



データレイクに関する5つの誤解

企業がデータの価値や重要性、およびデータから新たなチャンスや収益モデルが得られるという認識を高める中、すべてのデータを管理し、洞察を推し進める「最適な」手法を実現するテクノロジーがあふれています。

データレイクは、そのようなリファレンス・アーキテクチャーの1つとして注目されています。とはいえ、データレイクには誤解がいくつかあり、企業やIT管理者が選択肢を調べる際に、貴重な時間を無駄にしてしまう要因となっています。

1.

データレイクは単一のクラウドにしか導入できない → 誤解

データレイクは単一のロケーションやクラウドに限定されません。オンプレミス・デプロイメントにも制限されません。ハイブリッド・オプションであれば、複数のクラウドにわたってデータレイクを構築できます。データレイクは、テクノロジーから独立したリファレンス・アーキテクチャーです。これは、データを企業運営の中心として据え、大規模で多様な種類のデータへのアクセスを容易にし、ユーザーをセルフサービス分析で支援するためのアプローチです。

2.

Hadoopは唯一のデータレイクである → 誤解

データレイクという用語は、とかくHadoopやHadoop指向のオブジェクト・ストレージへと結び付けられがちですが、Hadoopを組み込まずに、データレイクを効果的に開発し、使用することもできます。例えば、異なるリレーショナル・データベース管理システムを基盤として、効果的なデータレイクを構築することも可能です。データレイクは多様なテクノロジーを組み合わせ、洞察獲得システム (SoI) を確立し、データサイエンティストがビジネス・ニーズに対処するために、俊敏にデータを調査できるようにします。

3.

データレイクにはあらゆるデータを放りこんでおけばよい。

ガバナンスは必要ない → 誤解

ソフトウェアとハードウェアはデータレイク・ソリューションの重要な構成要素ですが、同じく重要なのが、データのカタログ化、データの品質、およびデータのガバナンスと管理プロセスです。

データ・ウェアハウスによっては、大量のデータが閉じこめられた巨大なブラック・ホールになっていたりします。このように、適切なガバナンス・ポリシーを適用しないと、データレイクがデータの湖ならぬ「データの泥沼」になってしまうおそれがあります。

4.

データレイクの成功はアクセスの提供で測定される → 誤解

誰もがアクセスできる場所にデータを放りこむだけでは、真の意味での分析ソリューションとは言えません。分析ソリューションの目的は、納得のいくビジネス洞察へとつながるデータ分析を行い、新たな収益源、顧客維持モデル、あるいは製品拡大を見出すことにあります。

しかし、そのデータは信頼できるものであり、適切で、しかも、データの消費者すべてが利用できるものでなければなりません。データレイクには、暗号化されたデータをビジネスの場で理解されやすいビジネス用語に変換することができる、インテリジェントなメタデータ・カタログが必要です。このカタログはまた、構造化および非構造化情報資産の双方から得られるデータの出所を明確にして質を高め、ガバナンス・ファブリックを増強して、情報の保護、標準化、効率的な管理、信頼性を確保するのに役立ちます。

5.

データレイクはデータ・ウェアハウスの代わりである → 誤解

データレイクは、複数のエンタープライズ・データ・ウェアハウス(EDW)に加え、ソーシャル・メディアやIoTといった他のデータ・ソースも取り込みます。これらはすべてデータレイクに集約されて、ガバナンスを組み込めるので、社内のユーザーは誰でも、信頼できるデータを容易に見つけ出すことができます。

したがって、データレイクはEDW環境を補強し、データサイエンティストやアナリストがデータを容易に探索し、新たな視点、洞察、機会を発見し、イノベーションやビジネスの成長を加速させることを可能にします。



マルチクラウド・データレイク のメリット

間違った情報に翻弄されないください。ガバナンスが適用されたデータレイクにより、全社で幅広い構造化データおよび非構造化データにアクセスできると同時に、場所に関わらず、信頼性と安全性を確保できます。

ガバナンスが適用されたデータレイクは、ビジネス・ニーズに合わせて最適化すると、分析を加速させ、洞察の精度を高めることができます。その理由は以下のとおりです。



データは安全で信頼性の高いインフラストラクチャー基盤上にあります。



管理されたデータ・フィードは、データレイクに信頼できる情報を投入し、情報資産、そのメタデータ、ビジネス・コンテキストを文書化し、データレイクにリアルタイムでデータを流し込みます。



データの質、出所、系統が明確です。



データはビジネス用語で保存されているので、データサイエンティストは暗号化された用語の解読に時間を取られることなく、速やかに仕事に取り掛かることができます。



データは適切に分類・保護された上、ガバナンスが適用されています。

真実はここに

IBMのデータ管理ポートフォリオにより、お客様は、データレイクについてのよくある落とし穴や誤解を回避し、ガバナンスが適用されたデータレイクの導入を成功させるための重要な手順を、迷わずに踏んでゆくことができます。

IBMデータレイク・ソリューションを確認する—より多くのデータ・ソースから集められた様々な種類のデータを活用することで、よりスマートな意思決定を後押しします。



資料を読む

ガバナンスが適用されたデータレイクがどのようにして、重要なビジネス洞察を導き出すきっかけとなるかを探ります。

