



企業名

株式会社リクルートテクノロジーズ



場所

Tokyo Japan

区分
サービス

製品

Elasticスタック



全文検索、可視化

リクルートテクノロジーズ、全社検索基盤、 プッシュ通知基盤、サービスの可視化

Elasticsearchを採用することで、検索基盤の検索性能や更新性能の向上、プッシュ通知基盤の運用負荷の軽減、アドホックな可視化要望への対応などを実現した。

20

台

クラスタの数

2

億

ドキュメント総数

400

GB

総データサイズ

3つのサービス基盤でElasticsearchを活用

リクルートテクノロジーズは、IT・ネットマーケティング領域でリクルートグループのビジネスを進化させることをミッションに、次世代技術のR&D・新ソリューションの開拓やビジネスの実装に取り組んでいる。

リクルートテクノロジーズ インフラソリューション2部の中原裕成氏は、「新しい技術の“開拓、実装・展開、運用”というテクノロジーサイクルを回すことで、リクルート全体の利益に貢献することが役割です」と語る。

取り組みの一環として、(1) サイト内検索に活用するための全社検索基盤「QASS (Query Analytics Search System)」、(2) リアルタイムな検索要件に活用するための全社プッシュ通知基盤「Pusna-RS (パスナ-RS)」、(3) ポンパレやホットペッパーなど、リクルートグループが展開するサービスサイトの「サービスの可視化」という大きく3つの分野を実現するためのサービス基盤の実現にElasticsearchを活用している。



「QASSにElasticsearchを採用することで、更新性能や検索性能、サイトごとのカスタマイズ性の向上が可能になります。またPusna-RSにElasticsearchを採用することで、リアルタイムの更新やクラスタ運用の負荷軽減が可能です。さらにサービスの可視化にElasticsearchを採用することで、可視化の運用負荷の低減、アドホックな可視化要望への対応が期待できます」

- 中原氏

Elasticsearchで全社検索基盤を構築

リクルートでは、複数のサービスサイトを運営しているが、サービスサイトの検索エンジンとしてApache Solrを採用していた。各サービスサイトは、それぞれ専用の検索エンジンを持つ構成になっており、各検索エンジンが協調したり、データを連携したりする仕組みは持っていなかった。そのためすべてのサイトに同じドキュメントを登録する必要があり、同じ更新作業を検索エンジンの数だけ実施しなければならないなどの課題があった。

中原氏は、「サービスサイトと検索エンジンが1対1の構成のままでは、データの管理が複雑な状況でした。そこで全社統一の検索基盤を構築し、すべてのサービスサイトがそれを利用する構成に移行することで、最適な検索結果を提供できる環境の実現を目指しました。また検索品質の継続的な向上も必要でした。最適な検索結果の提供と検索品質の向上を実現するプロジェクトとしてQASSがスタートしました」と当時を振り返る。

QASSのアーキテクチャとしては、エンドユーザーが各サービスサイトで検索をすると、各検索サイトからQASSに検索クエリが投げられ、QASSから検索要件に応じた検索結果が返される仕組みを目指している。しかしApache Solrでは、求める検索性能や更新性能を得ることができなかった。中原氏は、「ドキュメント数が増加したときのパフォーマンスの劣化やノード数が増加したときの更新時間が課題でした」と話す。

またQASSプロジェクトでは、検索結果を分析し、その分析内容をもとにチューニングを行うことで、定期的に検索品質を向上させることを目的とした検索結果分析基盤をHadoopやSparkを利用して構築している。さらに検索結果を分析する専門のチームを編成することで、QASSから得られた分析結果をもとに、どのようにチューニングを行えば検索品質を向上できるかを日々検討している。

中原氏は、「QASSにより膨大なデータを収集できたので、ユーザーの行動（検索）に応じた機械学習を行い、自動的に検索品質を向上する仕組みを実現できました。また集約された検索結果の知見を、サジェスト基盤や検索ロジックの可視化、改善効果の可視化などの機能に利用しています。このように、QASSを中心とした検索エコシステムが確立されており、その中核としてElasticsearchが有効活用されています」と話している。



「以前からElasticsearchに注目していたのですが、ちょうど流行りはじめたという話を聞いたので、QASSの検索エンジンとしてElasticsearchを採用することを決めました。Solrにも、SolrCloudと呼ばれるクラスタリングによる負荷分散の仕組みがありましたが、ElasticsearchはApache ZooKeeperを使うことなく、単体でクラスタ構成を実現できるメリットがあり、運用負荷を低減することを評価しました。」

- 中原氏

Elasticsearchでプッシュ通知基盤を構築

Elasticsearchは、QASSだけでなく、全社プッシュ通知基盤であるPusna-RSにも採用されている。プッシュ通知とは、メッセージを送る相手がスマートフォンアプリを起動してなくても通知を送ることができる仕組みである。プッシュ通知は、開封率が高いことから、メルマガに代わる販促ツールとして注目されている。

プッシュ通知のメリットは、リアルタイムに情報を配信できること、ユーザーのアクティブ率を向上できること、休眠ユーザーを再起できることなど。一方、デメリットは、過剰なプッシュ通知によるユーザー離れのリスクや実装に工数がかかることなどだ。この実装に工数がかかるというデメリットを解消することを目的にPusna-RSが開発された。

Pusna-RSのすべての機能は、アマゾン ウェブ サービス (AWS) 上に構成されている。中核となるデータストアの部分は、Amazon DynamoDB (DynamoDB) とElasticsearchの連携で実現されている。中原氏は、「Pusna-RSでは、複雑なデータ構造が不要なので、DynamoDBを利用して、すべてのデータをキーバリューストアで管理しています」と語る。

DynamoDBの特長は、並列スキャンによる高速な全件データを抽出が可能なこと。ただしDynamoDBは、データ検索が得意ではないため、それを補完する検索エンジンとしてElasticsearchを利用している。Solrの利用も検討したが、リアルタイム更新が可能なことからElasticsearchが採用された。

中原氏は、「Pusna-RSでは、Elasticsearchを全文検索エンジンとは少し違う使い方をしています。具体的には、同じデータをDynamoDBとElasticsearchに保存しておき、単件、全件の抽出はDynamoDBから、条件指定や検索はElasticsearchからと、用途に応じて適したデータソースを使い分けています」と話している。

各サービスの可視化

QASSおよびPusna-RSでは、用途に応じてデータを整理、投入、可視化している。QASSでは、結果分析基盤から、検索結果の分析を行い、分析結果をElasticsearchに投入して、検索効果をKibanaで可視化している。一方、Pusna-RSでは、各サービスノードからfluentdでログをElasticsearchに転送し、そのログをベースに各サービスで利用している指標をKibanaで可視化している。

QASSの可視化では、検索基盤チームがKibanaのダッシュボードを利用して指標を可視化している。どのような検索ワードが使われているのか、アルゴリズムを変更したときに検索の順位がどのように変化したかを利用者側に提示し、そのフィードバックを検索基盤に生かしていくというサイクルを回すことで、自動的な検索品質の向上だけでなく、手作業による検索品質の向上にも可視化を活用している。

Pusna-RSの可視化についても同様に、プッシュ通知の状況やプッシュ通知によりデバイスが起動されたかどうかなどを、Kibanaで可視化している。これにより、プッシュ通知の効果的な時間帯や内容の解析や改善に活用している。中原氏は、「QASSやPusna-RSの可視化だけでなく、可視化で得た知見をAdobeのReports&Analyticsやウェブサーバーログ、ウェブビーコンなど、別のデータソースの可視化にも横展開しています」と話している。

Elasticsearchの運用ノウハウと注意点

Elasticsearchの導入により得られた運用ノウハウについて中原氏は、次のように語る。

「Elasticsearchには、さまざまな有効な機能がありますが、中でも (1) プラグイン機構の利用、(2) スナップショット機構の利用、(3) 環境にあわせたインデックス作成という3つの機能は特に有効でした」

(1) プラグイン機構の利用では、Elasticsearchに標準で搭載されているプラグイン機構を利用して、インデキシングの最適化を実現し、QASS独自のプラグイン機構で動的プラグイン更新を実現。サービスサイトの運営に影響を及ぼすことなく、A/Bテストなどを実施して、検索改善を実施できる仕組みを実現している。

また (2) スナップショット機構の利用では、スナップショット/リストアAPIをJenkinsで定期的に行うことで、サービス用クラスタの複製を作成し、検索改善を行っている。内部では、検索改善用のクラスタをデータサイエンティストに提供し、アルゴリズムの変更が検索結果にどのように影響するか、実データを使って検証している。

さらに (3) 環境にあわせたインデックス作成では、クラウド環境とオンプレミス環境で、異なるインデックスの作成方法を採用している。

「クラウド環境では、特性を生かしてBlue-Greenデプロイメントでクラスタごとにインデックスを再構築しています。これにより、実環境に負荷をかけることなくインデックスを再構築できます。一方、オンプレミス環境では、エイリアスを利用して、検索サービスに影響を及ぼすことなくインデックスをデプロイしています」(中原氏)。

Elasticsearch運用時の注意点について中原氏は、「Elasticsearchは、頻繁にバージョンアップがあるので、なるべく最新バージョンを利用するようにしています。バージョンアップのときには、2つのバージョンのクラスタを準備して、古いバージョンを切り離すという作業をしています。バージョンによりAPIやレスポンスが変更されているので、計画的に実施することをおすすめします」と話している。