

5 テクノロジーコンサルティング事業本部：データマネジメントプラットフォームグループ

# データ+AIの民主化を促進する データ分析プラットフォーム「データブリックス」

株式会社 NTT データ（以下、NTT データ）は、2021 年 5 月に Databricks Inc.（読み：データブリックス、本社：米カリフォルニア州）とソリューションパートナー契約を締結した。データブリックスは、「データ+AI の民主化」を目的とし、データエンジニアリング、機械学習、アナリティクスなどのあらゆるデータにまつわるユースケースに対応したオールインワンデータ分析プラットフォームをマルチクラウドで世界 9,000 社以上の企業に提供している。

## 「データ+AI の民主化」の課題を解決するデータブリックス

現在、多くの企業が DX を目指し、データの利活用に取り組んでおり、データからインサイトを得ることを目的に、DWH と BI により、データの可視化を行っている。しかし、可視化したデータから人間が物事を判断するには限界がある。そこで、世界の先進的な企業はデータから AI を作り、未来を予測し、高度な意思決定の実現や新しい商品サービスを生み出している（図 1）。

DX 白書 2023 によると、米国では AI などを活用した高度なデータ分析は

当然の取り組みになっているが、日本は一部の先進的な企業だけが取り組んでいる状況であり、より AI を活用した高度なデータ利活用が求められる。

AI の開発は DWH ではなく、データレイクと呼ばれるビッグデータを元のままのさまざまな形式で保存するプラットフォームで行う。そうすると、企業は DWH とデータレイクの 2 つのプラットフォームを運用する必要があるが、分散するデータの管理負荷やシステムコストが増加する問題が発生する。この問題を解決するのが、データブリックスが提唱する「レイクハウス」である。レイクハウスは、データエンジニアリン



株式会社 NTT データ  
テクノロジーコンサルティング事業本部  
デジタルテクノロジー&データマネジメントユニット  
データマネジメントプラットフォームグループ  
テクニカル・グレード 齋藤 祐希氏

グ、AI/ 機械学習、BI/ アナリティクスなどのさまざまなユースケースを 1 つのプラットフォームでカバーする。NTT データは、データブリックスをデータ活用領域の中核テクノロジーと位置付け、お客様の「データ+AI の民主化」の推進を支援する。

## モダン・データアーキテクチャ「レイクハウス」

レイクハウスは DWH とデータレイクの良いとこ取りがコンセプトの新時代のデータアーキテクチャである（図 2）。以下に、データブリックスのレイクハウス・プラットフォーム（以下、レイクハウス・プラットフォーム）の主な特徴を挙げる。

- ① ペタバイト級のデータや大量の機

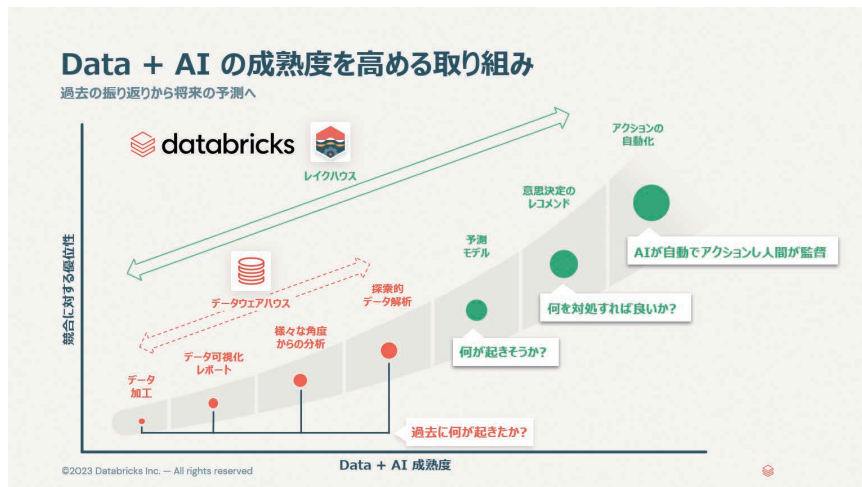


図 1 データや AI を活用していく上での成熟度

機械学習モデルをリアルタイムで処理が可能

・並列分散処理による高パフォーマンスな ETL (Extract Transform Load: 抽出 / 変換 / 格納) と機械学習の実行

・Time to Value/Market (価値の実現 / 市場への展開のリードタイム) の高速化

### ② データ管理のシンプル化

・非構造化 / 半構造化データに対し、構造化データライクに、信頼性向上や意味 (タグ) 付けが可能

・ストリーミングデータを、リアルタイムでバッチデータと統合して処理が可能

### ③ 機械学習ライフサイクルのシンプル化

・機械学習の実験から本番活用までのライフサイクルを一元的に管理可能

・単一の機械学習モデルのレポジトリ機能により、全モデルのホワイトボックス化が実現しデータサイエンティスト間のモデルの共有 / 再利用が実現

### ④ データチーム内の共同作業の効率化

・コメントやチャット機能などのモ

ダンなコミュニケーション機能や、各データやモデルのバージョン履歴管理などをブラウザーベースの Notebook 上で提供

### ⑤ SQL アナリティクス

・BI ツールから直接データレイクにアクセスし、ダッシュボードを作成することも可能

### ⑥ マルチクラウド対応で簡単にデータ分析環境を構築することが可能

・Microsoft Azure、Amazon Web Service、Google Cloud Platform をサポート

・データ分析基盤を数クリックで構築可能

## レイクハウス・プラットフォームのオープン性

レイクハウス・プラットフォームの特徴として「オープン」という言葉がよく使われ、いくつかのポイントがある。

まずデータブリックスは Apache Spark の生みの親が創立した会社であり、そのプラットフォームの根幹には複数の OSS が使われている。つまり、ベンダーロックインが少なく、広く一般的に使われているテク

ノロジーを利用できる。

次に、データのオープン性がある。レイクハウス・プラットフォームを利用するにはデータをデータブリックスに送る必要はなく、データはお客様のクラウド環境に格納したまま利用できる。データは資産と言われる時代で、資産を自社内に保持できる点は大きなメリットにつながる。

最後に、オープンソース AI モデルである。データブリックスは 2023 年 3 月に、ChatGPT に代表される大規模言語モデルの AI モデル Dolly を公開した。ChatGPT の魔法のような体験をより身近に感じ、AI の民主化を促進できる。

このようにデータブリックスはデータプラットフォームの域を超え、進化し続けている。

## データブリックスで誰もが AI を活用する世界へ

データブリックスの料金体系は従量課金であり、スモールスタートで開始できる。また、業界別の主要な分析ユースケースのテンプレート (ソリューションアクセラレータ) を活用することで、分析自体もクイックに開始できる。

上記に加えて、NTT データには、データブリックスに精通したエバンジェリストが日本最多数在籍しており、金融機関を中心とした導入実績から作り上げたりファレンスアーキテクチャを活用し、セキュアかつ全社レベルまで拡張可能なデータ分析基盤を構築するサービスを提供している。

NTT データはデータブリックスを活用し、GAFA などのビッグ・テックだけでなく全ての企業が当たり前のように AI を活用する世界を目指す。



図 2 レイクハウス・プラットフォーム