



CASE STUDY

Genomics England社、 Quantum ActiveScaleでペタバイト規模の ゲノム・データの保存、保護、アクセスを提供

Genomics England社は、2013年、希少疾患や一般的な癌を患う膨大な数の患者の全ゲノム・シーケンシングを実施する画期的な取り組みである「100,000 Genomesプロジェクト」の支援を目的として、英国保健省によって設立されました。2018年、プロジェクトが大幅に拡大されたため、Genomics England社はデータ・ストレージ・インフラストラクチャを拡張する必要がありました。Quantum ActiveScale™オブジェクト・ストレージを単一の統合環境の一部として使用することで、数百ペタバイトのゲノム・データの保存、保護、アクセスの提供が可能になりました。



導入
製品

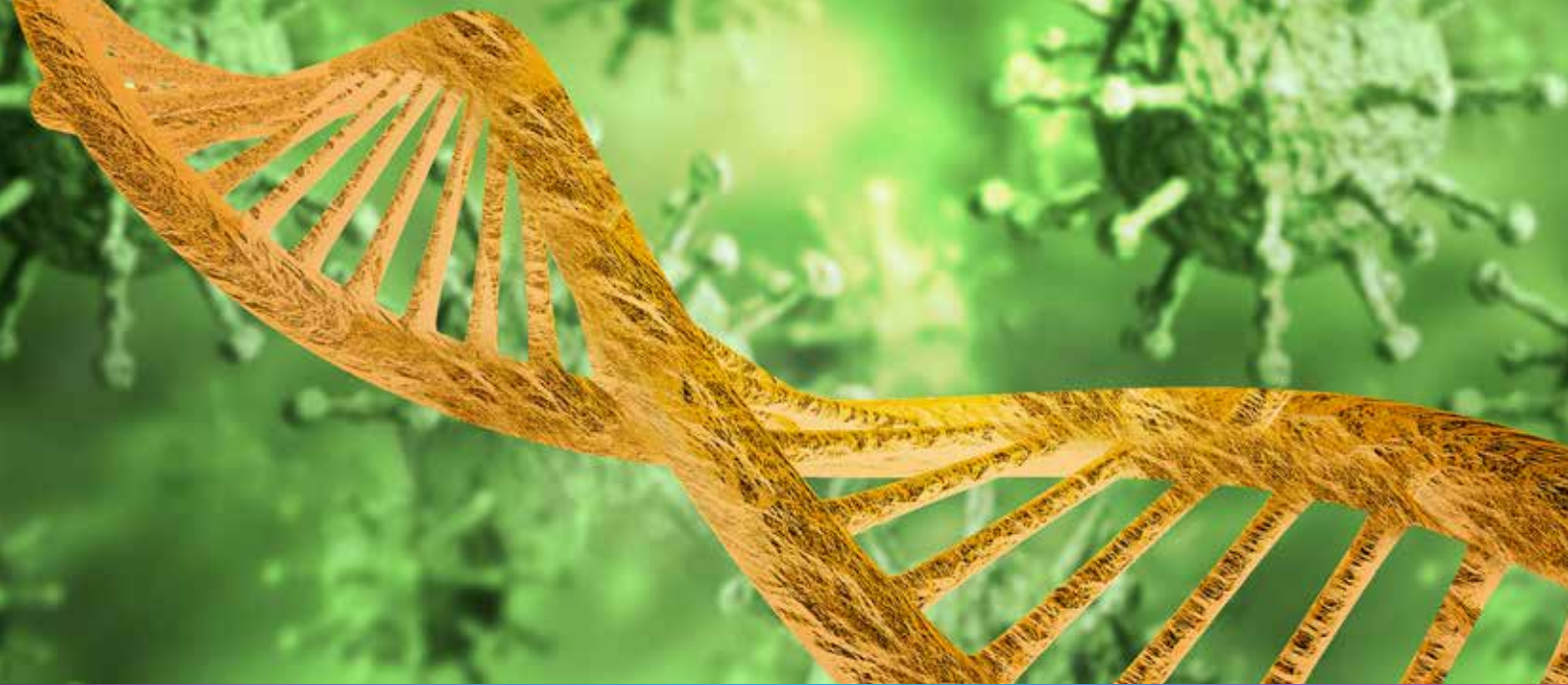
ActiveScaleオブジェクト・ストレージ



ActiveScaleシステムで私たちが気に入っているのは、その固有のアーキテクチャが、RAID置換技術というレイジャー・コーディングされたデータのインテリジェントで動的な配置によって支えられていることです。

David Ardley氏

Genomics England社テクニカル・デリバリー責任者



既存のNASソリューションよりもはるかにスケーラブルなもの、つまり数百ペタバイトに拡張できるインフラストラクチャが必要でした。

David Ardley氏 - Genomics England社テクニカル・デリバリー責任者



ソリューションの概要

- Quantum ActiveScaleオブジェクト・ストレージ
- WekaI0分散ファイル・システム
- Mellanox高速ネットワーク

主なメリット

- **以前のストレージ・ソリューションの容量の限界を克服し**、急増する大量のゲノム・データに対応するスケーラビリティを獲得
- COVID-19感染者のゲノム・シーケンシングの必要性など、**新たな緊急課題に取り組むための俊敏性を向上**
- 3つのデータセンターに分散した地理分散型環境で、**最重要データの保護と復元力を強化**
- 使いやすい効率的な環境で**データ保存のコストと複雑さを軽減**

Genomics England社は野心的な目標を掲げています。同社は、膨大な数の人々のゲノムをシーケンスし、治療の改善に役立つ新しいインサイトを生み出すと同時に英国のゲノミクス産業の発展を加速するために設立されました。

同社は発足から数年で10万件のゲノム・シーケンシングという当初の目標を達成できることを証明しました。また2018年には、5年間で最大5百万件のゲノム・シーケンシングという新たな目標が設定されました。

残念ながら、ゲノム・データの保存に使用されていた既存のネットワーク接続ストレージ (NAS) ソリューションではこの任務を全うできませんでした。21PBのデータを保存していたNASは、ノード拡張の限界に達していたため、それ以上拡張できませんでした。

NASの容量が限界に近づくと、パフォーマンスが低下しました。また、すべてのゲノム情報の復元力に確信を持てずにいました。データはテープにバックアップされていましたが、Genomics England社には、災害などで大規模な混乱が発生した場合にデータを迅速に復

元できる堅牢な災害復旧戦略がありませんでした。

急増するデータ量を保存および保護できるソリューションが必要でした。「既存のNASソリューションよりもはるかにスケーラブルなもの、つまり数百ペタバイトに拡張できるインフラストラクチャが必要でした」とGenomics England社テクニカル・デリバリー責任者のDavid Ardley氏は語っています。新しいソリューションは、世界中の3千名以上の研究者がデータにシンプルかつ柔軟にアクセスできるようにする必要もありました。

Quantum ActiveScaleオブジェクト・ストレージの使用

Genomics England社は、新しいストレージ・ソリューションの設計・実装について、英国を拠点とする独立系データ・サービス企業のNephos Technologies社に相談しました。NephosとGenomics England両社のチームが協力して、現在の要件を確認し、将来のニーズを予測してから、新しい多面的なソリューションを設計し実装しました。

ストレージ・ソリューションには、Weka I/O 高性能ファイル・システム、Mellanox 高速ネットワークキング、Quantum ActiveScale オブジェクト・ストレージが組み込まれています。これは、フラッシュ・ストレージと長期的データ・レイク・リポジトリとして使用される ActiveScale オブジェクト・ストレージを組み合わせた2層アーキテクチャになります。導入当初は1.4PBのフラッシュと40PBのオブジェクト・ストレージを搭載していました。

それぞれを個別に拡張可能な2つのストレージ層は、単一のハイブリッド・ストレージ環境として存在します。その結果、研究者は高度にランダム化された方式でデータをクエリするという柔軟性が得られます。

コロナ禍における新たな課題への取り組み

Genomics England社では、新しいストレージ環境を導入してから数年で、再び拡張する必要がありました。2020年初頭から始まったCOVID-19の感染拡大により世界中の医療科学会に新たな緊急課題が生じ、Genomics England社はどのような人がウイルスに感染しやすいかを特定するのに最適な立場にありました。同社は、最大2万人のCOVID-19重症患者および最大1万5千人の軽症患者のゲノム・シーケンシングに取り組みました。さら

に、COVID-19の調査で得たゲノム・データに研究者や新薬開発者がアクセスして使用できるようにする、次世代のゲノム研究プラットフォームを立ち上げました。

Genomics England社がCOVID-19の研究への関与を深めていたのとほぼ同時に、ActiveScaleソリューション・プラットフォームがクアンタムによって買収されました。弊社のチームは、Genomics England社のスムーズな移行を補佐しました。

Genomics England社は、クアンタムとActiveScaleシステムを全面的に信頼し、COVID-19の研究に対応するためのオブジェクト・ストレージ環境の拡張を進めました。100PB以上の容量の拡張が行われました。

RAID (Redundant Array of Independent Disks) により拡張はシームレスでした。「ActiveScaleシステムで私たちが気に入っているのは、その固有のアーキテクチャが、RAID置換技術というイレイジャー・コーディングされたデータのインテリジェントで動的な配置によって支えられていることです」とArdley氏は語っています。このデータ配置により、パフォーマンスと可用性を損なう可能性のあるシステムのリバランスが不要になります。

「ActiveScaleシステムのシンプルさ、使いやすさ、アーキテクチャが気に入っています。これらのシステムは、オンプレミスおよびハイブリッド・データ・ストレージ・ソリューションのエクサバイト規模の展開にシームレスに拡張できるように設計されています」

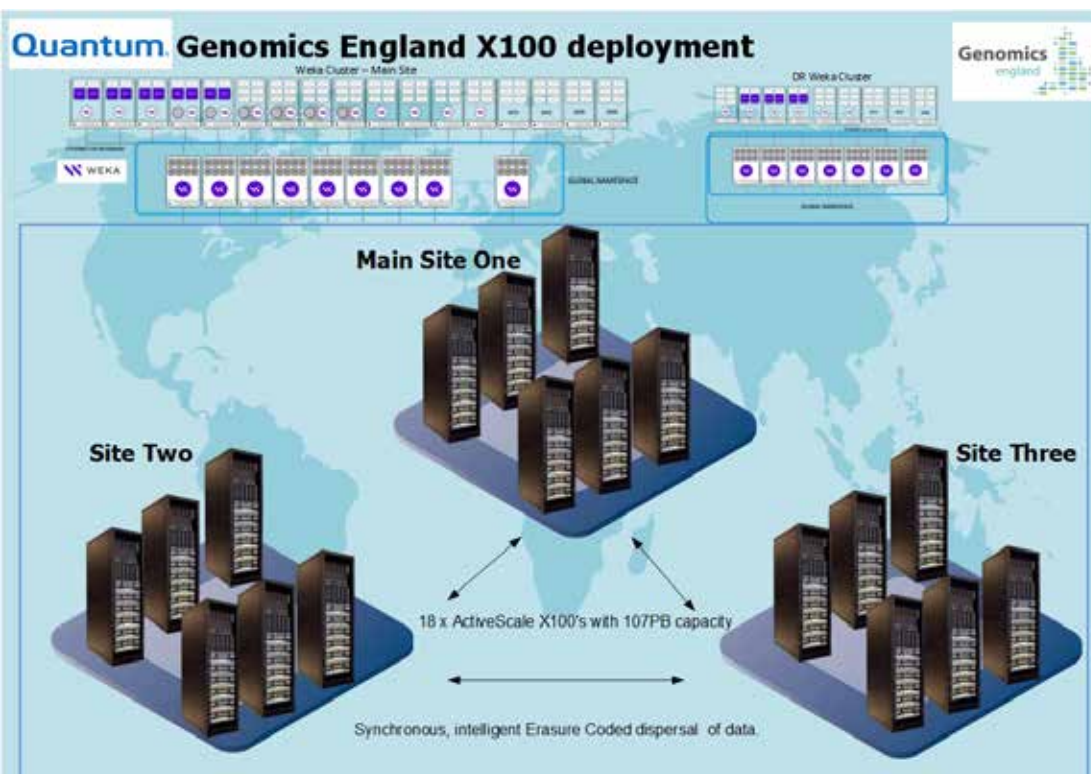
David Ardley氏
Genomics England社テクニカル・デリバリー責任者

Genomics Englandについて

Genomics England社は、希少疾患の患者とその家族、および一般的な癌の患者の10万件の全ゲノム・シーケンシングを実施する取り組みである「100,000 Genomesプロジェクト」の支援を目的として2013年に設立されました。

その後最大500万のゲノムに拡大されたこのプロジェクトは、英国の国立衛生研究所、NHS England (英国の国民保健サービスの一部)、およびその他の組織の資金提供を受けています。保健省が所有するGenomics England社は、より速く、より深いゲノム研究を可能にし、それを必要とするすべての人にゲノム医療を提供することを目標としています。

2020年、Genomics England社は、COVID-19感染者の全ゲノム・シーケンシングを実施するために、医師・科学者団体との協力を開始しました。その目的は、ウイルスに対する遺伝的感受性についての理解を深めることです。





重要なゲノム・データの保護

ActiveScaleオブジェクト・ストレージはデータを保護し、Genomics England社が重要な研究成果に対して必要とするデータの復元力を提供します。同社はActiveScaleの地理分散機能を利用しています。ActiveScaleにより、3つのデータセンターにデータを分散し、サイトの損失などの大規模な災害から完全にデータを保護しています。残りのサイトでは引き続き読み取りと書き込みのためにデータにアクセスでき、さらなるハードウェア障害の発生にも耐性があり、19x9 (99.9999999999999999%)のデータ耐久性を提供します。

さらに、この技術は、自動データ・スクラビング、検証、修復、最適化を通じた自己修復プロセスを実現することにより、高いデータ整合性の確保に有効です。

コストと複雑さを抑えたスケーラビリティの獲得

ActiveScaleにより、Genomics England社は以前のNASソリューションの容量の限界に直面することがなくなりました。同社は、ストレージの大規模なオーバーホールを行うことなく、オブジェクト・ストレージを拡張して、より多くのゲノム解析をサポートし、COVID-19の追加研究を行うことができました。将来的に、Genomics England社は、ActiveScaleオブジェクト・ストレージとAmazon S3準拠のパブリック・クラウド環境を容易に統合し、保護と拡張の柔軟性を追加することができます。

このスケーラブルなストレージ環境は、コストの削減にも役立ちます。Nephos社によると、Genomics England社のチームは、以前の環境と比較して、ゲノムあたりのストレージ・コストを75%削減しました。2023年までには96%のコスト削減が期待されています。

同様に重要なこととして、Genomics England社のチームは、複雑さを増すことなくこれらのメリットを体感できました。Ardley氏は次のように語っています。「ActiveScaleシステムのシンプルさ、使いやすさ、アーキテクチャが気に入っています。これらのシステムは、オンプレミスおよびハイブリッド・データ・ストレージ・ソリューションのエクサバイト規模の展開にシームレスに拡張できるように設計されています。」

当製品の販売に関するお問い合わせは取扱店までお願いします。

<http://www.quantum.com/jp/>

Quantum

日本クアンタムストレージ株式会社

TEL. 03-6890-3038

©2021 Quantum Corporation. All rights reserved. Quantum, Quantumロゴ, ScalarおよびStorNextはクアンタム社の登録商標です。その他すべての名称またはロゴは、それぞれの所有者の商標または登録商標です。本カタログに記載の仕様については予告なく変更することがあります。