

プラットフォーム構築から始める デジタルデータ活用



杉田直哉



下田崇嗣

CONTENTS

- I デジタルデータを取り巻く技術動向
- II デジタル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
- III デジタルデータ活用プラットフォームの始め方
- IV 自社に最適なデジタルデータ活用プラットフォームを見極める

要約

- 1 これまでの企業における情報システムを支えてきた技術とは別に、IoT（Internet of Things：モノのインターネット）やAI（人工知能）を中心とした新たな技術が登場しており、技術の成熟を見極めながら、あらゆる情報をデジタルデータとしてビジネスを高度化・創造していくことが経営課題となっている。
- 2 デジタルデータ活用プラットフォームを構築するにあたり、データ分析プロセス、扱うデータの構造、処理方式に着目してリファレンスアーキテクチャを体系化した。どの技術を活用してデータを活用すればよいのか、デジタルデータを活用するためのプラットフォームはどのように作るべきなのかについて野村総合研究所（NRI）の視点を提示する。
- 3 デジタルデータを活用するためにどういうプラットフォームを作るべきか、最初からあるべき姿を追求するために正解のない問いに悩んでいる企業は多い。特定の利用シーンをあえて想定せず、まずはセキュリティ機能と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として定義して構築することが効果的である。
- 4 データ活用プラットフォームを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抵抗を受けることになった。特にオープンソースやパブリッククラウドを用いた場合は、それぞれが重視する点を踏まえたセキュリティ対策を実現して初めて、効果的なプラットフォームの構築が可能となる。
- 5 デジタルデータ活用基盤の上で、複数のデータ活用テーマのトライ&エラーを繰り返しながら、自社にとって最適なデジタルデータ活用プラットフォームの要件を見極めていくというアプローチを提言する。

I デジタルデータを取り巻く 技術動向

1 企業における

これまでの情報系システム

これまでの情報系システムでは、社内の基幹システムに散在しているデータを収集して経営に活用していた。基幹システムから集められたデータは重複や誤記、表記の揺れなどを修正して活用できる形式に整形し、データウェアハウス（DWH：意思決定のために必要なデータを収集し、目的別に再構成して時系列に蓄積した統合データベース）と呼ばれるデータベースに格納される。その統合されたデータをビジネスインテリジェンス（BI：企業などの組織のデータを、収集・蓄積・分析・報告することで、企業の意思決定に役立つ手法や技術）のツールにより可視化して、意思決定のためのレポートを作成していた。

2 デジタルデータを取り巻く 新たな技術の登場

昨今ではデジタルデータ（ビッグデータ）の取り扱いを可能にするさまざまな技術が登場したことにより、企業が活用できるようになってきている。現在、モノのインターネット（IoT：さまざまなモノに通信機能を持たせ、インターネットへの接続や相互に通信すること）が注目を集めており、企業を取り巻くあらゆる情報を収集することが可能になっている。収集したビッグデータは、GPGPU（画像データ処理を行う集積回路を画像処理以外の目的に応用する技術。CPUと比較して並列作業を効率よく処理する何千ものコア

が搭載されている）などの高性能な計算リソースによって、AI分野で期待が高まっている機械学習や深層学習を現実に処理できるようになっている。それにより、精度の高い予測や人間に近い判断をビジネスに適用していくことが期待される。

3 デジタルデータを 転送・収集する技術

IoTで大量の接続デバイスからデータを収集していくことを想定すると、無線ネットワークの高速化・大容量化の進展は必須である。2020年の商用化を目標として、現在普及しているLTE-Advancedなどの4G規格のさらに先を目指した5G規格の仕様策定の検討が進んでいる。さらには、IoTによってデバイスを自動的に制御するなどのネットワークを利用するサービスがより多様化・高度化していくことを見据えた、モバイルエッジコンピューティング（MEC：エンドユーザやデバイスに近い無線ネットワークあるいはその近くにサーバーを設けてアプリケーションの処理を行わせる手法）と呼ばれる、デバイスとデータセンターの間に入る多段的なアーキテクチャの開発が進められている。

4 デジタルデータを 処理・保管する技術

IoTデバイスから集められたデータは、これまでの情報系システムよりも非常に大規模な容量となる。リレーショナルデータベース（RDB：関係モデルと呼ばれる概念に基づいてデータを扱うデータベース。現在最も普及している）に代表される既存の規格では、そのようなビッグデータを格納して保持するこ

とが困難になるため、NoSQL（リレーショナルデータベース以外のデータベースを指す）と呼ばれる複数のサーバーに分散してデータを保持・保管する技術が登場してきた。また、ビッグデータを処理する技術として代表的なApache Hadoop（大規模データの蓄積・分析を分散処理技術によって実現するソフトウェアフレームワーク）を中心としたオープンソース（OSS：ソースコードを広く一般に公開し、ソフトウェア自体の再頒布も自由に行えるような考え方に基づき公開されたソフトウェア）のコミュニティが中心となり、分散ファイルシステムのHDFSをはじめとするさまざまな技術が登場し、現在も発展を続けている。

5 デジタルデータを分析する技術

AI分野では、アマゾン・ドットコム、グーグル、IBM、マイクロソフトなどの大手のクラウドベンダーが自社のクラウドプラットフォームへの積極的な投資を進めている。各社は比較的汎用性のある基礎的な画像の解析、音声の解析、自然言語の処理、Siriのような会話インターフェースなどのAIエンジンを内蔵したサービスとして提供する。その一方で、ユーザーが自由にAIエンジンを構築することを支援する深層学習フレームワークもオープンソースで公開している。グーグルはTensorFlow、アマゾン・ドットコムはDSSTNE、マイクロソフトはCNTKと、各々が特徴のあるフレームワークを提供し、デファクトスタンダードを獲得するための競争が激しくなっている。なお、国内ではベンチャーのPFNからChainerというフレームワークが公開されている。

II デジタルデータ活用 プラットフォームの リファレンスアーキテクチャ

1 デジタルデータを活用するためのプラットフォーム

昨今のデジタルデータ（ビッグデータ）の世界では、データから情報、そして情報から知識を得ることが重要となっており、データを活用して新たな価値を創出することが競争力の源泉となるといわれている。一方、デジタルデータを取り巻く世界には第I章で述べたさまざまな技術が登場しており、ここ数年でデータ活用の世界に大きなブレークスルーをもたらしている。では、どの技術を活用してデータを活用すればよいのか。デジタルデータを活用するためのプラットフォームはどのように作るべきなのか。

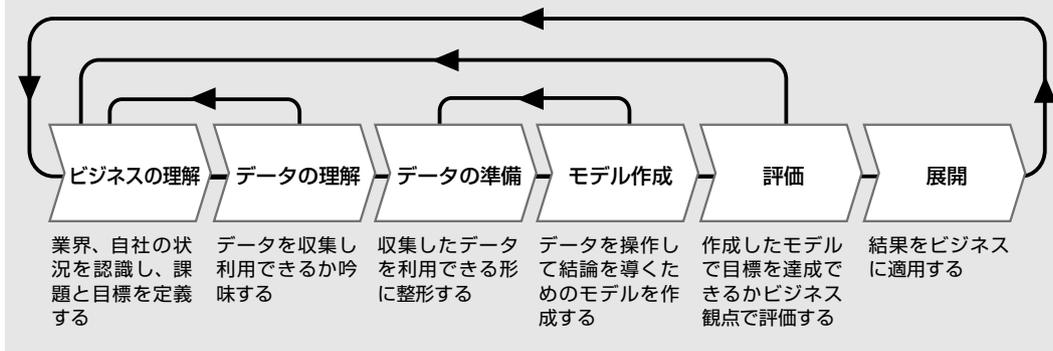
2 デジタルデータ活用 プラットフォームを検討する観点

デジタルデータ活用プラットフォームは、①データ分析に必要なプロセスの理解、②扱うデータの構造、③処理方式、の3つの観点で検討することが重要である。

(1) データ分析に必要なプロセスの理解

データ分析プロセスの参考となるリファレンスとして、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IBMに買収されたSPSS、NCR、タイムラークライスラー、OHRAがメンバーとなっているコンソーシアムが開発したデータマイニングのライフサイクルを整理したものである。このラ

図1 CRISP-DM データマイニングの標準プロセス



ライフサイクルには、ビジネスの理解、データの理解、データの準備、モデル作成、評価、展開、の6つのプロセスが存在する（図1）。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て整理できる。また、分解した階層ごとに必要な機能は、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。

(2) 扱うデータの構造

データには構造化データと非構造化データが存在する。構造化データとは、企業の顧客情報や販売データ、在庫データのようにデータ構造を定義してリレーショナルモデルをベースとしたデータベースに格納できるデータである。非構造化データとは、文書、画像、動画、音声のようなデータ構造の定義が困難なデータである。ビッグデータの世界で扱われるデータは、8割以上が非構造化データといわれており、必要なプラットフォームも、リレーショナルデータベースを利用している基幹システムとは異なるアーキテクチャを考える必要がある。

(3) 処理方式

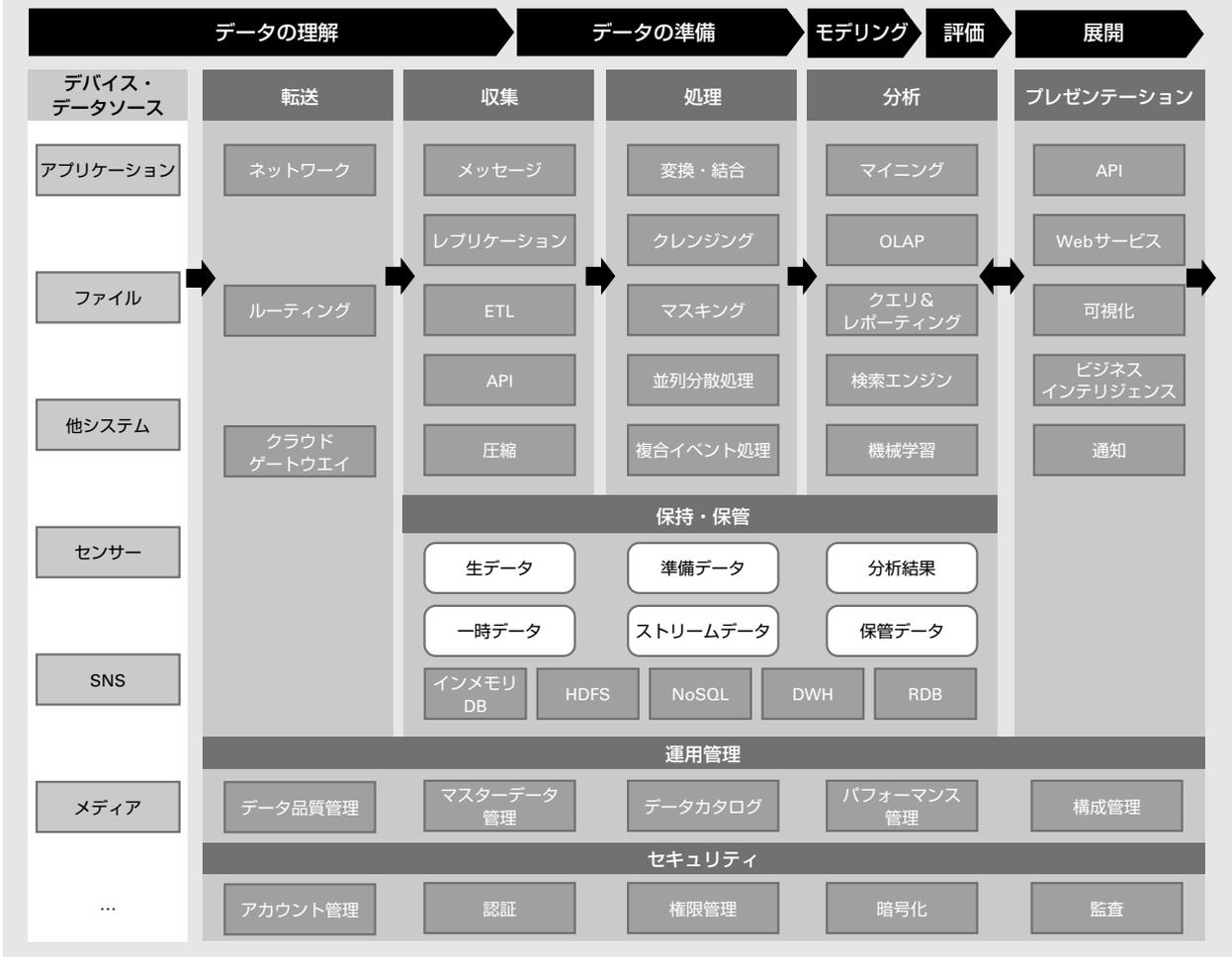
データの処理方式は2タイプに大別できる。1つは一定期間データを蓄積して処理するストック型の処理、もう1つは時系列に発生するデータを連続的に処理するフロー型の処理である。SNSやIoTで使用される各種センサーが生み出す大量データをリアルタイム分析するような処理は、フロー型処理に該当する。

3 デジタルデータ活用

プラットフォームのリファレンスアーキテクチャ

データ分析プロセス、データ構造、データ処理方式を考慮して、NRIではデジタル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を体系化（図2）している。リファレンスアーキテクチャは8つの階層で構成される。「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2つの階層である「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

図2 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



群で構成している。

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すべてをそろえる必要があるわけではない。各階層にある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が変わってくる。例えば、経営コクピットのような業務システムの構造化データを扱う分析であれば、ETLツール（Extract/Transform/Loadの略称、情報源からデータを抽出し、変換加工してターゲットにロードするツール）を使ってデータ収集、加工処理を行

い、DWH（目的別に再編成して時系列に整理蓄積した統合データベース）やRDB（データを複数の属性と値の組として格納したりレシヨナルデータベース）などに蓄積してBIツール（ビジネスの意思決定を支援する分析・可視化ツール）を使って分析、可視化することができる。また、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場合は、Sparkのようなインメモリ高速分散処理フレームワークや、複合イベント処理エンジン（CEP：Complex Event Processing、センサーデータなど刻々と発生す

る大量データをルールに基づき素早く分析、判断、処理するための技術)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て、必要となる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。

階層を構成する機能群は、ベンダー製品・ツールでそろえることも可能だが、昨今では

HadoopやR(統計解析言語・環境)に代表されるオープンソースや、アマゾン・ドットコムやマイクロソフトのようなクラウドベンダーが提供するサービスを使ってプラットフォームを構築しているケースが増えている(表1)。特に最近、ビッグデータ分析や、IoT、機械学習関連のクラウドサービスが次々と登場しており、利用に乗り出す企業が

表1 リファレンスアーキテクチャの階層と階層機能を実現する代表的なオープンソース、クラウドサービス

階層	概要	オープンソース (例)	クラウドサービス(例)	
			Amazon WebService	Microsoft Azure
1 転送	<ul style="list-style-type: none"> エッジデバイスの管理・監視、デバイス・データソース⇄クラウド間のシームレスな接続機能を提供する階層 	<ul style="list-style-type: none"> Nginx Squid SoftEther 	<ul style="list-style-type: none"> VPC ELB DirectConnect 	<ul style="list-style-type: none"> DNS LoadBalancer ExpressRoute
2 収集	<ul style="list-style-type: none"> 各種センサー、モバイルなどのデバイス、SNSやメディア、Webや各種アプリケーション、データベース、外部システムなどのデータソースが生成・蓄積しているデータ・情報を集める機能・インターフェースを提供する階層 	<ul style="list-style-type: none"> Fluentd Flume Embluk Kafka Talend 	<ul style="list-style-type: none"> SQS StorageGateway MigrationService Kinesis 	<ul style="list-style-type: none"> ServiceBus StoreSimple API App EventHub
3 処理	<ul style="list-style-type: none"> 収集したデータを分析できるように変換・保存する機能を提供する階層 データ変換、クレンジング、マスキングする技術や、大規模データ処理を実現する並列分散技術、複合イベント処理技術などが含まれる 	<ul style="list-style-type: none"> Hadoop Spark Presto S4 Stome 	<ul style="list-style-type: none"> EMR DataPipeline Lambda KinesisAnalytics 	<ul style="list-style-type: none"> HD Insight DataFactory StreamAnalytics
4 保持・保管	<ul style="list-style-type: none"> データを保持・保管する階層 大規模データ処理を実現する分散データストアや、高速処理を実現するインメモリDB、データ構造の柔軟性、スケーラビリティが高いNoSQLなどが含まれる 	<ul style="list-style-type: none"> PostgreSQL VoltDB HDFS Cassandra MogoDB 	<ul style="list-style-type: none"> RedShift S3 RDS DynamoDB ElasticCache 	<ul style="list-style-type: none"> Storage SQLDatabase SQLDWH DataLakeStorage RedisCache
5 分析	<ul style="list-style-type: none"> 蓄積・保持されたデータを分析する階層 マイニングや機械学習などの分析機能が含まれる 	<ul style="list-style-type: none"> R Python Mahout 	<ul style="list-style-type: none"> MachineLearning Elasticsearch MobileAnalytics 	<ul style="list-style-type: none"> MachineLearning Search MobileEngagement
6 プレゼンテーション	<ul style="list-style-type: none"> 分析結果を取得するインターフェースの提供、メッセージ通知、分析結果を活用した意思決定支援を行う機能を提供する階層 	<ul style="list-style-type: none"> Pentaho JasperReports SpagoBI 	<ul style="list-style-type: none"> QuickSight API Gateway ElasticBeanstalk 	<ul style="list-style-type: none"> PowerBI APIManagement WebApps
7 運用管理	<ul style="list-style-type: none"> データ品質、メタデータの管理、ならびに分析基盤の信頼性や可用性、インフラ構成、パフォーマンス管理などを行う機能を提供する階層 	<ul style="list-style-type: none"> Nagios Hinemos Zabbix 	<ul style="list-style-type: none"> CloudWatch CloudTrail Config 	<ul style="list-style-type: none"> ResourceManager ApplicaitonInsight AzurePortal
8 セキュリティ	<ul style="list-style-type: none"> データ保護の視点で、ID管理、アクセス制御、認証、監査などセキュリティ機能を提供する階層 	<ul style="list-style-type: none"> OpenLDAP LISM 	<ul style="list-style-type: none"> DirectoryService IAM Inspector KMS 	<ul style="list-style-type: none"> AzureAD Security Center KeyVault

増えている。手軽に導入して効果があるかどうか試せるためだ。回転寿司大手チェーンの株式会社あきんどスシローでは、店舗混雑時の待ち時間予測にAWSの機械学習を活用している。竹中工務店ではAzureの機械学習で建物設備のモニタリング、管理・分析などを自動的に行う次世代建物管理システムの構築・提供を行っている。

4 リファレンスアーキテクチャとデータ活用の進め方

データ活用は、最初から結果、効果が見えるわけではなく、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、データを活用できるプラットフォームを構築する場合、最初から製品・ツールありきで構築してしまうと、無駄が生じる場合が多い。事実、データ活用がうまくいかないユーザー企業からは、「大量のデータを格納できるDWHやサーバー機器を購入したが使われない箱になっている」、「高価なETLツール、BIツール、データ管理ツールなどを購入したが一部の機能しか利用しておらず無駄が多い」といった声を頻繁に聞く。扱うデータ、活用ニーズは、ビジネスや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。デジタルデータからインサイトを得るためには、完璧なプラットフォームを求めて製品・ツール導入に目を向けるのではなく、今すぐできるところからコストを掛けずにスタートして小さな成功を蓄積していくことが重要である。リファレンスアーキテクチャで全体像を見据えながら、まずはできるところから小さく始めてもらいたい。

Ⅲ デジタルデータ活用 プラットフォームの始め方

1 実証実験によるデータ活用のアプローチ

まだデジタルデータ利活用の取り組みを始めていない企業では、いきなり全社規模でのデジタル化戦略を策定しても、現場の実行力がついていかず前に進まない可能性が高い。デジタルデータ利活用を定着させる上では、現場レベルでの実証実験を通じてデジタルデータから価値を創造し、ビジネスアナリティクスの効果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。弊社が支援した大手製造業の事例では、実証実験として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。

基幹システムの世界では、グランドデザインを構想段階でしっかり検討しておかないと、後の工程に無駄が多く発生したり、柔軟に拡張できなくなったりすることが多々ある。そのような経験からか、どうしても最初にあるべき姿を描くことを求めてしまいがちである。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。データ活用の場合、ニーズは都度発生し、あらかじめ業務要件が定まっていることは多くない。どういう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。実証実験を進めていくにあたり、あるべき業務の姿から考えていくこれまでの基幹システムの発想では、デジタルデータ活用プラットフォームの検討はうまく進まない。

プラットフォーム構築の検討を開始した時

点では、

- ①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
- ②2～3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によらない部分にフォーカスし、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ジタル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として定義して構築することから始めた。

なお、プラットフォーム構築後に対象の社内利用部署、外部分析ベンダーともに変更になっている。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デファクトスタンダードを選定した。分析テーマについても、どのテーマから着手するかは最終決定には至らず、分析の前処理となるデータ加工の仕様なども決定していなか

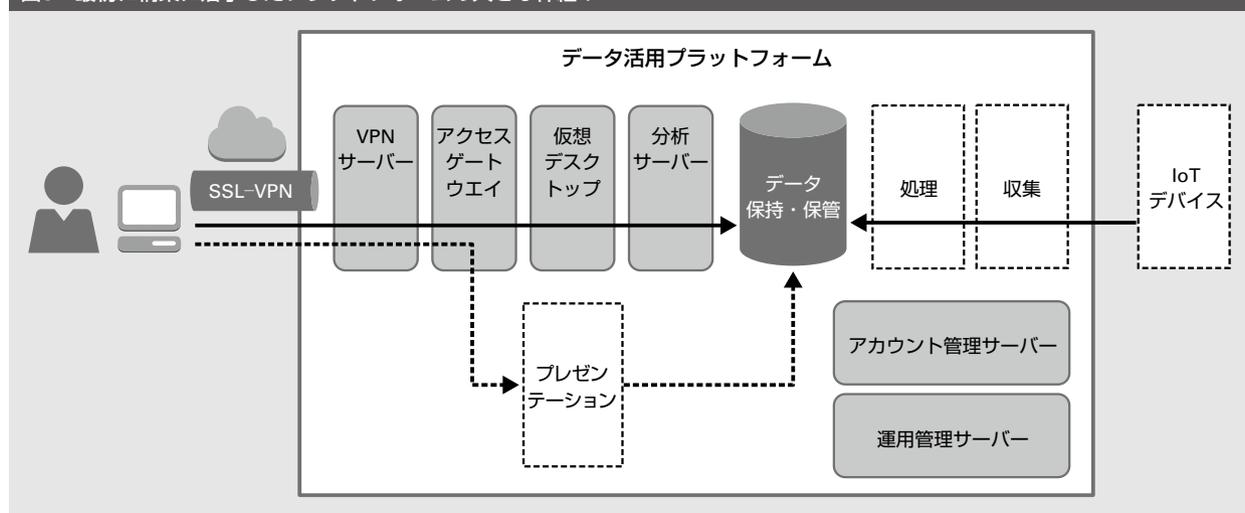
った。さらに、プラットフォームの構築後、分析の対象はまったく別のテーマに変更になった。

2 プラットフォームとして 最初に構築した箇所

IoTデバイスから新しくデータを収集することから始めると、分析可能なデータ量を蓄積するのに時間がかかってしまう。まず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を用意することに着手した(図3)。もし、構築の時点でIoTデバイスからデータをリアルタイムに収集することが決まっていたのであれば、データを受信して特定の用途に対応して加工するための「収集」と「処理」部分も開発対象に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ことはなかった。

必要最低限なプラットフォームをスモールスタートで構築するためには、初期投資を抑

図3 最初に構築に着手したプラットフォームの大きな枠組み



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がキーテクノロジーとなる。特定ベンダーのツールを入れると、そのツールに習熟したデータサイエンティストが必要となり、ベンダーロックインになる。多くの便利な機能を備えていることが多いが、最初から導入するのは早計である。プラットフォームの構築後いくつかの分析テーマを進めていくに従って、結果的に分析テーマや外部ベンダーが追加される度に追加要件が発生し、プラットフォームの環境に変更が入ることになったが、大きな枠組みが事前に用意されていたためデータ分析の進捗が阻害されることはまったくなかった。

なお、既に企業のITシステムでオープンソースやパブリッククラウドの利用が始まってから長い期間が経過しているため、それらの技術の採用は当たり前と感ずるかもしれない。ただし、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激しい抵抗を受けることになった。

3 データオーナーが 安心してデータ提供できる プラットフォーム

分析プラットフォームを構築後、データを提供してもらうためにデータオーナーにどのような環境かを説明することになる。ITに詳しくない人にとって、「クラウド」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き、敏感に反応してしまうキーワードである。「オープンソース」も無料で使用できてしまうことに不信感を持ち、スパイウェア

などのウイルスが仕込まれている危険を想定してしまう。自社内のIT環境におけるセキュリティ対策と対応させて、今回構築したプラットフォーム特有のセキュリティリスクとは何かを明確にすることが必要になる。リスクへの対策の結果、自社内のIT環境と同等以上にセキュリティが確保されることを説明しなければならない。

また、企業にとって競争優位に関わる分野であるほど、本来であればデータを有効に活用してより優位性を確保していきたいはずである。当然、競争優位に関わるデータは求められる機密性が高く、データの分析を行うにも厳しいセキュリティの確保が求められる。この事例においては、基本的なセキュリティ対策として、以下を実装した。

- ①通信データの暗号化：リモートデスクトップ接続の通信暗号化にSSL-VPNの方式を選択した
- ②保管データの暗号化：保管データは暗号化しないと格納できない設定とした
- ③保管データのマスキング：今回はIoTデバイスからのリアルタイムでの収集はしないため、対象外となった。必要に応じて機密情報をマスクしてからプラットフォームにアップロードする運用となった
- ④アクセス制御・アクセス権限の管理：分析テーマごとにデータの保管場所を論理的に分け、柔軟にアクセス権限を振り分けられるようにした
- ⑤操作証跡の取得：プラットフォームの入り口にユーザー操作をすべて記録する機能を備えたアクセスゲートウェイを配置した
- ⑥システム監視：管理者権限での不正ログ

インなどのセキュリティ・インシデントも監視するようにした

加えて、外部の分析ベンダーへの対策以上に、社員のユーザーの操作制限を厳しく制限するようにした。当然のことながら、分析ベンダーがデータを不正に持ち出せないように対策を施してはいたが、社員がプラットフォーム上にデータをアップロードできなければ想定以上に被害が拡大する可能性は抑えられない。日本年金機構における個人情報漏洩の事件などに代表されるように、企業における最大のセキュリティホールは無知なユーザーである。

4 分析ベンダーが利用しやすいプラットフォーム

データ分析ができる環境を用意しただけでは、必ずしも外部の分析ベンダーにそのプラットフォームを使ってもらえるとは限らない。分析ベンダーは自社の環境での分析を希望することが多く、説得が必要になる。データサイエンティストにとって、使い慣れたツールを使用する方が圧倒的に生産性は高くな

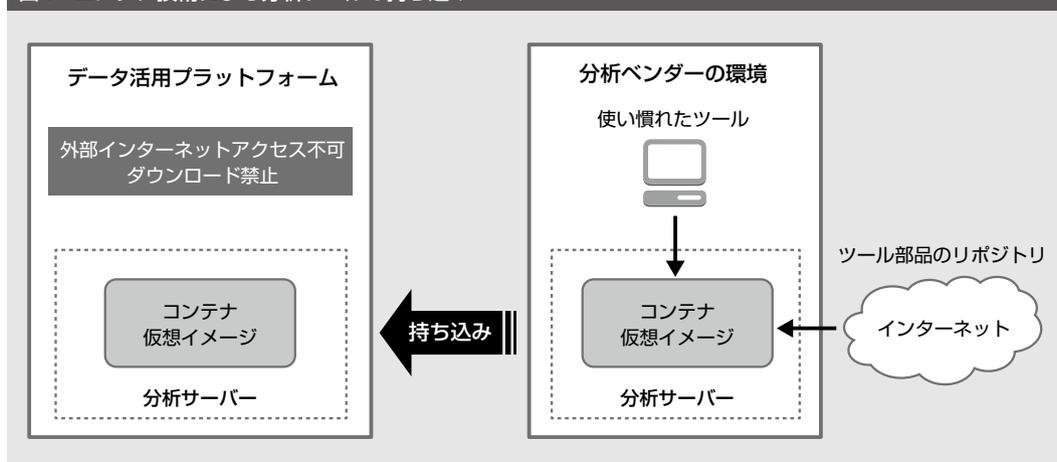
り、自分で開発したオリジナルのライブラリやフレームワークを活用している人も多い。そのため、セキュリティポリシーが守られる範囲で分析ベンダーが自分でツールを持ち込むことを可能にし、自由に分析サーバーをカスタマイズできるようにした(図4)。

ツールを持ち込むにあたっては、インターネット経由でのインストールを許可してしまうと、セキュリティ上の抜け穴となり好ましくないため、コンテナ技術を利用して分析ツールを持ち込むよう依頼した。分析に取り組む前に要件を決められないのと同様に、どのツール部品が必要になるかは分析を進めてみないと分からないことが多く、インターネット経由で必要な部品を都度インストールする形式が一般的な方法になっている。

分析ベンダーの自社環境において必要な部品をインターネット経由でインストールした後、イメージとして作成し持ち込んでもらうようにした。

また、複数の外部の分析ベンダーが利用する場合、分析ベンダー各社のプライバシーの確保が欠かせなくなる。特にデータ分析の過

図4 コンテナ技術による分析ツールの持ち込み



程での自社特有のノウハウやアルゴリズムなど知財に関わる部分は競合他社に見せたくないため、実際に分析アルゴリズムを配置して分析処理を行う箇所については、各分析ベンダーのプライバシーが確保されるマルチテナントに対応した環境でなければいけない。

IV 自社に最適な デジタルデータ活用 プラットフォームを見極める

1 プラットフォーム構築後の対応

要件を決めずに始めるとはいえ、最終的には自社に適切なプラットフォーム要件を見極めていくことが、デジタルデータの活用を定着させるためには大事である。BIツールのように使い方がある程度決まっているツールであれば、導入してユーザーに使い方をレクチャーすることでデータの可視化方法が提供される。一方で、機械学習でデータの予測モデルを構築するなど、より高度なデータ分析に取り組む場合は、データの性質に応じた手法の選定が重要になり、特定のやり方での統一は難しいことが多い。また、分析の処理にかかる計算リソースは、分析対象のデータ量に依存することが多い。ビッグデータというキーワードで呼ばれるように、データのサイズが大きくなっていくと、それを処理する基盤を変えていかなければいけない。分析テーマごとに、サイズに応じてデータベースを拡張したり加工する分散処理の基盤を用意したりすることが必要になる。経費を削減するためには、分析が終わったらデータベース自体を壊してしまうなど割り切った対応も必要である。

2 プラットフォーム構築による 社内での広がり

紹介した事例では、データを安全に自由に分析できる環境が作られると、それまで社内に眠っていた分析テーマ、分析人材、分析対象データが自然と集まってきた。技術的な面が要因となり、Excelで処理する限界を超えたデータ量のため分析できなかった部署の方は、ビッグデータを分析するための機能をプラットフォームに期待していた。セキュリティ面が要因となり、これまで社内ルールに則ってデータを外部に出せなかった部署の方は、外部の分析ベンダーに分析を依頼するための場としてプラットフォームに期待していた。

データ活用プラットフォームにアップロードできるようにするために、部署内の機密的な情報は外部の分析ベンダーに開示できるレベルまで最低限マスクすることが求められる。結果的に、複数の分析テーマを通してデータ活用プラットフォームに複数の部署のデータが蓄積していくようになり、複数の工程のデータを掛け合わせた新しい分析も可能になってくる。また、この事例ではプラットフォームを自分で手を動かして分析してみたい、新たな分析手法を試してみたいという人が社内からも続々と出てきた。

3 データを活用するための プラットフォーム

このように集まってきたリソースを活用して分析テーマを進めていった結果、分析モデルが構築されデータの活用効果が立証されると、業務アプリケーションとしてシステムを展開するフェーズに移ることになる。アプリ

ケーションも業務要件が決まっているわけではなく、使いながらどんな仕様が最も効果的か見極めていくことになる。こちらも同じようにまず枠組みを用意し、やりながら見極めていくことが望ましい。最近ではアジャイル開発技術の派生の位置付けで取り組む企業も増えてきているが、データ活用のアプリケーションもDevOps（開発担当者と運用担当者が連携して協力する開発手法）と同じように開発の工程と運用の工程が密接した関係になっている。使いながら「もっとこのような軸でデータを見たい」、「この分析モデルの精度は十分ではない」といった運用の中で課題が発生し、データ分析とアプリケーション開発を繰り返していくことになる。ウォーターフォールのように要件を先に決めていくやり方ではうまくいかない。

このような試行錯誤を繰り返す取り組みを

通じて、最終的にリファレンスアーキテクチャのカバーする範囲まで拡張していくことで、自社に最適なデジタルデータ活用プラットフォームを見極めていくアプローチを推奨する。

著者

杉田直哉（すぎたなおや）

ITアーキテクチャーコンサルティング部上級ITアーキテクト

専門はシステム化構想・計画策定、アーキテクチャーデザイン

下田崇嗣（しもだたかし）

ITアーキテクチャーコンサルティング部上級ITアーキテクト

専門はシステム化構想・計画策定、アーキテクチャーデザイン