



顧
お客様企業
日本経済新聞社

本社所在地
東京都千代田区

業種
新聞・出版・放送など

製品
Elasticsearch

ユースケース
全文検索

サブスクリプション
ゴールド

日本経済新聞社： ドキュメント総数1.6億件、 国内最大級のデータベース事業の検索を担う

日本経済新聞社のデジタル事業では、その全コンテンツについて検索エンジンを共通化し、これをクラウド上で運用するという「コンテンツプラットフォーム構想」を推進。この構想のもとで、検索エンジンに採用されたのがElasticsearchだ。

1.6

億件

ドキュメント総数

48

台

ノード数(最大)

3~4

名

運用チーム人数

日本経済新聞社について

1976年創刊の日本経済新聞をはじめとする新聞を中心とした出版や放送など幅広い事業を展開。デジタル分野でも、2010年に日本初の本格的有料電子版サービスとして創刊した「日経電子版」や、1984年にサービスを開始した「日経テレコン」など多くのサービスを手掛ける。

<https://www.nikkei.com/>

アクロクエストテクノロジー株式会社について

働きがいのある会社日本1位を3度受賞した「全社員で給与まで決める」オープンな社風が特徴のITベンチャー。Elastic Stackの導入／設計／運用を幅広く支援するサービスを展開。日本唯一のElasticコンサルティングパートナーであり、Elastic認定資格者数は国内最多。2021年、同社の3名が3種全てのElastic認定資格を取得。日本での「Elastic認定資格コンプリート」はElastic社以外では同社の3名のみ。

<https://www.acroquest.co.jp/>

国内最大級のビジネスデータベースサービスにおける検索基盤の刷新を構想

日本経済新聞の発行をはじめ、出版、放送など幅広い事業を手掛ける日本経済新聞社。同社にとって、グループ事業の成長エンジンと位置付けられているのがデータビジネスで、「日経テレコン」や「日経NEEDS」など長い歴史を持つデータベースを核として、数々のサービスを展開している。

これらデータベース事業の代表格である日経テレコンは、1984年にサービス開始と特に長い歴史を持つ。今では国内外750以上の新聞・雑誌の記事や専門情報が登録され、マクロ統計や企業情報、マーケットデータを提供している。東証一部上場企業の約7割が導入している国内最大級のビジネスデータベースサービスだ。

日経テレコンの機能の中でも、多くのユーザーが活用しているのは1999年から提供されている全文検索だろう。膨大なコンテンツの中から、横断的に必要な記事や情報を迅速に探し出すことができる、便利な機能だ。ただし、システム運用面では相当な負担でもあった。

「全文検索機能は、提供開始から最近までオンプレミスで構築・運用してきました。日経テレコンの検索エンジンだけでも物理サーバ約200台という規模で、数年おきに行う機器更新だけでも2年以上かけて順次入れ替える必要があり、多大なリソースを費やしていたのです。また検索エンジンのバージョンアップも、こうした更新のタイミングでないと実施できず、新機能の導入も遅れがちでした。開発も社外のベンダーに委託していたことから、スピード感の点でも課題があったのです」と、日本経済新聞社デジタル事業情報サービスユニット

ト サービス企画・開発グループの日當泰輔氏は説明する。

しかも、これまでの検索基盤はコンテンツやサービスごとに分かれていた。日経テレコンはテキスト系コンテンツ、日経 NEEDS は数値データが主体であるなど、サービスごとにコンテンツの方向性が違うため、それに合わせて個別に設計・構築するのが合理的だったともいえよう。とはいっても、結果としてサービス URL はもちろん検索 API も全く異なるものになってしまっており、横断的に利用することが困難だった。複数のデータベースを横断的に利用するような新サービスを企画しても、開発が難航しかねない。

そこで日本経済新聞社は、こういった検索基盤にまつわる数々の課題を一掃すべく、2018 年に「コンテンツプラットフォーム構想」を打ち出した。具体的には、データベース事業の全コンテンツについて検索エンジンを共通化し、これをクラウド上で運用するというものだ。運用性やスケーラビリティを向上させると同時に、横断的なデータ活用も容易にする。また同時に、外部へ委託していた開発を社内エンジニアによる内製に切り替え、開発スピードの改善も図った。

検索エンジンの共通化により、利用者からの要望や検索ログなどの知見が1カ所に集約されるため、課題が顕在化しやすくなった。例えば、検索結果のノイズは結果からわかりやすいが、検索の取りこぼしについては顕在化しにくい。日経テレコンは従量課金サービスであるため、取りこぼしをなくすことが売上アップに直結するのだ。こうした知見を基に、プラットフォーム側で一括して改善し続けることで、効率的かつ継続的に全サービスの検索体験を底上げすることが可能になった。

分散配置できコンテナイメージが配布されている Elasticsearch を採用

コンテンツプラットフォーム構想の検索エンジンに採用されたのが、Elasticsearch だ。日當氏は、このコンテンツプラットフォーム構想に合わせて配属された社内エンジニアの一人で、検索エンジン周りの担当としてプロジェクトの中でも重要な役割を担った。日當氏は検索エンジンの選定について、以下のように説明している。

「Elasticsearch 採用の理由はいくつもありますが、Docker イメージが配布されている上にクラウド関連のプラグインが豊富に用意されている点は大きなポイントですね。コンテナの立ち上げからクラスタへの追加までの作業のほとんどを自動化でき、人手をかけず容易に運用できます。また機能面では、『高速検索』『分散アーキテクチャ』『チューニングの幅』といった点特に重要視しました。分散アーキテクチャでノード追加が容易な上に、クラウド事業者のアベイラビリティゾーンをまたいだ配置で可用性の向上も図れます。これからの事業拡大を図る上で、継続的、安定的にサービスを提供できる点は重要です」

チューニングの幅が重要だった背景には、サービスとしての検索結果をできるだけ揃えたいという意図があった。長年に渡り多くのユーザーが利用してきたサービスで、ある時点から急に検索結果が変わるようでは、ユーザーにも混乱をもたらしかねない。しかも、日経テレコンや日経 NEEDS は出力した情報に応じた従量課金となっているため、検索結果が大きく異なるようではサービス収益にも影響してしまう。

“



コンテンツプラットフォームを刷新し、様々な新しい取り組みも可能になりました。Elasticsearchでなければ、今のような運用は不可能だったと思います。Elasticsearchの採用は正解でした。

デジタル事業 情報サービスユニット
サービス企画・開発グループ
検索エンジニア
日當泰輔 氏

「以前の検索エンジンの機能を踏襲し、結果もできるだけ揃えられるかどうかといった点は、選定段階から重視していました。例えば日経テレコンでは、同義語やシソーラス辞書、and/orなど演算子を使った検索、さらには検索キーワードのハイライト表示や類似ドキュメントの提示といった機能もあり、それらの機能も再現しなければなりません。システム構築そのものはあまり困難ではありませんでしたが、その後の検索結果のチューニングには少し苦労しました。Elasticsearchはアナライザにより検索結果を細かくチューニングできるので、既存の検索精度をほぼ再現できましたが、むしろ既存エンジンの結果が最適だと思えないケースもありました。そうしたとき、Elasticsearchが出てくる結果が妥当だと認めてもらうべく、サービス担当チームと交渉することが最も大変でした」（日當氏）

サーバ台数の大幅減と運用自動化でコスト削減、可用性も向上

Elasticsearchを用いて新たに構築された検索エンジンは、日経テレコンなどサービス側からクエリを受け付ける検索用ドメイン、CDN（コンテンツ配信ネットワーク）、APIゲートウェイなども含め、Amazon Web Services（以下、AWS）上に置かれている。サーバ台数は、検索エンジンの性能向上もあって従来の約200台から48台へと大幅に減少。AWSの機能によるコンテナオーケストレーションや、Elasticsearchのクラウド用プラグインの活用などで運用の多くを自動化、Kibanaダッシュボードを用いたサービス監視など様々な負荷軽減策も相まって、合計10クラスタを3～4名のチームで運用できるようになった。これらの効果を合わせ、年間ランニングコストは20%以上も低減できた。

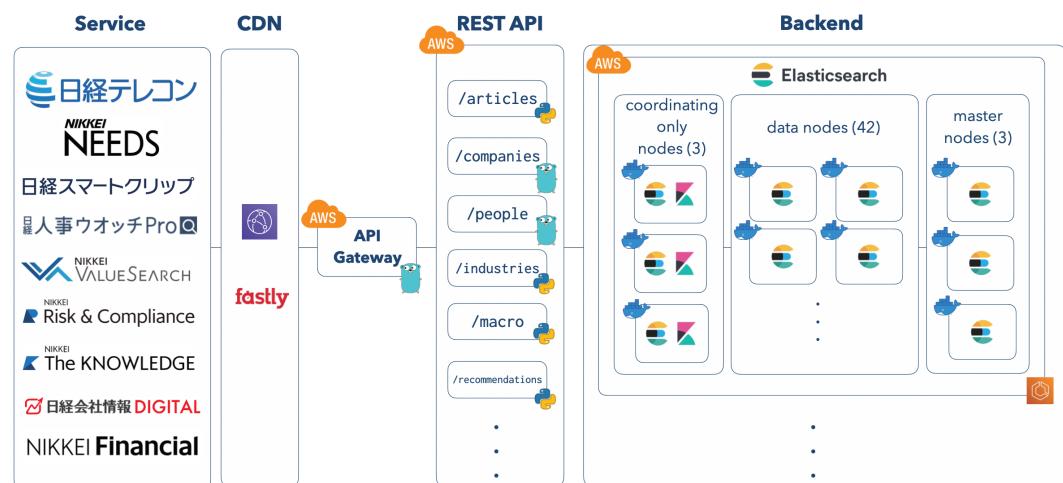
ソフトウェアのバージョンアップも、従来は機材更新に合わせて行うのが通例だったのに対し、今ではローリングアップデートが可能になり、数分もあれば更新できるようになった。今後は、Elasticsearch最新バージョンへの追隨を図っていくとしている。

アベイラブルゾーン（AZ）障害に備えElasticsearchのクラスタを複数AZに分散して配置するなど可用性に配慮した設計とした結果、実際にAZ障害が発生した際にもサービスにほぼ支障なかった。

「我々がアラートで気付いて状況を確認したときには、すでにレプリケーションからの自動復旧プロセスが始まっていて何も操作する必要もなく、ほどなくして復旧を完了していました。オンプレミス環境だった頃には、障害ノードを切り離して再起動、クラスタに復帰させるといった一連の作業を全て手動で行う必要があったことから考えると運用負荷は雲泥の差ですし、復旧までの時間も短縮されています」（日當氏）

知識も経験も豊富なパートナーが開発や運用を強力に支援

このコンテンツプラットフォーム構想や、それに関連したElasticsearch、Kibana、Beats、



図：日本経済新聞社が新たに構築したコンテンツプラットフォームの概要

Logstash で構成される Elastic Stack 活用にはパートナーの存在も大きかった。Elastic 認定エンジニア、アナリストが日本で最も多く在籍、企業としても日本で唯一のアドバンスドリセラー、コンサルティングパートナーであるアクロクエストテクノロジーだ。同社の支援について、日當氏はこう評価する。

「コンテンツプラットフォーム構想をスタートして以来、内製開発を主体としてプロジェクトを進めていますが、Elastic Stack の機能は膨大です。アクロクエストテクノロジーには、コンサルティングや運用診断などで協力してもらっていますが、Elastic Stack に関する知識や経験が豊富で、例えば解決しようとする課題に対して役立つ機能などは即座にアドバイスしてくれます。また、今後リリースされる新機能から、検索結果や性能の検証方法といった、ドキュメントにないようなことも知っているため、とても頼りになる存在です。サーバー台数を大幅削減できたのは同社のチューニング技術の成果が大きく、当初の設計ではインデックスが約 2TB あり、性能目標を達成するには 165 台が必要な状況でしたが、チューニングの結果、インデックスサイズを 60% 以上の削減に成功し、48 台で性能目標を達成できました。運用診断サービスも、監視項目ごとにわかりやすい指標が設けられていて全体像の把握が容易で、解決策も提案してくれるため、次のアクションにつなげやすいです」

Elasticsearch によるプラットフォーム刷新が、数々の新たな取り組みを可能に

日経テレコンなどのデータベースを受け継いだコンテンツプラットフォームには、2021 年 5 月時点で 1.6 億件ものドキュメントが登録されている。その数量は、記事などのコンテンツでは日本最大級の Elasticsearch データベースとみられる。しかもその量は今後さらに増え続け、2022 年末頃には 2 億件にもなると見込まれている。もちろん、コンテンツ量やアクセスの増加などに応じて、Elasticsearch のクラスタ台数も随時増加していくことになる。その判断には、先に説明した監視ダッシュボードや、運用診断サービスでの「クラスタ性能に関する中期的な推移」などの情報が役立てられている。

※ Elasticsearch 運用診断サービス：https://www.endosnipe.com/elastic_cluster_diagnosis

今回の構想では、多種多様なコンテンツを一つの検索エンジンから全文検索できるようにすることで、既存サービスの改良や、全く新しいサービス開発につなげることも大きな目的だ。例えば、取引先に対するデューデリジェンスに役立つ「日経リスク&コンプライアンス」も、本プロジェクトの成果を元に開発されたサービスの一つ。AI（人工知能）など先進テクノロジーを駆使して、この膨大なコンテンツを分析、取引先などのネガティブ情報をスクリーニングするなどの機能を提供する。

「日経テレコンに関してもリニューアル計画が進行中です。例えば、これまで登録ユーザーのみコンテンツを閲覧できる仕様でしたが、最近では一部コンテンツをオープン化しました。これにより外部の検索エンジンでも日経テレコンの記事がヒットし、より多くの方に触れるようになります。そして同時に、Elasticsearch への検索リクエストも増えることになりますが、オンプレミス時代とは違ってスケーラビリティが高いので、必要に応じて増強できます」（日當氏）

※ Open版日経テレコントップページ：<https://www.nikkei.com/telecom>

さらに機能面でも、例えば Elasticsearch を生かし、係り受け検索や横断サジェストなどの新たな取り組みを予定している。また、検索精度を向上させるために、クエリや結果などをログとして蓄積し、分析していくことも計画中だ。これら様々な新しい取り組みも、コンテンツプラットフォームを刷新し、Elasticsearch による運用があつてこそ可能になることだ。そうした面でも、Elasticsearch の貢献は大きいと言えるだろう。

●日當氏は、「第 38 回 Elasticsearch 勉強会」で、「日経テレコンを支える Elasticsearch」と題して講演されました。ぜひ資料をご覧ください。

<https://speakerdeck.com/hinatades/ri-jing-terekonwozhi-eruelasticsearch>

お問い合わせ

Email: elastic-japan@elastic.co

全文検索エンジンを提供する企業、Elastic は Elastic Stack (Elasticsearch、Kibana、Beats、Logstash の製品群) の開発元です。検索、ログ、セキュリティ、分析などのユースケースで大規模データをリアルタイムに処理するサービスを、オンプレミスと SaaS で提供しています。Elastic のコミュニティは 10 万人規模に成長しています。Elastic Stack は Cisco、eBay、Goldman Sachs、Microsoft、The Mayo Clinic、NASA、The New York Times、Wikipedia、Verizon を含む世界中の企業や組織で採用され、ミッションクリティカルなシステムを支えています。Elastic は、世界各国から社員が働く「分散型企業」として 2012 年に設立されました。詳しくは、elastic.co/jp/ をご覧ください。