

データパイプラインを Pub/Sub や Dataflow で構築し BigQuery へリアルタイムにデータを蓄積、Looker により顧客のデータ活用を促進



yappli

株式会社ヤプリ

<https://yapp.li>

〒106-6241 東京都港区六本木 3-2-1
住友不動産六本木グランドタワー 41階

2013年4月、東京都港区青山においてファストメディア株式会社を設立。2017年4月、株式会社ヤプリに社名変更。2020年12月、東京証券取引所マザーズへ新規上場。Mobile Tech For All(モバイルテクノロジーで世の中をもっと便利に、もっと楽しく)をミッションに、スマート デバイスに特化したインターネット事業を展開。あらゆる企業の DX と課題解決をサポートするとともに、アプリのテクノロジーに誰もが簡単にアクセスできる社会を目指している。主力製品のノーコードのアプリ開発・運用・分析プラットフォームである Yappli は、550以上のアプリへの採用実績がある(2020年12月時点)。

■ インタビュイー (写真左から)

プロダクト開発本部

- ・中原 隆文 氏
- ・松川 佳弘 氏
- ・阿部 昌利 氏
- ・古屋 陽介 氏

ノーコードで大企業向けモバイルアプリを開発・運用・分析できるプラットフォームを展開する株式会社ヤプリ(以下、ヤプリ)。分析機能と可視化の強化を目的に Google Cloud を活用したデータパイプラインを構築しています。このプロジェクトに携わったメンバーの皆さんに話を伺いました。

■ 利用している Google Cloud ソリューション

ストリーム分析

■ 利用している Google Cloud サービス

BigQuery、Cloud Run、Pub/Sub、Dataflow、Cloud Functions、Cloud Storage、Looker など

データウェアハウスとしての BigQuery のコストメリットや運用効率を評価

Yappli は、ドラッグ&ドロップだけのノーコードでモバイルアプリを開発、運用、分析できるプラットフォームです。迅速にモバイルアプリを開発できるのはもちろん、管理画面により簡単に運用が可能。データ分析により、Yappli で開発したモバイルアプリの利用状況を容易に把握することができます。また、年間 200 以上の製品アップデートがあることも特長の 1 つです。ヤプリでは、自社のユーザーのアプリやビジネスをグロースさせるためには、顧客が状況に応じて見たい情報を見られるようにすることや、ヤプリ側のデータドリブンなカスタマーサクセスの実現が必要であると考えていました。従来、Google アナリティクスを利用しデータの収集や分析をしていました。しかし、そのデータだけでは、ヤプリの考えるデータ活用を実現できないため、サードパーティ ツールに依存せずに自社でコントロールできる、まさに“かゆいところにまで手が届く”データ基盤が必要でした。データサイエンティストの阿部さんは、「顧客へ提供している分析ツールは、Google アナリティクスがメインだったので、より詳細な情報を提供できる

仕組みが必要でした。たとえば、ユーザー ID と個別のログ情報がひもづいていなかったため、特定のウェブサイトを参照した人にプッシュ通知を送ることができませんでした。そこで、分析結果を次のアクションにつなげるためのデータパイプラインの必要性を感じていました」と話します。

データパイプラインの構築にあたり、OSS を利用した自社開発をはじめ、さまざまなツールの比較検討を実施。検討の結果、データ分析基盤として、BigQuery を採用することが最適と判断。トラッキングからデータ転送、集計までの一連の流れには、Google Cloud のプロダクトを選定しています。選定理由を、SRE の中原さんは、次のように話します。

「分析基盤に関しては、集計速度が段違いに速く、ある程度のデータ量の分析であればストレージ単価に関してもメリットがあるので、BigQuery 一択だと感じました。Google Cloud のプロダクトを中心とした構成にしたのは、使い勝手のよさが決め手です。リソースを気にすることなく、オートスケールなどの機能で運用保守の効率化も期待できることを評価しました。」

Google Cloud のテンプレート活用でインフラ実装の工数を大幅に削減

データパイプラインの構築プロジェクトは、2020年1月からプロダクトマネージャー1名、サーバーサイドエンジニア2名、SRE1名で、要件定義や仕様の策定も含めて実装を開始。Google Cloud のフルマネージド

サービスを活用することで、約6か月でデータパイプラインを構築し、2020年7月から運用を開始しています。データパイプライン構築では、BigQuery を中心に、トラッキングサーバーとして Cloud Run、BigQuery

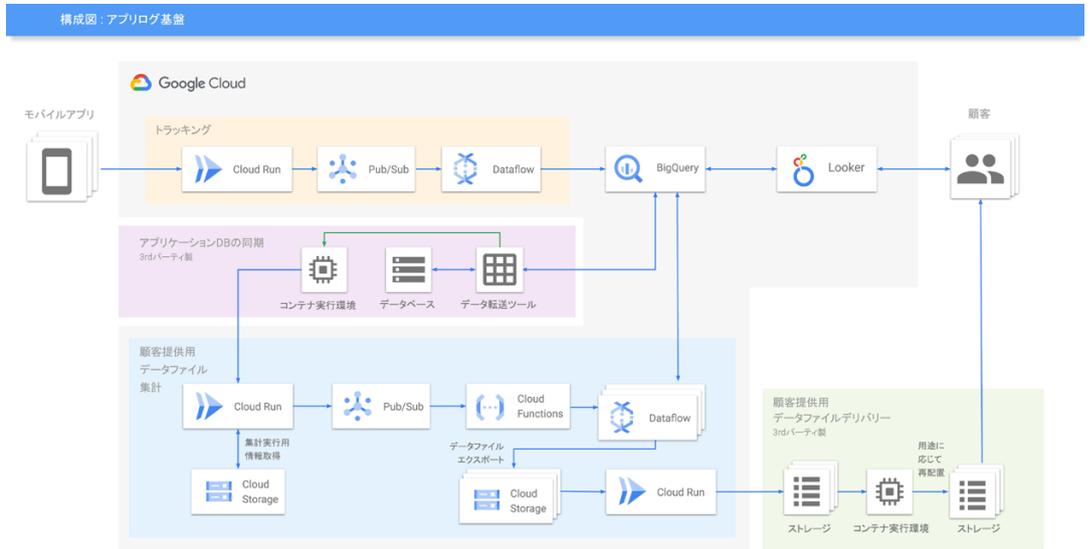
へのストリーミング挿入とデータ集計のためのバッチ処理の振り分けには Dataflow、集計データの格納先に Cloud Storage を利用しています。リードエンジニアの松川さんは、次のように話します。

「ストリーミング挿入は、テンプレートを利用してチューニングなしで簡単に実装できました。リアルタイム分析とバッチ処理分析も、Dataflow を使うことでシンプルな構成になっています。工夫しなくても実装できる点が Google Cloud のメリットです。」

Cloud Run 上に実装されたアプリケーション層自体も、リクエスト情報を Pub/Sub に対してメッセージキューする役割のみで、ストリーミング挿入の責務は Dataflow 側が担ってくれるため、開発工数を大幅に削減できました。削減できた時間は、オートスケールやパフォーマンスなどの事前チェックにかけられました。「パフォーマンス テストでは、疑似的に負荷をかけて、1 年後に想定されるキャパシティにも対応できることを確認しています。現時点で、データサイズはテラバイト オーダー。今後も毎月 1 割程度ずつ増えていく予定です。今後もデータの単調増加が続くこと

が想定されるため、BigQuery のパフォーマンスには期待をしています。」(中原さん)

現在、データ パイプラインは、大きな障害もなく、安定稼働しています。松川さんは、「Cloud Run は、リソースが必要になっても、お任せでスケールしてくれるし、コスト的にも必要な分しかかからないのでベターな運用ができています。インフラの実装に関しては、本当に楽をさせてもらいました」と話しています。



※カードを積み重ねている箇所はアプリごとにデータファイルや処理が分かれている

Looker で顧客向けダッシュボードの拡張性を担保

現状、リアルタイム分析はリリースされていませんが、今後は EC サイトで買い物かごに商品を入れたままサイトを離脱した人にプッシュ通知を送るなど、よりリアルタイム性の高い分析の提供を目指しています。一方、すでにリリースされているバッチ処理では、ユーザー情報や集客情報、行動情報などを毎晩集計し、翌朝 Yappli の顧客が分析できるようになっています。プロダクト マネージャーの古屋さんは、「プッシュ通知を送った結果、どれくらいのユーザーがアクションするかは、Yappli のお客さまが分析したい部分です。リアルタイム分析が公開されれば、Yappli にとって強力な武器になると考えています」と話します。

現状、データ分析では、Yappli の管理画面やデータポータル、Looker などの分析ツールを柔軟に使い分けることで、モバイル アプリ ユーザーの分析ができるようになっています。ユーザーの属性に関しては、Yappli の顧客がフォーム機能を使って作成した登録画面から入力された属性データが利用できます。アプリ側で把握できるのは、個人を特定できない属性データのみです。

ツールの使い分けに関して古屋さんは、次のように話します。「Yappli の管理画面である程度の定型データは参照できます。データポータルは、BigQuery に蓄積されているデータを、パッと可視化して簡単に共有できる

便利さがあり、カスタマーサクセス部門でも、業務の中でデータポータルを使っています。開発部門では、OSS のダッシュボードツールである Re:dash も使っています。」

Looker の採用理由を阿部さんは、「Looker はダッシュボードからドリルダウンして ID レベルのデータまで簡単に辿れたり、他システムへデータを連携したりと、お客さまの使い勝手の良さや将来の拡張性が最大の魅力です。お客さまごとに見たい数字や連携したいシステムは異なるので、1 つのツールで実現する上で Looker は最適なツールです」と話します。

今後のデータ分析に対する経営層の期待は、アプリの IR 資料でも言及されるくらい重要になっています。データ パイプラインの将来について阿部さんは、次のように話しています。

「データ パイプラインのゴールは、データを軸に据えた機能を提供することで、お客さまの DX を推進することです。そこで、より詳細なデータを自社で収集できるデータ パイプラインの構築が不可欠でした。今後は、機械学習や自然言語処理、画像認識などの技術を活用することで、データ分析を次のステップに進め、お客さまの DX をさらに推進するための支援をしていきたいと思っています。そのためのサポートを、Google Cloud には期待しています。」

Google Cloud を活用することで、ビジネスの将来に注力できるようになります。インフラストラクチャの管理やサーバーのプロビジョニング、ネットワークの構成などに起因する負担を軽減することができます。つまり、インベーターもプログラマーも、自分の本来の仕事に集中することができます。

お問い合わせはこちら
<https://goo.gl/CCZL78>

