DNA シーケンサー:化学処理した DNA サンプルに対して様々な分析処理を行うことで、DNA の塩基配列を自動的に読み取る装置。

*2 塩基配列:核酸を構成する塩基部分の配列

*3 PB(ペタバイト):約 1,000 兆バイト

■スーパーコンピュータシステムの特徴

1. ファイルシステムによるビッグデータの高速処理

今回導入されるスーパーコンピュータシステムでは、新型 DNA シーケンサーの登場により発生するビッグデータの高速処理を実現するために必要とされる、超高速アクセスが可能なファイルシステムを実装しています。

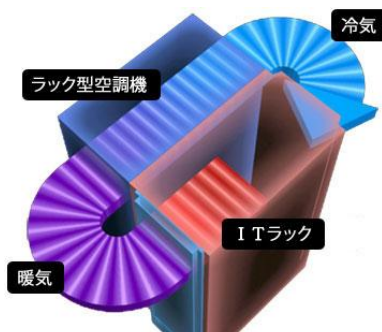
具体的な機能として、ひとつのデータを複数のボリュームに分散配置し、並列アクセスを可能としています。これにより、解析サーバは複数のファイルサーバに同時にアクセスして解析処理を行ったり、解析処理した複数のデータを並列に格納処理することができ、処理時間の短縮を実現しています。また、ファイルシステムと解析サーバとを接続するネットワーク装置の理論転送帯域は合計で 960Gbps を確保し、計算処理を行う上で機器間のデータ転送によるボトルネックを回避しています。

2. 高効率な空調機や、空調環境のコンサルにより PUE1.2 を実現し効率的な電力消費を実現

今回、本稼働したスーパーコンピュータシステムでは、世界トップレベルの省エネルギー性能を有するラック型空調機を採用しました。ファンによりITラック背面に排出される、サーバから発生した熱は、隣に置かれたラック型空調機により吸気され、前面へ冷気として排出されます。この局所冷却方式は一般的な壁設置型空調機における床下空調方式と比べ、熱だまりが起こりにくく、空調余裕率を抑えることが可能となります。また、日立が独自開発した3次元熱流体シミュレータを活用し、サーバ室内の空調環境についてコンサルティングを行う「AirAssist(エアアシスト)」サービス⁴の活用により、計算機と空調機器の最適なレイアウト設計を行いました。データセンターやサーバ室のエネルギー効率を示すPUE^(*)では、従来1.9であったのが、新システムでは1.2前後となり、効率的な電力消費を実現しています。

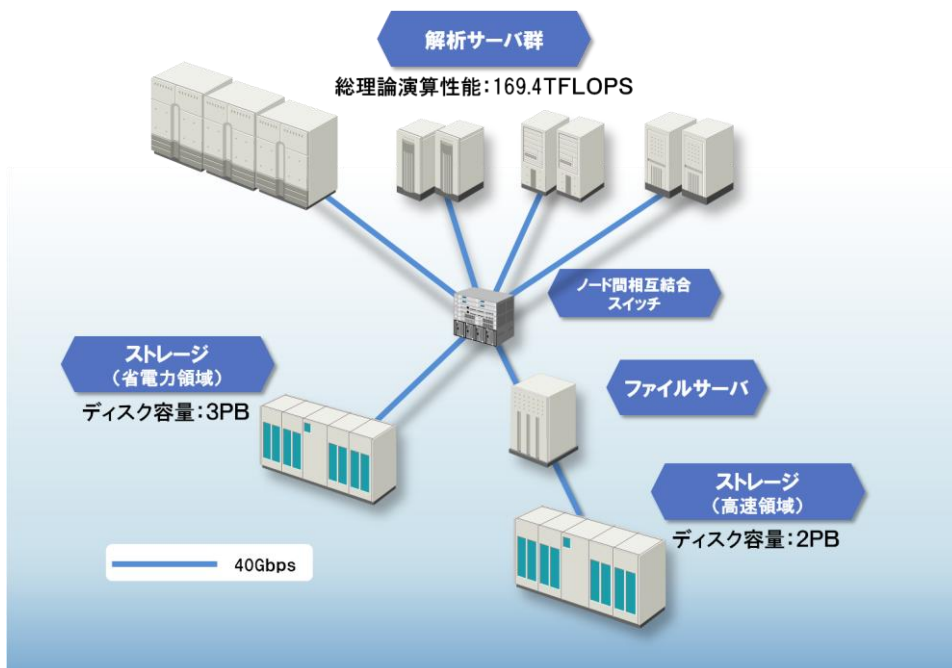
*4 PUE:IT 機器を含むデータセンターやサーバ室全体の消費電力を、IT 機器のみの消費電力で割った値。

最も効率が良いものはPUEが1.0になる。IT 機器を含むデータセンターやサーバ室全体の消費電力と、IT 機器の消費電力が等しくなり、データセンターやサーバ室に供給される電力が全て IT 機器が消費するため、無駄がない。



ラック型空調機のモデル図

■「スーパーコンピュータシステム」の概要



■ 国立遺伝学研究所のホームページ

<http://www.nig.ac.jp/>

■ 本件に関するお問い合わせ先

株式会社日立製作所 情報・通信システム社 公共システム営業統括本部

カスタマ・リレーションズセンタ [担当: 西本、佐々木]

〒136-8632 東京都江東区新砂一丁目 6 番 27 号 新砂プラザ

URL: <http://www.hitachi.co.jp/Div/jkk/inquiry/inquiry.html>

以上

このニュースリリース記載の情報(製品価格、製品仕様、サービスの内容、発売日、お問い合わせ先、URL 等)は、発表日現在の情報です。予告なしに変更され、検索日と情報が異なる可能性もありますので、あらかじめご了承ください。
