

東京海上日動のデータドリブンな意思決定を推進

## データガバナンス強化や機械学習など 様々なユースケースを ワンプラットフォームで実現する データブリックス



TOKIO MARINE  
NICHIDO

### 東京海上日動

#### 導入前の課題

- 大量のデータ分析をする機会が増え、既存の分析基盤では手に負えないケースが出てきた
- 複数のサービスを使いこなす必要があるため、利用に一定のハードルがある
- ガバナンスやセキュリティを担保しつつ、ユーザーがさまざまなユースケースに対して柔軟に対応できるプラットフォームへの移行が必要

#### データブリックスを活用した解決策

- ユーザビリティの向上でデータ処理や分析の生産性が向上
- 複数の機能をワンプラットフォームで実現
- 多くのユーザーが膨大なデータを柔軟に扱えるアクセス制御を実現。新しいユースケースや領域への活用を拡大

#### 超大量のデータ分析にも対応するインフラの伸縮性や、データガバナンス強化が必要に

1879年に日本初の保険会社として誕生した東京海上保険と日動火災保険が2004年に合併し、大手損害保険会社の1社として損害保険業界を強力に牽引する東京海上日動火災保険株式会社。自動車保険保有契約件数は1400万件を超える規模を誇っており、東京海上グループとして国内損害保険事業とともに、国内生命保険事業や海外保険事業、金融事業などグローバルにビジネスを展開している。

そんな同社において、IT戦略や企画立案などを行っているのがIT企画部だ。東京海上日動システムズ株式会社とともに、グループ全体のビジネスをITの面から強力に下支えしており、データ活用においては、2017年ごろからデータドリブンな意思決定を推進するデータ分析基盤の整備にいち早く取り組んできた。「損害保険事業に関連したことから、事故の未然防止につながるサービス提供はもちろん、防災やヘルスケアなど新たな領域への展開にデータを積極的に活用しています」とIT企画部 企画グループ 課長 石井 尚行氏は説明する。

そんなデータ活用に欠かせないデータ分析基盤は、IaaS上に環境を構築するところから始まり、各種マネージドサービスを組み合わせる構成へと進歩を遂げた。しかし、外部やグループ会社との協業が増加し、高度な分析ユースケースが増える中、それらに対応することに加えて、データガバナンス強化をはじめとした課題への対応



東京海上日動火災保険株式会社  
IT 企画部  
企画グループ  
課長  
石井 尚行氏



東京海上日動システムズ株式会社  
システムイノベーション推進本部  
システムイノベーション推進部  
課長代理  
久保 太樹氏



東京海上日動システムズ株式会社  
デジタルイノベーション本部  
データ活用部  
スペシャリスト  
岡野 翔一氏

も求められていた。「1回の分析で5000億レコードというこれまで扱ったことのない超大量のデータ分析が外部との協業で求められる場面もあり、既存の環境では手に負えない状況でした」と語るのは、東京海上日動システムズ株式会社 デジタルイノベーション本部 データ活用部 スペシャリスト 岡野 翔一氏だ。

インフラ面では、連携するツールのライセンス調達に時間がかかる他、複数のサービスを使いこなす必要があるため、利用する際に一定のハードルがあるなど、別の問題が顕在化した。「これまで以上にガバナンスやセキュリティを担保しつつ、BIだけでなくさまざまなユースケースに対して柔軟にスケールでき、ユーザーにとって使いやすいプラットフォームへの移行が望まれていたのです」と東京海上日動システムズ株式会社 システムイノベーション推進本部 システムイノベーション推進部 課長代理 久保 太樹氏は語る。

### ワンプラットフォームで環境が整備できる データブリックスを高く評価

そこで、既存で利用していたHadoopの代替を前提に、数年先まで十分戦えるデータ分析基盤となる次世代プラットフォームの検討が進められた。同社ではプラットフォームの選定にあたり、「データ量の増加や非定型データといった多様なデータへの対応」や「ユーザー数増加への対応」、「開発したAIの業務システムへの組み込み・継続的改善の向上」、「ガバナンス強化」、「コスト最適化」を重点的に評価することにした。

そこで、クラウド型DWH製品などを候補にPoCを実施した結果、同社が注目したのがデータブリックスだった。データレイクハウスとしてデータ加工からBI、AI/ML含めたユースケースが1つのプ

ラットフォームで実現でき、外部協業の拡大を支えるガバナンスの仕組みを備えた最適なアーキテクチャを高く評価したという。

「Databricks Notebookはデータサイエンティストが使うJupyter Notebookを拡張した機能であり、既存ユーザーも抵抗なく利用できます。また、大規模な分析には大きめのクラスタをデータサイエンティスト(ユーザー)自身が立ち上げることができ、データの処理も非常に高速です。これらにより、データ処理や分析の生産性が向上し、ユーザーにとっても使い勝手が向上します」と岡野氏。石井氏も「複数サービスを組み合わせると管理が複雑化し、ガバナンスを効かせることが難しくなる。複数の機能をワンプラットフォームで実現できるのはガバナンスの観点からも大きい」と評価する。また、「コスト面においても既存環境と比べると60%ほどの削減につながり、比較した他のソリューションと比べても2~30%のコストメリットが出ると試算しました」と久保氏。

結果として、既存の課題を大きく解消する次世代のデータ分析基盤として、データブリックスが提供する統合データ分析基盤「レイクハウス・プラットフォーム」が採用されることになった。

### 約750名が活用するデータ分析基盤の中核となる レイクハウス・プラットフォーム

現在は、レイクハウス・プラットフォームを中心にデータ分析基盤が整備されており、分析に必要なデータを業務システムや外部システムからクラウドストレージに蓄積し、分析用のDWHから目的ごとのデータマートレイヤーにデータを加工し展開、TableauやDatabricks Notebookなどを活用してデータ可視化やAIモデル構築を行っている。蓄積されているデータ量は530TB(圧縮

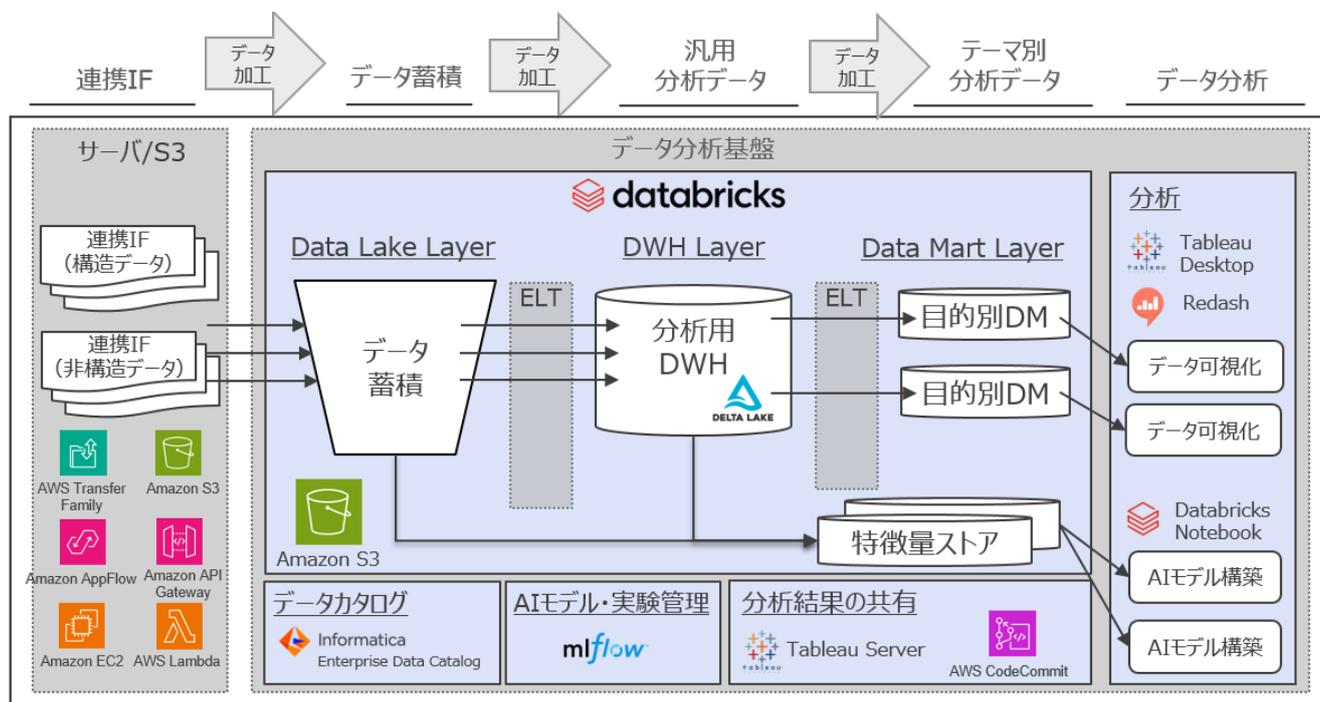


図:本事例のベースとなった Trusted Data Foundation\* powered by Databricks のリファレンスアーキテクチャ

\* Trusted Data FoundationはNTTデータが提供するデータ分析基盤構築サービスです。

後)を超えており、社内外合わせて750名ほどのメンバーがデータ分析基盤を活用、案件ごとに作られた分析ラボは40を超えている。このように、同社では多くのユーザーが膨大なデータをさまざまな目的で活用するための柔軟なアクセス制御を行っている。

活用領域については、損害保険に関連した保険料算出などの分析はもちろん、建設機械などに搭載されたドライブレコーダーから危険な挙動を判断してアラートを出すための事故防止支援や、災害時の避難指示や被害に関連したシミュレーションなどにも利用されている。

また、近年は画像や動画、音声、テキストなどの非定型データの活用にも積極的だ。例えば、損害サービス業務に関して音声マイニング技術を活用して通話内容を分析するAIを自社開発している。なお、現在は従来環境も並行稼働させているが、2023年度には全てのユーザーが活用可能になり、全ての切り替えを来年度中に実施する計画だ。

### 大量データの高速処理とセルフサービス化を実現、データの民主化に大きく貢献

レイクハウス・プラットフォームでの環境では、従来の環境に比べて大量データの処理が高速化されているという。「大量データを加工する場面で、以前は5時間以上かかっていたETL処理がわずか3分で実現できた例もある。Sparkの得意とするインメモリ処理が効果を発揮しています」と久保氏は評価する。

また、以前は分析に必要なリソースを増やすまでに早くても1~2日ほどは必要だったが、今は分析者自身がリソース変更を行えるようになった結果、わずか数分でスケール調整ができるようになっている。「性能が既存と段違いなため、革新的な環境だというコメントが現場から寄せられています。超大規模なデータをユーザーが自分の手で扱えるようになったことで、理想的なデータの民主化に近づいています」と岡野氏は力説する。

今回データブリックスの提案と導入支援を手掛けたNTTデータについては、「データ分析のパートナーとして我々のビジネスに長年寄り添っていただいています。データブリックスに対するNTTデータの存在感は大きく、データ分析の分野で今後もご支援いただきたい」と石井氏。NTTデータにはDatabricks Inc.の年次イベントにて世界で3名が選出されたAPJ Partner Champion of the Yearの1人が在籍するなど、グローバルでの存在感を示しており、高く評価されている。現行プラットフォームの抱える課題や中長期経営戦略を踏まえた将来像を理解し、最もフィットするソリューションとしてデータブリックスを提案した。また、PoC段階では顕在化した課題に対してデータブリックスを巻き込みつつ、より

幅広いテクノロジーの知見を生かした代替提案によってデータブリックスと3社で課題解決を実現するなど、高度な技術力と先見性についても高い評価を得ているという。

### 現場への利用促進とともに、外部との連携強化を進めたい

今後については、さまざまなクラウド上のファイルやテーブル、MLモデルのガバナンスを簡素化するUnity Catalogについて期待を寄せているという。「米国イベントでUnity Catalogの強化が示されたことで、データシェアリングなど外部協業なども加速していくことを想定しています。従来のデータリネージだけではなく、外部のデータソースやMLモデルを含めたEnd to Endのリネージ情報が取得できることで、我々が求めるガバナンスレベルにも沿った形で利用できるようになるはず」と久保氏。

また、さらなるデータの民主化を進める意味でも、ハンズオン研修など現場への利用促進を進めていきたいと岡野氏。「ユーザーにどれだけ使ってもらえるかに注力したいと考えています。また、分析して終わるのではなく、次に繋げていくためにも、分析したモデルをデプロイして既存システムから使えるような環境づくりに向けて、今後もサポートいただきたい」。

さらに、今回インフラとして先行実装できたデータ分析基盤を活かし、外部との連携強化を進めていきたいと力説する。「データブリックスを持っていることで、必要なデータをセキュアな形で提供していくことができることは、我々にとって極めて大きな武器です。外部パートナーと一緒にデータ分析をさらに進めていくことで新たな価値創造へ繋げていきたい」と石井氏に今後について語っていただいた。



(本資料に掲載されている社名および製品名は各社の商標および登録商標です)

株式会社NTTデータ

〒135-6033 東京都江東区豊洲3-3-3 豊洲センタービル  
Tel: 03-5546-8051 Fax: 03-5546-2405  
<https://www.nttdata.com/jp/ja/>

