

NTTデータが考える Modern Data Stackというトレンド

デザイン&テクノロジーコンサルティング事業本部
デジタルテクノロジー&データマネジメントユニット

株式会社NTTデータ

〒135-6033 東京都江東区豊洲3-3-3 豊洲センタービル
Tel: 03-5546-8051 Fax: 03-5546-2405
<https://www.nttdata.com/jp/ja/>

CONTENTS

CHAPTER.1 NTTデータとデータ活用プラットフォーム

CHAPTER.2 技術の進歩で何が起きたか？

CHAPTER.3 データ活用の現場で起きている課題

CHAPTER.4 レガシーなデータ活用プラットフォームを進化させる “Modern Data Stack”というトレンド

CHAPTER.5 NTTデータが考えるモダンなデータ活用プラットフォームの特徴と 当社の取り組み

CHAPTER.6 <組織・企業文化に関するトレンド> データの分散管理

CHAPTER.7 NTTデータが提供する 「価値向上の手法：Value Engineering」

CHAPTER.1 NTTデータとデータ活用プラットフォーム

20年来日本のデータ分析・活用の最先端で業界を牽引してきたNTTデータでは、大小さまざまなデータ活用プラットフォームを構築してきました。

昨今、DX・デジタルといった潮流と、クラウド技術をベースとした新しいサービスによって大きくその姿が変わりつつあります。

四半世紀に及びデータ分析/活用の最先端で業界を牽引

1990年前後のERPブーム、2000年前後のCRMブームの時代に、BI+データマイニングのサービス開始以降、クラウド技術、ビッグデータやAI&IoTといったブームの中で、様々な活動を当社は行ってきました。

2023年4月からは、NTTデータのDXコンサルタント、サービスデザイナー、データサイエンティスト、ITアーキテクト、顧客接点領域（Salesforce）のコンサルタント・エンジニア等の専門家600名を集め、デジタル変革を進める「デザイン&テクノロジーコンサルティング事業本部」を整備するなど、活動の幅は、日に日に広がっています。

2019年3月にデータ分析基盤ソリューション「Trusted Data Foundation®」の提供を開始（#1）

さまざまな業界向けに支援してきたデータ分析基盤の構想立案・構築、データマネジメント戦略の策定・推進、情報活用組織の立ち上げ・運営サポートなどの実績を活かし、データ資産を分析活用するための環境をオールインワンで提供するソリューションとして、2019年3月25日には、「Trusted Data Foundation®」を発表しました。

データ活用したいお客様が、自由自在にデータを調査・準備・分析可能であり、データ活用業務の高度化を実現する環境を構築したいといったご要望にお応えすべく、以下の2つのタイプのサービスを提供しました。

- ・ソリューションベースDSI：「多くの分析者が自在にデータを調査・準備・分析可能であり、全社的なデータ分析業務の高度化を実現する環境を構築したい」といったご要望に対応
- ・アナリティクス・マネージドサービス：「すぐに分析を始め、効率良く試行錯誤しながら方向性を見出し、今後のビジネスを成長させていきたい」といったご要望に対応

サービス提供開始後も、非常に多くのお客様に長期にわたり利用していただく一方で、当社も実際に様々な課題に直面して、データ活用プラットフォームの難しさに直面してきました。

一方で、日本よりデータ活用が進んでいるといわれているグローバルを見てみると同様な課題に直面し、すでにその課題の解決に向けた取り組みをしていたり、課題解決を目指したソリューションが提供され始めていたことも増えてきました。

「Snowflake」、「Databricks」、「dbt Labs」とパートナー契約を締結

データ活用のテクノロジー進歩において、Amazon Web Services、Microsoft Azure、Google Cloud Platformといったクラウドサービスが非常に多くのデータ活用関連のサービスを提供してきたことは間違いありません。そういったクラウドサービスを利用しつつ、2019年の「Trusted Data Foundation」の提供以降、いずれのクラウドサービスとも親和性が高いサービスでありながら、当社が抱えていた課題を解決するソリューションを継続的に探索してきました。各ソリューションが提供する世界観やビジョンに対して共感し、提供する機能を安心・安全にお客様に届けるために、「Snowflake」とは2020年2月、「Databricks」とは2021年5月、「dbt Labs」とは2023年7月に、パートナー契約を締結しました。こうして、データ活用領域において、新しい仕組みを取り込みながら、「Trusted Data Foundation®」とともに、当社は成長をしてきました。

データ活用プラットフォームにおける新たなトレンド

今、当社が目目している「Snowflake」、「Databricks」は、いずれも機能の中心は、データを蓄積することにあるソリューションであり、いずれも従来のデータストアに関する課題を大きくブレイクスルーしています。

以前は、データストアは、目的や利用用途によって製品を使い分ける必要があり、複数の製品を利用することで、SLAを満たす反面、非常に複雑な運用や、データ整備を含めたシステム開発によるコスト増、アジリティに関する問題が発生していました。クラウドベンダからのマネージド・データベースサービスの提供によって、構築期間や運用コストは一部削減できましたが、むしろ管理対象が増えてしまう事象も発生していました。

つまり、データ活用を取り巻く、ビジネスユーザ・システム管理者などの様々なステークホルダが、アジリティであったり、拡張性であったり、運用に対してまだまだ多くの不満が解消されていない状態であると考えます。

こういった不満を解消する可能性があるソリューションが出てきていますが、むしろ新しいサービスが多すぎて理解できない、従来のサービスが機能を拡張して、複数のサービスのすみわけがわからないという事象が発生しているのが、データ活用の領域の実態です。

このホワイトペーパーでは、こういった様々なビジネス的・技術的な背景や想定課題と現在のトレンドを改めて整理しています。皆様のデータ活用のおかれている状況について改めて、考えていただく機会となれば幸いです。

CHAPTER.2

技術の進歩で何が起きたのか？

データ資産からインサイトを獲得し、行動に移すというデータドリブンな活動を支えるプラットフォームには、小さく始め、成功体験を重ねるために、柔軟性とアジリティが従来に比べて大きく求められるようになりました。

この要望とクラウド技術は、非常に高い親和性があり、クラウドサービスを最大限利用したデータ活用プラットフォームが構築され、その利便性を上げるために、さらに様々な技術進歩が生まれています。

クラウドサービス関連技術の進歩と、その大原則について常に意識をする

クラウドのためのアーキテクチャー設計という原則

AWSがまだ日本リージョンに来て間もない、2011年5月19日に、「クラウドのためのアーキテクチャー設計 - ベストプラクティス -」(※1) というWebinarと資料が公開されました。

この資料には、クラウドサービスの特徴は、「抽象化されたリソース」、「必要なときに瞬時に調達」、「スケールアウト自由自在」、「従量課金」、「自動化」であり、クラウドのためのアーキテクチャー設計ベストプラクティスとして、「1. 故障のための設計(Design for failure)」、「2. 疎結合なコンポーネント」、「3. 弾力性の実装」、「4. 全レイヤにおけるセキュリティ担保」、「5. 並列処理の使い倒し」、「6. 異なるストレージの使い分け」の6つが原理原則として書かれています。

上記の発表から10年以上がたち、この原理原則を大きく意識したソリューションが増えてきました。さらに、そういったソリューションの多くが、昨今のDX・デジタルの流行もあり、データ活用領域のソリューションで多く見かけられるようになってきました。

DWHとデータレイクといったデータ蓄積技術の進歩を理解する

DWHとデータレイクの進歩とその融合

この四半世紀、当社とともに、DWHとデータレイクは技術の進歩と時代の要請に従って進化し続けてきました。オンプレミスにおいて、分散ストレージからオブジェクトストレージが主流になってきました。そして、DWHは、クラウドサービスとして提供されるとともに、オブジェクトストレージ等他サービスを統合して利用する機能が充実、データレイクはオブジェクトストレージへのアクセス手法が充実してきました。DWHとデータレイクの長所を組み合わせる両方を活用することは可能ですが、データの重複が生じ、アーキテクチャが複雑になり、どうしても運用負荷が高くなるため、DWHとデータレイクの両方の特徴を兼ね備えていることが求められるようになってきました。

ELTからELTへの変化から見たクラウドサービスをうまく使うということ

クラウドサービスのメリットを利用するために、ETLではなくELTを採用する

DWHにデータをロードする処理と言えば、ETL(Extract Transform Load)といった順序にバッチ処理をすることが、ベストプラクティスだと言われていました。以前は、データサイズが小さかったため、データを処理する場所にデータを移動させて変換することが普通でしたが、Hadoop全盛期に、データが巨大になるとデータの移動が大変なため、処理をデータが保存されている場所で変換するやり方(ELT)に、大量データでのバッチ処理におけるベストプラクティスが変わりました。

クラウド型のDWHでは、Hadoop同様に、大量のデータの処理が必要であり、かつDB上での処理が可能となっています。つまり、従来のベストプラクティスが、むしろアンチパターンになるということであるため、従来の設計思想で構築されていたソリューションにとっては、1から作り直す必要が生じる可能性があるくらいの変化が、技術の進化によって起きています。

技術を使う効果は何か？

クラウドネイティブ技術の導入効果

CNCF(Cloud Native Computing Foundation)の定義によると(※2)、「クラウドネイティブ技術は、スケラブルなアプリケーションを構築および実行するための能力を組織にもたらし、回復性、管理能力、および可観測性のある疎結合システムが実現する。これらを堅牢な自動化と組み合わせることで、エンジニアはインパクトのある変更を最小限の労力で頻繁かつ予測どおりに行うことができる」と記載されています。データ活用領域において、頻繁に変わる要件に対して、Agilityや柔軟性を確保するためにクラウドネイティブ技術を利用することは、重要になっています。

※1: <https://www.slideshare.net/kentamagawa/ss-8023416>

※2: <https://github.com/cncf/toc/blob/main/DEFINITION.md>

CHAPTER.3

データ活用の現場で起きている課題

クラウドサービスを用いて、データ活用プラットフォームを構築したものの、想定以上にデータ活用が進まないケースも存在しています。

ビジネス観点やユースケース不足、人材が不足しているという課題も大きいですが、データ活用プラットフォームには、運用性や拡張性、データの品質や利便性、利用するユーザの組織・文化との軋轢といった数多くの課題が発生しています。

データを集め、蓄え、活用するデータ活用プラットフォームの基本機能の課題は、如何に意味あるデータを早く集め、使いやすい状態にして、実際の業務適用していくか？つまり、「アジリティの向上」がテーマ。

データがいままで以上に業務で使われなければ、新しい価値は生まれない

クラウドでのデータ活用プロジェクトで、オンプレミスからクラウドストレージにデータをリフトし、クラウド型DWHなどにデータをロードし、BIツールから参照するというのがほとんどすべてのプロジェクトの第一歩となっているのではないのでしょうか？もちろんこの一歩目も大きな歩みですが、次の一歩に悩んでいるお客様も多いのが現実です。

データ収集対象から如何に早く・簡単にデータを収集するか？特に、SaaSサービスとの連携などでは連携方式、オンプレミスのシステムとは既存有識者との調整などで、データ収集・蓄積をどういった優先順位で、如何に低コストで、価値あるデータを集めるかの課題は多いです。これはデータ戦略の課題とも言えます。データ活用に価値がある自社のデータはどこにあるのか、2ndParty/3rdPartyのデータとどう組み合わせるかなども、活用視点で考えないといけません。さらに、どうしてもデータのことになると、業務知識が必要になり、中央集権的に集められた組織（共通のIT部門）だけでは、データの中身のことかわからず、業務チームに聞きながら対応することも多く、どうしてもボトルネックが発生していることが多いです。

データ活用の各種ツールを利用して、インサイトを得たとしても、それを利用して業務を実行したり戦略・戦術に変換していかないと、真の意味でのビジネス的な効果は大きく得られません。最後の業務適用についても、実際に、データ活用の効果を真剣に期待するお客様で大きな課題となっています。業務適用するには、そのデータの鮮度やリアルタイム性が課題になっている場合もあります。

データマネジメント領域での課題は、整備されたデータをどうやって活用しやすい状態にしておくか？つまり、「データの利便性向上」がテーマ。

データが今まで以上に使いやすい状態でない、データドリブンの文化が育たない

とにかくデータを集めたものの、データがどこにあるかわからない、データの品質が悪かったり、データとデータの紐づけが難しく、データが使いづらい状態では、データドリブンの活動を行おうとしたユーザの変化を取り逃がしてしまったり、想像以上にデータ活用に時間がかかってしまったり、場合によっては異なるアクションを起こすことになってしまいます。

データカタログの整備なども、いくつかの企業では始まりましたが、様々な理由でうまくいっていないことが多いのも現実です。その中で、大きな課題の1つとして、データマネジメントに関するカタログ導入を推進する組織と、業務部門との協力関係に課題があることが多いです。カタログ整備に関しては、プラットフォーム全体での相互協力が非常に大きな導入ポイントではないかと考えます。

データの品質や、データの名寄せなど、実際のデータに関する課題も多いです、比較的きれいな状態のデータモデルを定義し、利用者に利用しやすいように準備を試みているプロジェクトもあります。しかし、活用要件がない場合や、あったとしても将来の活用シーンは予想できないため、適時変更しながら、データスチュワードのようなアジリティを持った体制や組織で対応している場合もあります。ただ、これにも限界があり、成果にできるだけ直結する価値の高いデータをどのように整備するかについては、課題が多いテーマの1つです。

非機能領域での課題は、可用性や拡張性や性能に関する解決策がサービス側から提供されるようになってきており、プラットフォームそのものへのアクセスが容易か？と、か長期的な観点でシステム運用が楽か？つまり、プラットフォーム自体の「総合的な運用性の向上」がテーマ。

データ活用のプラットフォームの利便性・アクセス容易性や運用性も重要

皆さまのデータ分析環境は、そもそもログインなどが容易に可能でしょうか？昨今、各種クラウドサービスが利用されることが多くなっていますが、SSO(Single Sign On)でその各種サービスに入ることが可能でしょうか？その対象に、データ活用プラットフォームが入っていますでしょうか？改めて、そもそもデータ活用プラットフォームに多くのユーザを誘致できていないという課題もあります。

また、データ活用プラットフォーム全体の運用は比較的楽なものでしょうか？どうしても、新しいことを後から追加していくような、拡張的なプラットフォームを構築方針とすることが多く、運用容易性の向上が後回しになったり、バージョンアップなどのライフサイクルへの対応などで、運用現場が疲弊しているようなシーンも、徐々に増えてきています。

CHAPTER.4

レガシーなデータ活用プラットフォームを進化させる “Modern Data Stack”というトレンド

“Data Stack”

“Modern Data Stack”の説明に入る前に、“Data Stack”をまず説明します。Data Stackとはデータ活用プラットフォームを構成する独立した製品群を指します。データベースやBIツール等はもちろん、データのロードや加工処理をするETL製品や、ビジネスメタデータを整備して、データ活用を促進するデータカタログ製品などもわかりやすいと思います。データの収集・加工、メタデータ管理、データ品質管理、データパイプラインの管理やセキュリティといった、データアナリティクスに関わる独立した製品群の総称です。

典型的なデータ活用プラットフォームは、データ収集、加工、蓄積、分析、パイプラインといったレイヤーで構成されます。Data Stackはどれかのレイヤーに機能を特化していることが一般的で、データ活用プラットフォームを作る際は、必要なレイヤーを定め、ビジネス要件を満たすData Stackを当てはめて、プラットフォームを構成していく流れが多いと思います。

従来のData Stackの課題

近年、様々な企業がデータ活用に注目し、データ活用プラットフォームを求める動きが加速しました。データ活用プラットフォームを求める動きが活発化する中で、前章で説明したとおり、データ活用の現場で課題が次々に明確となっています。ここでは、これまでのData Stackとレイヤーの課題に着目して例をあげます。

・データ収集・加工レイヤー

- 前章で記載した通り、ETLの考え方が主流でした。しかし、クラウドが広がりを見せた後も、オンプレミスのETLの考え方から抜け出せない製品が多かったです。
- ✓ クラウド利用にも関わらず、ライセンス体系がオンプレ時代の考え方であるコア課金・ノード課金・コネクタ課金。
- ✓ 処理エンジンがIaaSに移行したのみで、ワークロードの分離ができず、バッチウィンドウを定め利用時間の制限、ワークロード毎にサーバリソースの割り振りを設定する等、複雑な設計や運用。
- ✓ 重要なデータエンジニアリングでは、バージョン管理が煩雑、データリネージュが管理できない、ドキュメントの整備が複雑。

・データ蓄積レイヤー

- オンプレミスのRDBMSをベースとしたサービスで作られることが多かったレイヤーであるため、クラウド特性を活かせないことが多いレイヤーでした。
- ✓ データレイク、DWH、データマートが乱立し、データがサイロ化。
- ✓ クラウドネイティブでないため、ディスクサイズやコンピューティングリソースの事前計画、運用管理が必要。

・データ活用レイヤー

- データからインサイトを得て、業務活用していくレイヤーであり、業務ユーザーが直接触る箇所ですが、プラットフォーム要件が優先され、ユーザーの声が届きにくい特性があります。
- ✓ 分析者増加に伴う、DaaS環境や共通サーバの払い出しによる、環境払い出しや環境のメンテナンス、運用コストの増大。
- ✓ リソースの共用によるノイズネイバーの問題。
- ✓ SoEとの分断による連携の難しさ。
- ✓ 制約の多さによるユーザービリティ低下（インターネットに接続が許されず、SaaSが使えない。必要なデータやツール、ライブラリにアクセスできない等）

・ガバナンス・セキュリティレイヤー

- 全レイヤーを横断するため重要なレイヤーですが、課題が多く、運用コストが増大するレイヤーとなっています。
- ✓ データ配置箇所の増大に伴うデータガバナンスの煩雑化。
- ✓ サービス増加により認証が分散してしまい、統一的なアクセス制御ができない。
- ✓ データサイロ化により、データアクセスの認可制御が困難。

従来のData Stackで要件を実現しようとした結果、技術制約等で実現が難しく、ユーザービリティが低下し、本来の目的であるデータ活用が進まないシステムになってしまうケースや、手続きが多くなり意思反映に時間のかかる仕組みになるケースが生じています。

ここにあげたものは一例ですが、皆様の中には、現在運用中のシステムで当てはまる項目が思いつく方もいらっしゃるのではないのでしょうか。

導入したデータ活用プラットフォームに課題が多いと、利用の拡大が難しくなります。データ分析の結果をビジネスに活かすことができず、ROIが低くなることもあるでしょう。どうか課題を解決しようと工夫を凝らした結果、ドキュメントでの管理、運用の増加につながり、システムの保守運用が耐えられないケースもあります。酷い場合は活動が停止するケースまであります。Data Stackの抱えている課題の解決が重要で求められています。また、どのような役割を持つData Stackを採用するか、どのようなレイヤーが必要か、これらの検討がとても重要です。

“Modern Data Stack”とは

ここまでData Stackという言葉と、Data Stackの抱える課題を紹介しました。では、“Modern”とは何を指すのでしょうか。

気をつけていただきたいのは、“Modern Data Stack”とは、何か特定のソリューションや技術、アーキテクチャを指す言葉ではありません。

前述したデータ活用現場での課題を解決するため、世界中のプロダクトベンダーが、次々にソリューションを改善し、新しいソリューション、新しい考え方を生み出しています。これらの改善や新規ソリューションには、後述する例のように、いくつか共通的なトレンドがあります。

従来のData Stackやレイヤーの課題を解決しようとする、Data Stack技術動向には共通的なトレンドがあり、Data Stack進化に関わるこのトレンドと、トレンドに沿ったData Stackを指す言葉が、“Modern Data Stack”です。つまり“Modern Data Stack”とはトレンドの総称なのです。

世界中のプロダクトベンダーは、自社の製品がデータ活用現場の課題をどのように解決するかを明示し、Modernであると主張することで他社製品との優位性を表しています。

- ・データ収集・加工レイヤー：SaaS化によるインフラ管理コストの低減や、課金体系変更により従量課金となり、柔軟なリソースのスケールが実現しており、アジリティの向上につながっています。当社パートナーであるInformatica社やdbt Labs社など、様々な製品がクラウドネイティブの考えで改善を進めています。

- ・データ蓄積レイヤー：Snowflake や Databricksのように、データレイクやDWH両方の特徴を併せ持ち、DataCloudやレイクハウスといった新しい考え方のサービスが台頭しています。これにより、データの一元化が可能となり、データのサイロ化の課題を解決しています。また、コンピュートリソースとストレージを分離することで、スケーラビリティの向上、ワークロードの分離が可能となりました。

- ・データ活用レイヤー：ユーザーが干渉し合わない分離した環境の提供が可能となっています。ユーザー毎に自由にライブラリ環境を整えられ、ユーザービリティの改善につながっています。リバースETLの考え方により、これまで分離されていたデータ活用プラットフォームと業務システムを相互に連携し、本来の目的であるデータ活用結果を、ビジネス価値に結び付ける製品も出てきており、重要となってきています。

- ・ガバナンス・セキュリティレイヤー：データカタログ製品によるデータのメタデータ管理が可能となってきています。運用負荷が高く課題は残っています。各データソースに統一的なアクセス管理を可能とする製品も出てきています。

こういった例に示した、Data Stackの進化のトレンドの総称が“Modern Data Stack”です。

自社のデータ活用プラットフォームのモダナイズ化のために求められること

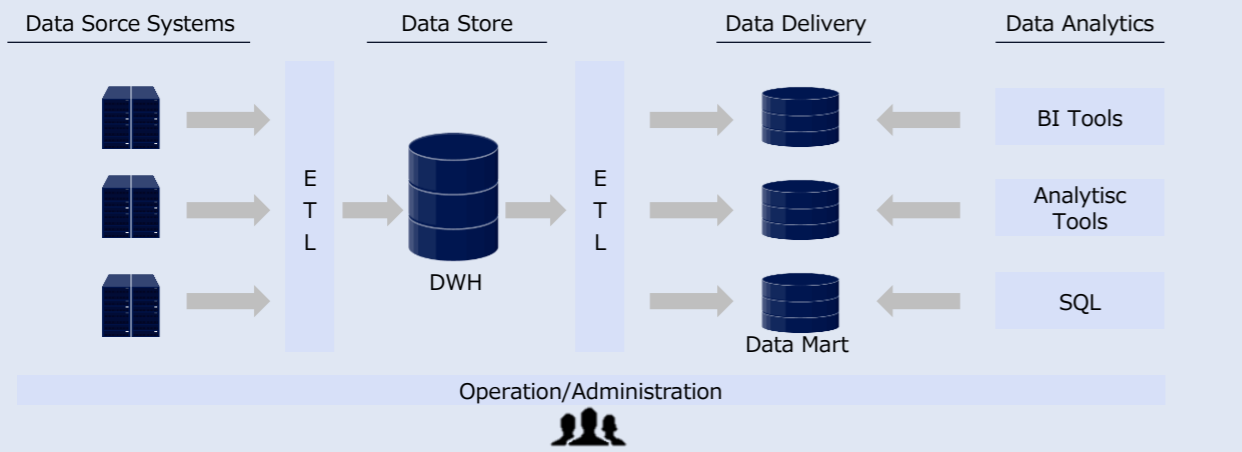
ある構成がModern Data Stackであると主張し、全ての企業に最適なただ一つのアーキテクチャを定めることはできません。

Modern Data Stackというトレンドからデータ活用プラットフォームの各レイヤーに求められる要件を具体化し、数あるData Stackの特性とのフィット感を見極めることが求められます。また、データ活用現場での課題解決と要件充足のために、Data Stackを選定・組み合わせ、Best of Breedでデータ活用プラットフォームを提供できる、体制、ナレッジ、プロダクトベンダーとのアライアンスの整備が求められます。

Modern Data Stackは進化し続けるトレンドであり、アーキテクチャが完成することはありません。自社の目指す姿をとらえ、今の実現できる技術と先を見据えたアーキテクチャを検討して改善を続けることが求められます。

図1 従来のデータ活用プラットフォームのレイヤー概念

データソースからデータ活用レイヤーまでデータの流れは一方方向となっており、データ活用が流れの終端となる構成。1990年代から変わっていない。



CHAPTER.5

NTTデータが考えるモダンなデータ活用プラットフォームの特徴と当社の取り組み

今後求められるデータ活用プラットフォームのレイヤー像

Modern Data Stackのトレンドを踏まえ、NTTデータが考えるデータ活用プラットフォームの特徴を説明します。

従来のデータ活用プラットフォームのレイヤーは、データソースがスタートにあり、データ分析やBIツールによる可視化といった活用レイヤーをゴールとし、一方通行のデータの流れる絵で描かれることが多かったです（図1）。これは1990年代に提唱されたDWHシステムの概念図から、今に至るまで大きな変化がありませんでした。これまでの一方通行の流れる構成では、データ活用プラットフォームの目的である、データ分析結果の業務適用に結び付けることができません。目的を達成するために必要となる考え方は「環流」です。以下、重要となるデータおよびOpsの環流について解説します。データ活用プラットフォームでは、データソースとして業務システムや外部ソースがあります。こうしたデータソースから取得したデータは、必要な形に加工整形され、データの分析や可視化を行って終わるのではなく、その結果を業務システムへのリバースや、外部公開といった使い方もされます。業務システムに分析で得られた結果を適用し、業務に変化があればそのデータを再度取得して分析することで、データ活用から得たインサイトの効果を測定するループが回ります。個々の利用者のデータ分析の結果をデータストアに書き戻し、それを他の分析者に共有するといったデータのコラボレーションも、データ活用の現場ではよくある要件です。

このように分析や可視化をゴールとするのではなく、データ蓄積レイヤー、データエンジニアリングを中心にデータが環流するという考え方が重要です。データ環流とともに、ビジネス価値を創出し続けるために、データ活用プラットフォーム上で様々なOpsを回すことが必須となります。

データ活用プラットフォームは、環流の中心となり、業務システムとの連携がより重要になってくるため、業務システムのサブシステムとしての位置づけではなく、業務の根幹を担う位置づけとなるでしょう。

NTTデータが考えるデータ活用プラットフォームの特徴

NTTデータはこれまで多くのお客様にデータ活用プラットフォームを導入してきました。この節では、お客様への導入実績を踏まえ、当社が考えるデータ

活用プラットフォームに求められる特徴を整理します。これらは今後現場で求められることであり、システム企画時に考慮すべき特徴です。

1. クラウドネイティブ

クラウドネイティブの考え方のData Stackを採用し、システム検討ができていくことが重要です。

- ・ ストレージの容量無制限、コンピューティングリソースのオートスケール、従量課金といった特徴を持ち、事前の細かいキャパシティプランニングが不要、柔軟性があること。
- ・ ユーザの増加やワークロード量の増加への対応、容易なリソースの増強、プロダクトベンダーとの契約プラン変更のような、リードタイムを伴う作業回避が可能で、拡張性があること。
- ・ 様々なワークロードからの、独立したアクセス（バッチやユーザ間の干渉がない）が可能であること。多様な働き方やグローバル展開している企業の増加等、一律に利用可能な時間帯を定めるのが難しいケースが増えており、限定しないサービス解放時間の設定や、ユーザ同士のリソースの奪い合いがない環境の実現が可能であること。
- ・ 障害が生じた場合に影響を抑え、データ損失を防ぐことが可能な、回復性があること。

2. 統一的セキュリティ・ガバナンス

Data Stackおよびデータの横断的な管理により、セキュリティやガバナンスを効かせることが重要です。

- ・ 増加していく異なるData Stackやデータソースに対して、統一的な認証認可による、データアクセス制御が可能であること。Modern Data StackではSSOによる認証はもはやデファクトスタンダードであり、セキュリティ向上と運用負荷低減を可能としています。
- ・ メタデータ管理の仕組みを組み込むことが可能なこと。データガバナンスの重要性が増しており、データカタログ製品導入の検討や運用方針の策定が必要です。

3. ビジネスアジリティの加速

作ったデータ活用プラットフォームで価値を創出し続けるためには、分析して終わりではありません。SoE領域と連携し、価値を出して意味のあるものとなります。継続的に価値を出し続けるためには、DataOps、MLOpsやDevOps等、目的に応じたOpsを回さなければなりません。

データ分析を行いながら、データを改善していく、分析結果は素早く業務システムに適用し、業務を改善していく。その結果をもとにさらにデータを改善して、業務へ活かす。データと業務連携のアジリティの高さを実現するのがModern Data Stackの大きな特徴です。

NTTデータの取り組み

データ活用プラットフォームを求める企業は、Modern Data Stackの考え方の広がりにより、選択肢や考慮事項が増え、自社の目的を達成するための最適解を見つけることが、益々難しくなっています。NTTデータでは、Modern Data Stackのトレンドを踏まえ、お客様のデータ活用要件を満たすデータ活用プラットフォームを実現するために、次の営みを行っており、Best of BreedでModern Data Stackの提供推進と拡大に取り組んでいます。

1. 技術調査と検証

当社ではグローバルな経営基盤を活かし、新しいData Stackの素早い調査と、技術検証を行っています。当社が考えるModern Data Stackの特徴を備えているのか、お客様へ導入する際に課題となりそうな事項はあるか、Data Stackがターゲットとする機能に市場のニーズはあるかといった様々な観点で、実際にプラットフォームを構築して検証を行います。検証結果をナレッジとして社内に蓄積・共有し、お客様要件を満たせるData Stackの、迅速な選択とアーキテクチャ提案を実現します。

2. プロダクトベンダーとのアライアンス

ナレッジ蓄積のみに留まらず、当社では、お客様への価値につながると判断したData Stackのプロダクトベンダーに一早くアプローチをし、アライアンスを構築し、お客様へのデリバリ体制の強化を行っています。

例えばSnowflakeとの国内唯一の資本業務提携の合意やDatabricksとのパートナーシップ締結を実行しています。また、データの環流で中心的なレイヤーとなるデータエンジニアリングでは、dbtと呼ばれるプロダクトに着目し、dbt Labsといち早くパートナーシップを結び、お客様のデータ活用のアジリティ向上につながるデリバリ体制を強化しています。

3. Trusted Data Foundation(TDF)の進化

検証の結果やアライアンスの変化に合わせて、TDFを継続的に進化させ、リファレンスアーキテクチャのアセット化を行い、お客様への迅速な提供を図っています。

Data Stackのモダナイズ化一例

最後にNTTデータが導入したModern Data Stackトレンドに沿ったデータ活用プラットフォームの一例（図3）を紹介いたします。

お客様は、IaaSベースのアーキテクチャで、定期的な高額コスト、ライセンス制約やIaaSベースであることによる拡張性のなさ、運用の煩雑さに悩まされていました。海外拠点もありましたがワークロードの干渉で利用拠点拡大の要件実現も難しい状況でした。

システム構成を一新し、APIによるデータの収集と、データストアのレイヤーにはSnowflakeを採用しワークロードの分離とスケラビリティ、コストの最適化を実現しました。活用レイヤーではTableauによるBI、SaaS採用でスケラビリティのあるデータ分析環境を整備し、ユーザビリティの向上を図っています。小さく始めて拡大していくことを大きな方針とし、必要な機能を選定しながら順次データ活用プラットフォームの拡大を行っています。

図2 今後求められるData Stackレイヤー

データの流れるは一方通行から環流となり、業務や分析の結果をもってデータの改善を行い、それをまた業務に活かしてサイクルが回せる構成が求められます。

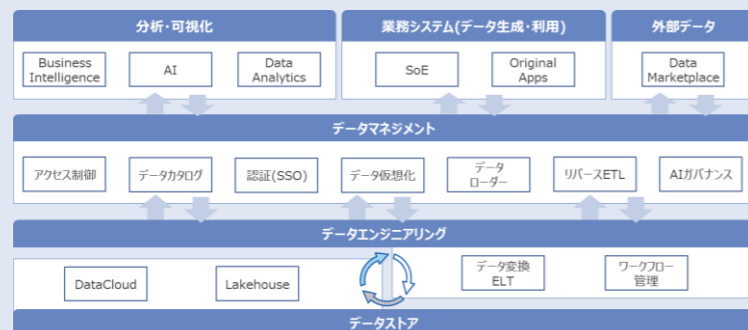
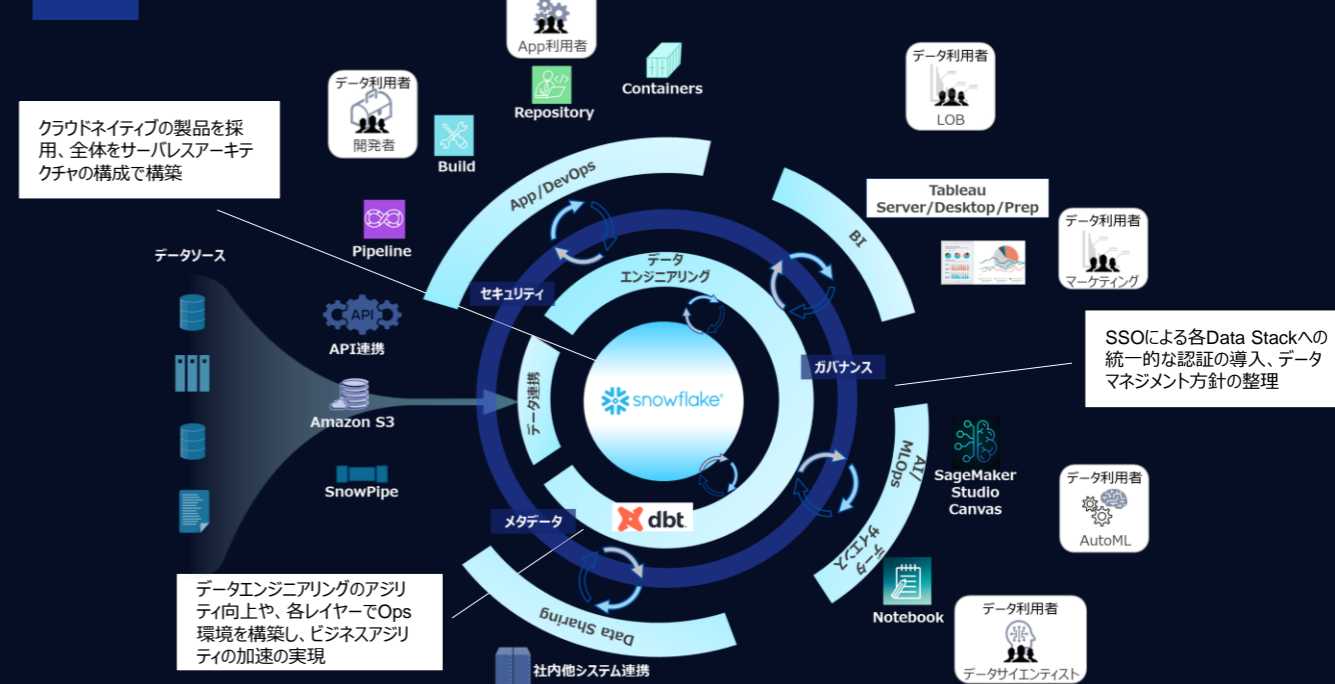


図3 モダナイズ化実績案件一例



CHAPTER.6

<組織・企業文化に関するトレンド>

データの分散管理



グローバルで今一番注目を浴びており、様々な議論・活動がなされているトレンドとしてデータの分散管理があります。この考え方を実現する手段としても Modern Data Stackは取り上げられ、その相性は非常によいとされています。

NTTデータでは、世界各国のグループ会社と共に、現地のお客様の生の声を集めて最新のトレンドについて研究しています。その中で注目されているトレンドの一つに、データメッシュというキーワードに代表されるデータの分散管理があります。現在、特に海外のお客様では、CXOレベルのお客様からデータメッシュに関するコンサルティング、ソリューション提案のご依頼を多くいただいております。すでに多くのプロジェクトが進行しています。日本においても欧米のお客様と同様の課題が見られるお客様は多く、今後大きなトレンドとなることが予測されます。

データメッシュの4つの基本原則

「データ活用の現場で起きている課題」でも触れられていたように、現在の企業は中央のプラットフォームと中央のデータ管理組織がデータ活用のボトルネックになっています。データメッシュはこの課題に対処するための組織論も含めたコンセプトです。

データメッシュは以下の4つの基本原則を掲げています。

- Domain Oriented Ownership
組織をドメインという単位に分け、ドメインごとに自律的にデータを管理し、他の組織に提供します。中央の組織に頼らずにデータを管理・提供することで、それがボトルネックになることを避けます。
- Data Product Management
データを製品として他の組織に提供します。データを提供するドメインは、データの連携先を顧客と捉え、データの説明責任と信頼性の確保に努めます。データの管理を各ドメインに任せることによって生じるサイロ化の問題を避けます。
- Self-Service Data Platform
各ドメインが独力でデータプロダクトを開発・運用できるよう支援する機能と、データ利用者がセルフサービスでデータを扱える機能を備えたプラットフォームを提供します。ITに詳しくない人でも使いやすいプラットフォームを提供することで、各ドメインが自分たちでデータを管理、利用することによるコストの増加、人材確保の難しさを緩和します。
- Federated Computational Governance
個人情報保護などの共通ルールやデータプロダクト開発のインセンティブを設計することで、エコシステム全体の最適化を行います。共通ルールは中央で管理・統制されるのではなく、各データプロダクトの中に実装されます。中央集権的な体制がないことでデータが管理されていない状態になることを避けます。

このようにデータを分散管理することで中央のプラットフォームと組織がボトルネック化することを避け、組織におけるデータ活用のスケーラビリティとアジ

リティを高めることができます。

データの分散管理を促進するMDS

データの分散管理を実現する上での大きな課題は、各ドメインがセルフサービスでデータの加工やデータの提供、コンプライアンスの準拠やメタデータの整備など、今まで中央のデータ管理組織がやっていた作業をやらなければならないことです。多くの企業では各ドメインに十分な数のデータエンジニアやデータスチュワードといったデータの専門家を配置することは難しいでしょう。この課題に対してMDSには大きなメリットがあります。

「Modern Data Stackトレンドの特徴」でも触れたようにMDSには以下の特徴があります。

- クラウドネイティブな仕組みにより事前の見積もりや細かいパフォーマンスチューニングが必要ない
 - ビジネスアジリティを確保するために、エンジニアリブで使いやすい
- もし前述したSelf-Service Data PlatformがMDSのツールを提供するのであれば、ドメインによる独力のデータプロダクトの開発と運用は、データの専門家なしでも非常にやりやすくなるでしょう。MDSはデータの分散管理によって生じる弊害を緩和し、企業におけるデータ活用のスケーラビリティとアジリティをさらに高めてくれるのです。

データの分散管理のコンセプトにおけるMDS導入のポイント

MDSはデータ活用にとって非常に強力ですが、さまざまな種類のMDSを無計画に導入するとつぎはぎだらけの管理されていないプラットフォームになってしまいます。これを防ぐために以下のポイントが重要です。

- ドメイン間のデータ連携に標準パターンを作る
 - 標準パターン毎に最適なMDSのツールを（理想的には）1つに絞る
 - データプロダクトのインターフェースにREST APIやSQLなどデファクトスタンダードなものを採用することで、データプロダクトの中で使っているツールがデータプロダクトの外から直接見えないようにする
 - Self-Service Data Platformがそのツールを各ドメインに提供する
- これにより組織内にさまざまなMDSツールが氾濫することを防ぐことができます。また、今後より優れたMDSツールが出てきた時に、それに乗り換えることが容易になります。

MDSはデータの分散管理によって生じる弊害を緩和し、企業におけるデータ活用のスケーラビリティとアジリティをさらに高める。

CHAPTER.7

NTTデータが提供する

「価値向上の手法：Value Engineering」

ビジネス成長に順応しながら、ビジネス価値を高めるデータプラットフォームを目指し

これまでの内容で、Modern Data Stackやデータメッシュといったトレンドをご理解いただけたことと思います。トレンドが生まれた背景やそれがもたらすベネフィット、いわばテクノロジーの本質を理解して、自社の課題解決に役立つか見極めるとともに、ビジネス価値を創出していくことが重要です。

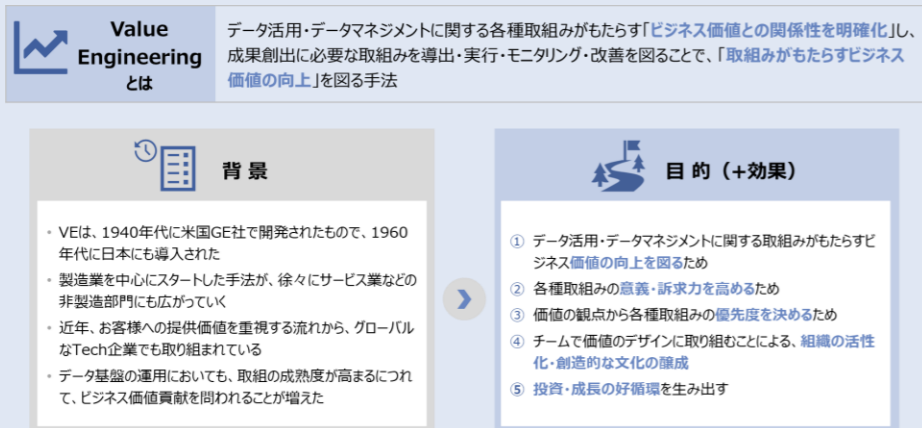
ただ、「データ基盤の運用によりビジネス成果を創出していきたくいけど、運用負荷が高まるのは嫌だ。」

データ基盤を運用し、利用者・データのニーズに追従していくと、みなさんが陥る課題だと思います。日々発生する運用課題と、利用者ニーズの充足、

マネジメント層からのビジネス成果創出期待への対応に苦労していないでしょうか。MDSは柔軟性やアジリティを向上させることによって、ビジネス成長を阻害しない、いわば「順応型」のデータプラットフォームです。それに対し、ビジネス価値を創出・向上させるアプローチとして、「Value Engineering」があります。

「Value Engineering」は、1940年代に米国GE社で開発されたもので、製品やサービスの価値向上をシステム化された手順によって行うものです。製造業を中心にスタートしたこの手法が、徐々にサービス業などの非製造部門に広がっていきました。近年のお客様への提供価値を重視する流れから、グローバルなTech企業でも取り組まれている手法です。(図4)

図4 Value Engineeringの定義と目的

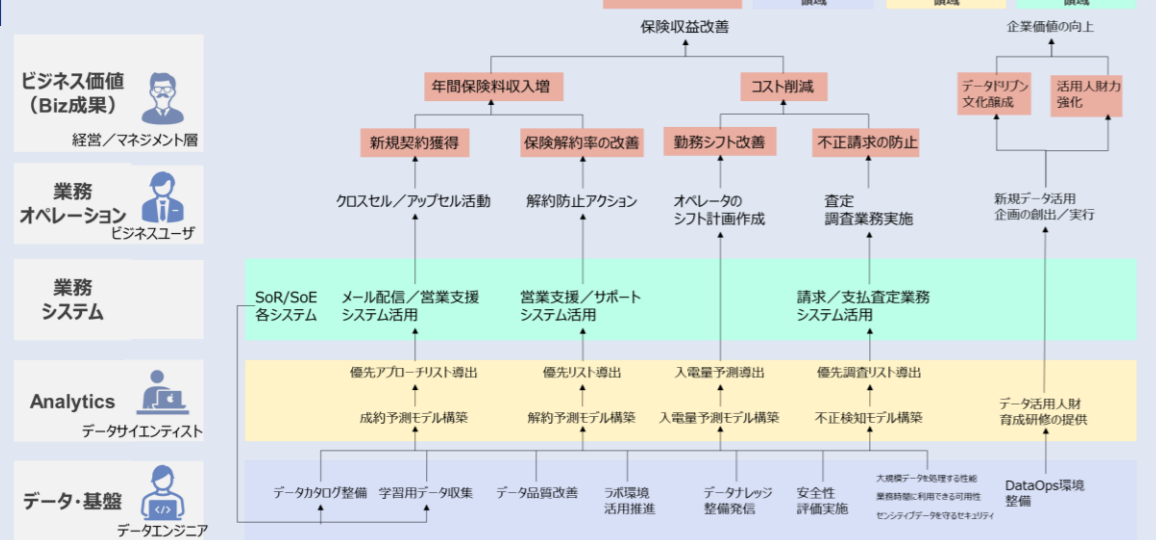


Value Engineeringとは、価値を創出・向上するための「タテ」のデザインである

ビジネス価値を創出・向上させるためには、「データ基盤」「Analytics」「業務・顧客接点システム」「業務オペレーション」が

有機的に連携する必要があります。(図5) 各レイヤでのあるべき姿をデザインする「システムアーキテクチャ」「Analyticsのユースケース」「ビジネスプロセス」「サービス」などのデザインは、いわば「ヨコ」のデザインであるのに対し、各レイヤを連携してビジネス価値をデザインすることは、いわば「タテ」のデザインといえます。

図5 価値創出・向上のための「タテ」のデザイン (Valueツリー)



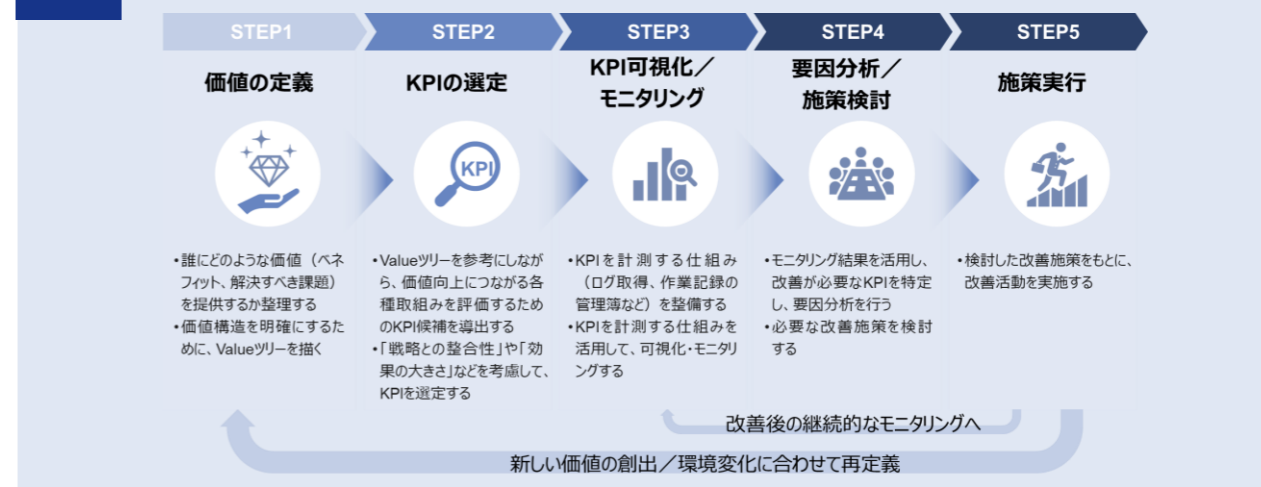
Value Engineeringの活動は、5つのSTEPから構成されます。効果的に進めるためには、継続的に価値向上を目指し、STEP1/STEP3へ戻る循環型の取組みとすることです。(図6)

- STEP 1 : 価値の定義
- STEP 2 : KPIの選定
- STEP 3 : KPI可視化/モニタリング
- STEP 4 : 要因分析/施策検討
- STEP 5 : 施策実行

データ戦略を描き、必要なデータ基盤を準備するまでたどり着けたとしても、ビジネス環境の変化に伴い、利用者・データ・活用テクノロジーが変化していくなか、運用を通じてビジネス価値を創出・向上させるためには、以下の「問い」に答えられることが重要です。

- 問1 : 創出・向上を目指しているビジネス価値は何か?
- 問2 : 連携するステークホルダとの関係性は良好か?
- 問3 : 取り組みを評価するKPIは何か?
- 問4 : 現在の状況と、注力すべき課題は何か?

図6 Value Engineeringの取組全体像



NTTデータにおけるValue Engineeringの取組み

お客様へ提供しています。(図7) そこでValue Engineeringの手法を活用し、各領域のプロフェッショナル人材が有機的に連携することで、お客様と共に創出すべきビジネス価値を見極め、成果創出を実現しています。

NTTデータでは、データドリブンでお客様のDXを推進するために「ビジネス」「IT・Tech」「データ&アナリティクス」「人財・組織」の4要素をバランスよく成熟させていくべく、デジタルサクセスの支援をお

図7 「デジタルサクセス」(全体イメージ)



おわりに

モダンな力で進化を続ける旅路へと



本ホワイトペーパーでは、データ活用プラットフォームに関連する「技術の進歩で何が起きているのか」「データ活用の現場で起きている課題」を交え、データ活用プラットフォームを進化させるための“Modern Data Stack”というトレンドや関連する当社の取り組みをお伝えしてきました。

いかがでしたでしょうか。皆様のデータ活用の置かれている状況について、改めて考えていただく機会となれば幸いです。

不確実性の高い世界で、企業の持続的な成長を実現するために、あなたのシステムにも“Modern Data Stack”というトレンドを取り入れ、進化を続けるチャレンジを当社とともに取り組んでみませんか。

注釈：

文章中の商品名、会社名、団体名は、各社の商標または登録商標です。

■ 執筆者：

株式会社NTTデータ

デザイン&テクノロジーコンサルティング事業本部

村山 弘城

亘 健典

樋口 翔

株式会社NTTデータグループ

技術革新統括本部

大山 真実

執筆協力：上記各組織技術者一同

■ 問合せ先：

株式会社NTTデータ

デザイン&テクノロジーコンサルティング事業本部

<https://enterprise-aiiot.nttdata.com/contact>