

HOKUSAI
SAILING SHIP

日本の研究開発を支える 大規模データ基盤を構築

導入機関：国立研究開発法人 理化学研究所

DDNのHPCストレージソリューションを活用し、 S3 APIをサポートしたプライベートクラウド型統合ストレージ環境を スーパーコンピューター「HOKUSAI-SS」上で実現

- ・データ解析
- ・クラウド連携
- ・スーパーコンピューティング

DDNソリューション

DDN ES7990X

DDN EXAScaler

DDN EXAScaler S3 Data Service

日本で唯一の自然科学の総合研究所として、物理学、工学、化学、計算科学、生物学、医科学など、多岐に亘る領域で研究を推進する国立研究開発法人理化学研究所（以下、理研）。研究成果を社会に普及させるため、大学や企業との連携による共同研究、受託研究等を実施しているほか、知的財産等の産業界への技術移転を積極的に進め、特定国立研究開発法人として、国際競争の中で革新的な研究成果を創出し、日本のイノベーションシステムを強力に牽引する中核機関としての責務を果たすための「科学道」を邁進しています。近年、様々な研究活動におけるデータ処理の必要性の増大、アプリケーション実行環境の柔軟性、そして研究環境の変化への柔軟な対応が求められる中、理研が構築・運用してきたスーパーコンピューターシステムも、従来とは異なるデータ処理・データ保存環境への対応が急務となっていました。

そうした要件に応えるものとして新たに構築されたのが、データ科学基盤「HOKUSAI SailingShip」（以下、HOKUSAI-SS）であり、そのデータ活用を支えているストレージシステムが、データダイレクト・ネットワークス（DDN）が提供する「DDN ES7990X」「DDN EXAScaler」、そして「DDN EXAScaler S3 Data Service（S3DS）」といったDDNのハードウェア、およびソフトウェアソリューションです。



パフォーマンス

- ・7台のアプライアンスを用いて総容量30PBのHPCストレージ環境を構築し、150GB/secのIO性能を実現



拡張性

- ・S3DSの活用により、クラウドストレージとのデータ相互利用が可能



柔軟性

- ・NFS、CIFSのほか、S3互換によるオブジェクトストレージにも対応

目的

- 高速なHPCストレージを提供すると同時に、将来的なクラウド上（S3）でのデータ活用や保存にも容易に対応可能な環境の実現
- 日々増加する膨大な研究データの処理と保存を可能とする、プライベートクラウド型の大容量ストレージ基盤の構築

HOKUSAI-SSは、理研がこれまで蓄積し今後も増加が見込まれる膨大な研究データの処理や利活用を促進することに加え、近年、研究データを広く社会と共有することで、効率的な科学の発展を目指す「オープンサイエンス」のためのデータ活用基盤としての役割も期待されています。理研 情報統合本部情報システム部研究基盤課 技師の検崎博生氏は、「これまで理研が構築してきたスーパーコンピューターは、バッチジョブシステムによる高性能高並列な計算環境を提供するコンピューティング・オリエンテッドなシステムでした。対して、HOKUSAI-SSはデータ・オリエンテッドなシステムであり、データ処理とデータ保存環境を用意するとともに、プライベートIaaSとして様々なシステムインフラを仮想化することで、各研究者がHOKUSAI-SS上にテナントを構築しデータ分析に利用できるようにしています」と話します。

HOKUSAI-SSが有する先進性の1つが、高速・大容量のHPCストレージ環境を実現すると同時に、オブジェクトストレージであるS3への対応も図られている点です。検崎氏は、「データファームにはPOSIXによる高速アクセスに加えて、多種多様なデータ処理やデータ保存環境を提供できるようNFS、CIFS、そしてS3互換によるオブジェクトストレージにも対応しています。S3への対応を図った理由は、オープンサイエンスの基盤として異なるサービス間でもデータのやり取りが容易に行えること、そして、将来的にさらなるデータ容量の増加に直面しても外部のクラウドサービスにアーキテクチャーはそのままに、データをオフロードできると考えたからです」と説明します。

「一方、これまで理研での利用経験が少なかったS3へすぐに全面移行するのは難しく、既存のプロトコルも同時に利用できる環境が不可欠でした。また、運用性を考慮した場合、オンプレミスのHPCストレージとS3環境を個別に構築して連携させるのではなく、両方が同時に使えるストレージシステムを求めていました」（検崎氏）

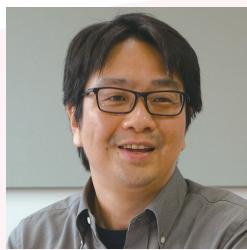
「HPCストレージ環境を実現すると同時に、将来的なクラウドストレージとの連携を図ることができました」



實本英之氏

国立研究開発法人 理化学研究所
情報統合本部 基盤研究開発部門
データ管理システム開発ユニット
ユニットリーダー

「DDNのHPCストレージは、高性能計算やホスティング環境における大規模データ共有の足回りを支えています」

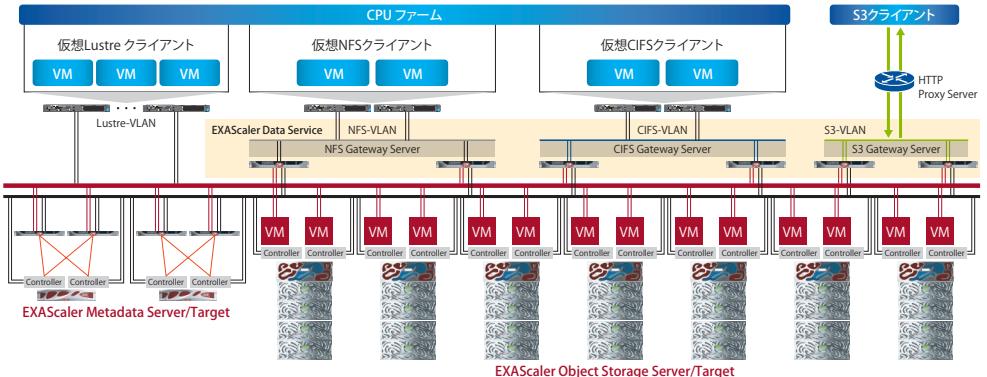


検崎博生氏

国立研究開発法人 理化学研究所
情報統合本部 情報システム部
研究基盤課 技師



プライベートクラウド型統合ストレージ環境の基盤となったDDN ES7990X



実現したソリューション

- これまで以上に高速かつ、大容量のHPCストレージ環境を構築。さらに、将来的なクラウドストレージの活用や、研究者間の容易なデータ共有も可能なS3互換のS3DSによるストレージ環境を実現
- 今後のデータ増にも対応可能な総容量30PBに及ぶストレージ環境を確保、LustreファイルシステムをベースにしたDDN EXAScalerとしても安定した稼働を継続

理研が掲げた要件に応えたのが、DDNのストレージハードウェア／ソフトウェアソリューションです。HOKUSAI-SSでは、DDN EXAScalerが稼働する、総容量30PBに及ぶ7セットのストレージアプライアンスDDN ES7990X、そしてDDN S3DSを用いたS3ゲートウェイを主軸とするシステムが構築されています。

2019年11月からの構築を経て、2020年10月から本番稼働を開始したHOKUSAI-SSですが、DDNのストレージソリューションは、安定した稼働を続けています。仮想化によって提供される各テナントの共有ストレージとして、ユーザーのオンデマンドクラスタでの高性能計算、また、ホスティング環境における大規模データ共有の足回りを支えています。今後も各研究所で進行中の様々な研究データを統一的に管理していくための大規模かつ高速・堅牢なストレージとしての役割が期待されています。

情報統合本部 基盤研究開発部門 データ管理システム開発ユニット ユニットリーダーの實本英之氏は、今回のストレージシステムの中核機能となるDDN S3DSの期待効果について次のように語ります。

「将来的なS3によるクラウドストレージの利用を想定する一方で、現時点ではどれくらい情報の参照量があるのか、どの程度の容量を確保すればよいのか見積もることは困難です。対して、オンプレミス環境でS3プロトコルが利用できるようになったことで、今後、クラウドストレージを利用する場合の性能策定フェーズにおいて、強力な効果を発揮すると期待しています」(實本氏)

このほかにも上位システムでS3プロトコルが利用できれば、S3DSにより既存のオンプレミスのHPCストレージ環境をクラウドストレージに移行させたり、S3互換クラウドストレージどうしを容易に接続させたりすることも可能になるなど、相互運用性の向上も期待効果として挙げられています。

「新たなデファクトスタンダードとなりつつあるS3はデータ公開に適したプロトコルであり、Nextcloudをはじめとするミドルウェア、およびユーザーが利用するソフトウェア内での対応が進んでいます。DDNのソリューションによりオンプレミスにS3互換ストレージ環境が構築できた価値は大きく、クラウド利用時の見積もりが可能なほか、送受信が大きいデータはオンプレミスに、小さいデータはクラウドに保存するといった運用も可能になるとと考えています」(實本氏)

今後の展開

現在、HOKUSAI-SSはユーザーの習熟や利用率向上を進めている段階にあり、研究所内の各研究部署が独自にサーバーを構築・運用管理を行わずとも済むよう、システム利用に向けた啓蒙活動を進めています。検崎氏は、「DDNからは常時、細やかなサポートを頂いていますが、今後も引き続き、ソリューションのさらなる効果的な活用方法についてのアドバイスや提案をお願いしたいと考えています」と話します。

そして、實本氏も次のように今後の展望とDDNに対する要望を語りました。

「世界を相手に日本が先進的な研究活動を推進していくためには、多くの大学や研究機関が相互に利用可能な、標準化されたデータ基盤が不可欠です。HOKUSAI-SSがそうした基盤の1つとなるよう、今後も機能強化を図っていきますが、DDNのストレージにもさらなる進化を期待しています」(實本氏)

DDNについて

データダイレクト・ネットワークス（英語表記 DataDirect Networks、略称 DDN）は、AI/DL/ビッグデータ向けストレージのグローバルリーダーとして、データ集約型の企業のビジネス課題を解決する包括的なストレージソリューションをグローバルに展開しています。

© DataDirect Networks. All Rights Reserved. DataDirect Networks、DataDirect Networks ロゴ、DDNは、DataDirect Networks の商標です。その他の名称およびブランド名は、各社の登録商標または商標です。

202108

株式会社データダイレクト・ネットワークス・ジャパン

■ 本社 〒102-0081 東京都千代田区四番町6-2 東急番町ビル8F
■ 大阪オフィス 〒530-0001 大阪府大阪市北区梅田3-3-45 マルイト西梅田ビル5F
■ 名古屋オフィス 〒451-6040 愛知県名古屋市西区牛島町6-1 名古屋ルーセントタワー40F

Team-JPSales@ddn.com
<https://ddn.co.jp/contact>