

Deloitte.

デロイトトーマツ



データアーキテクチャ・ データマネジメントのあるべき姿

真のデータ利活用によりデータドリブン
経営を実現するアーキテクチャの要点

デロイトトーマツ コンサルティング 合同会社
テクノロジー戦略・トランスフォーメーション

Introduction

真のデータ利活用によりデータドリブン経営を実現するアーキテクチャの要点

変化の激しい経営環境において、データ活用を目的にDWHやBIなど多額の投資が行われているものの、まだ目に見える十分な成果があげられているとは言いがたい。デジタルを武器にデータをビジネスに役立て、データを活用した意思決定の高度化を進めるためには、ITシステムとしてのデータプラットフォーム導入ではなく、データ活用のための意識や文化を社内全体に浸透させるための包括的な態勢整備が急務となっている。

はじめに	4
データを活用した戦略策定アプローチ	4
データプラットフォームにおける課題とあるべき姿	5
データレイクハウスアーキテクチャの獲得	6
データガバナンス基盤の整備	7
クラウド分散管理の仕組みの整備	8
データマネジメント組織の整備	9
まとめ	11

データを組み合わせ新しいインサイトや価値を導き出すデータの目利き力が企業価値の向上につながる

はじめに

IT・デジタル化が進み、さまざまな情報がデータとなり流通している。その量は加速度的に増大し、いたるところで蓄積されている。蓄積が進んだデータからは、情報源となる業務システム、機器や顧客などの状態がわかるだけでなく、他の情報と組み合わせることで、業務上の非効率、機器のトラブルや顧客の離脱要因といった、隠れた課題や機会を把握できるようになる。さらにAIや機械学習といった技術を重ねることで、将来予測も可能となった。情報は価値を秘めた資産となり、そこからいかに価値を引き出すかに多くの企業が試行錯誤しながら取り組んでいる。

近年のCOVID-19や世界的な政治・経済情勢の変動により、過去数十年間、企業の経済活動や交流の増加を促進してきたグローバル化のリスクや脆弱性が露呈し、グローバルなヒト・モノの物理的な移動が減少してきている一方で、情報・データの流通は逆に加速している。例えば出張のような物理的なアクティビティは、オンライン会議やビデオによる視察で実現できるようになった。また、ハードウェアの部品交換が必要だった場面も、ソフトウェアのアップデートによって取り扱われるようになってきている。このように、物理的な制限がある場合でも、ビジネスの活動はデジタルの領域で続行され、仮想空間でのデータのやり取りで代替されるようになっている。

この新たな時代におけるデータ活用については、両極化している。大手のプラットフォーム、例えばGAFA（Google、Amazon、Facebook、Apple）やBAT（Baidu、Alibaba、Tencent）は、あらゆる活動をデータとして把握しつつある。ひょっとするとこれらのプラットフォームを追従しようとする企業もあるかもしれないが、先行者に困り込まれた牙城を崩し、収益につながる規模を確保できることは困難と言わざるを得ない。また、これらの巨人たちの影響力はその大きさゆえに規制の対象ともなっており、プラットフォームとして新たに参入し、データを独占することを目指すことは非現実的である。むしろ、多くの企業にとっては、様々な情報源からデータを組み合わせ、新しい価値を生み出すことが重要である。そのためには、「データの目利き」が重要なポイントとなる。これは、情報の海から最も関連性の高いデータを見極め、それを利用して有効な意思決定を行う能力を指す。

このような背景から、データを価値ある資産として捉え、適切に活用するデータドリブン経営が求められるようになった。データドリブン経営は、「データをもとにした経営」を意味する経営用語で、これまでのような経験や勘に頼るのではなく、収集・蓄積されたデータの分析結果に基づいて戦略・方針を策定することを示している。データドリブン経営では、ITツールを活用しながら収集したデータを適切に分析していく必要がある。そのためには、これまで紙媒体で保持していた情報のデジタル化や社内のデータ分析基盤の整備を行い、効果的なデータ分析ができる状態にすることが不可欠となる。

ここで、データ活用のイメージをもつため、データ活用の必要性を認識し、それに取り組む具体的なケースをあげておく。

ケース1：外部からの要求への対応

2020年3月期より、有価証券報告書における記述情報の開示の充実に向けた取組みが本格化している。このような規制当局、投資家や顧客の期待など、外部からの要求に対して、企業はデータ活用を本格化する必要がある。具体的には、報告書への事業リスクの記述充実化のために事業リスクの数値化（デジタル化）や、各組織・事業間のデータの横串連携が求められ、常に最新の情報を経営層に提供する体制を目指していくことが不可欠となっている。

ケース2：経営ダッシュボードの実現

不確実性が高まる経営環境下で、ファクト(=データ)に基づいて意思決定を行うアプローチと同時に、データのリアルタイムな収集とその可視化は競争力の源泉となる。そのために、企業は財務データだけでなく、非財務データや、外部データも含めた多岐にわたるデータを活用し、様々なリスクシナリオを考慮し、経営判断に必要な各種指標値を把握可能とする経営ダッシュボードを構築することで、タイムリーな経営判断を可能としていけるようにすることが重要となっている。

ケース3：KPIの定義とアクションの実施

データ収集だけでなく、そのデータを基にKPIを設定し、経営アクションに結びつける取り組みが求められる。このために、異変を早期に察知するためのしきい値を定義し、全社共通のプラットフォームを整備することが重要である。また、KPIを定義する際は、業務の実態や戦略的な目的を反映させることが不可欠である。

データを活用した戦略策定アプローチ

従来、多くの日本企業では、意思決定は経験や主観に基づいて行われてきた。人海戦術で年単位のデータを集計し、経営企画部のスタッフが数週間かけて、過去の情報に依拠しつつ戦略を練るケースが多い。しかし、現代の高速なビジネス環境においては、このようなアプローチで検討した戦略は意思決定の段階で陳腐化する。これからの時代においては、不確実性に対応した複数のシナリオを予測し、意思決定し、さらにその結果をデータで可視化しながら継続的にアクションを調整していく、というアジャイルな戦略策定や意思決定プロセスへのシフトが求められる。担当者の経験に基づく、恣意性も含んだ要素抽出と推測というアプローチから、データに基づき価値あるインサイトを導出し、高頻度かつ短サイクルでビジネスをアップデートしていくアプローチへと変えていくことが不可欠になってきている（図1参照）。

このアジャイル型の戦略策定アプローチを成功させるためには、データの収集と分析の仕組みを整えるだけでなく、本質的かつ有効性の高いデータを正確に見極めることが肝要である。企業が過去のデータだけをもとにした分析のみを行うのではなく、企業外にあるさまざまなデータも組み合わせ活用し、先を読みアクションを取るというフォワードルッキングなアプローチが求められる。データを組み合わせることで新たな価値が生まれることを意識し、異なる情報源からのデータを組み合わせることによって、『データの目利き力』が企業価値の向上につながる。

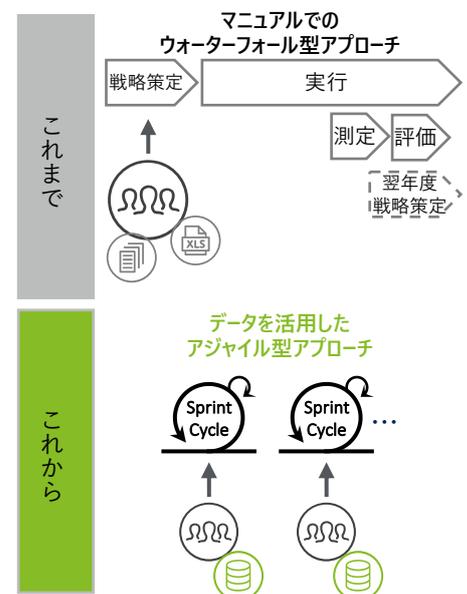


図1：データ活用アプローチ

ビジネス価値創出につながる データプラットフォームのあるべき 姿を定義する

データプラットフォームにおける課題とあるべき姿

ここから、データ分析基盤の整備の観点で、従来の日本企業における課題と、あるべき姿について述べていく。

①従来システムのデータ活用における課題

従来は、システムごとにデータが点在しており、把握できる活動は一部の事業や業務にとどまるほか、企業全体のポートフォリオが把握できず、適切でタイムリーな意思決定が困難なものとなっている事が多い。具体的には、各システムにデータが点在していること、データの意味定義・マスタデータがシステム毎に異なること、ファイルによるデータ連携により、リアルタイム性が欠如していることや、データの視える化については、巨大なおぼけエクセルやアクセスによるローカルでのデータ集計となっていることが多い（図2の左部参照）。

これらにより、計画を実行へ落とし込むことや、その計画に対する実績のモニタリングや分析ができ無いという課題、業務とシステムの個別最適化となってしまう全社横断的な視点でデータを捉えることができないという課題、業務の属人化によって集計作業に時間が掛かりタイムリーなレポートができないという課題や、コーポレートが、グループ子会社や事業部からデータを収集・分析する仕組みがないという課題が顕在化していることがほとんどである。

②日本企業でよくある取り組み

日本企業においてデータ活用を進めようとした際にデータプラットフォームの導入案件でよく見られるのが、ストレージベンダーやBIツールベンダーの提案してくるソリューションをそのまま採用する形での、全社のデータを集めてくる巨大な全社データベースやデータウェアハウス（DWH）という仕組みの構築である。具体的には、まずデータ分析やレポート要件に合わせたデータ項目のリストアップを行い、全社データベースやDWHのデータ構造を定型フォーマットのテーブル構造として定義し、データソースとなる既存システムからテーブル構造に合わせバッチ処理にて前処理を行いデータを収集するという仕組みを作るものとなる（図2の中央部参照）。

このような仕組みでは、都度テーブル設計やBI画面の開発が必要となるため必要ときにタイムリーにデータ活用を推進できないという課題や、システムごとに異なるデータの意味定義を解消できないため大量のデータを集めたにも関わらず活用できないという課題、そして、個々のシステムからは登録用のテーブル構造に合せて集計データのみ連携する事になりがちでDWH内で詳細データがないため詳細分析につながらないといった課題が発生する事が多い。

オンプレミスのDWHシステムは、その殆どが固定的なレポート帳票をBI画面で構築し、その帳票に必要な項目に従って固定的なテーブル設計を行うアプローチで構築されており、ITベンダーによるシステム構築にかかるコストを増大させる要因となる他、変更容易性に欠けるため、デジタル時代に求められるさまざまなデータをかけ合わせるというデータ活用アプローチにおいては不向きである。

③目指すアーキテクチャ

企業内外にあるさまざまなデータを組み合わせるデータ活用を行うために望ましいデータアーキテクチャは、データファブリック・アーキテクチャである。これは、従来の個々のデータベースやDWHとは異なり、複数のデータ環境間でデータの流動性を高めること、つまり、大きな全社データベースや全社DWHにデータを積み、データのサイズが大きくなるにつれてデータを移動・変換・集積することが複雑になり、データ活用そのものが困難になるという課題に対処することを目指したアーキテクチャである（図2の右部参照）。

データファブリックアーキテクチャの特筆すべき特徴は、メタデータとAIを活用して関連データを発見することに重点を置いていることにある。1つ目にはデータ民主化を進めるため、データサイエンティスト等のテクニカルな人材以外にも業務ユーザーがデータにアクセスできる仕組みとする。2つ目に、メタデータ管理や機械学習を用いて、サイロの解消、データ・ガバナンス手法の一元化、全体的なデータ品質の向上を実現できる仕組みとする。3つ目に、多くのデータ・ガバナンスのためのガードレールやアクセス制御機能にて、特定のデータが特定の役割の人だけに利用可能になる状態を確保し、優れたデータ保護を実現できる仕組みとする。

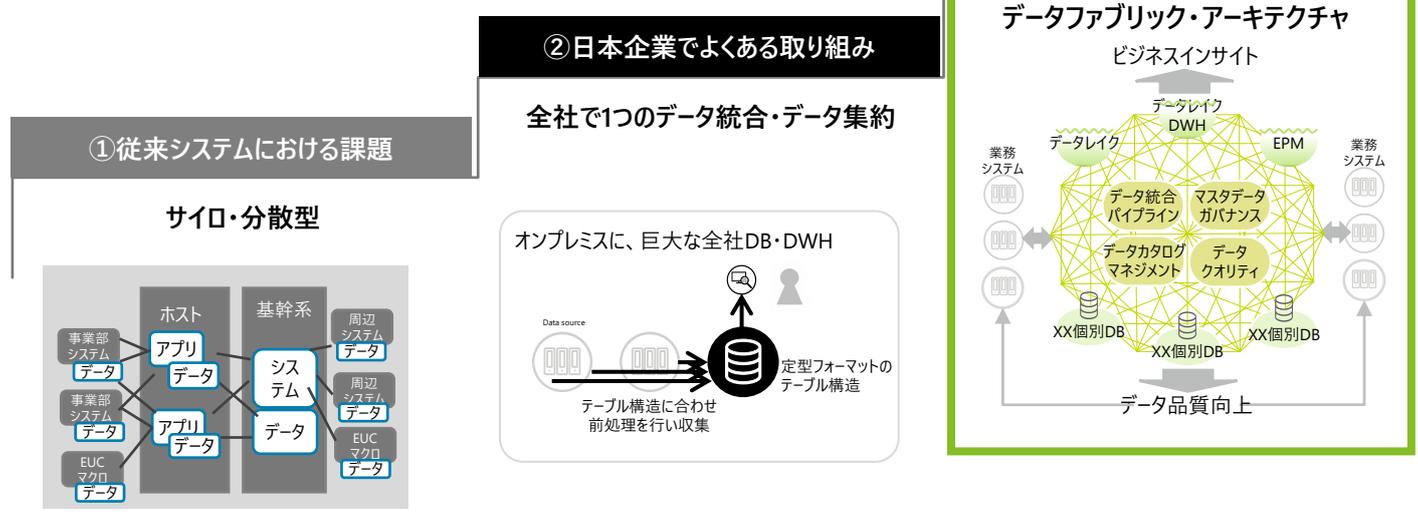


図2：データプラットフォームにおける課題と目指すアーキテクチャ

DWH依存から脱却し、AI活用に資するデータレイクハウスを整備する

ここから、真のデータ活用をすすめるべく日本企業が整備すべき要素を1つずつ見ていこう。まず1つ目は、データファブリックアーキテクチャの構成要素となるデータレイクハウスである。

データレイクハウスアーキテクチャの獲得

①データレイクテクノロジーの登場

先に述べたように、現在の日本企業で主流になっているデータ基盤のほとんどは、業務システムなどが生成する構造化データの扱いを得意とするDWHの整備が先行しているが、企業がBig Dataをもとにデータ活用していくためにはDWHでは実現できない幾つかの課題が顕在化してきている。

主な課題の1つ目は、従来のDWHでは、実は大量データの取り扱いが困難であることが挙げられる。従来のDWHは、オンプレミス環境に物理的なストレージを配置して実現することがほとんどであり、ストレージサイズ変更に対する柔軟性が無いことが課題となる。例えば、過去1年間としていた分析対象のデータを3年にしたいという要望や、新規に構築されたデータを分析対象に加えるといった、初期設計時以降に追加される要望への対応は、極めて困難なものとなる。また、主な課題の2つ目に、分析対象のデータが限定されることが挙げられる。DWHはあらかじめデータの設計が必要となるため、別システムで発生する新たなデータを格納することや、画像や動画、音声というような規則性を持たない非構造化データを扱うことが出来ない。これらの課題解決のため、「低コストで柔軟な容量を確保すること」と、「事前設計不要ですぐにデータ保存開始すること」が可能となるテクノロジーが必要となってきた。

ここで登場したのがデータレイクという仕組みである。データレイクという概念は、Pentaho社（オープンソースBIツールベンダー）のCTOのJames Dixon氏が、2010年10月New Yorkで開催されたHadoop Worldで、「ビッグデータを整理するための新しい概念として、「データレイク」ということを提唱したことが最初と言われている

(<https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>)。

データレイクの実装手法は各ITベンダやクラウドベンダにより様々あり、データレイク製品と呼ばれるソリューションは玉石混交となりつつあるが、データ活用のために重要となる技術的なポイントを2つ挙げておく。

1つ目は、規模にかかわらず、すべての構造化データ及び非構造化データを保存できるデータレイクを実現できることである。従来のリレーショナルデータベースやDWHは、データの分析や利用に合わせてはじめてデータ構造を定義しておき、そのデータ構造に合わせた形でデータを加工や変換してからデータを保持するという「スキーマオンライト」の考え方で構築される。これに対して、データレイクは、構造化データも非構造化データもそのままの形（生データの形）で保存しておくことができる仕組みを実現する。データを格納する際に、そのデータに事前定義されたスキーマは無く、データを読み込むときにはじめてスキーマ定義に合うかどうかを検証するという「スキーマオンリード」という仕組みにより、すべての構造化データ及び非構造化データを保存できるデータレイクを実現する。

2つ目は、API連携できるストレージ機能であることである。データレイクは生データを保持できる仕組みであることが求められるが、全てのデータをフラットテキストファイル化した上で保管する単純な物理ストレージのみでは実現できない。「スキーマオンリード」という考え方において、データ蓄積とデータ分析処理を分離した上で、ダッシュボード、ビッグデータ処理、機械学習など、さまざまな処理とAPIにてデータの中身を連携できる、クラウドネイティブなストレージ機能を用いて実現される。

これらの機能は、主にAmazon Web Services、Microsoft Azure、Google Cloudの3大ハイパーケーラー・クラウドサービスにおいて、それぞれ、Amazon Simple Storage Service(S3)、Azure Blob Storage、Google Cloud Storage(GCS)というオブジェクトストレージサービスを中心として実現されるようになった。これらのサービスは、先に述べた「低コストで柔軟な容量を確保すること」と、「事前設計不要ですぐにデータ保存開始すること」が可能であり、構造化データであるCSVファイル、半構造化データであるJSONやXMLファイルに加えて、動画、音声などの非構造化データも格納可能であるため、企業のデータプラットフォームとして導入することが広まり始めている。

②データレイクの階層構造

データレイク概念が登場した当初は、生データを生成されたままの形で格納するという使い方をすることから、データソースごとに文字コードや改行コード、日時の表現形式がシステムごとにバラバラとなり、分析に使用するために都度形式を揃える処理が必要とされ、分析作業の速度・効率が上がらないといった課題があがってきた。しかし、データレイクが登場してから10年を経る中で、このような課題に対する方策の検討が進み、データレイクに格納したデータを段階的に処理を施して内部的に複数の階層に分けて処理する構造がベストプラクティスとして定着してきた。各階層の役割について、次に示しておく。

- 第1階層（収集レイヤ）・・・様々なデータソースからデータを収集するためのハブ機能やアダプター機能を保持する階層
- 第2階層（蓄積レイヤ）・・・様々なデータソースから収集した生データをそのままの形で保存しておく階層
- 第3階層（一次処理・加工レイヤ）・・・すぐに検索が実行できるための変換処理を行う階層（例：西暦・和暦など表示形式変更、Null値の変換、大文字小文字の統一、テーブルの結合、大量のファイルのパーティショニング、ファイルフォーマットの変換など）
- 第4階層（活用・分析レイヤ）・・・BIツールやレポートサービス、機械学習サービスと連携し、データ活用を行うための階層

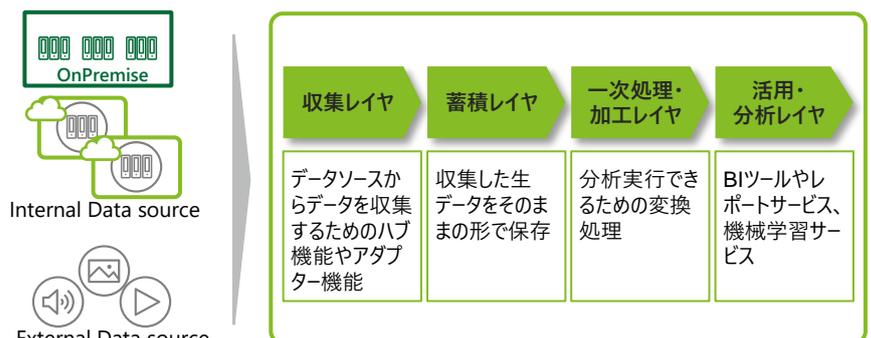


図3：データレイクの階層構造

③モダンデータアーキテクチャ（データレイクハウス）

データレイクを扱うようになった企業においても、データアーキテクチャにおける現状の課題として、業務システムなどが生成する構造化データの扱いを得意とするDWHの整備と、大規模かつ非構造化データ活用のためのデータレイクの整備が分断管理されていることが挙げられる。この状況は、AI活用の足かせとなっている。DWHに保存される過去の蓄積データを元にAI分析をしたり、データレイクに保存される鮮度の高い多種多様なデータをもとにAI分析したりすることが想定されるが、本来的には多角的なデータを元に価値創出することが目的となるため、利用するデータは構造化／非構造化を問わず用いるべきであるが、DWHとデータレイクの分断により、それが困難となっているのだ。

ここで、近年脚光を浴びてきているのが、DWHとデータレイクの両方のいいとこ取りをした「レイクハウス」というサービス・製品である。DWHの分析力と、データレイクの拡張性の両方を持ち合わせた新しい仕組みであり、定型／非定型など多種多様なデータを格納できる他、高度分析サービスとAPIによる連携を強化したものとなっている（図4参照）。データレイクハウスにおいては、全てのデータを1つのデータストアに集約管理することを目的とせず、データレイク、DWH、リレーショナルDB、NoSQLなどの非リレーショナルDB、機械学習など目的に応じたデータストアに分散配置した上で連携させ、全体としてスケラブルでセキュアでデータプラットフォームを実現するものとなる。

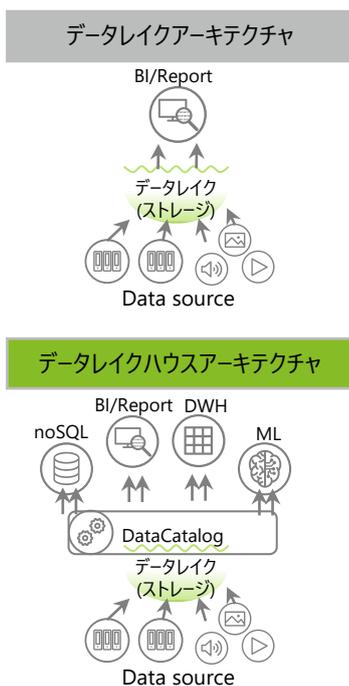


図4：データレイクハウスアーキテクチャ



図5：データガバナンス基盤に求められる仕組み

データスワンプ化するのを防ぐために、データガバナンスの共通基盤を整備する

データ活用をすすめるべく日本企業が整備すべき要素の2つ目は、データガバナンス基盤である。

データガバナンス基盤の整備

データレイクは構造化データや非構造化データといった形式を問わず、あらゆる情報をそのままの状態保持できることが最大のメリットとなるが、データが無秩序に保管された状態（データスワンプ＝データの沼）となってしまう、情報がどこにあるのか把握しきれなくなるリスクがある。各部門がそれぞれのルールでデータを保存してしまうなどにより、企業全体としてデータスワンプに陥ることのないよう、統制（ガバナンス）を実現する仕組みが必要となる。

データガバナンスという、一般的には、全てのデータを一箇所に集めて中央集権管理をすることで実現しようとする考え方が取られがちだが、これはオンプレミス時代の考え方であり、クラウドベースのデータレイクハウスにはそぐわない考え方となる。データレイクハウスアーキテクチャにおけるデータガバナンスに関するベストプラクティスは、すべてのデータそのものを一元管理するのではなく、データの意味定義を一元管理する事を目的とし、4つの機能で実現される（図5参照）。

①データパイプライン管理

1つ目がデータパイプライン管理である。これは、ソースシステムからデータを移動し、要件に基づいてデータを変換し、ターゲットシステムにデータを格納するために必要な手順を実装するための機能である。従来、システム間のデータ連携について個々のシステム間で1対1でファイルにてデータ転送する仕組みとす

ることが多いが、これでは開発対象や管理対象が膨大になってしまう他、どのソースシステムからターゲットシステムへデータを移動し格納しているかについて、全体を把握することが困難になってしまう。データパイプライン管理を共通基盤として整備することで、全体のデータ連携に対する管理統制を実現し、データリネージ（データの作成から現在に至るまでの全履歴）を追跡可能にする。

②データカタログ管理

2つ目がデータカタログ管理である。データカタログは、組織内のデータに関する情報を集約し、一元化して管理するためのデータの索引である。実態データそのものを管理するのではなく、組織内のデータについて、データの種類、属性、取得方法、更新頻度などを文書化し、管理することで、図書館における索引管理と同様に、どんなデータがどこにあるかを容易に検索・参照できるようにし、どのような時にどのデータソースを用いればよいかを判断可能とする。

データカタログとして管理すべき情報は大きく3つに大別される。1つ目は技術的メタデータという情報で、DBやテーブルのスキーマ、テーブル、列、ファイル名、レポート名などが含まれる。2つ目は、ビジネスメタデータという情報で、業務や事業における分類、使用適性、評価などをコメントとして記述するものである。3つ目がオペレーショナルメタデータという情報で、更新タイミングや、データアクセスの詳細などを含む。

データカタログを管理するという概念は昔からあったものの、企業全体のデータカタログをマニュアルで作成管理することは時間とコストがとても掛かるものであった。しかし、機械学習の発達により、既存システムのテーブル情報をディスクバリエーション機能にて収集したり、データパイプライン管理と合わせてデータカタログの整備が高精度で自動化できるようになるなど、技術の進歩により、ビジネスで実用可能なレベルで実現できるようになった。

クラウド分散管理の仕組みの整備

③ マスタデータ管理

マスタデータとは、業務で扱う基本データを指し、例えば顧客マスタや製品マスタなど様々な種類がある。多くの企業において、部署ごとのシステムでデータを管理しており、本来同一であるべきデータが異なるシステムで異なるデータの持ち方となっていることが多く、全貌を正しく把握することが難しかったり、企業全体としてマスタデータに一貫性がないことにより、データ活用の作業効率が悪くなり生産性の低下につながる。マスタデータ管理（MDM：Master Data Management）機能にて、複数システム間で異なるマスタデータの値の関連付けをしたり、複数のシステムで唯一の正本となる値にマスタデータを統一することで、企業内に散乱していたデータを統合し、精度の高いデータ活用につなげることを可能とする。

④ データクオリティ管理

データ活用だけでなく、業務においてデータを扱う際に、データの統合や移動でエラーが発生したり、不正確なデータが含まれていることがある。例えば、顧客データのレコードにおいて、顧客IDカラムが空欄になっていてNot Null制約に違反するといったことや、電話番号カラムにおいてハイフンの有無や桁数などフォーマットの異なるデータが含まれることや、会社名カラムにおいて同一の会社にもかかわらず「株式会社」と「(株)」という表記が混在しているなどである。データ品質を確保するためには、完全性、適合性、関連性、一貫性、重複度などの観点から課題の内容を正確に把握し、各課題に応じたクレンジングを行う必要がある。しかし、データ品質は高ければ良いというわけではなく、品質を高めればそれに伴いコストも高くなることに留意し、データ品質基準、コストのバランスをとる必要がある。

① データ境界ガードレール

データレイクに格納されたデータには、個人情報や業務上の機密情報など、すべてのユーザに自由にアクセスさせてはならないデータが含まれる場合が多い。このため、データレイクへのアクセスには、適切な権限を適用し管理することが求められる。ここで重要となるのが、ガードレールという概念である。ガードレールとは、企業内のクラウドプラットフォーム活用の際に厳しい制限を設けるという考え方ではなく、自由な利活用を推奨しつつ望ましくない利用のみを制限あるいは検知できるようにし、開発のアジリティを損なうことなくガバナンスをコントロールするという考え方である。ガードレールはセキュリティの基本的な考え方にに基づき、予防的ガードレールと発見的ガードレールの2つに分類される。予防的ガードレールとは組織として実施して欲しくない操作をあらかじめ制限しておくものであり、発見的ガードレールは、予防的なガードレールで完全には対応できないリスクやケースバイケースで判断する必要があるリスクに対して検出し、可視化するものとなる。

データレイク構築で求められることは、信頼できるアイデンティティのみが、期待されるネットワークから信頼できるリソースにアクセスすることを確実にするデータ境界(Data Perimeter)のガードレールである。データ境界を確立するためのガードレール設定は主に3種類挙げられる。

- リソースポリシー・・・アクセスされるリソース側にて、どのような時に、どのリソースに、どのようなアクションを、誰に実行される事が可能かを設定する
- NWエンドポイントポリシー・・・どのサービスにどのNW経由でアクセス許可するかを設定する
- サービス制御ポリシー・・・組織（OU：Organization Unit）単位でどの環境（管理単位）にアクセス可能とするかを設定する

② 管理単位と分散管理の仕組み

ここで、ハイパースケーラークラウド特有の考え方となる“管理単位”について述べておく。クラウドコンピューティングサービスを利用するにあたり、最も基本的な検討項目は「管理単位」と呼ばれ、AWSではアカウント、Azureではサブスクリプション、Google Cloudではプロジェクトと呼ばれる。「管理単位」は、マルチテナント管理を実現する上で不可欠となる仕組みで、ハイパースケーラーの仮想化技術を用いて、IPアドレス体系や仮想ネットワーク設計などのネットワーク、アクセスポリシーなどのセキュリティ、サービスレベルと運用・監視、クラウドプロバイダからの請求単位について完全に分離される仕組みとなっている。この仕組みにより、オンプレミスでは分離できなかった、ひとつのDC内における「IPアドレス体系や仮想ネットワーク設計などのネットワーク、アクセスポリシーなどのセキュリティ、サービスレベルと運用・監視」を完全に分離して管理でき、セキュリティと運用性の向上が期待できる。

企業全体としては1つの管理単位にシステムを複数構築するのではなく、システムごとに異なるサービスレベルやセキュリティ、ガバナンスを混在させないようにする。まず、共通要素となる基礎構成や最低限遵守すべきガバナンスルールを自動的に適用させるランディングゾーンを用いた共通管理の仕組みを構築する。そして、「統合管理・監視」から管理単位を個々のシステムへ払い出し、どのシステムでも共通的に使う管理要素はクラウド管理者で一括で適用し、一定のガバナンスが効いた状態でクラウド利用を実現する（図6参照）。

クラウド上のデータ分散管理に対するセキュリティと運用性能向上

データ活用をすすめるべく日本企業が整備すべき要素の3つ目は、クラウド上での分散管理の仕組みを用いた、セキュアでスケーラブルなデータレイクハウスである。



図6：管理単位とクラウド分散管理の仕組み

③セキュアでスケーラブルなデータレイクハウス

データレイクハウスは、データレイクを実現する高度なストレージサービスを中心としつつ、各種データ分析（DWH、BI、機械学習など）を行うためのサービスを用いて、データ活用を実現する仕組みである。ここで気をつけるべきことは、データ活用者は、データレイク上のデータを直接参照することはほとんどなく、データ分析サービスを用いてデータ活用を行うことである。このため、データレイクハウスを構成する各サービス間でのデータの移動が発生することになる。

- 1つ目は、データレイクに蓄積されたデータを目的に応じた分析サービスに移動する“内から外への移動”
- 2つ目は、分析サービスで生成したデータを一元的にデータレイクで保存するために移動する“外から内への移動”
- 3つ目は、各種分析サービス間で直接データを移動する“サービス間の移動”である

これらの複数サービス間のデータ移動を加味しながら先に述べたデータ境界を実現しつつ、様々な分析を高速に処理していくというセキュアでスケーラブルなデータレイクハウスを実現するための最新のベストプラクティスは、ハイパースケーラークラウドの技術をベースにして、データ境界を実現した上で、オンデマンドで処理能力を変更し「必要な時だけ必要な処理能力を得る」仕組みで実現することである（図7参照）。

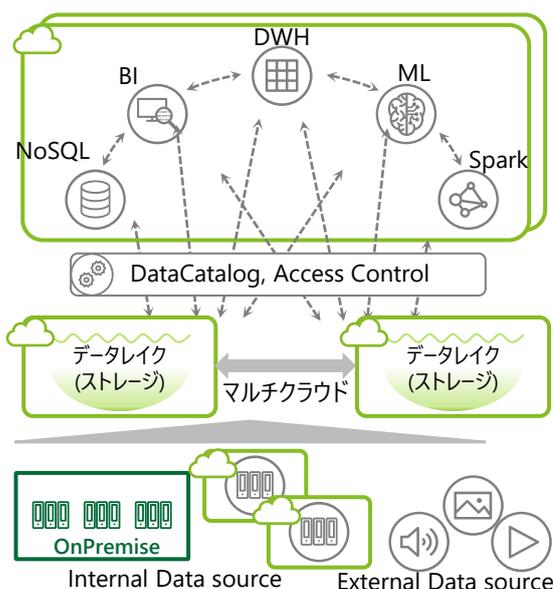


図7：セキュアでスケーラブルなデータレイクハウス

データ境界の実現については、どのデータにアクセスしてよいかを“管理単位”ごとに分け、その“管理単位”の中で分析サービスを構築し、データレイクを保持している“管理単位”へアクセスさせるアーキテクチャとすることである。直接どのユーザがどのレコードにアクセスしてよいかという制御を精密に管理していくことは非現実的であることから、実際にデータを蓄積するデータレイクを保有する“管理単位”に対して、分析サービスを保有する“管理単位”のどのサービスからアクセスを許可するかという考え方でデータ境界を実現することが望ましい。

また、データレイクについては、たとえばグループ企業内や取引先などとデータレイクのデータを相互に活用し合う、という利用形態に発展することが想定される。ケースによっては、AWSのデータレイク上のデータを、Azureで動作している別のデータレイクのデータとして共有するといったことも想定される。この際、オンプレミスのDWHにおいてよく取られがちな手法となる、共有のためにデータを転送し、受け取り側でデータを再ロードする手間をかけることなく、データの相互共有を実現することが望ましいが、データがハイパースケーラークラウド上であれば、各サービスのAPI連携によりデータ移動を実現することが可能となる。このような巨大なデータの相互交換を実施し、そのデータ分析に対する処理性能を上げるためにも、ハイパースケーラークラウド上でのスケーラブルなサービスを用いることが不可欠となる。

データプラットフォームとデータマネジメント体制の整備を両輪として進めることでデータ活用を進める

データマネジメント組織の整備

①データマネジメントの定義（DMBOK2より）

データを「経営戦略を決定する上での重要な資産」と捉え、意思決定のために常時利用可能な状態とするためには、仕組みとしてのデータプラットフォームを整備してデータをただ収集するだけでは、データの価値を発揮することは出来ない。データ本来の資産価値を見出すためには、適切な管理・運用を行う「データマネジメント」が必要となる。DAMA International（国際データマネジメント協会）が発行している「データマネジメント知識体系ガイド 第二版（DMBOK2）」によると、データマネジメントは次のように定義されている。

- データとインフォメーションという資産の価値を提供し、管理し、守り、高めるために、それらのライフサイクルを通して計画、方針、スケジュール、手順などを開発、実施、監督すること。（『データマネジメント知識体系ガイド 第二版（DMBOK2）』, DAMA International, 日経BP社, 2018, p.40）

つまり、データを登録・更新・活用することからそれらの業務を遂行するために必要な活動のすべてを含むものとしてデータマネジメントが定義されている。

DMBOK2では、広範なデータマネジメントにかかる概念を11個の領域に分解して整理されているので、以下にリストアップして紹介する。

1. データガバナンス・・・データマネジメントを統制するための活動
2. データアーキテクチャ・・・データ活用の目的と、データとビジネスの紐づけにかかる戦略と計画策定
3. データモデリングとデザイン・・・データ要件を洗い出し、データ間の関連定義を行う活動
4. データストレージとオペレーション・・・データ基盤について構築・運用する活動
5. データセキュリティ・・・データアクセスに対する権限管理を行う活動
6. データ統合と相互運用性・・・各システムやデータストア間のデータ連携・統合を行う活動
7. ドキュメントとコンテンツ管理・・・各種データのライフサイクル管理を行う活動
8. 参照データとマスターデータ・・・組織横断で一貫したマスターデータを維持管理する活動
9. DWHとBI・・・データ蓄積・分析に必要な技術環境を構築し維持する活動
10. メタデータ管理・・・データソースのメタデータやその業務用途等のカタログの維持管理を行う活動
11. データ品質・・・データ品質統制の標準や仕様を定義し品質レベルの管理を行う活動

②データマネジメント体制

データ活用に向けた取り組みを進めるために特に必要となるのがデータマネジメント体制の整備である。データ基盤があればあとは自発的に業務部門や事業部門のユーザが自発的にデータ活用を進めていく、ということは実際には期待できないため、ITやデジタル部門と、業務部門・事業部門の双方を巻き込んだ、全社横断的なデータマネジメント体制とすることが必要となる。このデータマネジメント体制においては、新しい価値提供やビジネス貢献につながるためのデータ利活用とデータ品質維持を牽引することが求められる。

DMBOK2の11領域のうち、データ基盤の構築や運用にかかるような取組みやデータモデリングなどは、ある程度IT部門中心に進めることはできるかもしれないが、データの意味や中身に関係してくる、マスターデータ、メタデータ（カタログ）、データ品質などについては、IT部門だけで進めようとすると早晚行き詰まり、データの主観部門となる業務部門・事業部門を巻き込んだ取組みとしておくことが不可欠となる。

データマネジメント体制において求められる具体的な検討事項としては、データ活用の目的定義や、事業部門向けのデータ活用シナリオの検討の他、活用対象とするデータの明確化やデータカタログの整備、データ活用に関する管理方針の策定が挙げられる。また、長期的に組織におけるデータ活用の取り組みを浸透させていくために、ポリシーへの準拠状況のモニタリングや、データ利用に関するトレーニングの企画・実施も行っていくことが望ましい。データ基盤を用いてデータ活用を行うことを全社に広げていくには、社員の意識への働き方も求められ、社員の意識や文化の変革に踏み込んだ取組みも必要となる。

このようなデータマネジメント体制については、発足当初はバーチャル組織としてスタートしつつ、いずれは実組織化を考える事が望ましい。

データマネジメント体制については、役割を整備し、チームとして全社のデータ活用を支援・推進していくことが成功の鍵となる。必要となる7つの役割について次に示しておく。

1. **チーフデータオフィサー**：データ活用の戦略策定、意思決定、理解浸透を牽引する
2. **データマネジメントオフィス**：戦略立案や実行、データ活用にかかる社内教育の企画や、関係部署との調整を行う
3. **データスチュワード**：データカタログや品質管理や、部門横断でのマスターデータの管理を行う
4. **データオーナー**：自部門の保有する個々のデータの所有者として、対象データの品質担保を行う
5. **データアナリスト**：業務上の意思決定に向けたデータ分析や活用を行う（セルフBIの利用を含む）
6. **データサイエンティスト**：データを活用した研究／高度なデータ分析や、データアナリストの支援を行う
7. **データプラットフォームアーキテクト**：データ活用にかかる共通基盤の整備を行う

③データ活用プロセスと役割

データ活用プロセスについては、一番最初のプロセスとして「戦略・計画策定」を行った後、「データ基盤の構築」フェーズと、「データ連携・蓄積」フェーズ、そして「データ分析」フェーズに大別できる。それぞれのフェーズでどの役割がどのようなことをするべきかについて、以下に述べておく。

「データ基盤の構築」フェーズでは、全社的なデータ活用戦略や計画策定を行う。この取組みは、①チーフデータオフィサーと、②データマネジメントオフィスを中心を実施するものとなる。並行して、事業部ごとのデータ活用戦略や計画策定を行う場合は、必要に応じて①②のサポートの元、基本的には事業部における④データオーナーが中心となり定義する。

「データ基盤の構築」フェーズでは、データレイクや、分析用のDWHやデータマートの作成を行う。この取組みでは、データ基盤などのプラットフォーム構築は⑦データプラットフォームアーキテクトを中心に進めつつ、どのようなDWHやデータマートが必要かというデータの項目設計については、事業部門側の⑤データアナリストや高度の専門家となる⑥データサイエンティストが検討する。

「データ連携・蓄積」フェーズでは、既存システムからデータレイクへデータ連携を行いつつ、マスターデータの管理、データカタログの登録、連携元のデータの品質向上の取り組みを進める。個々のシステムにおけるマスターデータ管理や品質向上は、原則としては既存システム側にて事業部の④データオーナーにて実施し、全社としてのデータカタログやマスターデータの管理方針、並びに整備がきちんと進んでいるかの管理把握については、③データスチュワードが対応する。

「データ分析」フェーズにおいては、データレイクに集まったデータに対して各種データ分析サービスを用いてデータの抽出や分析を行う。個々のデータ分析の目的に応じてどのようなデータ分析サービスを用いるかについての検討は主に⑤データアナリストや⑥データサイエンティストで実施し、データ境界に基づき必要なデータ分析サービスを整備するのは、②データマネジメントオフィスとともに⑦データプラットフォームアーキテクトにて実施する。⑤データアナリストからの要請に基づいて、どのようなデータがどのシステムに保持されているかやデータレイクのどのエンティティとして保持されているかについてのカタログ情報の提供については、③データスチュワードで実施するものとする。

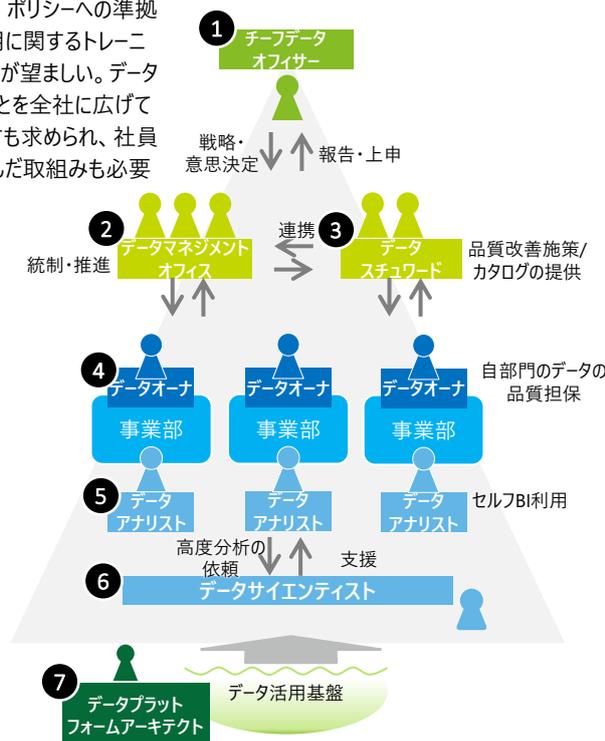


図8：データマネジメント体制イメージ

データ活用を実現するためには、ビジョン、アーキテクチャ、ガバナンスの三位一体となった取り組みが不可欠となる

まとめ

データ活用に向けては、データ基盤があればすべて解決とはならない。データ活用の号令のもとにデータレイクのような器だけを用意した結果、データ活用が全く進まないという事例は枚挙に暇がない。データ活用を実現するためには、全社としてのビジョンに加えて、どのようなデータプラットフォームを構えるかというアーキテクチャと、全体の統制をどう効かせるかというガバナンスの3つの領域が合わさってデータマネジメントを確立することで、その歩を進めることが可能となる。

しかし、ビジョン、アーキテクチャ、ガバナンスの検討を進めるに当たり、多くの企業が陥りがちな課題がある。1つ目は、事業部の視点を取り込まずに事業

部の協業体制が得られないこと。2つ目に、業務や事業部門のデータ分析要件の整理がなされるまで検討が始まらないこと。3つ目は、業務や事業を取り巻く環境変化に加えて、テクノロジー進化の激しいクラウドデータレイクハウステクノロジーを用いる事が不可欠の中で、検討そのものから構築まで外部のSierに丸投げしてしまうこと。4つ目に、ビジネス、アプリ、データ、クラウドプラットフォームにまたがる幅広い知見を持ち、全体のビジョン、アーキテクチャ、ガバナンスの三位一体の整合を取りながら進められる有識者が社内不在であることなどがあげられる。

デロイトトーマツでは、日本企業がこれらの課題に陥ることの無いよう、構想策定から、データレイクハウスの構築、企業内にこれらを手の内化するためのデータマネジメント体制の組成まで、End to Endで支援するサービスを提供している。最初の6ヶ月で将来に向けたビジョンの明確化と、アーキテクチャとガバナンスの青写真を描き、実行計画まで落とし込みを行い、実行フェーズについても成熟度と進捗度に応じて柔軟に伴走をしていくことが可能である。

特に、効果的にデータ活用を進めるためには、「構想初期フェーズからの事業部の巻き込み」、「プラットフォームファーストでスモールスタートしながらデータ基盤の整備」、「データ活用にかかるアーキテクチャとガバナンスの手の内化」、「全体目線を持ったアーキテクトの参画」が不可欠となる。これらの4つのアプローチを取り入れることで、絵に描いた餅に陥らない、実現性のあるデータ活用へ導くことを可能とする。

全ての検討を自組織のリソースだけで行うことは現実的ではなく、外部リソースの活用も視野に入れて検討することも良いと考えられる。外部の支援サービスを用いつつ、それらを手の内化することを見据え、外部業者に丸投げするのではなく、伴走してもらう中でそのノウハウを自組織に貯めながら、推進していくことが重要である。迅速性・柔軟性を兼ね備えたアーキテクチャを手にし、デジタル時代を勝ち抜くために、必要な取り組みを推進していただきたい。

プロフェッショナル



佐藤 岳彦
Technology Strategy & Transformation
執行役員 マネージングディレクター

外資コンサルティングファームを経て現職。
官公庁、金融、製造業を中心に、IT構想策定、全社IT/dXアーキテクチャ策定、大規模ITプロジェクトのマネジメント等、テクノロジーコンサルタントとしてクライアントの変革を支援。
全社アーキテクチャ、クラウド、セキュリティに関するエキスパート。

メンバー

秋田 修吾
Technology Strategy & Transformation
シニアマネジャー

佐藤 佑哉
Technology Strategy & Transformation
シニアマネジャー

土田 泰徳
Technology Strategy & Transformation
シニアマネジャー

南野 香澄
Technology Strategy & Transformation
シニアマネジャー

関根 淳
Technology Strategy & Transformation
マネジャー

白石 智一
Technology Strategy & Transformation
マネジャー

須藤 正人
Technology Strategy & Transformation
マネジャー

※デジタルトランスフォーメーションを指す用語として、デロイト トーマツ グループでは、デジタルを導入することを主目的とした変革(DX)ではなく、デジタルを道具として駆使し、ビジネス自体を根本的に変革していくBusiness Transformation with digital = “dX”であると定義しています。

Deloitte.

デロイトトーマツ

デロイトトーマツグループは、日本におけるデロイト アジア パシフィック リミテッドおよびデロイトネットワークのメンバーであるデロイトトーマツ合同会社ならびにそのグループ法人（有限責任監査法人トーマツ、デロイトトーマツリスクアドバイザー合同会社、デロイトトーマツコンサルティング合同会社、デロイトトーマツファイナンシャルアドバイザー合同会社、デロイトトーマツ税理士法人、DT弁護士法人およびデロイトトーマツグループ合同会社を含む）の総称です。デロイトトーマツグループは、日本で最大級のプロフェッショナルグループのひとつであり、各法人がそれぞれの適用法令に従い、監査・保証業務、リスクアドバイザー、コンサルティング、ファイナンシャルアドバイザー、税務、法務等を提供しています。また、国内約30都市に約2万人の専門家を擁し、多国籍企業や主要な日本企業をクライアントとしています。詳細はデロイトトーマツグループWebサイト、www.deloitte.com/jpをご覧ください。

Deloitte（デロイト）とは、デロイトトウシュトーマツリミテッド（“DTTL”）、そのグローバルネットワーク組織を構成するメンバーファームおよびそれらの関係法人（総称して“デロイトネットワーク”）のひとつまたは複数を指します。DTTL（または“Deloitte Global”）ならびに各メンバーファームおよび関係法人はそれぞれ法的に独立した別個の組織体であり、第三者に関して相互に義務を課しまたは拘束させることはありません。DTTLおよびDTTLの各メンバーファームならびに関係法人は、自らの作為および不作為についてのみ責任を負い、互いに他のファームまたは関係法人の作為および不作為について責任を負うものではありません。DTTLはクライアントへのサービス提供を行いません。詳細は www.deloitte.com/jp/about をご覧ください。デロイト アジア パシフィック リミテッドはDTTLのメンバーファームであり、保証有限責任会社です。デロイト アジア パシフィック リミテッドのメンバーおよびそれらの関係法人は、それぞれ法的に独立した別個の組織体であり、アジアパシフィックにおける100を超える都市（オーストラランド、バンコク、北京、ベンガルール、ハノイ、香港、ジャカルタ、クアラルンプール、マニラ、メルボルン、ムンバイ、ニューデリー、大阪、ソウル、上海、シンガポール、シドニー、台北、東京を含む）にてサービスを提供しています。

Deloitte（デロイト）は、監査・保証業務、コンサルティング、ファイナンシャルアドバイザー、リスクアドバイザー、税務・法務などに関連する最先端のサービスを、Fortune Global 500®の約9割の企業や多数のプライベート（非公開）企業を含むクライアントに提供しています。デロイトは、資本市場に対する社会的な信頼を高め、クライアントの変革と繁栄を促し、より豊かな経済、公正な社会、持続可能な世界の実現に向けて自ら率先して取り組むことを通じて、計測可能で継続性のある成果をもたらすプロフェッショナルの集団です。デロイトは、創設以来175年余りの歴史を有し、150を超える国・地域にわたって活動を展開しています。“Making an impact that matters”をパーパス（存在理由）として標榜するデロイトの45万人超の人材の活動の詳細については、www.deloitte.com をご覧ください。

本資料は皆様への情報提供として一般的な情報を掲載するのみであり、デロイトトウシュトーマツリミテッド（“DTTL”）、そのグローバルネットワーク組織を構成するメンバーファームおよびそれらの関係法人が本資料をもって専門的な助言やサービスを提供するものではありません。皆様の財務または事業に影響を与えるような意思決定または行動をされる前に、適切な専門家にご相談ください。本資料における情報の正確性や完全性に関して、いかなる表明、保証または確約（明示・黙示を問いません）をするものではありません。またDTTL、そのメンバーファーム、関係法人、社員・職員または代理人のいずれも、本資料に依拠した人に関して直接または間接に発生したいかなる損失および損害に対して責任を負いません。DTTLならびに各メンバーファームおよび関係法人はそれぞれ法的に独立した別個の組織体です。

Member of
Deloitte Touche Tohmatsu Limited

© 2024. For information, contact Deloitte Tohmatsu Group.



IS 669126 / ISO 27001



BCMS 764479 / ISO 22301

IS/BCMSそれぞれの認証範囲はこちらをご覧ください
<http://www.bsigroup.com/clientDirectory>