

AI時代に求められるクラウド分析基盤と 統合データ・プラットフォーム

AIを導入し競争力を強化する動きが加速し、AI活用の鍵となるデータの在り方が企業の競争力に直結する状況となっています。しかし多くの企業では、データが散在している、知見をどのように得るのかわからないなど、データ活用における課題・障壁があります。

そこで本稿では、データを企業で最大限に活用するための、データ管理からAI/機械学習の業務活用までを統合的にサポートするプラットフォームの要件を改めて整理し、データから価値を引き出すための2つのソリューション「IBM Watson Studio」と「IBM Cloud Private for Data」について解説します。

▶▶ 1. データ活用の取り組みの変遷

これまでは業務処理の単なる記録であったデータが、現在は企業内で共有されることで活用の場が広がり、データは企業の第4の資産と考えられるようになってきました。さらにデータとAIとの融合を行うことで、新たなビジネス・モデルや企業価値の創出につながる取り組みが加速しています(図1)。

多くの企業がAI活用によってビジネス・モデル変革を進め、そのベースとなるデータ活用に注力する一方で、世の中の80%の時間がデータの検索に費やされています[1]。利用するためのデータ自体が手に入らない、手に入ってもデータの整備がうまくいかない、ユーザーのデータ活用が

うまく進まないなどの課題が明らかになってきています。

そこで、データから価値を引き出してビジネスの差別化に生かすためには、どのようなデータ活用プラットフォームが求められるのかを考察していきます。

▶▶ 2. AI-Readyなデータ・プラットフォームの要件

ビッグデータやAI活用など、昨今のデータ要件の変化に対応するためのデータ・プラットフォームに必要な要件を、機能的側面と非機能的側面から整理します。

2-1. データ活用のステップとプラットフォームの機能要件

データからうまく価値を引き出すためのデータ活用のステップを、「データ整備」「分析・可視化」「業務適用」「継

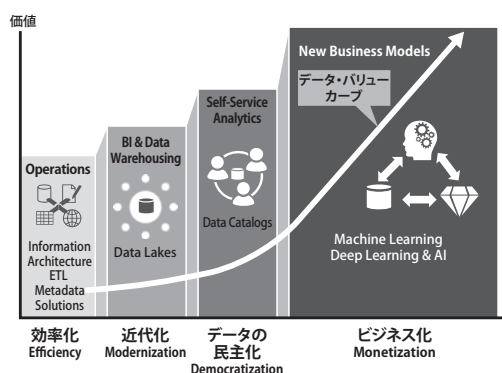


図1. データ活用の取り組みの変遷

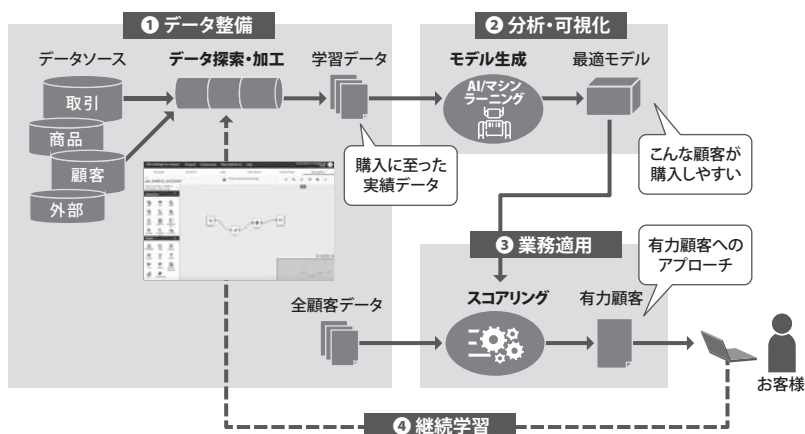


図2. データから価値を引き出すためのステップ例

続学習」の4つに分けて要件を整理します。データ活用の4つのステップについて、企業データから機械学習を用いて有力顧客を抽出するユースケースを図2に例示します。

①データ整備

企業内外にデータが散在している、データのフォーマットが統一されていないなどの理由から、データ活用プロジェクト全体の8割の工数は、データ収集から加工までの作業にかかると言われてしています。効率良く、スピーディーに、低コストで、質の高いデータを整備することが、このステップにおいて強く求められます。

②分析・可視化

このステップでは整備・加工したデータを基に非定型分析を行ったり、ダッシュボードなどに可視化したり、学習データとして機械学習機能で学習させ予測モデルを作ったりします。以下の点が求められます。

- 高度な分析スキルがなくとも精度の高い予測モデルが作れるような機械学習機能を装備し、多くのユーザーがデータから価値を創出できること
- 各担当者が利用しやすいツールを使い、生産性を上げられること
- AIプロジェクトに関わるメンバー各自が開発したデータや分析モデルなどのアセットを共有・再利用しながらも、権限管理などのガバナンスを効かせられること

③業務適用

このステップでは作成したモデルをビジネス・プロセスに組み込みます。データ活用は「分析をする」ことが目的ではなく、「価値を高める」ことが目的です。分析のアウトプットを業務に組み込み、顧客にサービスを届

けることで、データがビジネス価値となります。そのためには全社を俯瞰し、従来の業務をどう刷新していくかを考え実装する必要があります。つまり、分析のアウトプットを業務に実装するための仕組みが求められます。

④継続学習

上記に示す3つのステップを効率的に循環させ、継続的に精度を上げることも、データ活用をビジネス・プロセスに組み込む仕組みとして重要です。

2-2. プラットフォームの非機能要件

昨今のデータ活用要件の変化から導かれる、ビッグデータを収集・蓄積し最新のデータ活用を支援する基盤の非機能的な側面を図3に整理します。第四次産業革命の中核となるIoT、AI技術の活用に向けて[2]、ウォーターフォールで要件定義を経て設計・実装される従来型のデータウェアハウス(以下、DWH)の発想を脱却した、アジリティーや拡張性を重視したオープンなデータ・プラットフォームが求められています。

本章で述べたデータから価値を引き出すための要件・課題に対応するために、IBMは2018年3月にIBM Watson Studio(以下、Watson Studio)、2018年5月にIBM Cloud Private for Data(以下、ICP for Data)の提供を開始しました。2つのソリューションはともに、本章で述べたデータ・プラットフォームの要件を実装しています。提供形態は異なり、Watson Studioはクラウド上のフル・マネージドなSaaSソリューションとして、ICP for DataはオンプレミスやIaaSに導入可能なコン

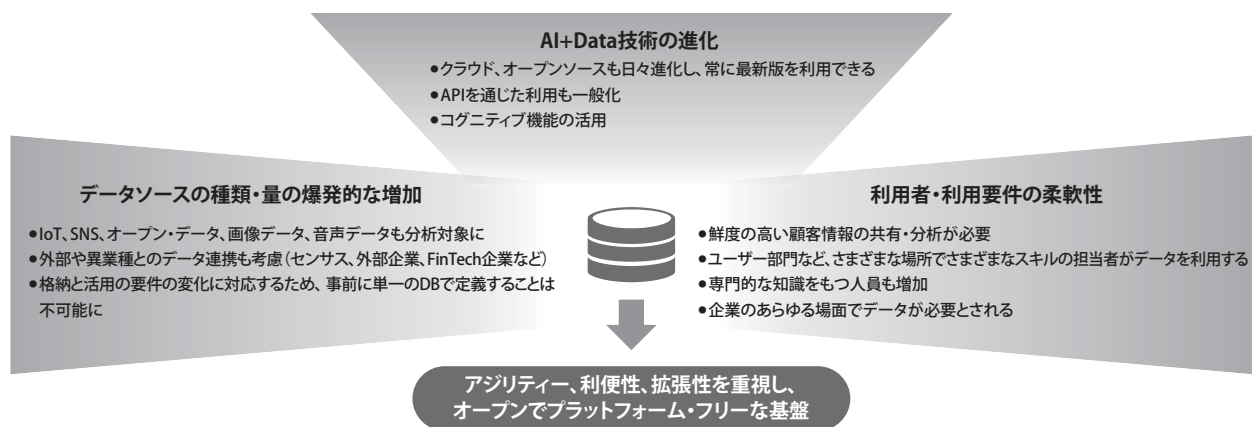


図3. データ活用要件の変化とプラットフォームの要件

テナ・ベースの統合ミドルウェアとして提供されます。
それぞれのソリューションを次章から解説します。

▶▶ 3. AIのためのクラウド分析基盤: Watson Studio

データ・AI活用によりビジネスの成功に導くためには、「データ基盤」「分析・AI基盤」「人工知能」の三者を相互に連携させ、継続学習のサイクルを確立することが重要なポイントとなります。IBM Watson(以下、Waston)ではそのための関連ソリューションを用意しており、Watson StudioはAIのためのクラウド分析基盤として位置付けられます。

また現在では、プロトタイプとなるクラウド環境を構築して、すぐに手に入るデータを基にクラウド上で試行錯誤しながらデータ分析を行うアプローチが一般的になっています。そのような環境を分析ユーザーのためのサンドボックスと呼ぶこともあります。小規模な分析プロジェクトから開始し、ステップを踏んで大きな仕掛けとするアプローチが必要となってきました。

図4では、クラウド分析基盤としてのWatson Studioに関連するサービス群を詳細化しています。その中で中心となるサービスの概観を解説します。

● 分析とAI開発

企業内でデータ・サイエンティストが使っているツールがそれぞれ違っていても、同じ環境で分析結果を閲覧、再利用しながらチームで分析できる仕組みとなっています。

例えば、データ・サイエンティストの利用の多いPython、Rなどでの分析が可能のように、オープンソース・ソフト

ウェアとしてPython Notebook、R Studioのツール利用が可能です。IBMソフトウェアとして利用者の多いSPSS Modelerの機能をWatson Studio内で利用することも可能です。さらに、分析スキルのあまり高くない業務ユーザーもマシン・ラーニング(以下、ML)のGUIから入力データを指定して機械学習を行うことも可能なため、データ・サイエンティストにとどまらず、ビジネス・ユーザー主導での分析に利用されます。

また画像認識が可能なWatson Visual Recognition APIやCore ML用の画像認識用モデルのエクスポートも利用できるため、分析プロジェクトだけでなく、AIプロジェクトの基盤としての利用も想定しています。

● エンタープライズ・カタログ

エンタープライズ・カタログ(Knowledge Catalog)は、企業内の部門間にまたがるデータの管理・可視化・意味付けに役立ちます。メタデータは「データのためのデータ」という意味で、テーブルのカラムやCSVファイルのヘッダー部分などデータに関する意味付けの部分を指し、データそのものではなくメタデータを管理しておくことで、実体のデータをコピーすることなく管理可能です。Knowledge Catalogのデータソースとしては、IBMのデータソースのみならずサードパーティーのデータソースにも接続可能です。部門ごとにバラバラのデータソースを利用して分析を行っている場合は、データを一元化し、意味内容を把握した段階で、分析で検索可能にするためのタグ付け機能を利用できます。利用するためのデータはデータ加工機能で加工を行え、分析の作業工数でもっとも大きいと言われるデータ整備の作業工数

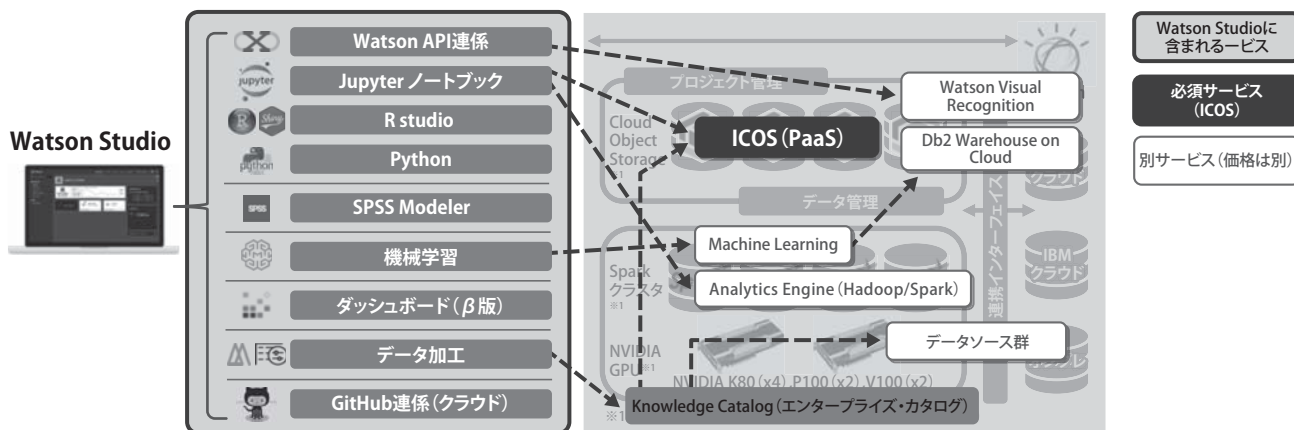


図4. Watson Studioに関連するサービス群

を削減可能です。

● データ蓄積

IBM Cloud Object Storage (ICOS) に大量データを格納して、大量データを処理できる仕組みがあること、アカウントを作ってインスタンスを作成するだけで手軽に始められることがクラウドで分析環境を実装する上での大きなメリットとなります。

▶▶ 4. 統合データ・プラットフォーム: ICP for Data

4-1. ICP for Dataが提供する機能

ICP for Dataでは、図5に示すとおり、2章で述べたデータ整備から業務適用までのフルサイクルを包括的にサポートするコンポーネントが装備されています。利用イメージとしては、ICP for Dataを既存のデータソースに接続させ、そのデータを加工したりMLで分析したりして、アウトプットをアプリに適用するような一連の流れで利用できます。

また各機能は従来のIBMミドルウェアやオープンソース・ソフトウェアが拡張し実装されています。例えばデータ蓄積には、ノン・チューニングなインメモリーDWHとして高速化・簡易化された「Db2 Warehouse」がコンテナ・ベースで同梱されます。プレパレーションとしては、多くの企業で活用されるETL「DataStage」が利用ユーザー部門向けにライトなGUIで最新化され、コンテナ・ベースで同梱されます。したがってIBMミドルウェアを利用中の企業では既存資産を生かし、最新化されたICP for Dataを利用できます。また他社システムを利用中の

企業でも、ICP for Dataを組み合わせ利用可能です。

次にICP for Dataに装備される主な機能のうち、エンタープライズ・カタログ、品質管理、プレパレーション機能の概観を記載します。

● エンタープライズ・カタログ

カタログの目的は大きく2つあります。

1つ目は、データ利用ユーザーが、適切なデータを必要なタイミングで探し、取り出しやすくすることです。カタログは、ビジネス用語でもデータを検索し、必要なデータの所在や来歴(どのような加工フローで作成されたか)を確認できます。そのためには、事前にビジネス用語とデータ定義(テーブル名・カラム名)のマッピングが必要です。そういった作業の負荷を軽減するために、AI技術が内蔵されていく方向にあります。例えば、過去のマッピング定義を学習データとして機械学習で自動マッピングさせることができます。

2つ目は、全社でのコラボレーションを促進することです。各メンバーが加工/生成した学習データや、分析モデルをカタログを介して共有・再利用することで、重複した加工作業や分析作業を削減し、企業全体の生産性を向上します。

● データ品質管理

データ品質の状況を可視化します。例えば分析結果に影響を与えるデータの欠損値の有無を確認できます。このようにデータの正確性をあらかじめ調査し管理しておくことで、各データ利用ユーザーが、都度データ品質を確認する手間を省きます。また過去の利用者によるデータへの評価を登録でき、データの正確性や信頼性を利用

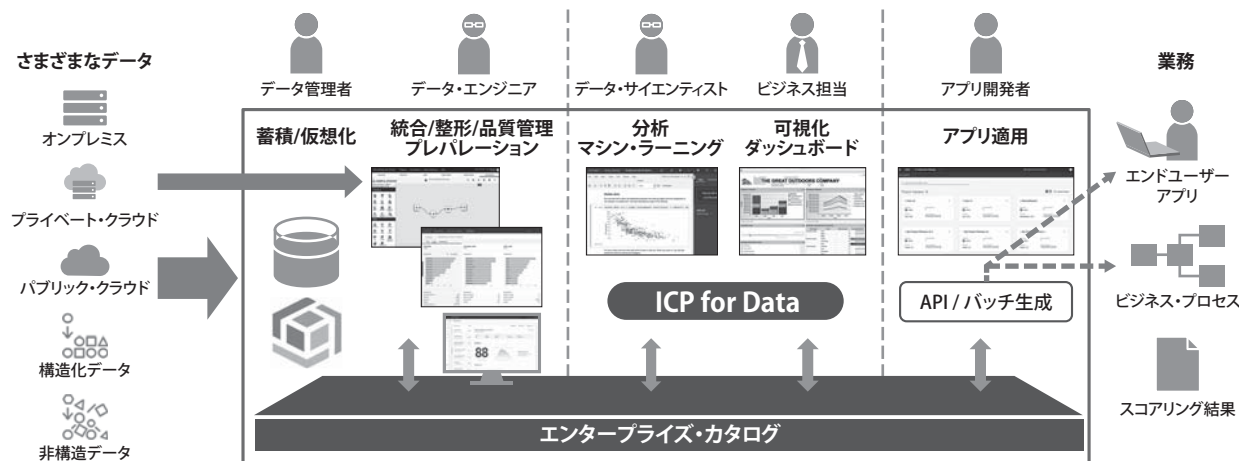


図5. ICP for Dataの概観

前に確認することが可能です。

● データ・プレパレーション

データ利用ユーザー自身が、データ管理者に依頼せずとも、データを加工することが可能です。担当者間のやりとりを減らし、データ活用のスピードを加速させ、生産性を向上します。

分析系の機能は、3章で説明したWatson Studioに含まれるものと同等のため、本章では割愛します。

4-2. ICP for Dataの構成

図6にICP for Dataの基盤構成を示します。ICP for Dataは、「IBM Cloud Private」(以下、ICP) [3]の上位層として構成されます。ICPは、Kubernetesベースのコンテナで実装された統合クラウド・プラットフォームで、基盤の導入・管理のワークロードを大幅に削減します。コンテナベースであるため、クラウド上で構築したDWHやETLをオンプレミス環境に移行しやすくてできるなど、ハイブリッド環境での可搬性も向上します。

またICPやICP for Dataは、オンプレミスのサーバーや、クラウドのIaaS上に構築可能で、セキュリティー・レベルを高められるため、個人情報を含めたデータの管理にも適しています。

▶▶ 5. Watson StudioとICP for Dataの位置付けと組み合わせ

2つのソリューションは、2章で述べた分析、データ・プラットフォームの要件をともに実装しています。選択肢として、スピードを重視したプロトタイプでのAIプロジェ

クトにはWatson Studioを、セキュリティーを重視した統合データ・プラットフォームの構築にはICP for Dataをといた使い分けが可能です。2つの組み合わせも可能で、オンプレミスとクラウドの適切なデータを取り出してETLバッチも活用しながらデータを整形しAI-readyな状態にするICP for Dataと、そのデータをAIに適用するWatson Studioといった連携も考えられます。

またソリューション間で以下のような連携のシナリオが考えられます。

①エンタープライズ・カタログに登録されたAIやデータのアセットを相互連携

- オンプレミス上の資産に登録したICP for Dataのカタログから、クラウド上のWatson Studioのカタログに、アセット情報をインポート可能です。オンプレミスでカタログに登録したビジネス用語を、クラウドのプロトタイプ環境にインポートするなどの利用シーンが考えられます。このように、それぞれカタログを組織横断の資産共有・再利用の場として活用できます。

②分析モデルや学習データを相互連携 (以下は将来的なロードマップも含まれます)

- プライベート環境の全社データを利用し、ICP for Dataのマシン・ラーニングでモデルを開発します。そのモデルをWatson Studioのスコアリング・エンジンで、クラウド・アプリに適用します。
- クラウド上で、Watson Studioのディープ・ラーニング・サービスでモデルを開発します。そのモデルをICP for Dataのスコアリング・エンジンで、オンプレミス環境のアプリに適用します。

2つのソリューションはIBM内の統合された開発拠点で開発されており、今後も相互連携を強化する方向です。

▶▶ 6. Watson StudioとICP for Dataの利用パターン

さまざまな業界において、デジタル変革推進とトップライン向上のために、データ活用を高度化するニーズがますます増えています。またIoTデータや非構造データなど、これまで利用できていないデータを含めて活用

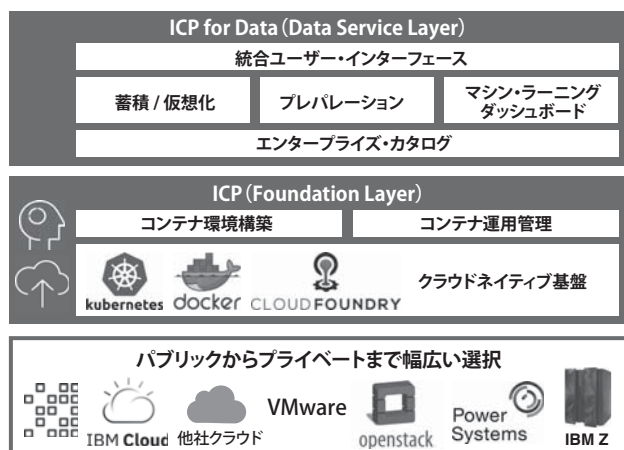


図6. ICP for Dataの基盤構成

するニーズも増えています。

それらのニーズに統合的にアプローチする際に、Watson StudioやICP for Dataを利用できることを、前章まで解説してきました。図7に、国内外の事例を基に汎化した利用パターンを示します。ビジネス担当者やデータ・サイエンティストがデータから新たな洞察を得るための試行錯誤ができるサンドボックス環境から、全社データ活用促進のためのデータ探索・加工・分析のサイクルを実装する環境や、高度な分析や自動化の仕組みを提供する環境まで、さまざまな業界で取り組み事例が創出されています。

このようにデータの蓄積・可視化が中心であった従来のデータ基盤に対し、図7に示すような分析・AI活用の要素を段階的に強化し、次世代の統合AI分析基盤へと発展させることができます。

7. おわりに

本稿では、企業のデータ活用を支援するAI、分析、データ・プラットフォームの在り方から、Watson Studio、ICP for Dataの最新情報を解説してきました。2つのソリューションは、今後も拡張され最新のテクノロジーを実装していくため、実装時点の最新情報をご確認ください。例えばWatsonでのAIモデルの管理については、AI OpenScaleサービスで、AIモデルの精度評価と開発のライフサイクル管理が行えることが予定されています。企業横断のAIプロジェクトにおいて、整備・開発したデータとAI関連のアセットをフルサイクルでより高度に管理・再利用する役割が期待されます。

IBMは、デジタル変革を加速する企業や組織に向けて、Watson StudioやICP for Dataを駆使し、単体ソリューションの組み合わせでは実現できないアジリティで、AIとデータを統合した次世代プラットフォームと包括的な支援を提供します。企業全体のデジタル変革に向けて、ビジネス部門とIT部門とが一体となって新たな価値創出を推進するための一助となれば幸いです。

[参考文献]

- [1] FROM DATA TO DISRUPTION: INNOVATION THROUGH DIGITAL INTELLIGENCE
IBM-sponsored report by Harvard Business Review Analytic Services, 2016,
<https://hbr.org/sponsored/2016/12/from-data-to-disruption-innovation-through-digital-intelligence>
- [2] 内閣府経済財政政策-日本経済2016-2017, 第1節 第4次産業革命のインパクト
<http://www5.cao.go.jp/keizai3/2016/0117nk/nk16.html>
- [3] IBM Cloud Private, <https://ibm.com/jp-ja/marketplace/ibm-cloud-private>



日本アイ・ビー・エム株式会社
IBMクラウド事業本部
Watson & Cloud Platform テクニカルセールス
シニア・アーキテクト

久保 俊彦
Toshihiko Kubo

1999年日本IBM入社。電子カルテシステムの開発プロジェクトへの従事を経て、Information Architectとして、金融・製造・流通業の基幹系/情報系システムのコンサルティング、アーキテクチャ設計を、海外SMEと連携し推進。現在は金融機関のビジネス/IT部門とともに、企業横断でのデータ・AI・クラウド活用による価値創出を推進している。



時光 さや香
Sayaka Tokimitsu

2002年日本IBMにSEとして入社。金融・製造・流通など様々な現場で8年間オープン系アプリケーションのデータ設計・データ関連コンサルティング、情報系システム構築を担当。7年前からは技術営業として、データ分析製品の選定支援を行う。現在はWatson Data & AIを中心としたクラウド・サービスの選定支援を行っている。

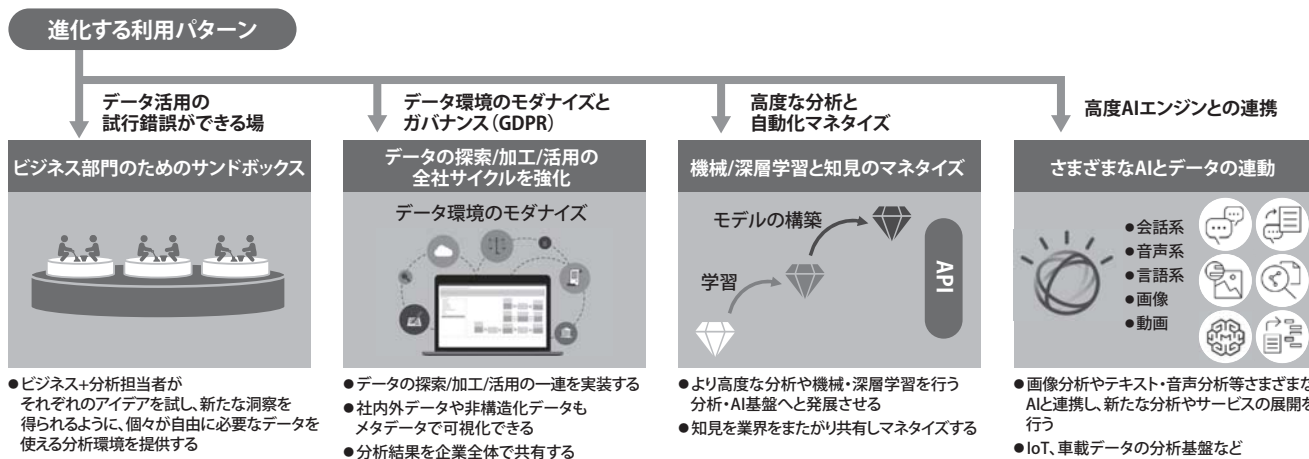


図7. Watson Studio/ICP for Dataの利用パターン