

迅速なデジタルデータ活用のための 段階的プラットフォーム構築

昨今、あらゆる情報をデータとして活用し、ビジネスに生かすことが経営課題となっている。そのためにはデータを分析・活用するプラットフォーム構築が不可欠だが、多くの場合業務要件が定まっておらず、構築に苦勞するケースが多い。本稿では実証実験の例を参考に、データ活用プラットフォーム構築のアプローチを探る。

野村総合研究所
システムコンサルティング事業本部
ITアーキテクチャーコンサルティング部
上級システムコンサルタント

しもだ たかし
下田 崇嗣

専門は基盤を中心としたシステム化構想



野村総合研究所
システムコンサルティング事業本部
ITアーキテクチャーコンサルティング部
上級システムコンサルタント

すぎた なおや
杉田 直哉

専門は基盤を中心としたシステム化構想



進む情報のデジタルデータ化

近年、デジタルデータを取り巻く技術が数多く登場し、これまで収集できなかったあらゆる情報がデータ化され、企業がそれを活用するという新たな潮流が起きている。

IoTデバイスからのデータ収集では、高速で大容量のモバイルネットワークの次世代5G規格や、物理的なコンピューティングリソース配置の最適化を図る技術の導入検討が進んでいる。また、データを分析するAIの技術では、IT大手クラウドベンダー（米国のAmazon.com社、Google社、IBM社、Microsoft社など）が自社のプラットフォームにおける実装を進めている。

データ活用プラットフォームの リファレンスアーキテクチャー

デジタルデータの活用には、基盤となるべきプラットフォームの構築が必要である。プラットフォームの検討には、まずデータ分析に必要なプロセスの理解と、扱うデータの構

造、処理方式の検討が重要となる。

データ分析プロセスの参考として、CRISP-DM（Cross-Industry Standard Process for Data Mining）と呼ばれる標準プロセスが存在する。これは、ビジネスの理解、データの理解、データの準備、モデリング、評価、展開といったデータの発生から価値を生み出すまでの流れをデータのライフサイクルとして捉えたものである。プラットフォームに必要な機能は、このライフサイクルを意識して階層別に分解し、整理することができる。また、分解した階層ごとに必要となる機能は、扱うデータの構造と処理方式で洗い出すことが可能である。

データ構造には大きく2タイプあり、1つは基幹システムに蓄積される構造化データ、もう1つは文書、画像、動画、音声のような非構造化データである。ビッグデータは、8割以上が非構造化データといわれておりプラットフォームも基幹システムとは異なる基本設計を考える必要がある。

データの処理方式も2タイプに大別することができる。1つは一定期間データを蓄積し

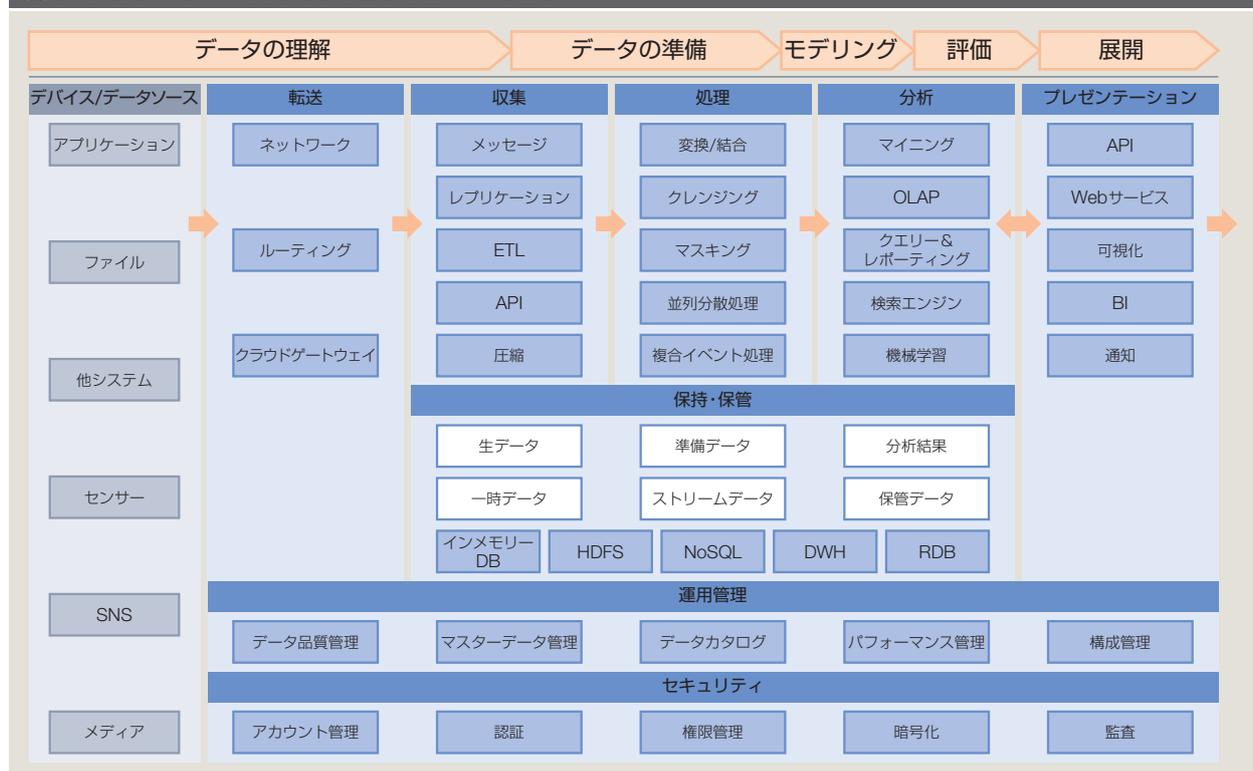
て処理するストック型、もう1つは時系列に発生するデータを連続的に処理するフロー型である。IoTが生み出す大量のデータをリアルタイムに分析する処理はフロー型処理に該当する。

野村総合研究所（NRI）ではデータ活用プラットフォームのリファレンスアーキテクチャーを体系化している（図1参照）。リファレンスアーキテクチャーとは、ITの一般的な用途に対応する標準的な使用方法とシステム構成をまとめたものである。NRIのリファレンスアーキテクチャーは8つの階層で構成されているが、この中の「転送」「収集」「処理」「保持・保管」「分析」「プレゼンテーション」の6階層はデータ分析プロセスを基に階層分けを行い、各階層にはデータ特性、データ処理方式を考慮した機能要素が含まれる。

残りの「運用管理」「セキュリティ」はプラットフォーム全体で必要となる機能群で構成されている。

データ活用プラットフォームを構築する際、各階層にある機能要素全てをそろえる必要はない。その機能は代表的なものであり、データ構造やデータ処理方式により必要な機能や組み合わせ方は変わってくる。業務システムの構造化データを扱う分析であれば、ETL（Extract/Transform/Load）ツールを使ってデータ収集、加工処理を行い、DWH（データウェアハウス）やRDBなどに蓄積してBI（Business Intelligence）ツールを使って分析、可視化することができる。大量の生データを高速にリアルタイム分析してアクションにつなげたいような場合は、Sparkのようなインメモリー高速分散処理フレーム

図1 プラットフォームのリファレンスアーキテクチャー



ワークや、CEP（複合イベント処理）エンジンなどの活用が考えられる。どのようなデータをどのように処理するかを見極め、必要な技術を組み合わせてプラットフォームを素早く構築していくことが重要となる。

データ活用プラットフォーム構築の実証実験

デジタルデータ活用を定着させる上では、現場レベルでの実証実験においてデータから価値を創造し、経営の意思決定に役立てるための仕組みを既成事実として積み上げていくアプローチが必要不可欠である。

ここではNRIが支援した大手製造業の事例を紹介したい。A社では、実証実験としてデータ活用を素早く開始することを優先し、まず大きな枠組みとなるプラットフォームを用意する進め方を採用した。

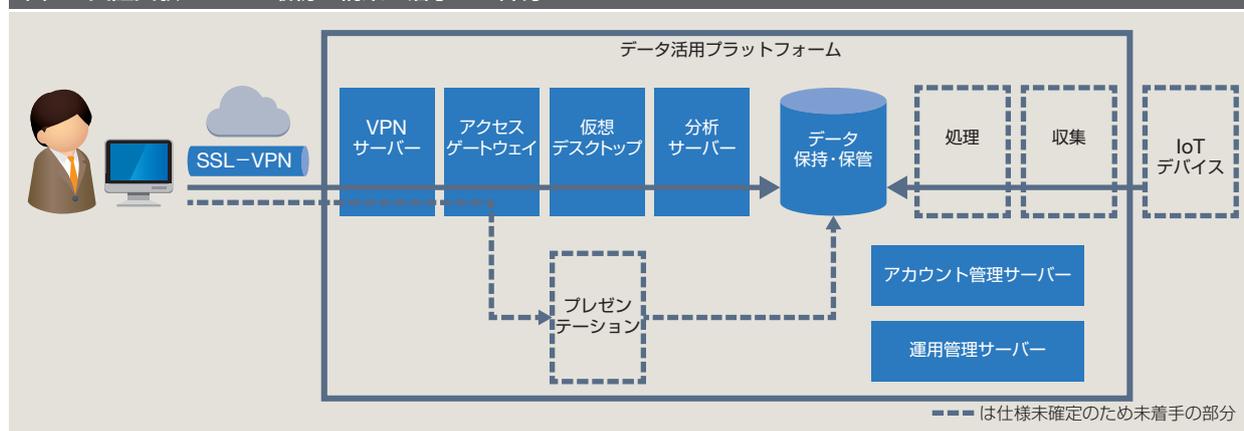
基幹システムの場合は「あるべき経営戦略やあるべき業務」が存在しているが、データ活用の場合、ニーズは都度発生するため、あらかじめ業務要件が定まっていないことが多い。このためどのようなプラットフォームを

作るべきか、正解のない問いに悩んでいる企業は多い。

A社でプラットフォーム構築の検討を開始した時点でも、業務要件として「いくつかの社内部署、複数の外部分析ベンダーが利用する」「2～3の分析テーマ候補」の2点だけが決まっているような状況だった。そのため、業務要件や分析のユースケース（システムが外部に提供する機能）以外の部分にフォーカスした。具体的には導入企業のデータガバナンスとセキュリティポリシーをベースとしてデータ活用に必要なセキュリティ機能を選定し、利用ユーザーや分析サーバー、分析データ量の増加に対応する拡張性を考慮したプラットフォームの大きな枠組みだけを要件定義として構築することから始めた（図2参照）。構築時点では誰が利用するか判明していなかったため、ツールやサービスも極力標準のものを選定した。

最初に着手したのは、既存データをアップロードして蓄積する「保持・保管」部分に加え、それを自由に操作して分析するための仮想デスクトップ環境である。これはIoTデバイスから新しくデータを収集することから始

図2 実証実験において最初に構築に着手した部分



めると、分析可能なデータ量を蓄積するまで時間がかかってしまうためである。もし、構築の時点でデータをリアルタイムに収集することが決まっているのであれば、特定の用途に対応したデータを受信して加工する「収集」と「処理」部分を開発対象として含めようと考えていたが、結局分析テーマに先行して仕様がまとまるようなことはなかった。

結果、A社は、確立した基盤構築には至らないまでも、このような実証実験で迅速にプラットフォームを構築したことにより、製品競争力の向上につながる分析を早期に実施することができた。

プラットフォーム構築におけるセキュリティ対策

必要最低限なプラットフォームをスモールスタートで始めるためには、初期投資を抑えやり直しや変更がしやすいオープンソースのソフトウェアやパブリッククラウドを使うべきである。なお、特定ベンダーのツールは、多くの便利な機能を備えているが、そのツールに習熟した人材が必要となり、ベンダー依存になる可能性もあることから、最初から導入するのは早計と考える。

そのオープンソースやパブリッククラウドにも問題点がある。実際に導入した事例では、データオーナーとなるユーザー部門や外部の分析ベンダーから導入に激しい抵抗を受ける場面があった。その理由はセキュリティへの不安である。ITに詳しくない人にとって、「クラウド」や「オープンソース」という言葉は依然としてセキュリティリスクの不

安を招くキーワードである。そのため、今回構築したプラットフォームが自社内のIT環境と同等、またはそれ以上にセキュリティが確保されることを説明する必要がある。

A社の事例では、以下のような基本的なセキュリティ対策を行った。

- ・通信データ、保管データの暗号化
- ・保管データのマスキング
- ・アクセス制御、アクセス権限の管理
- ・操作証跡の取得
- ・システム監視
- ・厳重な社員の操作制限

スモールスタートと柔軟なアプローチが効果的

デジタルデータ活用は、試行錯誤を伴うため投資対効果が見えにくい。そのような状況で、最初から製品・ツールありきでプラットフォームを構築してしまうと、無駄が生じる場合が多い。扱うデータ、活用ニーズは、ビジネスの変化、組織の変化に応じて変動する。その変化に柔軟かつ素早く対応し続けることができなければ失敗に終わってしまう。

まずは十分なセキュリティを確保した上でさまざまなデータ活用にトライし、活用の効果を実証しながら柔軟にプラットフォームの修正を繰り返していくべきである。その活動を通じて、自社にとって理想的なデータ活用プラットフォームとはどういうものかを見つけていくアプローチが効果的である。

リファレンスアーキテクチャーで全体像を把握した上で、まずは小さく始めるスモールスタートに着手してもらいたい。 ■