

# 購買ビッグデータの高速処理、データガバナンス、リアルタイム処理、少ない運用工数などの観点で、「レイクハウス・プラットフォーム」を採用

## TOSHIBA

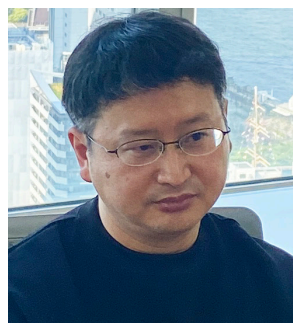
### データを価値ある形に変えてよりよい未来を、ともに創り続ける

東芝データは、「データを価値ある形に変えてよりよい未来を、ともに創り続ける」というパーパスを掲げ、データのカで社会課題を解決することを目指した事業を行っている。実社会で日々生まれるデータを価値ある形に変えて社会に還元することで、豊かな未来を創造するデータ循環型のエコシステムの構築に取り組んでいる。

### 電子レシートサービス「スマートレシート®」の活用推進

主要事業の1つがグループ企業である東芝テックが開発・運営し、東芝データが運営を支援している電子レシートサービス「スマートレシート®」のレシートデータ活用である。店舗に設置されたPOSレジが生成する購買データを、「スマートレシート®」会員本人の同意に基づき東芝データが統計処理し、メディア・広告代理店向けを中心に、購買動向把握や広告効果の検証といった様々なデータサービスへ活用する。同会員は、店舗会員/ポイント/地域ウォレットアプリなどと連携しているため、各アプリの利便性が上がる一方、メディアや広告代理店は、広告放映後の集客・実売への効果測定の可視化が高精度で可能となる。収集/分析されるデータは、会員が保有、リアルタイム取得、流通チャネル/決済手段横断、商品明細情報、購買証明といった5つの特徴がある。

「スマートレシート®」のデータ活用のユースケースは幅広く、異業種商品の売れ筋データを活用した新商品開発への活用や、消費に関する家計や経済のリアルタイムな状況把握としてオルタナティブデータの提供への寄与、ID連携により購入した食材からレシピを提案する新サービスの提供に加え、地方自治体と連携した交付金事業やポイント管理事業などが挙げられる。これらはPOSシステムのリーディングカンパニーである東芝テックのアセット基盤を中心に用いるため、集まる購買データは加盟店の業種・業態を超えた幅広いものとなる。



技術部フェロー 兼 技術第二グループ長  
博士(工学)  
城田 祐介 氏



技術部  
若松 友裕 氏

### レイクハウス・プラットフォームの選択肢がなければ、データ分析環境をスムーズに構築することは難しかった

東芝データ設立にあわせて、事業を支える社内データ分析基盤の構築が急務であり、いくつかの要件が存在した。1つ目は、半構造や非構造データを含めたビックデータを高速かつスケーラブルに処理することが可能であること。2つ目は、データを管理している自社システム環境からデータを移動させることなく分析機能を導入できること。3つ目は、少ないリソースで分析環境の構築/運用がまわせること。これらの点を総合的に検討した結果、データブリックスのレイクハウス・プラットフォームの採用を決めた。

レイクハウス・プラットフォームの主要コンポーネントであるSparkは、ビックデータを高速に処理可能で、処理の規模に応じて自動でスケールが可能である。また、アーキテクチャ上の特徴としてコントロールプレーンとデータプレーンが分離しており、データブリックス環境でマネージされたコントロールの下、自社環境に閉じたデータプレーン内でデータ管理してデータ処理を実行する。そのため、データのセ

セキュリティやプライバシーに配慮したシステム構築がしやすくなる。加えて、データブリックスはSaaS型のマネージドサービスであるため、分析環境のインフラ自体は容易に設定が可能だ。クラウドベンダーのネイティブサービスを組み合わせることで機能的には同様の分析環境は構築可能であるが、限られた期間と人的リソースで各サービスを個別に組み上げて運用するのは現実的ではないと判断した。改めて振り返ると、データブリックスのレイクハウス・プラットフォームの選択肢がなかったと仮定すると、現状のデータ分析環境をスムーズに構築することは非常に難しかったと感じている。また、会社の指針としてベンダーロックインを避ける必要もあったため、レイクハウスの主要コンポーネントが全てオープン(OSS)である点も評価が高かった。データブリックスの活用範囲は、基盤データのETLや各種データ分析にと

どまらず、東芝データの全社員がKPIデータを共有する経営ダッシュボードのような用途にも拡大している。

今後に向けては、鮮度の高いデータをリアルタイムで分析するストリーミングパイプライン構築への「Delta Live Tables」の活用や、データガバナンスを強化するための「Unity Catalog」の導入といった新機能の採用を進めている。また、購買データのドメイン知識を活用した高度なデータ分析を推進するために、データブリックス上でAI/機械学習や自然言語処理技術の活用にも挑戦している。

#### アーキテクチャー図

