

デジタルトランスフォーメーションを支えるデータ活用基盤

森田 登*
Noboru Morita
鈴木雅也*
Masaya Suzuki
鈴木利幸*
Toshiyuki Suzuki

Data Utilization Platform Supporting Digital Transformation

*三菱電機デジタルイノベーション(株)

要 旨

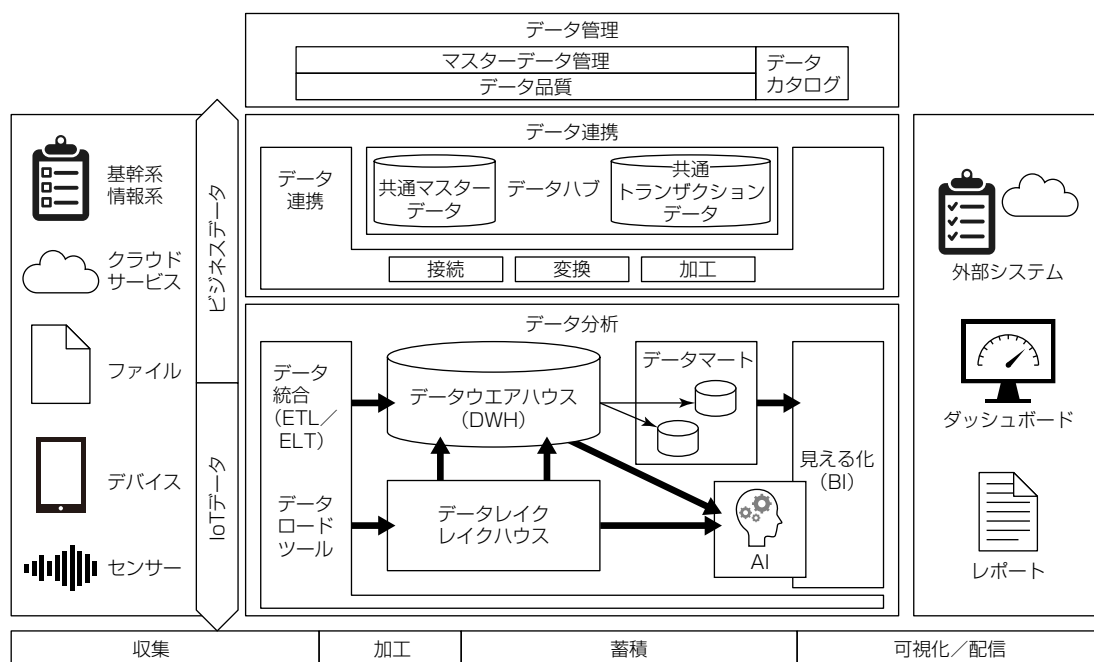
デジタルトランスフォーメーション(DX)の実現には、迅速かつ柔軟な意思決定を行う必要がある。そのためには、企業内データの収集・加工・統合を担って、継続的に高品質なデータを提供する基盤(データ活用基盤)が不可欠である。この基盤の構築は、目的や得られるメリットを常に明確にして段階的に進めるとよい。構成要素としては、共通マスターデータと共通トランザクションデータの整備、データハブによる疎結合連携等の基本的なアーキテクチャーの採用が標準になりつつある。開発時は、用いるツールやサービス、方式の特長を十分理解して選択し、常に目的を見失うことなく進めることが肝要である。

今後、AI技術の発展によってデータ活用の幅が広がって、なお一層、高品質なデータの維持管理が必須になると考えられる。それに伴って、データ活用基盤の役割と重要性は継続的に増していく。

1. ま え が き (1)

現在、多くの企業で、生き残るためにはDXを推し進めることが必須になってきている。DXの目的は、デジタル技術を活用して新たなビジネスモデルや価値の創出を行うことであり、業務効率化を目的とした従来のデジタル化だけでは不十分である。そこでは、データに基づいた意思決定が前提になり、正しいデータを自在に活用できる環境が必要になる。

本稿では、そのデータを活用できるようにするための基盤(データ活用基盤)(図1)を構築するときのポイントと、昨今のAI技術の発達を踏まえた動向について述べる。



IoT : Internet of Things, ETL : Extract Transform Load, ELT : Extract Load Transform, BI : Business Intelligence

図1-データ活用基盤

2. データ活用基盤の必要性

企業内のデータを活用するためにデータを収集して分析を進めようとしても、ソースとなるデータはそのままの形では集計・分析できず加工が必要になる。さらに、活用する必要があるデータの範囲の広がりとともに、その複雑度は増している。そのため、データの収集と加工は、変化に対応できる継続的な仕組みなしでは実現できないのが現実である。データ活用基盤はこの課題に対応し、継続的なデータ活用環境を実現させるために必要な機能を具備したシステムである。

3. データ活用基盤の構成要素⁽²⁾

データ活用基盤の主な構成要素と、その役割を表1に示す。全ての要素が必要というわけではなく、用途によって重要な要素を優先的に実現させていくのが一般的である。

表1-データ活用基盤の主な構成要素

分類	主な構成要素	役割
データ管理	マスターデータ管理	重複/異口同音を排除し、統合された真のマスターデータ
	データカタログ	データの出自、意味、場所の一覧を提供する
	データ品質	表記等の標準化、誤り等の修正をして、データの品質を確保する
データ連携	データハブ	複数のシステムを疎結合で連携し、拡張性・保守性を確保する マスターデータやトランザクションデータを共通のインターフェースで提供する
	ETL/ELT	ソースとなるデータを収集・加工し、分析用のデータベースを構築する
データ分析	データウェアハウス	分析用に蓄積したデータベース。主に構造化データを大量に保持する
	データマート	繰り返し実施される集計・分析用にあらかじめ作成したデータベース
	データレイク	多様なデータを主に生データのまま保持するデータベース
	BI	データを可視化して、意思決定を支援する

4. データ活用基盤構築のポイント

データ活用基盤構築時のポイントは多くあるが、本稿では特に昨今、重要度が増している事項に重点を置いて、次に述べる。

4.1 全般

データ活用基盤は、それ自体がビジネスメリットを生むものではないため、導入前に目的を設定することが大切である。何を実現したいのか、どのような課題を解決したいのか、目的を明確にすることが、導入を進める上での羅針盤になる。

目的はできるだけビジネスメリットに直接つながる具体的なレベルとして、かつ成果が迅速に得られるようスモールスタートすることが、データ活用基盤を成功させるポイントである。

全社的なデータ活用基盤を導入しようとする最初からスコープが大きくなりすぎて、成果が得られるまでに長期間かかるケースは多い。優先度が高い目的から始めて、データ活用基盤によって得られる成果を速やかに享受しながら段階的に進めることができれば、社内での合意も得やすく、プロジェクトが進めやすくなる。

データ活用基盤を導入すること自体が目的にならないよう、目的は明文化し、常に関係者で共有して進めていく。

4.2 データ管理

この節では、共通マスターデータと共通トランザクションデータの管理について述べる。

4.2.1 共通マスターデータ

データ分析では、ほとんどの場合、マスターデータを軸としてトランザクションデータの集計を行うことになる。したがって、マスターデータはデータ分析の視点や切り口になる重要なデータである。しかし、企業内で使われるマスター

データは、同じ対象を表すものであっても、表現が異なっていたり、タイミングによって情報の鮮度や粒度が異なっていたりすることが多く、このような不整合はデータ活用の支障になってしまう。そのため、企業内で散在しているマスターデータを統合・標準化し、共通のインターフェースを介して各システムに提供できるように共通マスターデータを設ける。これによって、複数のシステムを横串にしたデータ分析が可能になる。

4.2.2 共通トランザクションデータ

トランザクションデータは、企業の事業活動の記録である。社内に存在する様々な活動履歴を統合し、共通フォーマットに変換して分析することで、組織全体の動きを把握できる。この仕組みを実現するためには、統一された共通トランザクションデータを設ける必要がある。

しかし、マスターデータもトランザクションデータも、その統合・共通化は難易度が高いものもあるので、メリットの出るものから段階的に共通化していくのがよい。

4.3 データ連携

この節では、データ連携の仕組みについて述べる。

4.3.1 データハブ

複数のシステムの連携を、お互い1対1の密結合で実装した場合、そのインターフェース数は連携するシステム数の増加に伴って複雑さが増して、いわゆるスパゲティ状態^(注1)になりやすい。その問題を回避し、各システムが各々のタイミングで連携することを可能にするために、データハブは、ハブ&スポーク構造で各システムを疎結合で連携する仕組みを実現する(図2)。

(注1) 構造が複雑に絡みあって、全体の流れや依存関係が把握しにくくなった状態のこと。

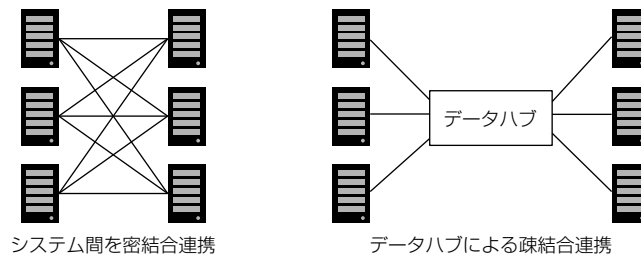


図2-密結合連携と疎結合連携

4.3.2 ELTによる構築手法

クラウド技術の進化によって、柔軟に拡張できるプロセッサ性能とストレージ容量を利用できるようになり、データウェアハウス構築でELTと呼ばれる新たなアプローチが可能になった。従来の手法(ETL)では、ソースデータを収集・抽出する段階で、分析要件に応じた設計を事前に行い、その設計に沿って構築を進めるのが基本であった。一方、新しいアプローチでは、まずソースデータを加工せずに収集・取り込みして、その後、強力なデータベース処理能力を活用して、分析要件に合わせたデータウェアハウスを構築する。この手法の利点は、詳細な分析要件の反映をデータ抽出・加工フェーズ以降に行えるため、要件変更への柔軟な対応が可能になる点である。ただし、従来のETLによる手法と比べると、データの二重保管が発生しやすく、プロセッサやストレージの利用コストが増加する可能性がある。そのため、ELTとETLの両手法の長所・短所を踏まえて、構築するシステムの要件に最適な方法を選択することが重要である。

4.4 データ分析

この節では、分析の際に用いるデータを保管するデータレイクや、データ分析に用いるツールについて述べる。

4.4.1 データレイク

データレイクを構築する目的は、主に二つである。一つは、活用用途未定のデータ(構造化/非構造化問わず)を将来の活用に備えて、生データのまま保管しておくことであり、もう一つは、非構造化データ(画像、音声、テキスト等)を集積してそれらを活用することである。データレイクは、多様な生データをため込む器であるため、その管理が更に重要になる。管理が不適切な場合、把握できないデータが蓄積されるだけで、将来的にも活用されないものになる。活用できないデータの集まりは、無用なストレージコストの増大を招きかねない。4.3.2項で述べたELTによる構築でも、一旦、収集するデータはデータレイクの位置付けになるため、ここでのデータの管理が重要になる。

4.4.2 BI

実際にエンドユーザーが分析等に利用する画面やレポートの作成を支援・提供する役割がBIであり、市販のサービスやツール(以下“BIツール”という。)を活用するケースがほとんどである。多くのBIツールが市場に出回っており、その選択肢は広い。概して、BIツールは二つの相反する役割を持つ。一つは、利用者の思いどおりの分析を容易に行える使いやすさの提供であり、もう一つは開示するデータや作成した画面・レポートに対して統制を利かせるための管理機能の提供である。BIの主たる目的は前者であるが、システム規模が大きくなると後者の重要度も増す。どちらの機能に注力しているツールかを見極めて、求める要件に合わせてBIツールを選択する必要がある。また、昨今のBIツールは多機能で扱いやすい反面、目的が曖昧なまま開発を進めてしまうおそれがあるため、留意が必要である。既存の分析画面やレポートの移行では、無条件に再現を試みることは避けて、個々に目的の再確認や今後の要否を棚卸しする必要がある。新規開発分も含めて、各画面・レポートの各々の目的を精査して、“何を判断するためのものか”を明確にしながらか開発することが重要である。

4.5 AI技術の活用

この節では、AI技術を用いたデータ活用の動向を述べる。

4.5.1 データ活用の民主化

目的の検索集計を生成AIによる自然言語を用いて可能にするBIツール等が広がりつつある。これによって、データ活用のための敷居が低くなり、データ活用の浸透を飛躍的に進める可能性がある。BIツールを利用する上で、クエリ言語や専用のGUI(Graphical User Interface)操作等が利用拡大の支障になっているケースでは、生成AIの活用は有効な手段になり得る。また、利用者が求めるデータの場所やお勧めのデータや集計軸を生成AIが提案してくれる機能も実装されつつある。AIは、データ活用の民主化(データ活用を専門としない人によるデータの活用)を推進する技術としても注目されている。

4.5.2 非構造化データの活用促進

従来のデータウェアハウスは、基本的に構造化データを対象としている。一方、AI技術の進化によって、大量の非構造化データを利用して学習し、分析・判断・予測等に活用できるようになった。データ活用基盤で、データレイクに保管した非構造化データに対して、AIの活用で分類や構造化などが可能になり、構造化データ中心の従来のデータウェアハウスとデータレイクの統合が可能になっている。これはレイクハウスとも呼ばれて、既に具体的な機能を提供するサービスが利用されてきている。企業が持つデータの80%以上を占めると言われる非構造化データの活用は、ますます進むと考えられる。

5. む す び

企業活動で必須になるデータ活用を支える基盤は、技術の発展とともにその役割も拡大している。今後、データ活用基盤は、関連部門に自動的に有益な情報を発信したり、業務を自律的に実施するAIエージェントにデータを提供したりする役割も担うと考えられる。その前提になる高品質なデータの提供と、データの管理はますます重要度を増していく。

三菱電機デジタルイノベーション(株)は、25年以上にわたってデータ活用ソリューションを提供し、データを活用するための基盤構築や運用支援で1,000社を超える実績を積み重ねてきた。培った経験とノウハウを生かして、今後も顧客の価値創出や社会課題の解決に貢献していく。

参 考 文 献

- (1) 独立行政法人 情報処理推進機構：DX動向2025
<https://www.ipa.go.jp/digital/chousa/dx-trend/tbl5kb0000001mn2-att/dx-trend-2025.pdf>
- (2) 森田 登, ほか：デジタルトランスフォーメーション(DX)を推進するデータ活用基盤, 三菱電機技報, 96, No.5, 206~209 (2022)

