

PoC(Proof of Concept: 概念実証)を経て、 データブリックスの統合データ分析基盤の採用を決定。



株式会社エブリー（以下エブリー社）は、『動画を通じてもっと楽しく、もっと充実した毎日に』をミッションに掲げ、2015年に設立された急成長企業である。国内最大級のレシピ動画メディアである『DELISH KITCHEN』、小さな子どもがいるママ・パパ向けの育児動画メディアである『MAMADAYS』、そして、男性向けのニュース&エンタメ動画メディアである『TIMELINE』の3つが、主な事業の柱となっている。

『DELISH KITCHEN』では、日々の献立立案サポートとして、ユーザーの振り舞いデータを元に、ユーザーへのレシピの推奨、食品メーカーとの共同プロモーション動画の作成、そして、スーパーマーケットの集客プランの立案/実行まで行っている。『MAMADAYS』では、各分野の専門家が監修したコンテンツ提供を軸に、育児に関する課題解決だけでなく、育児記録ができる機能も提供している。『TIMELINE』は、フェイスブックやインスタグラムなどのSNSを中心に展開し、ビジネス、テクノロジー、スポーツなどの分野に関する情報を動画で提供している。



株式会社エブリー
執行役員 CTO
開発本部長
梶原 大輔 氏



株式会社エブリー
開発本部/システム開発部/
データ&AI マネージャー
島田 雅年 氏

採用したプロダクト



ハイライト

データ量の増加に
応じて、
処理時間を一定に
保つことが可能！

インフラ管理が不要
であるため
大きな工数削減を
実現！

ETL処理上で
エラー発生時の
原因追求が
容易に！

課題

データの急激な増加により、分析にかかる処理時間も増加。
機械学習や統計的な手法も視野に、新たなプラットフォームの必要性が高まる。

エブリー社において、データは非常に重要な経営資源であり、全てのサービスの根幹となっている。収集したデータから新たな洞察を獲得し、ユーザーや法人クライアント企業へ、タイムリーに提供することが求められている。しかし、約2年前から、データの急激な増加により、分析にかかる処理時間

も増加してしまい、必要な処理が納期までに完了しない懸念が大きくなっており、新たなプラットフォームの必要性を感じていた。加えて、機械学習や統計的な手法の活用も視野に入れていた。PoC(Proof of Concept: 概念実証)を経て、データブリックスの統合データ分析基盤の採用を決めた。

比較検討

採用した主な理由は、フルマネージドでインフラエンジニアが不要な点。

データブリックスの採用の過程で、Amazon EMRやDataproc (Google Cloud Platform)も比較した。データブリックスを採用した主な理由は、フルマネージドでインフラエンジニアが不要な点、並列処理が前提で処理速度が高速な

Spark をベースとしている点であった。導入前はEMRで処理していたが、ミドルウェアのサポートなど将来性に不安がありデータブリックスの採用を決めた。

ビジネスインパクト

データ量の増加に応じてクラスターを増加することにより、処理時間を一定に保つことが可能に。運用面でも満足している。

データブリックスは期待通りのパフォーマンスを発揮している。Sparkをベースとしているため、データ量の増加に応じてクラスターを増加することにより、処理時間を一定に保つことが可能である。また、運用に関しても満足している。エプリー社では、DIKW (Data Information Knowledge Wisdom)というモデルを元に、ETLやデータレイクを担当するデータエンジニア、データの探索的分析を行うデータサイエンティスト、データをビジネスの文脈に変換するデータアナリスト、という3つの職種を設定している。データブリックスのNotebookベースの開発モデルは、これらの3職種間の協業をサポートし、データチームの生産性向上に寄与している。当時Sparkを実装するためには、IDE上でコードを作成しコンパイルをする必要が

あったが、データブリックスはその必要が無かった。また、仮にデータブリックス Notebookが無い場合は、Jupyter Notebookや Apache Zippelinなどで自前で構築する必要があり、インフラ管理工数が必要となることが想定された。データブリックス Notebookは、運用時や開発時にテキストベースでのコミュニケーションが可能であるため、インフラ管理が不要な点を合わせると、大きな工数削減につながっている。想定していなかったデータブリックスの利点としては、ETL処理上で、エラー発生時の原因追求がしやすい点がある。ETLを実行した結果がジョブとして記録が残り、エラーがNotebookのセル単位で表示されるので、バグの追跡が容易にできるためだ。

今後の期待

新サービスを適宜追加をしていく予定であるが、データブリックスの統合データ分析基盤を活用し、柔軟かつ迅速に対応していきたい。

データブリックスの導入を機に、データの持ち方も進化をしている。現在はデータウェアハウス、データレイク、データマートの3層構造となっている。Delta Lakeベースでデータレイクを構築している為、データが入る段階である程度の構造化が実現している。データウェアハウスに関しては、各事業ごとの状況に合わせて、異なるテクノロジーを採用している。また、重要度の高い『DELISH KITCHEN』においては、データブリック

スがOSSとして提供するDelta Lakeを採用してデータパイプラインを構築し、SQLエンジンとしてTreasure DataとRedshiftを利用している。

今後も各サービスにおけるデータの増大は必至であり、新サービスも適宜追加をしていく予定であるが、データブリックスの統合データ分析基盤を活用し、柔軟かつ迅速に対応していきたい。