

BigQuery を中心にデータ収集からデータの可視化や活用まで Google Cloud でデータ分析基盤を刷新



飲食店情報サイト「ぐるなび」を中核に、「食でつなぐ。人を満たす。」という存在意義のもと、「飲食店経営サポート企業」として事業を展開している株式会社ぐるなび。ITと人の力を融合した有益なサービスの提供を目的に、ITを活用した事業基盤の1つであるデータ分析基盤を Google Cloud に刷新。このプロジェクトについて、データ・AI戦略室 データ戦略グループの担当者3名に話を伺いました。

■利用している Google Cloud ソリューション

データ ウェアハウスのモダナイゼーション

■利用している Google Cloud サービス

BigQuery、Cloud Composer、Cloud Storage、Looker

データ分析基盤の刷新でユーザーの利便性を向上し、飲食店とのマッチングを強化

「ぐるなび」は、月間ユニークユーザー数が約4,400万人(2020年12月時点)、会員数約2,134万人(2021年7月1日時点)、有料加盟店舗数が54,342店(2021年6月末時点)の飲食店情報サイト。2021年7月より、楽天デリバリー事業、およびテイクアウト事業を承継し、食の総合サービスへの進化の一步も踏み出しています。

次の課題は、ユーザーの利便性、利便性を向上させ、飲食店とユーザーのマッチングを強化すること。レコメンド機能などで情報をパーソナライズ化し、ユーザーに合った選択肢を提供することで、飲食店への送客につなげていくことが重要です。その一環として、2010年より活用しているデータ分析基盤を Google Cloud で刷新しました。

グループ長の小山内氏は、「2010年ごろからオンプレミスのDWH(Data Warehouse)にアクセス解析データや社内データを統合し、キャンペーン分析やプロダクトPDCA改善などに活用していました。しかしアクセス解析ツールから、毎日Raw dataを抽出し、DWHにロードする作業のランニングコストがかかることが課題でした」と話します。

そこで2014年後半、Google アナリティクスプレミアム(現在のGoogle アナリティクス 360)の導入をきっかけに、BigQueryを採用。2015年には、マーケティング用のオンプレミスDWHもBigQueryに移行しています。小山内氏は、「BigQueryは、高速なクエリレスポンスとスケーラビリティ、スライドやスプレッドシートとの連携によりビジネス部門へのデータ提供を加速できることなど、

オンプレミスのDWHのパフォーマンスとは比較できないレベルで快適なことから、すぐに手放せなくなりました」と話します。

「2019年の組織変更によるデータ戦略グループの新設を機に、BigQueryで運用していたマーケティング用DWHと別部門で運用していたオンプレミスのDWH環境をクラウドに統合していくことを決めました。重複していた機能の最適化やオンプレミスで運用していたETLツールの保守サービス期間が終了するなどの複合的な理由により、データ分析基盤刷新プロジェクトがスタートしました。」(小山内氏)



小山内氏

中島氏

ぐるなび

株式会社ぐるなび

<https://corporate.gnavi.co.jp/>

〒100-0006

東京都千代田区有楽町1-2-2 東宝日比谷ビル 6F

1996年に飲食店情報サイト「ぐるなび」を開業。詳細なメニュー情報や今日のおすすめ情報等を事前に確認してから飲食店に行くという外食のスタイルを定着させました。現在ぐるなびは「食でつなぐ。人を満たす。」という存在意義(PURPOSE)のもと、事業を推進しています。今後も「飲食店経営サポート企業」としてさらなるサービスの拡充を図っていきます。

■インタビュー

データ・AI戦略室 データ戦略グループ

- ・グループ長 小山内 涼子 氏
- ・BIチームリーダー 中島 正統 氏
- ・BIチーム エンジニア 濱田 大樹 氏

データ分析基盤の運用課題を解決できるツールが Cloud Composer

新しいデータ分析基盤は、ストレージ サービスを利用したデータレイク（オンプレミス上の複数のデータベースおよび NAS 等のデータを統合）と BigQuery で構成されています。

BigQuery で必要なデータは、データレイクから Cloud Storage を経由して取り込み、まずは DataLake 層に格納しています。DataLake 層には Google アナリティクス 360 からエクスポートされたデータ等も格納されています。さらに、データ活用の迅速化や最適化を実現するため、格納したデータを加工して DWH、DataMart の各層を作成しています。この一連の処理を制御するため、Cloud Composer を利用しています。

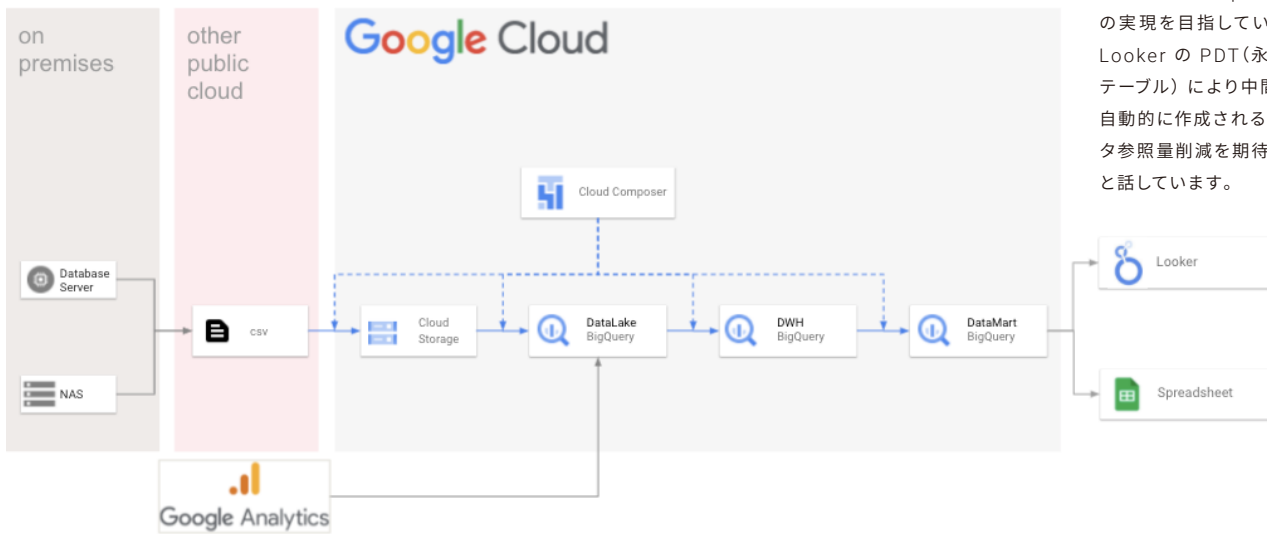
作成されたデータは、Google スプレッドシート、Looker による可視化やアドホック クエリによる分析等に活用されています。また、図には表現されておりませんが、データポータルも利用しています。

Cloud Composer を採用した理由を、BIチーム リーダーの中島氏は、次のよう

に話します。「現行 ETL ツールは、機能を活用しきれていないため、ライセンス費用が割高となっていました。また定期的なバージョン アップが必要なことや、CUI (Character-based User Interface) が使えず調査、作成、修正に時間がかかる、変更履歴や経緯が把握しにくいといった課題がありました。課題解決に向け、マネージド サービスであること、コードベースの管理ができること、コストや運用工数の削減、ワークフローと ETL が利用できること、という評価基準でツールを選定した結果、すべての課題を解消し、評価基準を満たすことができたのが Cloud Composer でした。」

また、Looker の選定理由を中島氏は、「現状のダッシュボードは、Script + SQL とスプレッドシートで運用していましたが、同じ SQL を重複して使っていたので、SQL が冗長的で、スパゲティ構造になっており、改善が必要でした。また、ダッシュボードを作成するまでの一連の開発工数、運用工数の削減も必要でした。こうした課題を解決するために、DWH の設計や LookML の活用による

DRY (Don't Repeat Yourself) の実現を目指しています。また、Looker の PDT (永続的な派生テーブル) により中間テーブルが自動的に作成されることで、データ参照量削減を期待しています」と話しています。



データ分析基盤の開発や運用工数・コスト削減が Cloud Composer 導入の最大の目的

Cloud Composer を採用した効果の中島氏は、次のように話します。「以前は、ジョブが失敗したときの障害対応に多くの時間がかかっていました。Cloud Composer の導入に伴い、ジョブの失敗をビジネス チャットに通知する機能を盛り込むことで、障害発生から検知までの時間が短縮されました。また、通知にログを確認できる URL を含めることで、原因究明がスムーズになりました。さらに、失敗したジョブは自動で再実行されるため、一時的なエラーであれば特に何も対応せずとも解消されます。これにより運用工数が大幅に削減できました。」

ライセンスコストの削減も効果の 1 つ。BIチーム エンジニアの濱田氏は、「Cloud Composer は、Google アカウントがあれば誰でも使えるので、何人でも使用しても追加費用がかからず、アカウント不足の問題が解消されました。また、環境の維持にかかる費用も現行 ETL ツールと比べて安く、ETL に関わるランニングコストが半分程度に削減できました」と話します。

Cloud Composer は、環境の構築と削除が容易なため、まず最小のスペックで立ち上げて検証した上で、最適なスペックまで柔軟に拡張していくことができます。これにより、必要な性能をみたした上で最小限の費用で運用することができるため、コストパフォーマンスに優れています。

また、BigQuery のメリットについて濱田氏は、「インデックスを考慮しなくても、

高速に結果が返ってくることに感心しました。複雑なクエリを実行するとなかなか結果が返ってこないこともあるのですが、BigQuery ではその心配はほとんどありません。また、他社のクラウド オブジェクト ストレージから Cloud Storage へのデータ転送サービスが提供されているため、Cloud Storage を経由した BigQuery へのデータ転送が簡単かつ高速におこなえる点も魅力の一つです」と話しています。

同社のデータ分析基盤は、2022 年夏ごろまでにオンプレミスを廃止し、すべてをクラウドによる運用に移行する計画。運用監視の効率化を期待し、データ分析基盤の運用はすべて Cloud Composer から SQL やデータ転送の仕組みを使うことを目指しています。BigQuery Omni によるマルチクラウド管理や Looker を活用した全体最適も進めていく計画です。

さらに今後の計画を小山内氏は次のように話します。「今後はデータガバナンスや Data Catalog を利用したメタデータの管理をどのようにするかを検討していきます。また、データ戦略グループには、機械学習を用いてプロダクトの価値向上に取り組むデータサイエンスチームもあり、現在 MLOps に取り組んでいます。今期中には Google Cloud で構築したプロダクトをリリースする計画なので、Google Cloud のサポートには今後も期待しています。」

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から
<https://goo.gl/CCZL78>

