株式会社リクルート

リクルート、大規模データ処理を支える共通基盤に AWS を活用、最大 30,000CPU コアの処理を実現し、ビジネスと開発の加速に貢献

概要：リクルートグループにおいて、マッチング&ソリューション事業を展開する株式会社リクルート。同社は、社内組織が利用するデータ処理基盤『Crois（クロイス）』を、アマゾン ウェブ サービス（AWS）のマネージドサービスを活用して構築しました。最大 30,000CPU コアの大規模処理に対応した『Crois』は、リクルートの不動産情報サイトのレコメンドシステムなど、多くのプロジェクトで利用され、ビジネスと開発の加速に貢献しています。

ビジネスの課題：全社で利用可能なデータ処理基盤『Crois』を開発

『リクナビ』『じゃらん』『SUUMO』など、カスタマーとクライアントを結びつけるマッチングサービスを提供するリクルート。さまざまな規模のサービスを運用するうえで、広告データ、ログデータ、属性データなどを利用したデータ分析は欠かすことができません。そこで同社のデータ推進室は、社内組織や社内プロジェクトで利用可能なデータ処理基盤として『Crois』を開発しました。

「2017 年にアドテクノロジーのデータ処理基盤のプロトタイプとして Version1 を開発したのが始まりです。その後、進化を重ねながら 2019 年に現在の形となる Version2 をリリースしました。リクルートが 2021 年に事業会社と機能会社を吸収合併して新たな組織として発足することになり、各社のデータ分析環境を調べてみたところ、類似する処理や要求事項があったことから、全社で共通的に利用可能な基盤プロダクトとして運営を始めました」と語るのは、プロダクト統括本部 プロダクト開発統括室 データ推進室 データテクノロジーユニット ユニット長の阿部直之氏です。

『Crois』は、機械学習等を活用したデータ処理を実行する基盤として、分析ジョブを管理するワークフロー機能を備えています。データ処理のプロセスは、再利用可能な「モジュール」としてカタログ化。汎用性の高いモジュールは、事業部を問わず共通で利用でき、ユーザーでも自作のモジュールを登録して共用することも可能です。

「コンセプトは、“データソリューションキッチン”です。データを食材、モジュールをレシピに見立て、レシピをもとに事業組織がデータソリューションという料理を作ることをイメージしました。これにより、事業組織のデータサイエンティストやデータエンジニアは、分析基盤を個別に作ることがなくなり、同時にエンジニアリングにおいて高いスキルを持たないサイエンティストやプランナーでも、エンジニアの手を借りることなくデータ分析に集中することが可能になります」（阿部氏）

ソリューション：AWS Batch と Amazon ECS on AWS Fargate で大規模処理を実現

現在、『Crois』が提供している機能は、モジュールの実行基盤、コンテナイメージのカタログ、ワークフローエンジン、権限管理システムがあり、それらはすべて API や Web UI を介して操作することができます。

コンテナの実行基盤には AWS Batch と Amazon ECS on AWS Fargate を採用しています。機械学習のような CPU、メモリー、ディスクを大量に消費するタスクや GPU インスタンスを利用するタスクでは AWS Batch、データウェアハウスにクエリーを実行するといった軽量なタスクには短時間でスケールができる Amazon ECS on AWS Fargate を利用します。

ワークフローエンジンには AWS Step-Functions を採用し、ジョブ実行の制御をすべて Step Functions へ移譲することにより『Crois』本体から切り離すことでスケーラビリティを確保するとともに耐障害性を実現しています。

導入効果
- **1,200**：1 日当たりの機械学習のジョブ実行数
- **30,000CPU コア**：最大処理能力
- **約 50**：利用されているプロジェクト数
- **2 分の 1**：レコメンドシステムの開発リードタイムの短縮
- **年間 約 10 人月**：レコメンドシステムで削減された工数

ご利用中の主なサービス
- AWS Batch
- Amazon ECS on AWS Fargate
- AWS StepFunctions
- AWS IAM
- AWS KMS

企業名：株式会社リクルート
業種：サービス
国名：日本
設立：2012 年 10 月 1 日
資本金：3 億 5,000 万円
ウェブサイト：
https://www.recruit.co.jp/

「社内向け大規模データ処理基盤『Crois』に AWS のマネージドサービスを活用することで、リクルートのビジネスに求められる高いアジリティとスケーラビリティを実現しています」

大石 壮吾 氏
株式会社リクルート　プロダクト統括本部 プロダクト開発統括室　データ推進室長

大石 壮吾 氏
株式会社リクルート
プロダクト統括本部
プロダクト開発統括室
データ推進室長

阿部 直之 氏
株式会社リクルート
プロダクト統括本部
プロダクト開発統括室
データ推進室
データテクノロジーユニット
ユニット長

鶴谷 誠文 氏
株式会社リクルート
プロダクト統括本部
プロダクト開発統括室
データ推進室

宇野 稔章 氏
株式会社リクルート
プロダクト統括本部
プロダクト開発統括室
データ推進室

企業概要

株式会社リクルート

1960 年、「大学新聞広告社」として創業。2012 年に持株会社体制へ移行し、マッチング&ソリューション事業を統括する株式会社リクルートと 7 つの事業会社に分社化。2021 年、事業会社と統合して新生リクルートが発足。現在、「Follow Your Heart」「まだ、ここにない、出会い。より速く、シンプルに、もっと近くに。」をビジョン・ミッションにクライアントとユーザーを結ぶマッチング事業を展開する。

AWS を採用した理由を、プロダクト統括本部 プロダクト開発統括室 データ推進室の宇野稔章氏は次のように語ります。

「当初は開発メンバーが数名と少なかったため、運用負荷を抑えられるマネージドサービスを利用することがポイントになりました。数百インスタンスを同時に実行するスケーラビリティを考慮すると、コンテナをシンプルに拡張でき、運用負荷がかからない AWS Batch と Amazon ECS on AWS Fargate の利用が有効でした」

権限管理システムでは、AWS Identity and Access Management (IAM)と AWS Key Management Service (AWS KMS) を利用し、ルートプロジェクト単位で IAM ロールと対になる KMS キーを発行することで、高いセキュリティを確保しています。

「データを暗号化せずに保存すると、アクセス権限のある開発者は中身を閲覧することができてしまいます。そこで、中間ファイルやアクセスキー等はすべて AWS KMS で暗号化し、開発者であっても関連する情報は見えないようにすることでセキュリティとプライバシーに配慮しました」(宇野氏)

『Crois』の開発・運用は、Version1 の時代から社内の開発チームにより、すべて内製で行われました。数名でスタートした開発チームも、ジョブの実行数が増えるにつれて 10 数名規模となっています。

「開発時は、AWS の豊富なサービスの中で、何がベストプラクティスなのか、どうすれば効率的に開発できるのかを常に意識しながらメンバー間で議論しています。困った時は AWS のソリューションアーキテクトに随時相談に乗っていただいています」(宇野氏)

導入効果：1 日あたりのジョブ実行数は約 1,200、約 50 のプロジェクトで活用

2019 年にローンチされた『Crois』は、社内プロジェクトですぐに利用されるようになりました。次第に実績が評価され、現在は 1 日あたりのジョブ実行数は約 1,200 に達しています。稼働率も 99.9% を維持しています。プロダクト統括本部 プロダクト開発統括室 データ推進室の鶴谷誠文氏は「2021 年に事業会社が統合し、データ推進室が事業領域ごとのデータ戦略・立案を行う領域特化型のユニットと領域横断で横串の支援を行う専門職種のユニットが交差するマトリクス型の組織になって以降、事業組織間において口コミで評判が拡がった結果、自分たちの事業でも『Crois』を利用したいという声が増えています」と語ります。

『Crois』上では、大規模な処理を複数同時に実行することが可能で、2022 年 12 月時点で最大 30,000CPU コアの処理に対応しています。

「リクルートの中には、大規模な機械学習を実行する部署が多く、1,000 並列、1,500 並列で処理したいといった要求もあります。『Crois』はこのような要求にも対応できるスケーラブルな基盤となっています」(宇野氏)

現在、リクルート内では約 50 のプロジェクトで『Crois』が利用されています。大規模処理の代表的なユースケースが、不動産情報サイト『SUUMO』のレコメンドシステムです。

「『SUUMO』のレコメンドシステムは、長年内製のバッチワークフローを利用していたため、技術的負債が積み上がり、開発・運用効率が低下していました。『Crois』を導入することで、汎用モジュールの利用やユーザーが自作したモジュールの再利用などにより開発効率が上がり、リードタイムは 2 分の 1 に短縮されました。管理運用も AWS のマネージドサービスに任せておけるので、年間で 最大 10 人月程度の運用工数の削減が見込まれます」(鶴谷氏)

「事業組織にとっては、データ処理基盤のような足場の機能を自分たちで開発することなく、本来コミットすべきマッチング精度を上げるような取り組みに集中できることが一番のメリット」と阿部氏が語るように、『Crois』にかかる期待はますます大きくなっています。

「利用案件が増えていく中で、今後はより多くの計算リソースと、高速な実行基盤が必要となることが予測されます。AWS には引き続き協力をいただきながら、スケーラブルな基盤に進化させていきます」(阿部氏)

アマゾン ウェブ サービス ジャパン合同会社
〒141-0021　東京都品川区上大崎3-1-1 目黒セントラルスクエア　https://aws.amazon.com/jp/
Copyright © 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.