

BigQuery の導入でデータサイエンティストが “価値ある仕事” に集中できるように



SONY

ソニーネットワーク
コミュニケーションズ株式会社

<https://www.sonymetwork.co.jp>
〒140-0002
東京都品川区東品川 4-12-3
品川シーサイド TS タワー

1995年11月、ソニーコミュニケーションネットワーク株式会社として設立。翌年から『So-net』としてISP (Internet Service Provider) 業務を開始。現在は、ISP 事業のほか、法人向けのクラウドを軸にした IT ソリューションや、インテグレーション事業など、多彩な事業を展開。クラウド&アプリ事業部門ではソニー製テレビ向けアプリ『Video & TV SideView』など、ソニー製品・サービスに対するクラウド開発・運用や、アプリケーション開発を行っている。

ソニーネットワークコミュニケーションズと言えば、インターネットプロバイダ (ISP) 事業『So-net』が有名ですが、実は、ソニー製品・サービスのクラウド・アプリ開発・運用でも大きな役割を果たしています。そんな同社のデータ分析に、今、Google Cloud Platform (GCP) が活躍。80:20ルールなど、同社が抱えていた悩みを、GCP が、BigQuery がどのように解決したのかを聞いてきました。

(利用している Google Cloud Platform サービス)
BigQuery、Cloud Dataflow、Cloud Pub/Sub、Container Engine など

ソニーネットワーク
コミュニケーションズ株式会社
クラウド&アプリ事業部門
クラウド開発運用部 2 課 課長
川田 雅人氏

脱「80:20 ルール」に向けて GCP を導入

ソニーネットワークコミュニケーションズ株式会社 クラウド&アプリ事業部門は、ソニー製品・サービスに対するクラウド開発・運用や、アプリケーション開発を行う部署。データビジネスの生産性を向上させるために、そこで得られたビッグデータを活用しデータ分析を通じて、製品・サービス内容や、ユーザーエクスペリエンス (UX) を向上させることも大きな役割の 1 つとなっています。「こうしたデータの分析をおよそ 4 年前から本格化させているのですが、オンラインサービスやアプリの普及に比例して、月間 100 億レコードを超えるまでに増え続けているデータ量をしっかり分析できる環境作りが急務となっていました。データの質・量を増やしつつもクエリは速くすることでデータ運用を効率化し、価値ある分析業務に集中できる仕組みを構築することによって、ビジネスの PDCA サイクルのパフォーマンス向上を図りました。さらに、データサイエンティストが満足する分析環境であるだけでなく、彼ら以外のビジネス担当者などにも分析業務のすそ野を広げていくことが、ビジネス基盤としての必要条件ととらえ、チャレンジしたいと考えていました。」

と、語るのはクラウド&アプリ事業部門クラウド開発運用部 2 課課長の川田 雅人さん。ちなみにそれまで同社では、他社クラウド

プラットフォーム上で、Hive や Spark などといったビッグデータ処理ツールを利用していたのですが、その環境では「80:20 ルール」の呪縛から逃れられないことも問題になっていたと川田さんは語ります。



「データ分析の業務において問題となる 80:20 ルールとは、作業時間の 8 割が分析ではなく、その準備に割かれてしまう状況のこと。データを集めたり、それを変換・加工したりする前準備にほとんどの時間を費やしてしまい、価値のあることには 2 割の時間しか割けていないんです。データ分析チームは、データサイエンティスト、データエンジニア、ビジネス担当から構成されています。私自身、データエンジニアとして長らく、この問題に悩まされていたのですが、マネージドサービスである BigQuery が、我々をこうした準備やインスタンスの容量・性能設計業務から開放、運用にパラダイムシフトをもたらしてくれました。」(川田さん)

この成功を受けて同社では、データの効率的運用や処理時間の短縮、そして費用削減などを目的に、2016 年より GCP を本格導入。BigQuery を中心に、Cloud Dataflow や Cloud Pub/Sub、そして、Google Container Engine (GKE) など

といった Google のテクノロジーを駆使して、複数アプリ・サービスの商用データ分析基盤を構築します。

「これはいわゆるプライベート DMP (Data Management Platform) なのですが、ユーザーセントリックに分析し理解すること、セグメントを作りビジネスに活用すること、これらの目的のために、必要な全てのデータは BigQuery に投入し管理します。BigQuery の安価なストレージとクエリ課金モデルによって、目的に必要なボリュームのあるデータを躊躇なく扱えています。データをシンプルに 1 つの場所に入れられることで、データ管理、分析、集計、BI (Business Intelligence) ツールも含めて環境が一元化され、データ分析チームの作業効率は大きく向上。高速で、必要なデータをすぐに取り出せる BigQuery だからこそ実現できたと考えています。」(川田さん)

GCP の成長が、我々のビジネスも成長させていく

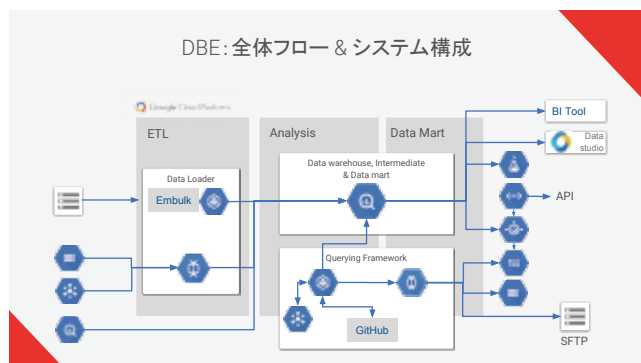
また、マネジメントサービスを導入したことで、これまで頭を悩ませていた監視・メンテナンス業務からデータ分析チームが開放されたということも大きかったそう。雑事に謀殺されず、データの利用や、データ構造の最適化に注力できるようになったことも大きな変化だったと川田さんは言います。

「BigQuery の強みは、GCP 内外に保存された生データをどんどん投げ込んでいけることなんです。それでは最高のパフォーマンスは引き出せません。やはり目的に合わせてデータ構造を最適化する必要があるんです。これについては今年 3 月にサンフランシスコで行われた Google Cloud Next '17 でも熱い議論が交わされていましたね。我々の場合、BigQuery 内のストレージに、ほぼ生の蓄積データとそれぞれの目的に特化したデータマート層、その間に分析しやすく結合・変換・補完などがされている中間層といったデータ構造を導入することで効率性、生産性を高めています。これによって専門的なデータサイエンティストが最高のパフォーマンスを引き出せるようにしたほか、簡単な SQL なら書けるというマーケティング者にとっても使いやすい環境を構築することができました。」(川田さん)

そしてもう 1 つの工夫としてクエリ開発&実行フレームワークも構築。クエリとスケジュール、出力先を指定できる簡単なテンプレートを作ることによって、その中身をデータサイエンティストが自在にカスタマイズできるのはもちろん、SQL を少し書けるという程度の人がフォーマットをほんの少しだけチューニングして使うということもできるようにしました。これによって、誰でも簡単に、分析結果をすぐに商用サービスに適用できるようにしているそうです。



今後の展望としては、Cloud Datalab や、Cloud ML (Machine Learning) などといった機械学習系のフレームワークとの連携を既に始めているほか、現在は非公開ベータ版として提供されている次世代のインテリジェントデータサービス Cloud Dataprep による、さらなる省力化にも期待しているという川田さん。「今、我々の業務が、GCP に支えられているという実感があります。GCP が成長すると我々のビジネスも成長していくのです。これからもぜひ一緒に進化していきたいですね。」(川田さん)



Google Cloud Platform

Google Cloud Platform 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から
<https://goo.gl/cJbFKS>



GCP の詳細については、右記 URL もしくは QR コードからアクセスしていただくか、同ページ「お問い合わせ」よりお問い合わせください。
© Copyright 2017 Google
Google は、Google Inc. の商標です。その他すべての社名および製品名は、それぞれ該当する企業の商標である可能性があります。

