

データ分析とマーケティング業務を一気通貫に実行できる「ソフトバンクデータ基盤」の再構築を目指す

SoftBank

通信事業というコアビジネスに加え、金融、コンテンツなど事業を多角化

ソフトバンク株式会社は、ソフトバンクグループ株式会社の傘下の企業で、移動通信および固定通信サービスやインターネット接続サービスの提供に加え、携帯端末の販売事業などを展開している通信事業者である。昨今では、金融事業など様々な事業ドメインに進出しているコングロマリットである。通信事業というコアビジネスに加えて、ヤフーやLINEを中心としたコンテンツビジネス、PayPayなど新領域のビジネスの3つの事業の柱がある。

データ戦略部は、コアとなる通信事業を中心として、全社横断でデータを提供している。全国で展開しているソフトバンクショップやユーザー向けのコールセンター、また法人営業ビジネスから獲得するデータなどを、一元的に管理・運用を行っている。同部門のミッションは、社員へデータ分析環境を提供することによる業務効率化、経営層へのリアルタイムの情報提供による迅速な意思決定支援、そして精緻な顧客プロファイルデータに基づくお客様への最適な提案による顧客体験の向上の3つである。これらのミッションを達成するために、「ハイブリッド戦略」をデータ基盤のアーキテクチャーとして採用している。具体的には、オンプレミスの環境に通信の秘密に関するセンシティブな情報をホストし、クラウド上にはそれ以外のデータをホストしている。前者には高度な機密性、後者には拡張性が重要視される形である。

圧倒的な処理パフォーマンスと拡張性の高さからデータブリックスを採用

2019年にデータレイク処理基盤と分析処理基盤を目的として、データブリックスの採用を決めた。採用にあたっては、複数の選択肢を比較検討したが、Sparkを活用した圧倒的な処理パフォーマンス、従来のHadoopベースの資産をそのまま有効活用できる点、そして容易にスケール可能な拡張性の高さの3つの理由でデータブリックスが最も優れていると判断した。

データレイク処理基盤においては、日々1,600以上のインターフェースからデータを受け取り、生データの加工やデータマートの作成処理をデータブリックス上で行っている。具体的には、2,000以上のジョブ



ソフトバンク株式会社
テクノロジーユニット
コーポレートIT本部
データ戦略部 部長
安芸 洋一 氏

の管理と共に、1時間当たり約25億のレコードを処理し、後工程の分析ツールやBIツールへと分析結果をフィードしている。

分析処理基盤は、主にデータサイエンティスト向けに提供している基盤である。データブリックスの活用している機能は前述のデータレイク処理基盤と同様であるが、アウトプットが分析モデルという点が異なっている。分析モデルの具体例としては、モバイルユーザーの解約抑止向けのモデルや、顧客満足度の指標の一つであるNPS (Net Promoter Score)に関連したモデルが挙げられる。

現在は、データ分析とマーケティング業務を一気通貫に実行できる「ソフトバンクデータ基盤」の再構築を行っている。環境としては、IT領域とユーザー領域として物理的にも論理的にも分けており、IT領域では3つの機能拡張/処理改善を実施している。1つ目は、データブリックスが提供するDelta Lake (デルタ・レイク)を活用することにより、クエリ性能が50%~70%ほど改善され、生産性の向上が実現している。2つ目は、ガバナンスの観点で、Overwatchを導入し、ジョブトラブル解決の効率化につながっている。3つ目は、ETL処理のデータブリックスへの一本化である。これにより、データパイプラインの交通整理がなされ、結果として障害ポイントの極小化につながっている。

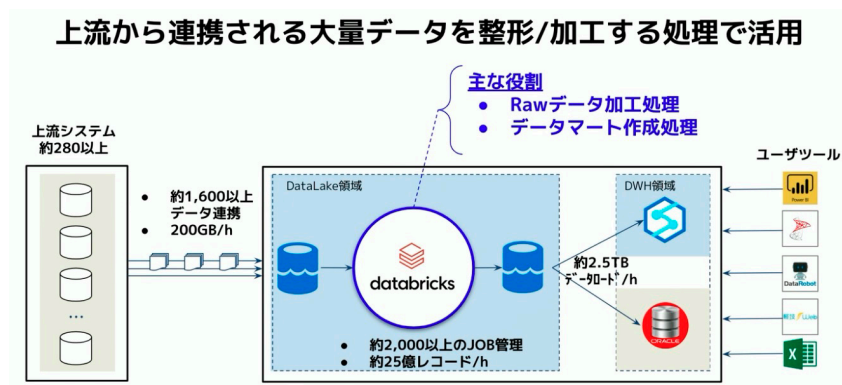
また、ユーザー領域では新たなツール提供として、業務部門へのデータプレバレーション環境の提供を予定している。従前では、各自のローカル環境でエクセルファイル上で分析を実施、そして部門導入したBIツールによる可視化といった状況であったため、ガバナンスやセキュリティの懸念が挙がっていた。これに対し、部門横断的に一貫した分析体験を提供するため、Databricks SQLの採用を決めた。これらの取り組みの大きな意義としては、SSOT (Single Source Of Truth) の実現である。データのありか、データ分析の環境、データ分析のアウトプットの全てが、一つの場所で実現することにより、データサイロ、プロセスサイロを打破し、複雑性の回避によるTCOや障害の増加を回避しつつ、一貫した分析体験を全てのユーザーが享受することが可能となる。

現存のDWHを データブリックスへ統合し、 さらなるSSOTの実現を目指す

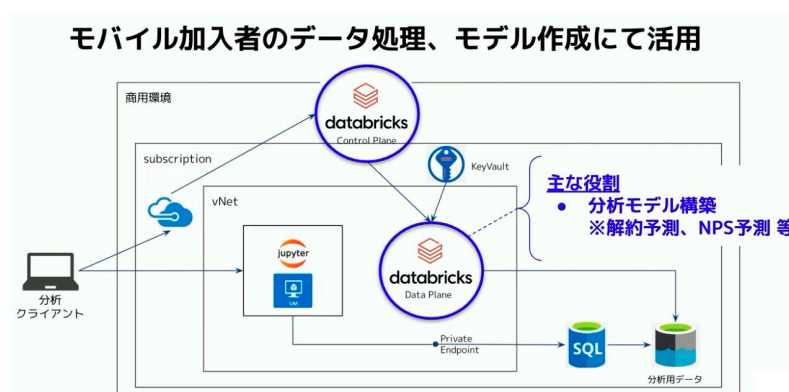
今後の分析環境のロードマップとしては、さらなるシンプル化を予定している。目指す全体像としては、一つのデータ分析基盤の上で、データエンジニアリング、データサイエンス、機械学習の全てのユースケースをサポートすることだ。この実現に向けて、データブリックス

の様々な新機能の実装をする予定だ。具体的には、Delta Engineの採用にさらなる処理パフォーマンスの向上、Delta Live Tableによるデータパイプライン処理の効率化によるデータ品質の向上、そして、Unity Catalogによるデータガバナンスの向上である。また、その先には、現存のDWHをデータブリックスへ統合することによって、データ品質と鮮度を向上しつつコストを削減し、さらなるSSOTの実現を目指す。

データレイク処理基盤のアーキテクチャー図



分析処理基盤のアーキテクチャー図



さらなる性能/品質の強化

