

データ分析環境の刷新にあたり、ビッグデータサービスを徹底比較。 パーソルキャリアが見据えるデータ利活用のビジョンと 「Azure Databricks」を採用した理由とは



業務部門とシステム部門の 連携により、データ分析環境の 刷新プロジェクトが始動

パーソルキャリアでは、転職サービス「doda(デューダ)」をはじめ、ハイクラス転職サービス「doda X」やエグゼクティブ人材専門採用支援サービス「EXECUTIVE AGENT」など、転職を希望するカスタマーとキャリア採用、求人掲載を希望するクライアント企業に向けたサービスを展開しています。サービスのオンライン化など、デジタル技術を活用した新たな価値の創出にも積極的で、各事業によって蓄積される膨大なデータをどのように活用していくのかは、近年における重要なテーマとなっていました。

こうした状況のなか、2018年に構築したデータ分析環境を用いて効果的なデータ利活用を模索してきた会社ですが、利用者数の増加に伴い、運用しているクラウド DWHで、キャパシティとワークロードの運用管理に手間がかかるといった課題が顕在化。その解決に向けて、データ分析環境の刷新に着手します。このプロジェクトは、業務部門とシステム部門の連携により推進され、以前よりデータ分析・活用を行ってきた P&M 本部におけるソリューション選定の背景として、大規模・高負荷分析のニーズに耐えるソリューションおよびレコメンドシステムのインフラとして、システムの選定や要件定義が進められました。パーソルキャリア株式会社 P&M本部 データビジネス部 ビジネスプランニンググループ シニアエンジニア(データ)の小林 裕也氏は、同社の事業で蓄積されたデータが持つ価値が高まったことで、今回のプロジェクトがスタートしたと語ります。

「ビジネスを展開していくなかで蓄積されたデータの分析・活用は、以前から行ってきましたが、ここ数年で使えるデータの量や質が大きく変化し、たとえば Web 上の施策から得られたデータを事業戦略の判断にも利用できるようになってきています。このようにデータの重要性が高まったことから、現状のクラウド DWHに限界が見えてきてしまい、データ基盤の整備が急務となりました」(小林 氏)。

システム部門として本プロジェクトのマネジメントを担った、パーソルキャリア株式会社 テクノロジー本部 インフラ基盤統括部 システ



ム共通BITA部 分析基盤グループ シニアエンジニアの鈴木 裕之氏も、データが増大していくなかでプラットフォームの強化は不可欠だったと振り返ります。

「当時のデータ分析環境はサービス展開、データ増大に対してスケールアウトして対応してきました。しかし、データベースに対する同時アクセス数増加によるレスポンス劣化や、変化していく分析ニーズに合わせたスケールを常時維持しようとする、オーバースペックとなり、大きなコストロスになってしまいます。こうしたスケール面での課題は、自由なデータ分析・活用を妨げるだけでなく、システム部門の工数も増大するなど、円滑なビジネスに影響を与えかねません。ユーザーはライト層とヘビーユーザーであるデータアナリスト層の2つに分けられますが、たとえばデータアナリスト層が利用した際は、高負荷処理を可能とし、同時に利用するライト層にも影響がないようなアーキテクチャを構築するなどの、抜本的な見直しが必要と判断しました」(鈴木 氏)。

主要なビッグデータサービスを 比較検討した結果、 Azure Databricks を中心とした レイクハウス アーキテクチャを採用

前述したとおり、本プロジェクトはパーソルキャリアの事業部門・システム部門が密接に連携しながら進められました。「これまでのデータ分析環境では、ライト層も含めて多くの利用者がいたため、まずは

そのなかで重要なプロジェクトの一部を切り離して、システムの選定や性能検証を行うことにしました。今回のスコープとしては3,000～6,000あるデータテーブル数のなかから、P&M本部が業務として扱う400程度のテーブル数をターゲットとしています」と鈴木氏。今までパフォーマンス面で問題を抱えていたP&M本部の業務がスムーズに行えるようになることを第一目標とし、数十億のレコードを使った分析を、リアルタイムで応答できるパフォーマンスを実現できるアーキテクチャを条件に、同社の要件にマッチしそうなクラウドサービスから、RFPを展開して提案を依頼したと経緯を語ります。

「選考のポイントは大きく3つあります。業務的に目的を達成できること。保守・メンテナンス性で、マネージドなサービスや自動で復旧できるバッチ処理など、我々が保守運用していくにあたってのメンテナンス面が容易であることと、アドホック利用のような分析ニーズにあわせてシステムコストコントロールができること。そして、利便性を損なわず、よりニーズに答えられる柔軟なアーキテクチャを導入することを大事にしました」(鈴木氏)。

こうして始動した本プロジェクトは、2021年10月～12月までの3カ月でRFPが作成され、年明けに各社に配布。2022年2月末に各社のプレゼンを受けて比較検討を行いました。このなかで、マイクロソフトのパートナー企業として提案を行ったのが株式会社 Colorkrew(カラクル)です。Azureに関する知見が豊富な Colorkrew では、Azure Databricks を中核とし、Azure DataLake Storage Gen2、Azure Data Factory や Azure DevOps Servicesを活用したデータ分析環境を提案し、100億レコード以上に及ぶデータに対する性能検証を実施。この内容を魅力に感じた パーソルキャリアでは、総合的な比較検討の結果、Azure Databricks の採用を決定しました。開発ベンダー側のプロジェクトマネージャーとして今回の取り組みに参画している株式会社 Colorkrew 原 裕一郎氏は、性能検証の内容について次のように説明します。

「実際に現行のデータ分析環境で使われているクエリのなかでも、特に重たいクエリを使ってどれだけ性能が出るかを、複数の側面から計測しました。クエリ自体がどれだけ高速に処理できるかはもちろん、多くのユーザーが併行して利用した際に問題なく処理できるかなど、実際の業務に近いシナリオで検証を行っています。さらに性能とコストのバランスに関しても検証し、結果を提示しました」(原氏)。

鈴木氏は「データアナリストのかなり尖った分析に対して瞬発力に対応することと、ライトな利用者にはコスト含め性能をセーブできる運用を実現したいと考えていました。アーキテクチャについて比較検証を行い慎重な検討を行いました。その結果、もっとも要件にフィットしたのが Azure Databricks でした」と Azure Databricks を軸とした提案を採用した理由を解説。「パワーユーザーのニーズだけを反映させると他社のサービスという選択肢もありましたが、ライト層の利用比重が大きい自社の分析環境においては、安定的にパフォーマンスを出せるようにコントロールできる Azure Databricks が最適と判断しました」と語ります。

ユーザー側の立場で参画している小林氏も「P&M本部は分析スキルにとがった人員が比較的多い組織ですが、それでもBIツールを活



パーソルキャリア株式会社
P&M本部 データビジネス部
ビジネスプランニンググループ
シニアエンジニア(データ)
小林 裕也氏



パーソルキャリア株式会社
テクノロジー本部 インフラ基盤統括部
システム共通BITA部
分析基盤グループ
シニアエンジニア
鈴木 裕之氏

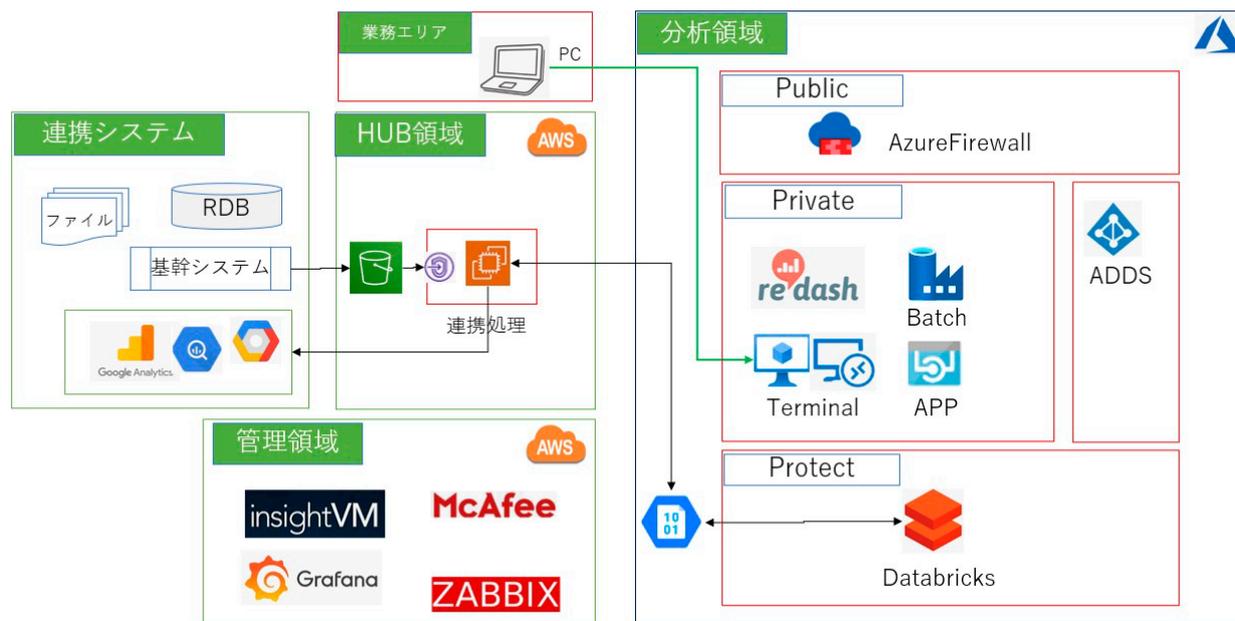


株式会社 Colorkrew
原 裕一郎氏

用するユーザーがほとんどで、アドホックに分析SQLを実行するユーザーは決して多くはありません。そのような状況下でも大規模・高負荷なデータ分析ニーズは増え続けており、それを支えるデータ基盤側の運用に対してもさらなる安定性と高負荷処理が求められてきています。今後さらに加速していく事が見込まれていくなかで、大規模・高負荷の並列処理パフォーマンスとランニングコストの両軸から Azure Databricks は、非常に魅力的な選択肢となりました」と採用の経緯を説明します。

進行中のプロジェクトに Azure Databricks の最新機能を取り入れ、先進的な データ分析環境を構築

2022年3月から比較検討を行い、Azure Databricks を軸としたレイクハウス アーキテクチャを採用したパーソルキャリア。開発ベンダーである Colorkrew と密接にやり取りしながら同年4月から3カ月をかけて要件定義を行い、7月からシステム開発をスタート。12月までに結合テストを行い、2023年に入ってからシステムテストを続けており、同年6月の稼働を目指しています。バッチ及びデータベース開発



の PL として本プロジェクトに携わったパーソルキャリア株式会社テクノロジー本部 インフラ基盤統括部 システム共通 BITA 部 分析基盤グループ リードエンジニアの玉井 孝平 氏は、システム開発において苦労したポイントについて次のように語ります。

「長らく成長した現行分析システムを刷新するとあって、仕様検討する範囲も広がりました。多種多様なデータを扱っていて、処理の汎用化設計には苦労しました。巨大なロジックにならないようマイクロサービス化することや、Azure 製品を活かすことを意識しました」(玉井 氏)。

鈴木 氏も「今回の刷新プロジェクトでは、昔のシステムをそのまま Azure Databricks に移行するのではなく、新しい分析軸や、履歴データと最新データをコントロールする仕組みなど新しい機能やロジックも積極的に導入しています。また現行システムは 2018 年から運営しており、時代の変化とともにデータのセキュリティ管理も対策してきました。今回のプロジェクトではコンプライアンスの強化も重要な目的としてシステムを構築しています」とシステム構成について説明します。

開発を担当した Colorkrew の原 氏も「今回のプロジェクトでは業務面・システム面で先進的な内容が盛り込まれており、コストやスケジュール業務要件があるなかで、どう実現していくか気を遣いました」と振り返り、さらに成長を続ける Azure Databricks の最新機能を取り入れるかどうかの判断にも苦労したと語ります。

「今回採用した Azure Databricks は成長中のソリューションで、開発を進めている間にもコアな部分を含めて新たな機能が追加されていました。そのなかには、今後5年10年にわたり Azure Databricks でのデータガバナンスの標準となる Unity Catalog も含まれており、それらをプロジェクト途中で取り入れるかどうかは難しい選択でした。スケジュール的に厳しい面もありましたが、将来を見据えて、で



パーソルキャリア株式会社
テクノロジー本部 インフラ基盤統括部
システム共通 BITA 部
分析基盤グループ
リードエンジニア
玉井 孝平 氏

きるだけ妥協せずに対応するようにしました」(原 氏)。

パーソルキャリアでは、こうした Colorkrew のサポート姿勢と、レスポンスの速さを高く評価しており、システム稼働後も保守業務を担当してもらう予定といたします。

リアルタイムでのデータ連携を実現し、顧客体験のさらなる向上を目指して取り組みを続けていく

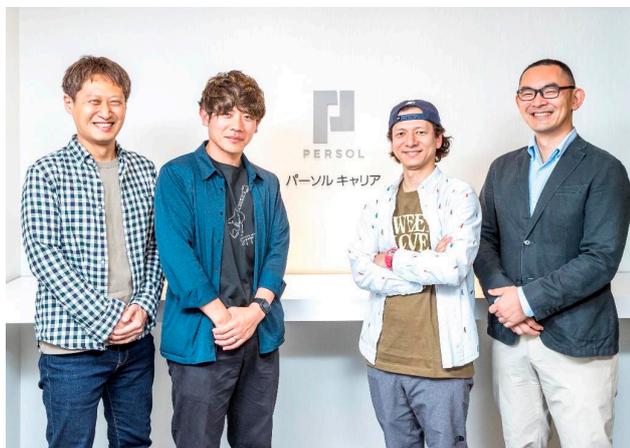
先に述べたとおり、現在はシステムテストを実施している段階で、その後はチューニングを進めて一般公開を目指していく予定となっています。「CRM のメルマガやレコメンドへの利用を想定しており、安定稼働を優先して開発を進めています」と鈴木 氏は現状を解説。小林 氏も「実際のパフォーマンスに関しては現在チューニングを進めているところですが、すでに負荷をかけても快適かつスピーディに動くことが確認できています。こうした変化も踏まえて、使う側としても活用の幅を広げていかなければならないと感じています。これまでは、分析環境のキャパシティを考えてブレーキをかけていた部分もあり、

今後はそうした意識を変革して、より効果的なデータ利活用につなげていきたいと思っています」と力を込めます。また鈴木氏は、Azure Databricks を採用したことで、アカウント管理や情報セキュリティ監査の面でも大きなメリットが得られたと語ります。

「通常のデータベースでは、踏み台のアカウントと DB のアカウントを別々に管理する必要がありますが、今回は Azure AD で紐付けできたため、一元的に管理できるようになっています。また当社のデータ分析環境では、データを入れる、または外に出す場合に情報セキュリティ監査部門のチェックロジックが働いていますが、これまではすべて手作業で行っていました。業務改善のため、コンプライアンスプロジェクトとして、データ分析環境構築と同時に進行了。これにより、運用・保守・メンテナンスの面で大幅なコスト削減が実現できることが見えてきています」(鈴木氏)。

データ分析環境が安定稼働した後の展望としては、リアルタイムでのデータ連携可能なデータハブ構想を実現させていきたいと鈴木氏。次のように今後の「データハブ化構想」について語ります。

「現在は1日1回のデータ更新ですが、今後は本プロジェクトの始動当初から実現したいと考えていた“リアルタイムでのデータの連携”に注力し、システム連携を強化していきたいと考えています。今回構築したプラットフォームではパーソルキャリア全体のデータを安全に管理していきたいと思っています」(鈴木氏)。



小林氏も「現状の分析・レポートではデイリーで満足している面もありますが、プラットフォームとしてリアルタイム性が付加されるのであれば、データの利活用は大きく変わると思います。現在のレコメンデーションはデイリーで回していますが、これを午前・午後でモデルを変えられるようにすれば、顧客体験の向上が実現できると思います」とリアルタイム性の向上による効果を期待しています。

また今回のプロジェクトは、パーソルキャリアが目指すマルチクラウド化を実現するための取り組みとしても有効と鈴木氏は語り、Azure 利用の知見を得たことで、システム基盤として Azure を選択するうえでのハードルを下げるのができたと喜びを口にします。今回の開発においては Azure DevOps が採用されており、同社ではこの経験も踏まえて、内製強化を進めていきたいと考えています。

パーソルキャリアでは、今後も先進的な技術の活用に取り組んでいく予定で、マイクロソフトが提供する統合データガバナンスサービス「Microsoft Purview」や、ChatGPT などを活用した AI サービスを簡単に利用できる「Azure OpenAI Service」などに注目しているといいます。同社が取り組むデータ利活用と、そのなかで使われるマイクロソフトのソリューションには、今後も注視していく必要がありそうです。

“現状の分析・レポートではデイリーで満足している面もありますが、プラットフォームとしてリアルタイム性が付加されるのであれば、データの利活用は大きく変わると思います。現在のレコメンデーションはデイリーで回していますが、これを午前・午後でモデルを変えられるようにすれば、顧客体験の向上が実現できると思います”

小林 裕也 氏：P&M 本部 データビジネス部 ビジネスプランニンググループ シニアエンジニア（データ）