

Cloud Run で機械学習活用をスモールスタート ゲーム内チャットの迷惑ユーザー判定自動化などを実現



KLab株式会社

<https://www.klab.com/jp>

〒106-6122 東京都港区六本木 6-10-1
六本木ヒルズ森タワー

2000 年 8 月に設立(当時の社名は株式会社ケイ・ラボラトリー)。当初は携帯コンテンツ事業を主軸としていたが、2009 年から携帯電話、スマートフォン向けのゲーム開発や運営に進出。現在は国内外に多くのゲームタイトルを提供している。従業員数(正社員数)は 559 名(2020 年 12 月末時点)

携帯電話向けゲーム市場黎明期から数々のヒットタイトルを生み出し、現在も有名アニメや漫画などをモチーフとしたモバイル オンラインゲームの企画や開発、運用で知られる KLab (クラブ)。そんな同社がここ数年行っている機械学習を活用した取り組みについて、同社のデータ基盤運用に携わるリード エンジニア、高田 敦史さんにお話をお伺いしました。

エンジニアリング本部 開発推進部
データ基盤(2021 年から機械学習)グループ
高田 敦史 氏

(利用している Google Cloud サービス)

Cloud Run、Cloud Functions、Cloud Composer、BigQuery など

Cloud Composer で効率的に Google Cloud プロダクトを連携

今回お話をお伺いする高田さんは、KLab が提供する多数のゲームアプリのデータを集約して、データ分析や運用に携わるスタッフに提供する「データ基盤(2021 年から機械学習)グループ」に所属。2 年ほど前からは、同社の持つビッグデータに機械学習を導入していくという取り組みを行っています。その成果としてまず実現したのが、それまで人力でやっていたゲーム内チャットの迷惑ユーザー判定自動化です。

「グローバルで配信しているタイトルのチャット機能において、海外の悪質な業者による自社製品の宣伝スパムに悩まされていました。当時はそれを人力で検知、排除していたのですが、そこに英語力のあるメンバーを取られてしまうのはもったいない。そこでこれを機械学習を駆使して自動化できないかと考えました。実のところ、この問題自体はそこまで喫緊の課題ではなかったのですが、機械学習を使ってデータを活用していくという目的にピッタリで、スモールスタートで始めたいというニーズにも合致していました。」

なお、KLab では当時からデータ分析基盤に Google Cloud、特に BigQuery を重用。そのため、この取り組みも元となるデータが格納されている Google Cloud のプロダクトを使って始めるのが自然な流れだったと言えます。この際、いくつかの狙いから、サーバーレスかつ、アプリケーション実行環境をコンテナ化して構築できる Cloud Run を選びました。「数ある Google Cloud プロダクトの中から、当時まだ β 版だった Cloud

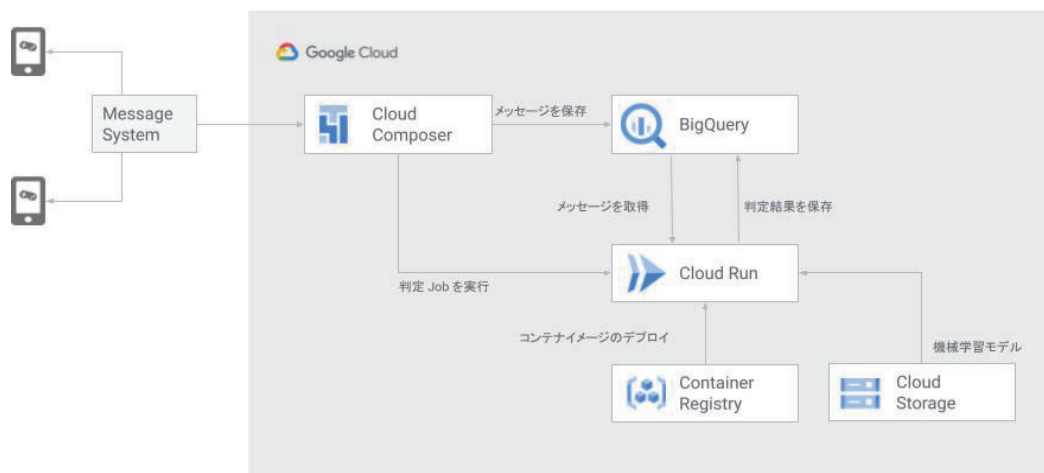
Run を選んだのは、第一にフルマネージドなサーバーレス プラットフォームでコストを抑えられるから。迷惑ユーザー判定は 1 日 2 回、数分で終わるバッチを起動するだけなので、常時起動のサーバーを立てるのはもったいないと判断しました。その上で同じサーバーレスの Cloud Functions を選ばなかったのは、コンテナベースの Cloud Run の方がライブラリの導入が比較的自由にできるから。チャット内容の分析という特性上、自然言語処理が絡んでくるので MeCab や fastText といった優秀なライブラリを使いたかったんですね。」



開発にかかった期間はトータルで約 2 週間。まずはローカルでモデルを作成し、検証、それを Cloud Run に載せ、BigQuery からデータを取って

くるような仕組みを付け加えるかたちでスピーディーに開発できたと言えます。「特に大きなトラブルもなく、ほぼマニュアル通りに開発できました。この際、特に重宝したのが Cloud Composer です。今回の取り組みでは Cloud Composer が各機能をつなぐハブのような役割をしていて、Cloud Run をキックしたり、いわゆる ETL ツールと呼ばれるような、データを抽出、加工、保存するといった一連の流れを Cloud Composer で全て管理しています。流れとしては、まず Cloud Composer が社内にあるチャットシステムの API を叩いてデータを取得し BigQuery に保存。その後、Cloud Run を立ち上げて、BigQuery に保存されたデータを取得して迷惑ユーザー判定を行い、その結果を BigQuery に戻すというかたちになります。」この際、何かエラーが起きた際のリカバリが簡単なことも Cloud Composer のメリットだと高田さんは言います。

「従来の環境ではエラーで処理が停止した際は、依存関係のあるタスクを所定の順番で再実行しなければならず、大きな負担になっていました。しかし、Cloud Composer なら、再実行コマンドを実行するだけで済むのがうれしいですね。」この取り組みによって、それまではほぼ人力で行われていた迷惑ユーザーの検知を大幅に効率化することに成功。コスト面も「ほとんど無視できるレベル」に抑えることができたそうです。高田さんのチームでは、その後もゲーム運営チームなどからの要望を受けて、アプリレビューや SNS のコメントを取得し、ポジティブ、ネガティブのタグ付けをし、件数を可視化するという仕組みを実現しています。さらに、自然言語処理の絡まない bot 判定などの処理を Cloud Functions で実施することにもしているそうです。



サービス規模拡大に柔軟に対応できるのも Cloud Run の美点

スモールスタートで始めつつ、約 2 年かけて、運営業務負担を軽減するさまざまな仕組みを作りあげてきた高田さん。今後は AI プラットフォームなども駆使して、さらに効率的なシステム構成に進化させていくことも考えているそうです。

「SNS のコメント取得が顕著なのですが、昨今、ユーザー数の増加に合わせて徐々に処理が重くなっているという現状があります。また、適用するタイトルをもっと増やしていけるよう、これまでよりも大きなパフォーマンスをだせるような構成に変えていくことを検討中です。具体的にはたとえば、機械学習の推論機能部分だけをデプロイできるようにする AI Platform Prediction を使って、重い推論処理をそちらに流してやるという構成を考えています。」

さらにもっと規模が拡大した際に、そのまま Google Kubernetes Engine (GKE) などに載せ替えられるのも Cloud Run の良いところだと高田さんは言います。

また、データ基盤や、そこに蓄積されたビッグデータを機械学習を駆使して有効活用するという今回の取り組みに、Cloud Run や Cloud Composer、

BigQuery といった Google Cloud のフルマネージドなプロダクトを活用したメリットについて、高田さんは次のように語ります。

「データ基盤は、基本的にだんだんとデータの流量が増えていって、ある日突然、バッチ処理などが動かなくなるという課題があるのですが、KLab のデータ基盤に関しては重い処理は基本的に全てクラウドに載せ、Google Cloud の優秀なオートスケールに任せる方針でやっているの、以前と比べてそうした心配はだいぶ減りましたね。」

こうした実績を踏まえ、最近ではデータ基盤以外の部分にも Google Cloud を使ってみたいというエンジニアが増えてきているのだとか。

「最近では、データ基盤だけでなく、ゲームのインフラにもクラウドを積極的に使っていきたいという声が多く聞かれるようになりました。ただ、データ基盤と、大規模なユーザーを対象としたゲーム基盤では求められるノウハウが全く異なります。そろそろ我々も GKE や Cloud Spanner などを使って、大規模なシステムを構築する経験を積んでいくべきだろうと考えています。実際、社内で行われている勉強会などでも、そういう空気感を強く感じるようになっていきます。」

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、イノ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>

