

FinTech時代、銀行系システムはどうあるべきか(4): 銀行がFinTech時代を勝ち抜くために行う データ活用の事例とアーキテクチャ、テクノロジー

本連載では、銀行系システムについて、その要件や歴史を整理しつつ、スマートフォンを使う銀行取引やブロックチェーンなど、新しい技術が及ぼす影響を考察していきます。今回は、銀行を取り巻く環境の変化を整理し、顧客理解をより深める「データ分析」について詳細に見ていきます。[野波衆太郎, 久保俊彦, 日本アイ・ビー・エム株式会社]

本連載「FinTech時代、銀行系システムはどうあるべきか」では、銀行系システムについて、その要件や歴史を整理しつつ、スマートフォンを使う銀行取引やブロックチェーンなど、新しい技術が及ぼす影響を考察していきます。

前回の「FinTech時代の銀行に求められるSoE/SoRアーキテクチャとAPI管理とは」から、銀行オンラインシステム発展の歴史や伝統的なアプローチを踏まえて、FinTech時代に向けて、銀行システムアーキテクチャに求められる要件について整理しています。今回は、銀行を取り巻く環境の変化を整理し、顧客理解をより深める「データ分析」について詳細に見ていきます。

さて、「良いバンキングサービスは必要とされているが、銀行が必要とされているわけではない」とは、Citiグループリテール部門チーフマーケティングオフィサーのヘザーコックス氏が伝えた衝撃的なメッセージです(参考)。

●今回の主な内容

- **銀行を取り巻く環境の変化**
 - 顧客の変化
 - テクノロジーの変化
- **顧客理解をより深めるデータ分析ユースケース**
 - 顧客の行動にフォーカスする顧客分析
 - マイクロセグメンテーションを基にしたアソシエーション分析
 - 従来の分析手法との違い
- **ビッグデータ活用に求められる銀行情報系アーキテクチャとは**
 - SoR + SoE + Sol で実現するデータ利活用アーキテクチャ
 - クラウドの活用例
- **テクノロジーの適用例と、その価値**
 - データレイク実現の主なコンポーネント概観
 - コグニティブ技術の適用について
- **まとめ**

顧客の期待	銀行に求められる対応
Find me	新たな顧客セグメントを見つけるために分析し、可視化していく
Ask me	商品、サービス、そしてソーシャルの側面から顧客にアプローチする
Advise me	顧客には、顧客を理解したエキスパートによってコンタクトする
Know me better	顧客の Wants と Needs を完全に理解してサービスをオファーすること
Stick with me	顧客の生活に結び付いたデータで洞察を深める
Excite me	予想しないタイミングで期待以上のサービスを提供する
Compare me	類似した属性の人 (People like me) が何をしているかを分析して提示する
Trade with me	商品サービスの優遇と引き換えに、さまざまな顧客データや新しいアイデアを共有してもらう
Educate me	オンライン教材や、「今の売れ筋」など旬な話題を提供する
Alert me	リアルタイムの予測分析により、パーソナライズされたコミュニケーションを行う
Let me choose	ニーズに合う複数の案を提示し、顧客に選択肢を与える
Protect me	多要素でのセキュリティ対策によって顧客を守る

表1 顧客理解、顧客経験深化の観点

書籍『Bank3.0 脱・店舗化するリテール金融戦略』（ブレット・キング 著、上野博 訳、東洋経済新報社 刊）によれば、「価値のバランスは消費者側に移り、顧客が主導権を持ち、選択権を握っている」と語られています。

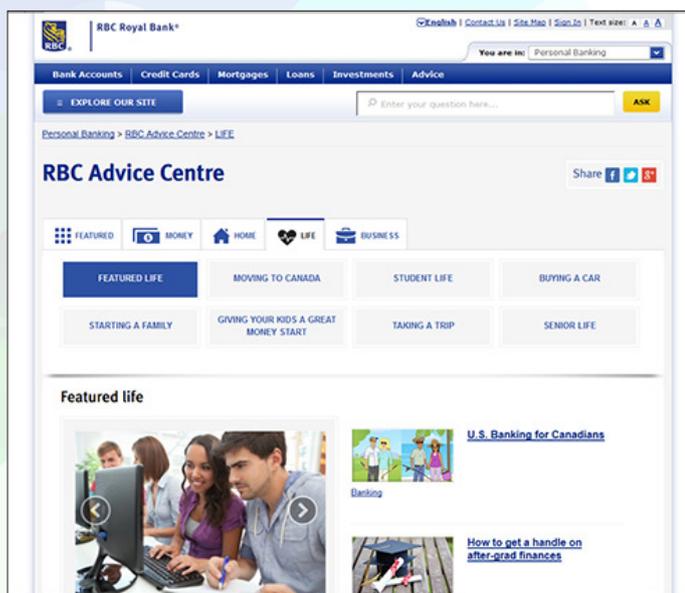
銀行は顧客中心主義へのブレークスルーを現実のものとして迫られているといえるでしょう。

顧客の変化

今後、銀行はどのようにして顧客と向き合い、顧客の理解を深めていけばいいのでしょうか。顧客が銀行に求める期待は多岐にわたっています。表 1 に示すような顧客理解、顧客経験の深化は、グローバルの先進的な銀行では実際に取り組みが進められています。

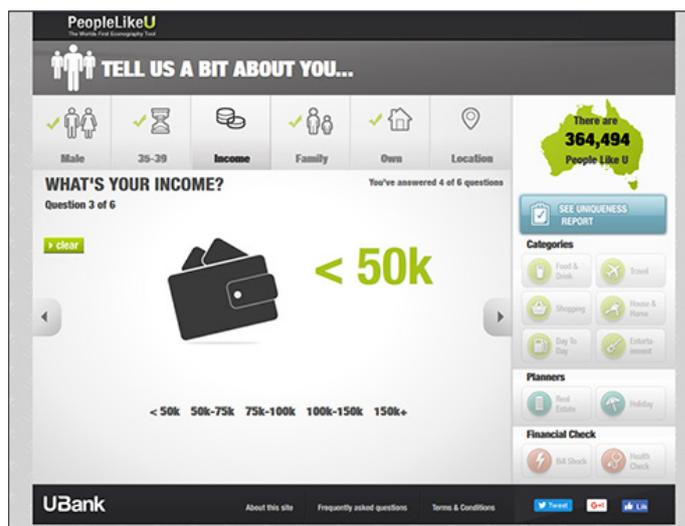
例えば、上記の「Advise me」では、Royal Bank of Canada が RBC Advise Centre で顧客データの分析に基づくエキスパートのアドバイスサービスを提供し、顧客満足度が 5 位から 1 位に向上したそうです。

また「Excite me」では、TD Bank がパーソナライズされた特別な顧客体験を提供する ATM サービスを実現しました。このサービスでは ATM は「Automatic Teller Machine」ではなく、「Automated Thanking Machine」と呼ばれています。



RBC Advise Centre

さらに「Compare me」では、UBank が「People Like U」というサービスで自分と類似属性の人の投資傾向情報を提供し、そこからさまざまなサービスや商品への送客を促しています。



People Like U

これらは全て、銀行の内部や外部のデータを徹底的に活用し分析した結果、実現されたものです。

テクノロジーの変化

こうした新たなデータ活用の取り組みには、テクノロジーの進化が大きく貢献しています。Harvard Business Review で示された「Analytics 3.0」では、データ活用テクノロジートレンドの変遷を大きく 3 つのステージに分けて説明しています。

• Analytics 1.0

数十年前からの基本的なデータ分析の取り組みでよく見られる状況としては、データ分析よりも準備に多くの時間を要したり、小規模の分析モデルを作るのにコストと時間を大量に要したりするなどの課題に数多く直面していました。また、予測やシミュレーションなどの高度な目的ではなく、データの可視化や課題の整理などを主に行っているのが「Analytics 1.0」のステージです。

• Analytics 2.0

2000 年代に入ると、扱うデータ量 (Volume)、データ処理の速度 (Velocity)、データの種類 (Variety) が飛躍的に増加し、いわゆる「ビッグデータ」が現実のものとなってきました。

このステージでは、データ分析により多くのリソースを使えるようになり、さらに多種多様な大量データの関連性に着目して分析を行えるようになります。また、分析モデルはピ

ビジネスオーナーのニーズに基づいて反復的に洗練させながら、レポートやチャート整理ではなく、予測に基づくアクションにつなげることを重視するようになります。

・ Analytics 3.0

データ分析ソリューションが意思決定の現場で使われるようになり、ジャストインタイムのデータ分析に基づいて成果に結び付くビジネスプロセスに組み込まれていきます。継続的なビジネス成果のためにリアルタイムからヒストリカルなデータ分析をミックスした新たなソリューションが実現します。

顧客理解をより深めるデータ分析ユースケース

顧客の行動にフォーカスする顧客分析

「顧客の理解を、さらに深めるデータ分析」とは具体的にどのようなものか見ていきます。

従来の顧客分析では顧客の取引データや属性データ（年齢、性別、家族構成、職業、年収、リスク許容度など）を主に活用してきました。これに加えて顧客のニーズをより深く、早く知るために、顧客の日々の行動や生活スタイルを表すデータを活用します。

顧客の行動データとしては、例えばチャネルシステムの利用状況（営業店、ATM、コールセンター、Webなどをいつ、ど

のくらいの頻度で、どこで使用しているか）や、コンタクト結果のテキストデータなどを活用します。さらに顧客特定が可能な場合や顧客からのオプトイン（利用許諾）があれば、外部 SNS のデータや提携先サービスのデータ、アカウントアグリゲーションサービスから得られる他行データや、Open API で接続される FinTech 企業のデータなども有効です。こうした銀行内外のビッグデータを活用してこれまでにない視点で顧客の理解を深めることができます（図 1 参照）。

マイクロセグメンテーションを基にしたアソシエーション分析

これらのデータを基に、「いつ、どこで、何を、どのように消費し、何に興味を持ち、どこに投資、蓄積しているのか」といった切り口で顧客全体を細分化します。この分析によって、顧客属性データや「取引データ × 行動データ」からなる粒度の細かいセグメンテーション（マイクロセグメンテーション）に顧客を分類できます。顧客は静的な属性変化だけではなく日々の行動の変化によっても、このマイクロセグメント間を移っていくことになります。このマイクロセグメント間の移動をトリガーにして、その後の金融消費行動やライフイベント、離反行動などを予測することが可能になり、その予測に基づいて顧客へ事前のアプローチを行えるのです（概要は図 2 を参照）。

図 1 顧客の行動にフォーカスする顧客分析の全体像

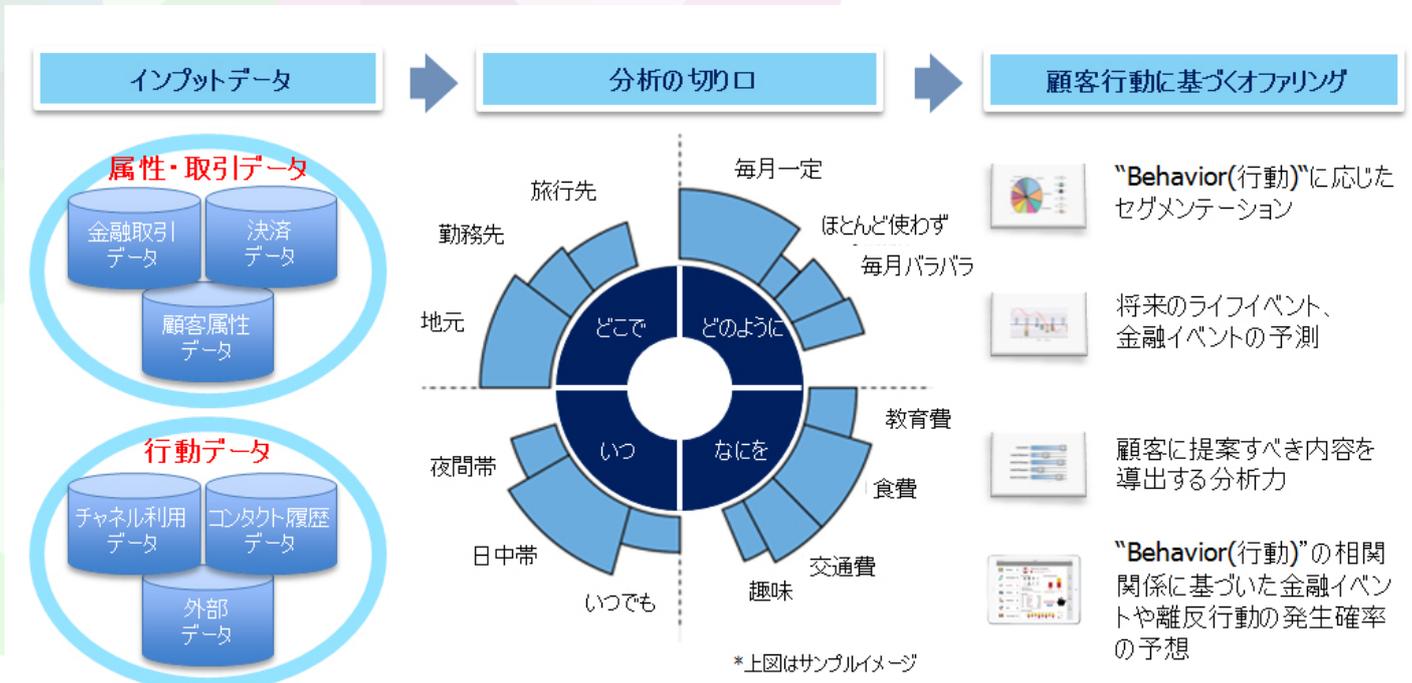


図2 行動ベースの顧客分析の流れ

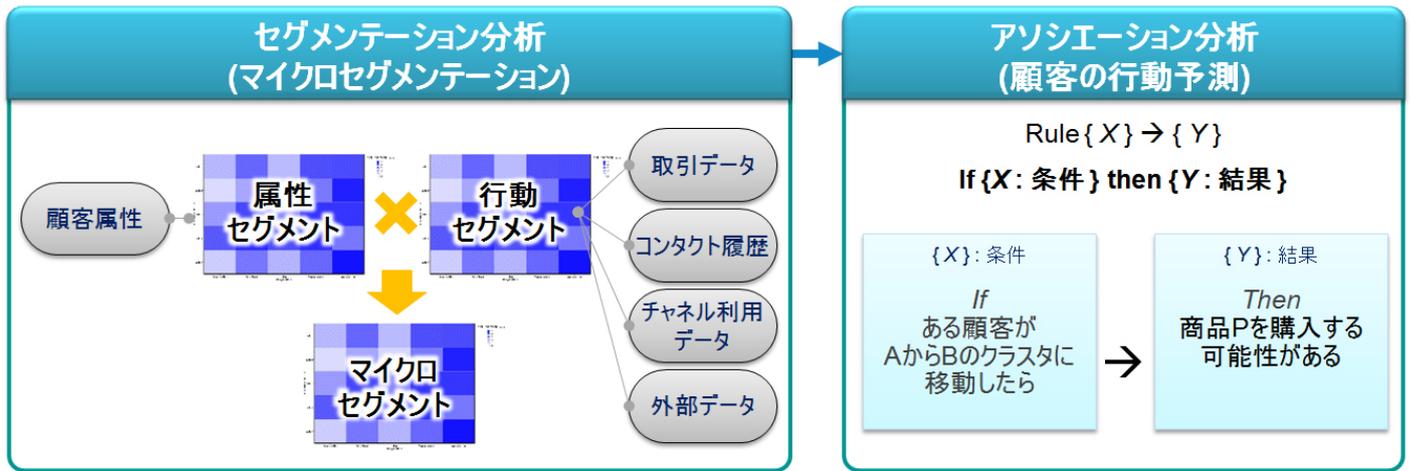
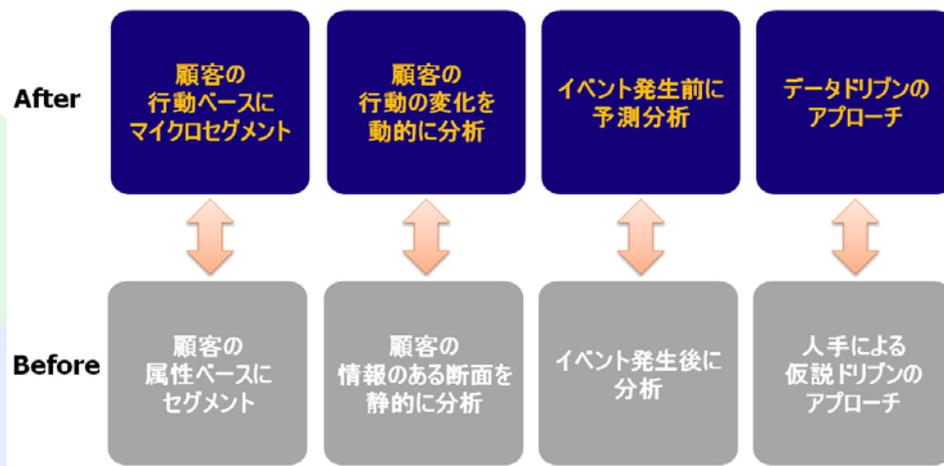


図3 従来の分析手法との違い



従来の分析手法との違い

こうした分析手法の従来との違いを整理します (図3参照)。顧客接点や顧客行動のデータを活用してマイクロセグメントに分類することで、より顧客を正確に深く理解できます。また固定的、断片的な属性変化だけではなく行動の変化も捉えることで、より動的な洞察を行え、その結果として顧客のライフイベントや金融イベントの予測ができるようになります。そして、これらの分析は「顧客属性や取引金額の多寡などを基に人が仮説を立てる」アプローチではなく、「データを主体としデータが示すものから顧客を理解しよう」とするデータドリブンのアプローチです。

ビッグデータ活用に求められる銀行情報系アーキテクチャとは

このような取り組みを進める上で、銀行の情報系システムはどのようなアーキテクチャを考えていくべきでしょうか。これを考える上では、本連載の前回記事にもある通り SoR

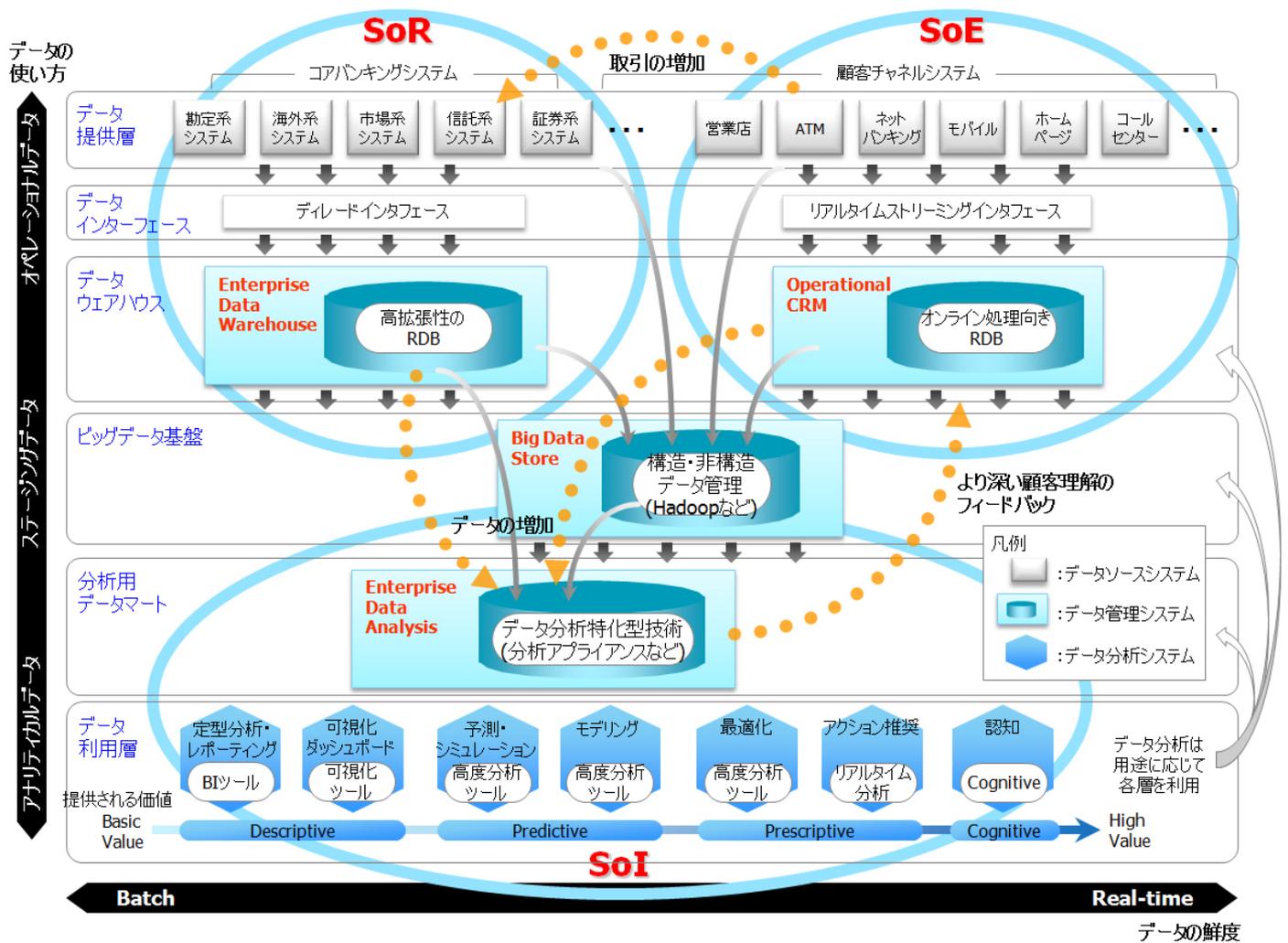
(Systems of Record) と SoE (Systems of Engagement)、そして Sol (Systems of Insight) の3つの領域を考慮する必要があります。

SoR + SoE + Solで実現するデータ利活用アーキテクチャ

顧客の属性情報や取引データは主に勘定系などの基幹システム群 (SoR) から提供されます。一方、前述したような顧客行動の情報を得ようとする、顧客との接点や関わりが強いシステム群 (SoE) から得られるデータが重要となります。これらのデータは銀行チャネル (営業店や ATM、Web、コールセンター、モバイルなど) での顧客行動データが主体となり、SoR にはない情報が多く含まれています。また、構造化データだけではなく、テキストやログ、音声や映像などのマルチメディアデータを含む非構造化データが対象となります。

SoR の構造化データと SoE の非構造化データを併せて Sol の領域に提供します。双方のデータを活用して顧客の分

図4 SoR + SoE + Solの銀行情報系アーキテクチャ概要図



析を行い、その結果を顧客チャネルにフィードバックしていきます。図4はこの効果的なサイクルを回すためのシステムアーキテクチャです。

このアーキテクチャではSoRとSoEからの構造/非構造データを併せて提供するビッグデータ基盤が重要な構成要素となります。データは可能な限り消去せずに蓄積したいというデータ利用者の根源的な望みを満たし、かつSoRの構造データだけではなくSoEの大量の非構造化データをいかに効率良く管理、提供していくかを考えると、図4のビッグデータ基盤は1つの解になります。

図4で挙げているApache Hadoop（以下、Hadoop）は、昨今急速に技術が向上し実際に活用されるようになってきました。データ管理に関して非常にコストパフォーマンスに優れたHadoopをビッグデータ基盤に適用することで、増大し続けるデータを低コストで管理でき、Sol領域に必要なデータを提供できます。

Hadoopのこうした活用例は銀行ではまだ多く見られま

せんが、みずほ銀行での取り組みはその一例です（参考）。Hadoopの技術発展によって今後エンタープライズレベルでの適用例は一層増してくるでしょう。

Sol領域では複雑なデータ抽出や非定型的なデータ処理、高度な計算処理が多用されるため、より分析処理に特化したデータベース技術と分析ツールの適用が望まれます。データ分析処理の形態は、さまざまな用途によって複雑さも変わってきますが、データ分析による提供される価値の違いから下記のように用途の分類を考えることができます。

- **Descriptive** : レポート、可視化などの用途。定型的な分析が主体で、起きている事実の整理を行う
- **Predictive** : 予測やモデリング、シミュレーション、アラートなどの用途。非定型的な分析が主体で、現状のデータから将来予測をベースに「What if」のシミュレーションや注意喚起の判断を行う
- **Prescriptive** : 最適化やアクション推奨の用途。どうすれば効果を最大化できるかをさまざまな変動要素を踏ま

えて分析する

- **Cognitive** : 人の行動や考えを認知して導く。自然言語やマルチメディアデータを含む大量のデータから回答を導き、学習を繰り返しながら回答精度を向上させる

クラウドの活用例

昨今、データ管理や分析の場もクラウド環境が有力な候補になっています。図4のアーキテクチャ論理構造でも下層部分は、もちろんクラウドの適用候補になります(図5参照)。高い柔軟性を持つクラウド環境でデータを提供し、クラウド上の新しい技術、ツール、アルゴリズムを使うことで、新たなビジネスモデルの試行や検証に取り組みやすくなります。

クラウドを活用する上で、クラウドにどうデータを配置するかが1つの大きな検討ポイントです。CSCC (Cloud Standard Customer Council) のPDFレポートの表5ではクラウドへのデータ配置判断の優先検討要素として、柔軟性、必要システムリソース、データボリューム、ネットワーク帯域、データ近接度合い、管理統制が挙げられています。クラウド環境はマルチベンダーで提供されることも考えられるため、インターフェースのオープン性も重要になります。

テクノロジーの適用例と、その価値

前述の通り、データ活用によって新たな洞察を得て実際にビジネスに生かしている事例が増えています。「データサイエ

ンティストに限らず、マーケターや営業といったユーザー部門などの方々が、必要な時に、必要なときに、(社内・社外を問わず)必要なデータを使って、迅速に洞察を得る」ためのシステムアーキテクチャと、それを実現するプラットフォームやテクノロジーが求められています。こうしたプラットフォームに必要とされる要素を本節で述べます。

データレイク実現の主なコンポーネント概観

ニーズに応じて迅速にさまざまなデータを分析/活用するためのプラットフォームとして、「Data Lake (データレイク)」という考え方が、2010年に米国で提唱されました。データレイクを実現する主なコンポーネントのうち、図6に示す5つの領域について解説します。

なお、図4で示した銀行情報系アーキテクチャと図6に示すコンポーネントをマッピングすると、表2のようになります。

・セルフサービスとユーザーインターフェース (図6-①)

ユーザーが必要に応じてデータを入手し、ビジネスの目的に応じて自身でデータを分析、可視化するためのユーザーインターフェースです。セルフサービス機能を強化したBI(ビジネスインテリジェンス)ツールや分析ツールがこのコンポーネントで利用されます。

また1人のユーザーのセルフサービスを促進するだけでは

図5 データ分析でのクラウド利用

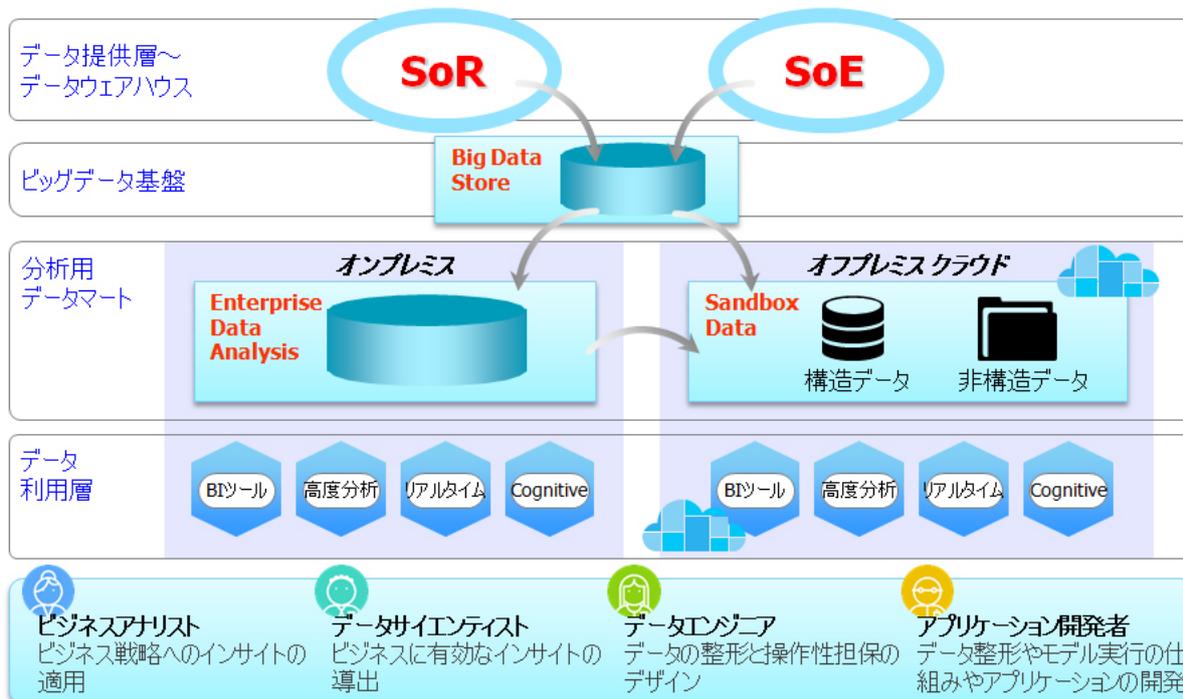
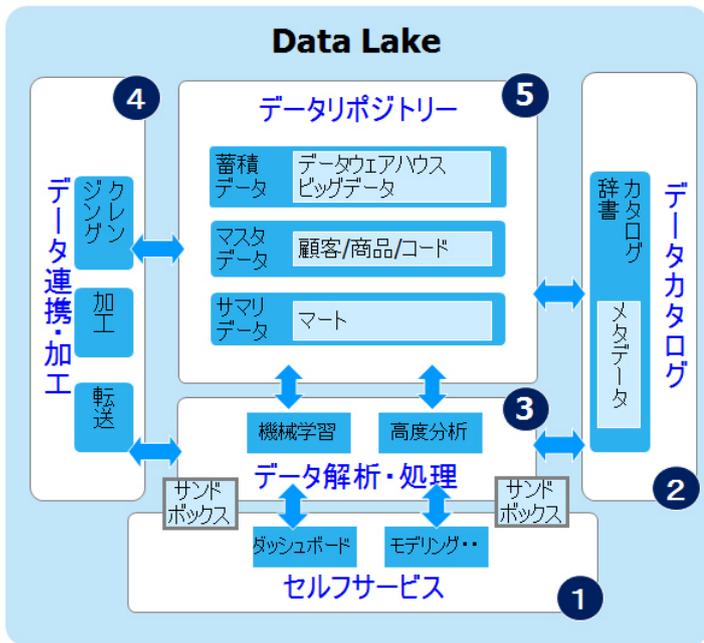


図6 データレイクを構成するコンポーネント



- 1 ユーザーがセルフサービスでデータ可視化・活用するためのユーザーインターフェース
- 2 ユーザーがデータを探索するためのカタログ・辞書
- 3 データから洞察を得るための分析エンジン
- 4 他システムとデータを連携するインターフェース
- 5 利用形態やソースデータの形式に応じて整備されたデータ蓄積基盤

図6 データレイクのコンポーネント		図4 銀行情報系アーキテクチャ
① セルフサービス	～	データ利用層
② データカタログ	～	データ利用層
③ データ解析・処理	～	データ利用層
④ データ連携・加工	～	データインタフェース
⑤ データリポジトリ	～	データウェアハウス、ビッグデータ基盤、分析用データマート

表2 図4アーキテクチャへの図6コンポーネントのマップ

なく、チームでのコラボレーションを促進するために、以下の価値を提供することも重視されています。

- ②～⑤の他のコンポーネントとシームレスに接続し、データの探索、収集、分析、デプロイまでを一気通貫で作業できること
- 上記作業を推進する上で、チームメンバー間での協業を促進するために、異なる言語での開発モジュールやデータセットを、統合されたプラットフォームで共有できること

これを実現するソリューションとして、Jupyter Notebook、GitHubといったオープンソースの技術が開発され、それらをエンタープライズ環境で活用するための拡張機能を追加したソリューションを、各ソフトウェアベンダーが提供を開始しています。

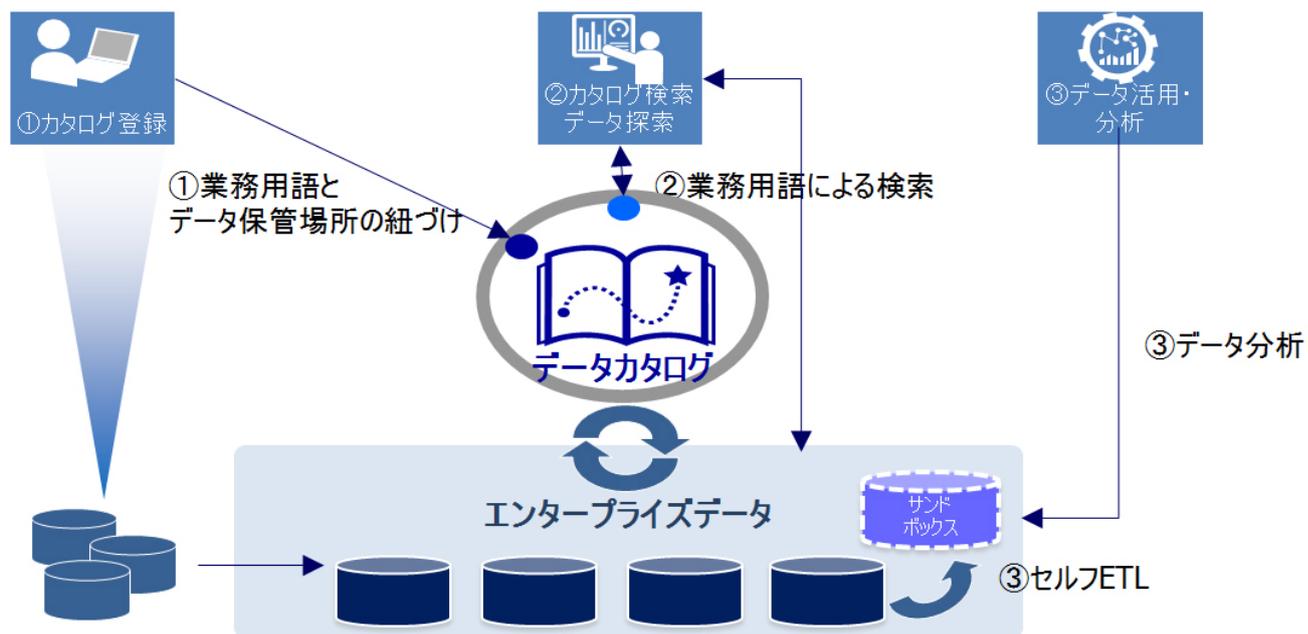
・データカタログ (図6-②)

ユーザー自身が必要なデータを探索し、取捨選択するために、データカタログを提供するコンポーネントです。このデータカタログには、メタデータを登録、蓄積、探索する機能が実装されます。

メタデータとは、データの定義情報であり、ビジネスユーザーが使えるようにするための用語 (ビジネスメタデータ)、そのデータがどのソースから来ているかを示す情報 (システムメタデータ)、アプリケーションやジョブの実行に関連する情報 (オペレーショナルメタデータ) が含まれます。またビジネスメタデータとシステムメタデータのひも付け情報もデータカタログに登録されます。

このデータのカタログ化で、ユーザーがアクセス権のあるデータにセルフサービスでアクセス、入手できるため、データ活用までの期間が短縮されます。また、組織横断的にデー

図7 データカタログを利用したデータ探索、分析、活用の流れ



カタログに格納されるメタデータとは？

メタデータには3種類あります。

1	2	3
ビジネスメタデータ	テクニカルメタデータ	オペレーショナルメタデータ
ビジネス用語(業務用語)、 ビジネスルール、管理方法、 ビジネス定義など	サーバー名、データベース名、 テーブル名、列名など	アプリケーションを実行時の情報: 頻度、レコード数など

タが可視化され、他部門のデータの利用が容易になるメリットもあります。

図7に利用イメージを示した通り、ユーザーはカタログを介してデータを探索し、探索したデータをセルフサービスETL（データの抽出、加工、ロードを担うソリューション）機能により、ユーザーが利用しやすいサンドボックスに抽出します。

・データ解析・処理（図6-③）

データから洞察を得るための分析エンジンが配置されるコンポーネントです。ここで利用される新しいテクノロジーとして、マシンラーニング（機械学習）と Apache Spark（以下、Spark）を紹介します。

データ分析においては、「ビッグデータ活用に求められる銀行情報系アーキテクチャとは」の章で述べたビジネスゴールやユースケースに基づく、探索的なアプローチが重要であり、これらの反復作業が示唆や洞察の獲得には欠かせません。一方、市場では多数の分析モデルやアルゴリズムが開発されており、上記ビジネスゴールと入手できるデータセットから、最適なアルゴリズムを選択する必要があります。

これらをスピーディーに効率良く進めるために、機械学習を中心とした手法である程度自動化できるようになっています。機械学習では、多数の分析モデル/アルゴリズムとデータセットとを照合し、その中から最適なものを自動的に導出します。

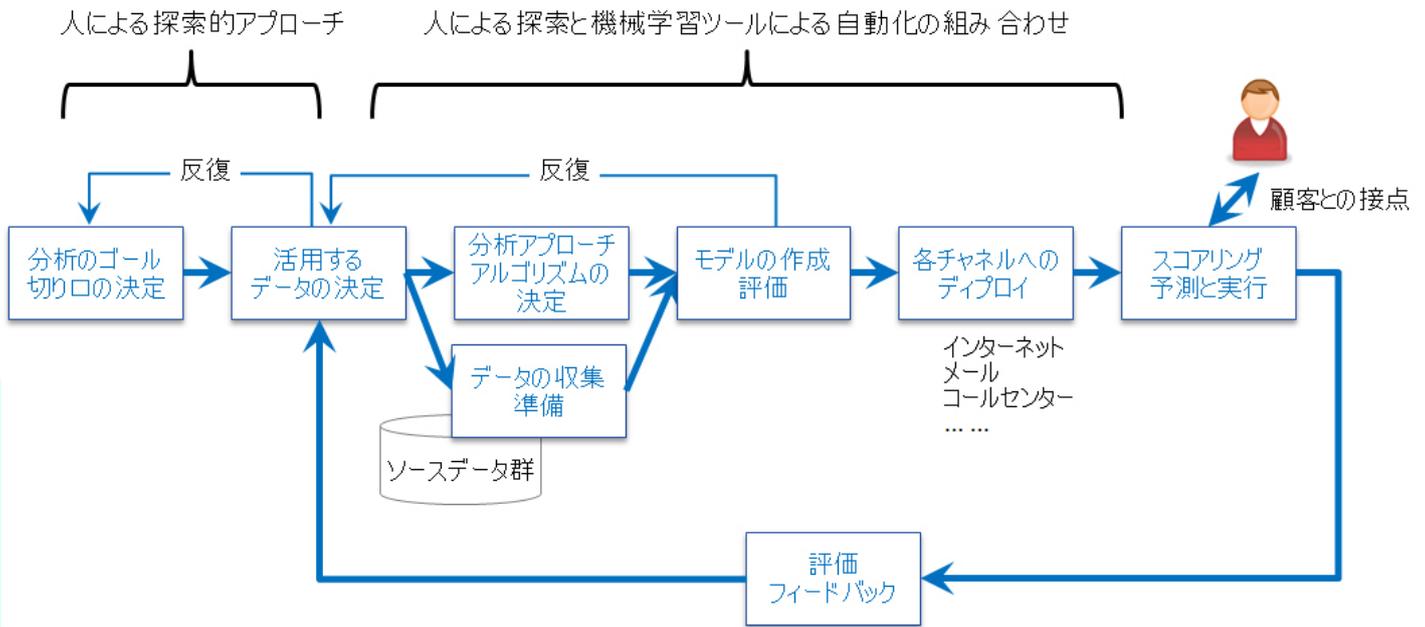
人手による探索的なアプローチと、機械学習によるアルゴリズムの自動選択とを組み合わせることで、より最適な洞察を得るようにデザインすることが必要です。

図8に示すように、この連続線上で、作成したモデルの各チャンネルへの展開、スコアリングの実行から評価に至るまで、ある程度自動的に行われるプラットフォームが必要とされています。

上記のアプローチを実行するプラットフォームを支えるエンジンとして Spark を活用するケースが増えていきます。Spark は、ビッグデータのキーテクノロジーとして、オープンソースの Hadoop を補完する位置付けで注目されており、データ解析や計算処理を、分散されたサーバのインメモリで並列実行するソフトウェアです。

Spark を銀行の情報系において活用する利点としては、下記が挙げられます。

図8 データ分析、データ収集、モデルのデプロイのステップ



1. 大量データ処理の高速化
2. 開発生産性の向上
3. アルゴリズムの高度化
4. 開発モジュールのプラットフォーム間の可搬性

大量データから高速に知見を得るために、生産性の高いプログラミング言語である Python、R、Scala など で開発した 機会学習のモジュールを、上記インメモリ処理の特性を生かせる Spark のプラットフォーム上で実行するような利用形態が考えられます。

・データ連携・加工 (図6-④)

分析の下準備として、データを抽出し形を整え、他のデータとの整合性をとる作業が必須です。データ活用全体のプロセスの中で、このデータ準備に最も多くの時間が費やされているといわれています。また前述の機械学習エンジンによる分析の自動化を試みる場合にも、インプットするデータの前処理や加工が重要です。

このコンポーネントでは、ETL、レプリケーションや仮想化ソリューションが選択肢として挙げられます。これらソリューションから、データ鮮度 (リアルタイムまたはバッチ)、データ加工の複雑性、ユーザーのセルフサービスかシステム開発を許容するか、などの要件を加味して最適なソリューションを選択します。

このコンポーネントでは、データが流入する過程で瞬時に分析し洞察を得ることで、意思決定やアクションのスピード

を高速化するためのテクノロジーとして、ストリームコンピューティングが使われるケースもあります。インターネットバンキングにログインした顧客の属性やセグメントから、取るべきアクションを機械学習モデルから瞬時に判断し、インターネット上でそのアクションを実行するようなケースでは、ストリームコンピューティングが活用されます。

・データリポジトリ (図6-⑤)

分析に活用するためのあらゆるデータを蓄積するコンポーネントです。各データリポジトリにアクセスする業務のデータ活用、蓄積に関する要件と、データリポジトリに適用される各テクノロジーの特性とのマッチングにより、最適なテクノロジーを選択します。例えば、従来リポジトリとして使用されているリレーショナルデータベースと、昨今の活用が進んでいる Hadoop の特性の違いを概観すると図9の通りです。

図4で示したアーキテクチャのビッグデータ基盤での処理に代表されるように、大量な非構造データを構造化し扱うための前処理、あるいは非構造データから洞察を得るための分析処理のために、Hadoopが使われるケースがあります。処理の一例としては、ソーシャルメディアに投稿されたメッセージなど大量なテキストの中から、関連する文脈や単語を探し出す処理が挙げられます。このケースでは、テキストデータの蓄積とテキストマイニングを行うプログラムの実行環境として、Hadoopが利用されます。

前述した Spark は、Hadoop 上の処理性能を上げたり、開発者の生産性を上げたりする目的で開発されました。従って、Spark との親和性が高い Python や R といったプログラミング言語で、Hadoop 上のデータを扱うアプリケーションが開発されることが多くなっています。また SQL on Hadoop のテクノロジーもあり、SQL に慣れた人が Hadoop 上のデータを扱えるようになっていきます。

なお、こうしたプラットフォームの技術要素は従来型の BI やデータウェアハウスを置き換えるものではなく、それを補完するものです。また、単一のプロジェクトとして集約型のプラットフォーム上に実現するものではなく、既存の拡張として取捨選択し実装するものです。従ってスモールスタートによる既存システムの拡張を可能としています。例えば、データカタログとサンドボックスのみ従来の環境に追加し、ユーザーが利用しやすい基盤を整備するといったアプローチが可能です。

コグニティブ技術の適用について

さらに昨今のコグニティブや AI 技術によりデータ活用の高度化が促進されるでしょう。例えば、以下のような目的で、すでにソリューションが実装されていたり、開発が進んでいたりします。

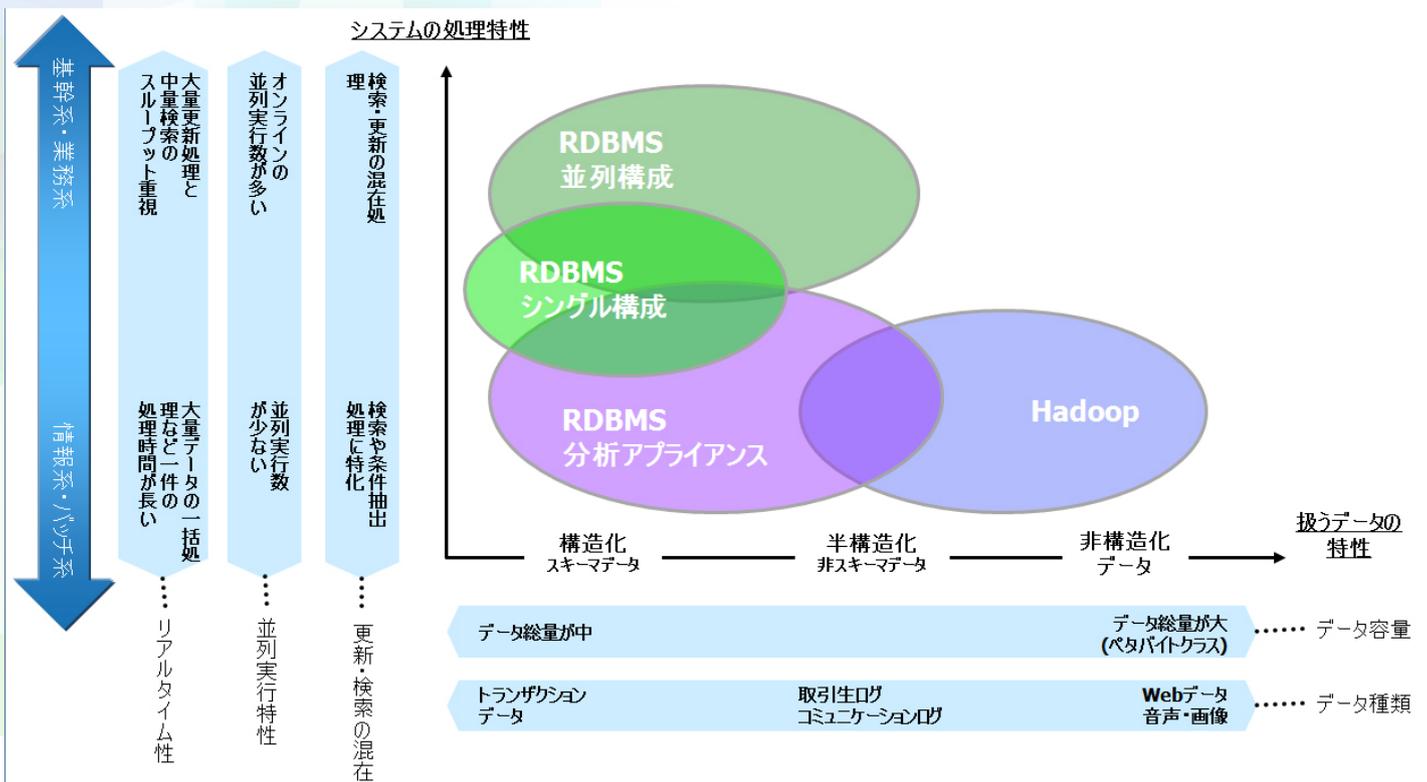
- 経営者や営業担当者らが自分自身でデータの可視化やダッシュボードの生成をする支援のために、コグニティブ技術が活用される（図 6-①セルフサービス）
- 見たい情報を自然言語で入力することで、データを探索し、データの型式に応じた可視化を自動的に行う
- ビジネス担当者が自然言語でのデータ探索を実現するために、システムに登録されているメタデータから、そのデータが持つビジネス上の意味に読み替えるのに、コグニティブ技術が活用される（図 6-②データカタログ）
- データ分析自動化の促進のために、得られるデータから適切な洞察を得るためのアルゴリズムの自動選択に、コグニティブ技術が活用される（図 6-③データ解析）

このように、企業ビジネスへのデータ活用の方針を考案し、適切なアプローチとアーキテクチャを判断する方たちの作業を補完する位置付けで、コグニティブ/AI 技術が活用され始めています。

まとめ

FinTech 時代の今、データ活用を進めるに当たり、「長期的な計画を立てつつ、少しずつ取り組み、素早く展開する」ことが重要となります。データドリブンのビジネス貢献に必

図 9 データリポジトリ、テクノロジーの特性の概観



要な人材やプロセス、テクノロジーに対して長期的なビジョンを持ちつつ、「業務や部門単位などで小さく始めて、その成果を実証していく」ことが重要視されていると考えます。

また、その短期的な実証の成果に対する可能性が周囲のステークホルダーに認知されるよう可視化しながら、大きく育てていくことが必要になります。本稿に記載したビジネス、データ、アーキテクチャ、テクノロジーを上流から下流まで総合的にデザインする組織、人材の育成が求められるでしょう。

次回は、FinTechで活用検討が進む ブロックチェーンについて

今回は、顧客インサイトを深めるためのデータ分析アーキテクチャおよびテクノロジーについてご紹介しました。顧客が主導権を握る FinTech 時代にデータ分析をどう活用していけばよいのか、そのための要素技術とその進化や方向性について説明しました。

次回は、FinTech 領域で活発に検討が進んでいるブロックチェーンを紹介します。