

研究開発意思決定をより一層加速するため創薬研究基盤としてのゲノム解析基盤を強化。膨大なオミクスデータを蓄積し自由自在な解析を可能に

「がんに強みを持つ先進的グローバル創薬企業」をビジョンに掲げる第一三共株式会社（以下、第一三共）では、ゲノム解析を中心としたオミクスデータ解析に積極的に取り組んでいます。競合他社に勝る研究開発を推進するためには、膨大なシーケンスデータや解析データを格納し、迅速に解析するための解析基盤が必要とされていました。この課題解決のために同社が選択したのが、IBMのSoftware Defined Storageを活用した階層型ストレージシステムと、GPUを搭載したIBM Power Systemsの組み合わせです。これらのソリューションによって創薬の未来が大きく拓かれようとしています。

【導入製品・サービス】 ● IBM Elastic Storage Server (ESS) ● IBM TS4500テープ・ライブラリー
● IBM Power System S822LC for HPC (Minsky)



課題

- 次世代シーケンサーによるゲノム解析に伴い、データ量が飛躍的に増大。格納するストレージが逼迫し、データ解析にも時間がかかっていた

ソリューション

- 高速ストレージ基盤「ESS」とテープ・ライブラリー「TS4500」による階層型ストレージシステムと、AIやディープラーニングに最適化された高速サーバー基盤「Minsky」を導入

効果

- 効率性と拡張性を備えたデータ解析基盤上で高速処理が可能になったことで、あらゆる角度から全ゲノム解析ができるようになった

【お客様課題】

逼迫したストレージ事情に対処するために
毎週手作業でテープにデータを転送

新薬を開発する上で、遺伝子やタンパク質などのオミクスデータ解析の重要性は急速に高まっています。「がんに強みを持つ先進的グローバル創薬企業」を目指す第一三共も、オミクスデータ解析への取り組みを強化しています。

その研究開発の一翼を担うのが、グループ企業の第一三共RDノバーレ株式会社（以下、第一三共RDノバーレ）です。同社は革新的医薬を開発するための技術基盤プラットフォームを運用し、高品質な研究開発プロセスの構築に取り組んでいます。

第一三共RDノバーレのトランスレーショナル研究部部長の菅野 道裕氏は「これから重要になるのは、基礎研究の成果を創薬に応用するための橋渡し研究（トランスレーショナル・リサーチ）と、薬の患者さんへの効果を解明することに役立つバイオマーカーの研究です。そこではコンピューターを駆使して情報を科学するインフォマティクスが必要になります」と現在の創薬におけるITの重要性を強調します。

これらの研究活動の出発点となるのがDNA塩基配列を高速に読み取るシーケンサーを使った解析です。第一三共RDノバーレでは2014年に次世代シーケンサーを導入し、ゲノム解析に取り組んできました。しかし、この取り組みが新たな課題をもたらしました。同社のトランスレーショナル研究部の主任研究員を務める小野 祥正氏は「旧式のものに比べて一気に1億倍ものデータが取れるようになって、これまでと同じデータ格納方法では早晚ストレージ基盤が限界を迎えることは明らかでした」と話します。

さらに、「以前は外部に解析を委託していましたが、社内にシーケンサーを導入したことで、日常的に利用できるようになり、シーケンサーの稼働率が高まって、当初の想定以上のデータ量に膨れ上がりました」（小野氏）。

一度の解析で発生するデータ量は約100GBで、データ処理するためにはその2倍の容量が必要です。当初30TBの容量を想定していましたが、実際には毎月2TB増えていくため、ディスクはすぐにいっぱいになってしまいました。

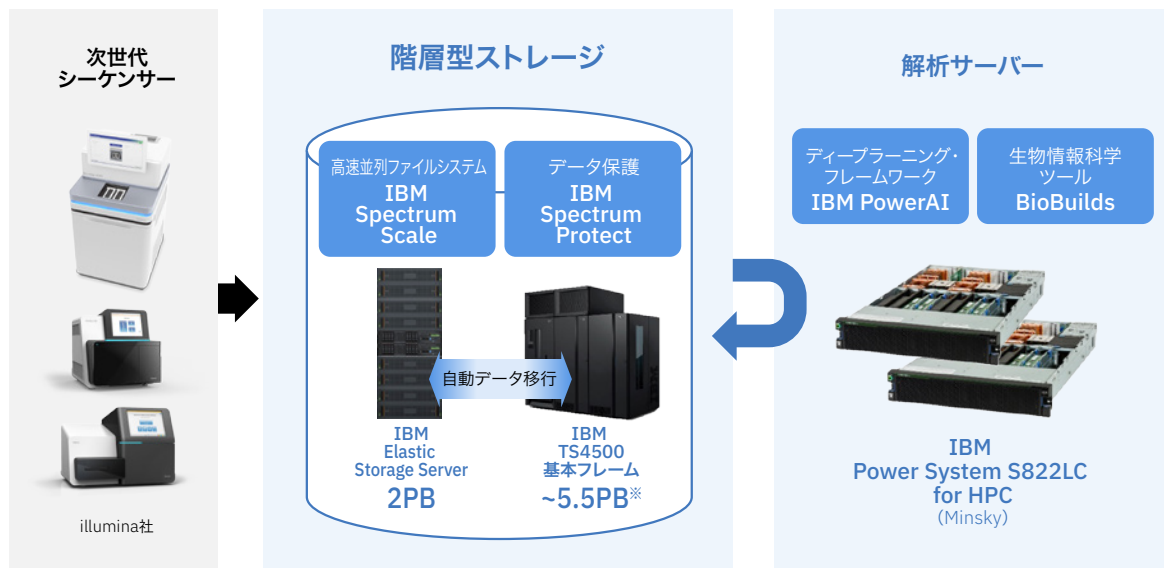
しかし、絶対的なディスク容量の不足という事態には運用で対処するしか方法がなく、データがたまと、その都度、手動でテープにアーカイブするという方法で対応していました。

IBMのソリューションで構築した新解析基盤は非常にパワフルです。今後、創薬のための研究スピードが加速すると期待しています。



第一三共RDノバーレ株式会社
トランスレーショナル研究部 部長
菅野 道裕氏

新ゲノム解析基盤イメージ



※ 拡張フレームを1台増設することにより最大10PBを追加可能

【ソリューション】

階層型ストレージと高速サーバーで 効率的で拡張性のある解析基盤を構築

こうした課題はすぐに第一三共のIT企画部と共有され、解析基盤リニューアルの検討が始まりました。「がんの薬を開発するためには、これからもサンプルは増え続けていきます。優先して解析基盤の課題解決に取り組むことを決め、2017年6月に数社にRFP(提案依頼書)を提示しました」と第一三共のIT企画部企画グループ主査の中野 暢也氏は語ります。

IT環境全体としては「クラウドファースト」を戦略に据えている会社ですが、ゲノム解析のデータは大容量で社内のネットワークへの影響が大きく、また研究のスピードをアップさせるため、オンプレミスで利用する判断が下されました。新しい解析基盤で重視されたのは、容量とデータの転送速度、そして解析フェーズへの移行のしやすさです。

会社からの提案依頼に対して数社から提案がありましたが、最終的に選定されたのは、IBMの階層型ストレージ「ESS」とテープ・ライブラリー「TS4500」、そしてGPUを搭載した高速サーバー基盤「Minsky」でした。中野氏は「パートナーの選定では、ハードウェアの面だけでなく、ライフサイエンスの知識も重視しました。IBMのソリューションは、トータルの質の高さと、オンプレミス環境とクラウド環境との間でデータを透過的に扱えるといった拡張性にも強みがありました」と話します。

IBM Elastic Storage Serverは、HPC分野で20年以上に渡り進化を続けてきた分散ファイル管理ソフトウェアIBM Spectrum Scale(旧GPFS: General Parallel File System)により、高速性・柔軟性・拡張性に優れたストレージ基盤を実現します。今回導入した構成は、2PB超の容量、23GB/s以上の転送速度で、膨大なオミクスデータへの高速アクセスを実現します。また、解析結果などのアーカイブデータは、エンタープライズ・テープ・ドライブを搭載するIBM TS4500テープ・ライブラリーとの間で透過的にポリシーに基づく自動階層管理を行い、効率的かつ経済的に保管されます。

ユーザーである小野氏は導入実績を重視しました。「IBMのソリューションは東京大学医科学研究所や海外の多くのゲノム研究機関に導入されており、大容量データを高速で処理できることは実証済みです。事実上、他に選択肢はないのではと考えていました」(小野氏)。

IBM POWERプロセッサとNVIDIA GPUがNVLinkにより緊密に結合されたMinskyは、これまで同社が解析用に使ってきたソフトウェアにも対応済みで、既存の解析用サーバーと高速で接続できる点や、AI活用のためのフレームワークを装備している点も高く評価されました。「今はデータ・ドリブですが、今後はAIドリブになることが見えています。その面でも将来性があると判断しました」と中野氏は語ります。

新しい解析基盤の構築は急ピッチで進められました。2017年9月に導入プロジェクトがスタートし、10月上旬にハードウェアを設置して、解析データをすべてESSに移行して、11月中旬には利用を開始しています。導入から利用開始までわずか2カ月。「日本IBMを中心にチームとしてうまく回すことができました」と中野氏は導入プロジェクトを振り返ります。

【効果/将来の展望】

ゲノム解析に集中できる環境が整い 新薬開発のスピードアップが図れる

新解析基盤の効果はすぐに現れました。データをESSに格納し、Minsky上で解析することで、解析スピードは一番速いプロセスで10倍になり、平均しても3倍から4倍になったと言えます。菅野氏は「IBMのインフラは非常にパワフルです。今後、創薬のための研究スピードが加速すると期待しています」と評価します。

ストレージの制限から解放されたのも大きな成果です。小野氏は「第一歩として最初から拡張性のあるシステムで解析基盤を構築したことで、データ容量についてのストレスから解放されました。精神的、作業的に楽になると同時に、調整することなく、試したいことを

拡張性のある解析基盤を構築したことで、データ容量についてのストレスから解放され、調整することなく、試したいことを試せるようになりました。



第一三共 RDノバール株式会社
トランスレーショナル研究部
主任研究員
小野 祥正氏

IBMのソリューションは、
トータルの質の高さと、
オンプレミス環境と
クラウド環境との間で
データを透過的に扱え
るといった拡張性にも
強みがありました。



第一三共株式会社
IT企画部
企画グループ 主査
中野 暢也氏

試せるようになりました」と語ります。

がんという病気の性質上、新薬開発のためには、ゲノム全体を解析する全ゲノム解析が必要になりますが、新解析基盤によって、全体でも一部でも自由に解析手法を選択することができるようになりました。「複数のアプリケーションを利用することもでき、解析のための環境が整ったと思います」(小野氏)。

中野氏は「今回の解析基盤は搭載されたソフトウェアのバージョンアップにより、将来的には科学的根拠を検証するバリテーション・システムとしても機能することが期待されます」と指摘します。階層型ストレージの膨大なスペースにすべてのログが残っているので、データの正しさが保証され、将来的には新薬の申請や臨床の報告のデータとしてもそのまま利用できるのが大きな強みです。

「それだけにセキュリティーが担保された中で、この解析基盤を活用することが重要になります。ソフトウェアをバージョンアップしてセキュリティー機能を強化するとともに、社員のITリテラシーを向上させるためのトレーニングも合わせて実施していく予定です」(中野氏)。

また、将来的にはこの解析基盤を中心に拡張していくことで、研究開発の可能性が大きく広がっていくことも想定されています。中野氏は「今後は外部のノウハウを取り入れて利用環境を整備することで、新しい解析基盤の良さを引き出したい。汎用的な量子コンピューターの開発でリードしているIBMには大変期待しています」と語ります。

「がんに強みを持つ先進的グローバル創薬企業」になるためには、ゲノム解析に基づいた新薬の開発体制の確立が必須です。今回構築した新たな解析基盤は、同社の2025年のビジョンの達成に大きく貢献できるものと期待されています。



Daiichi-Sankyo

第一三共株式会社

〒103-8426 東京都中央区日本橋本町三丁目5番1号
<https://www.daiichisankyo.co.jp/>

三共と第一製薬の経営統合によって2005年に誕生。「革新的医薬品を継続的に創出し、多様な医療ニーズに応える医薬品を提供することで、世界中の人々の健康で豊かな生活に貢献する」ことを企業理念としています。2025年ビジョンとして「がんに強みを持つ先進的グローバル創薬企業」を掲げ、先進技術の導入による研究開発の加速を図り、革新的医薬品の継続的な創出に取り組んでいます。



©Copyright IBM Japan, Ltd. 2018

〒103-8510 東京都中央区日本橋箱崎町19-21

このカタログの情報は2018年5月現在のものです。仕様は予告なく変更される場合があります。記載の事例は特定のお客様に関するものであり、全ての場合において同等の効果が得られることを意味するものではありません。効果はお客様の環境その他の要因によって異なります。製品、サービスなどの詳細については、弊社もしくはビジネス・パートナーの営業担当員にご相談ください。IBM、IBMロゴ、ibm.com、GPFS、IBM Spectrum Protect、IBM Spectrum Scale、Power、POWERおよびPower Systemsは、世界の多くの国で登録されたInternational Business Machines Corp.の商標です。他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。現時点でのIBM商標リストについてはwww.ibm.com/legal/copytrade.shtmlをご覧ください。