

データ活用に 「データ仮想化」という“新戦略” DX時代のデータマネジメント、 その在り方を探る

企業活動の中で生まれる膨大なデータ。それを“資産”と捉えてDX推進の原動力やビジネスの新たな価値創出につながる芽にすべく、データ活用やAIの利用に取り組む企業が増えている。

蓄積したデータをBIツールやAIに丸投げすればいいかというと、そう簡単な話でもない。機械で読み取れるようにデータを整える、膨大な情報を高速かつ効率良く利用できる体制を整えるなどの「データマネジメント」ができていないと、データもAIも真価を発揮できない。

これまでデータマネジメントというと、データの保存やストレージの容量、I/O（入出力）などのハードウェア面が議論の中心だっ

た。しかし各所に分散している多種多様なデータを1カ所に集めて整理し、巨大なストレージに格納するのは効率性や活用時のリアルタイム性の観点で現実的ではない。個人データや機密データのように移動が難しいデータもある。

そこで、新しいデータ戦略として「データの仮想化」に注目が集まっている。データの保存場所やフォーマットを変えずに、仮想的に統合することでデータの活用を促進する技術だ。先端技術に敏感な企業が期待する新しいデータマネジメントの技術について、この分野に詳しいデル・テクノロジーズとスタートアップの米Starburst Dataを取材した。

Starburst Data x Dell Technologies 戦略的なグローバル・パートナーシップを締結



Starburst

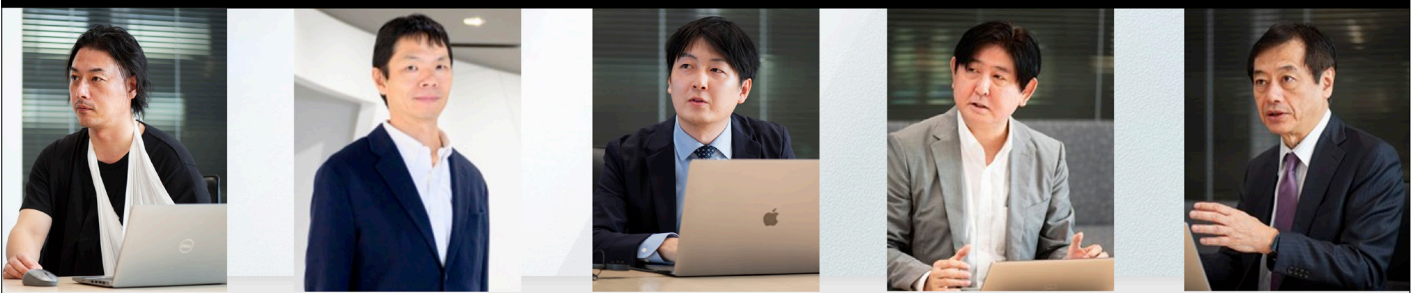


DELL Technologies

デル・テクノロジーズを通じ、ソフトウェア製品、プロフェッショナルサービス、ソリューション提供が日本国内で可能

事前検証済み設計ガイド、ベンチマークテスト結果、国内PoC環境、ハンズオン環境のご提供が可能

マルチクラウド・データ・アクセス、データ仮想化、新データレイクハウスソリューションを日本国内で実現



“DX時代”に求められる データマネジメントの姿

多くの企業が本腰を入れているDXやAIの活用で成果を挙げるには、何よりもまず「データ」が重要になる。データそのものは豊富にあるため、ここでいう“データの重要性”とは、どのようにしてデータを活用できる形にするかという点だ。「サイロ化」というキーワードを出せば、データ活用の難しさに共感できる読者も多いだろう。

従来のAI開発は、散在している大量のデータを収集し、大きなデータレイクに格納して管理する方法を採用することが多かった。目的に応じて利用しやすい形にデータを加工して、AIのモデル開発や学習に使うのだ。

この方法は、モデル開発など特定の場面でデータを使えば十分なケースに向いている。頻繁にAIをアップデートしたり、社内の情報を参照して成果物を出したりするようなリアルタイム性が求められる使い方には適さない。BIツールで経営データを随時チェックする「リアルタイム経営」を実現するにも、データを集めて整理して……という作業をしているのでは経営スピードに追い付けない。

「私たちが考えるデータマネジメントとは、さまざまな場所から生まれる大量のデータをリアルタイムに収集し、安全性やコンプライアンスなどを適切に管理した上でAIやアプリケーション、データサイエンティストなどが活用できるようにする仕組みです。部門横断

でアクセスできる、あらゆるデータが扱える、それが可視化されている、運用負荷が小さいなど、“DX時代”に即したデータマネジメントとそのためデータ基盤が求められています」——こう説明するのは、デル・テクノロジーズの堀田鋭二郎氏だ。

堀田氏は続けて、このAIブームでデータマネジメントに注目する企業が急増していると話した。しかし大量のデータ

を扱い切れない、サイロ化されていて活用どころではないなどの課題が山積しており、いま企業が運用している仕組みや技術では期待するパフォーマンスを出せない。

そうした課題を解決するために、データ仮想化の技術が一つの解になるとデル・テクノロジーズの小野良夫氏は説明する。

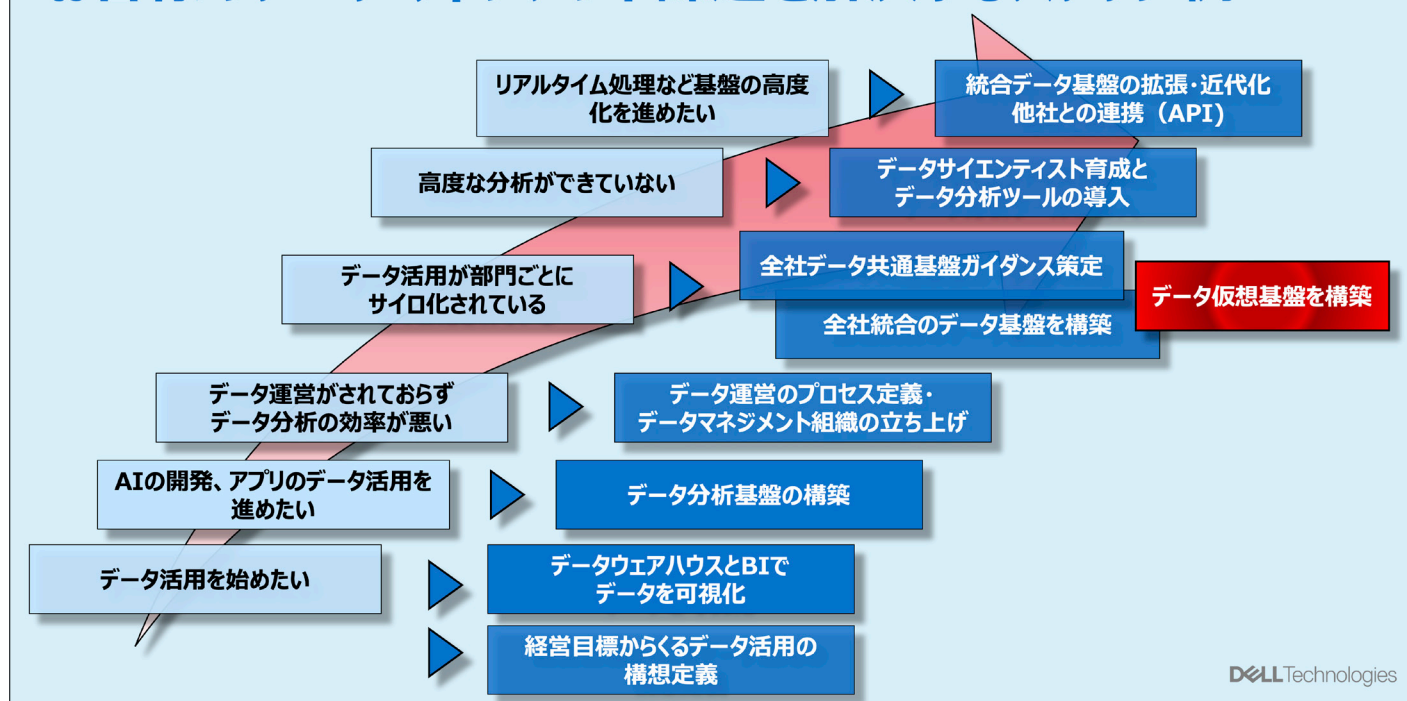
「データ仮想化とは、データ自体は複製したり移動させたりせずに統合管理する技術です。仮想データ層を1つ設けることで、複数のアプリケーションやユーザーがシームレスにデータにアクセスできるようになります」（小野氏）

データ仮想化という世界観をさまざまな企業で実現するため、デ



デル・テクノロジーズの堀田鋭二郎氏
(データワークロード・ソリューション本部 本部長)

お客様のデータマネジメント課題を解決するステップ例





デル・テクノロジーズの小野良夫氏（データワークロード・ソリューション本部 DA/ AI 担当スペシャリスト）

ル・テクノロジーズはデータアナリティクス領域に強みがある Starburst Data と戦略的なグローバルパートナーシップを締結。Starburst Data の技術とデル・テクノロジーズのハードウェア技術やコンサルティングの知見を組み合わせることで、検証済みの信頼できるデータ仮想化ソリューションを提供している。



Starburst Data の蛭原裕弥氏（ソフトウェアエンジニア、Trino メンテナー）

分散処理技術「Apache Hadoop」のクラスタ上で「Apache Hive」を使っていたユーザーの支持を得ていると Starburst Data の蛭原裕弥氏は紹介する。

「Starburst はクラウドをまたいだデータ活用も可能です。各クラウドにつないでいる Starburst クラスタを連携させることで低遅延なデータ活用を実現できます。データを複製して移動するわけではなく、要

求があったときに参照するだけなのでネットワークトラフィックへの影響も最小限に抑えられます」（蛭原氏）

データを移動しないという特性から、国をまたぐデータ活用でも活躍する。データ保護規則「GDPR」によって EU 域外への個人情報の移転を制限されるなど、データの収集や保管が難しい場合でも影響を受けずにデータを活用できる。

米国のある大手メディア企業は、これまでデータを抽出して加工し、読み込む一連の流れに約 18 カ月かかっていた。しかし Starburst を使ったデータの仮想的な統合に約 5 週間で成功し、ビジネスを中断することなく価値ある洞察を得られた。この取り組みによって約 2 億ドルの収益が生まれたと蛭原氏は紹介する。

データ仮想化のカギ「Starburst」何がすごいのか

Starburst Data は元 Teradata のメンバーを中心に、旧 Facebook 社でビッグデータ向けの分散クエリエンジンを開発していた 3 人のエンジニアなどが立ち上げたスタートアップだ。その技術力とデータ仮想化ソリューション「Starburst」で注目を集めている。

Starburst は、「Oracle Database」や「Salesforce」、オブジェクト形式のデータなど多種多様かつ巨大なデータを分析しやすくするツールだ。データの移動を最小限にしてコストを抑えつつ、しかも分析速度が向上するというメリットがある。大規模データの



Starburst の概要。データ仮想化を強力に推進できる（図版提供：デル・テクノロジーズ）

データ仮想化は “足元”のストレージ基盤も重要

Starburst が真価を発揮するには、“足元”に当たるストレージ基盤も重要な要素だ。Starburst はデータの保存場所こそ問わないものの、パブリッククラウドのストレージサービスはデータのロード速度やネットワークへの負荷、データ量が増えるほど増大する費用などの課題がある。そもそも個人情報や機密情報など、クラウドに置けないデータもある。

データマネジメントの効果を高めるには、クラウドやアプリケーションとの親和性が高く、大量のデータを確実に保管できるストレージ基盤が適している。Starburst との連携の観点で推奨しているのが、オブジェクトストレージ「Dell ECS」だとデル・テクノロジーズの矢ヶ部謙一氏は解説する。

ECS は Elastic Cloud Storage を指す。Dell ECS は、事実上の業界標準になっている Amazon Web Services (AWS) のストレージサービス「Amazon Simple Storage Service」(Amazon S3) の仕様に準拠しており、AI やアプリケーションなどからデータを参照しやすい構成を採用している。

「Dell ECS は SQL や NoSQL、ストリーミング、データアナリティクスなどのパフォーマンスも重視するワークロードで活躍しています。しかし、データ仮想化の領域では性能が良くて容量が大きいだけでは不十分で、データマネジメントの観点が重要です。その点、Dell ECS はデータマネジメント機能が豊富なので高い効果を

発揮できます」(矢ヶ部氏)

Apache Iceberg 形式で Dell ECS に保存されたデータは、そのまま「Starburst クエリエンジン」を使ってクエリできる。また、「Starburst マテリアライズドビュー」やキャッシュなども Dell ECS に保存可能だ。

AWS のユーザー認証機能「IAM」に対応し、アクセス権限を効率的に管理できるため情報セキュリティ的にも安全性が高い。その他、Dell ECS の設置拠点が複数サイトに分散しているためアクセスが容易なグローバルネームスペースに対応しているなど、Starburst を含むデータ仮想化のニーズに適した設計が Dell ECS の特長だ。



デル・テクノロジーズの矢ヶ部謙一氏
(UDS 事業本部 SE 部 マネージャ)

データマネジメントの道を共に歩む パートナー

データ仮想化は、企業がデータマネジメントを進める上で期待できる技術であり、Starburst と Dell ECS の組み合わせはモダンなワークロードにおけるデータ活用を促進するソリューションと言える。しかし、データマネジメントの取り組みを進めるとデータがどの

データマネジメント コンサルティング

IT基盤のみならず、人材育成、プロセス定義で貴社のデータ活用を促進します



DELL Technologies

ような状況にあるのか、どう活用したいのかなどの条件によって運用方法が変わる。また、データ仮想化はデータマネジメントという大枠の一部であり、全体を最適化するには広範な知見が必要なので企業が単独で進めるには限界がある。

そこでデル・テクノロジーは、データ仮想化を含むデータマネジメントの推進やデータ基盤の構築、データサイエンティストやデータ管理者の人材育成、社内ルールの策定などを包括的にサポートするプロフェッショナルサービス「データ・コンサルティング・サービス」を用意している。

同サービスはシステムを構築して終わるのではなく、顧客が真にデータを活用できるようにすることが目標だ。そのため、データのレビューや OJT での人材育成など、デル・テクノロジーが蓄積してきたノウハウを惜しまず提供している。

「データ活用を成功させるには、その企業のビジネスを分かっていることが重要です。お客さまご自身がデータや AI を活用できるようになることで、高いビジネス効果を発揮できます。当社が全てを請け負うの



デル・テクノロジーの山口浩直氏（執行役員 サービスビジネス営業統括本部 サービスプリセールス本部 本部長）

ではなく、お客さまの社内に知見やノウハウを蓄積していただきたいのです。そのためにデル・テクノロジーのサービスをご利用いただくことで、戦略的にデータ活用を進めていただくことが可能となります」——デル・テクノロジーの山口浩直氏はこう締めくくった。

データ活用が重要課題に 次世代の基盤「データレイクハウス」を 見すえた取り組みを

データ活用は、いまや企業における最重要課題の一つになった。従来のデータ基盤は、いま求められているレベルでデータ活用を支える仕組みにはなっていない。データ仮想化技術を取り込み、あらゆるデータに対応できる新しいデータ活用基盤が望ましい。それは、従来のデータウェアハウスとデータレイクの双方の特徴を併せ持つ「データレイクハウス」だ。

デル・テクノロジーと Starburst Data のソリューションを活用すれば、データ活用の基盤となるデータレイクハウスを実現できる。その先、将来的には場所も種類も形式も一切問わない「データファブリック」や、あらゆる関係者が自由にデータにアクセスできる「データメッシュ」を目指してデータマネジメントを成長させていくのが理想だ。そこでもデル・テクノロジーが強力なパートナーになってくれる。データを活用したいと考える企業は、一度相談してみてはいかがだろうか。