

生成 AI の価値を最大限に引き出すための データ基盤

アマゾン・ウェブ・サービス・ジャパン合同会社

2024/05/16



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Agenda

- 10:00-10:10 生成 AI の力を引き出すためのデータアーキテクチャ
- 10:10-10:35 生成AIデータ基盤としてのAWSマネージドデータベースサービス
- 10:35-10:40 休憩
- 10:40-11:10 生成AIデータ基盤としてのAWSマネージドアナリティクスサービス
- 11:10-11:50 生成 AI アプリケーションでデータを活用



生成 AI の力を引き出すためのデータアーキテクチャ

程 家

SENIOR SOLUTIONS ARCHITECT

2024/05/16



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



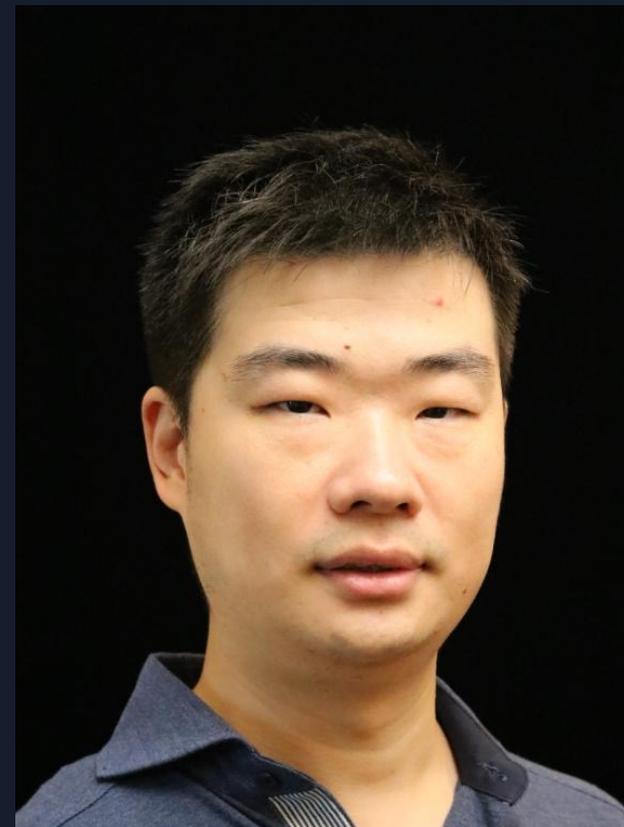
自己紹介

名前：程家 (てい か)

所属：アマゾン ウェブ サービス ジャパン
シニアソリューションアーキテクト

経歴：外資系ソフトウェアベンダーにて技術コンサルタントを従事

現在は、サービス業のお客様を中心に技術支援を行うとともに
AWSが提供するマネージド型データベース関連の技術支援





生成 AI アプリケーション

= LLM ?



生成 AI アプリケーション

ストレージ
構造化データと非構造化データ

運用データベース
SQL、NoSQL、ドキュメント、グラフ、ベクトル

分析とデータレイク
検索、ストリーミング、バッチ、インタラクティブ

データ統合
取得、変換、ストリーミング

データガバナンス
カタログ、品質、プライバシー、アクセス制御



What every CEO should know about generative AI

May 12, 2023 | Article



「...ジェネレーティブAIの生命線は、特定のビジネスコンテキストや問題のために磨き上げられたデータへのスムーズなアクセスです。データを効果的に処理して、すぐにアクセスできるようにする方法をまだ見つけていない企業では、生成AIを微調整して、潜在的に変革をもたらす可能性のある用途をさらに引き出すことはできません...生成AIから得られるビジネス価値と競争上の優位性に基づいた、明確なデータおよびインフラストラクチャ戦略が不可欠です。」

Source: McKinsey & Company, "What every CEO should know about generative AI," May 2023 : <http://tinyurl.com/4v28a4f9>.



データは差別化要因となる



汎用的な
生成 AI



自社のビジネスと顧客を
理解している生成 AI



よりパーソナルな バーチャルエージェント

目的

オンライン旅行代理店において、
パーソナライズされた旅行日程を作成する

データ

顧客プロフィール

 過去の旅行歴

 ウェブ履歴

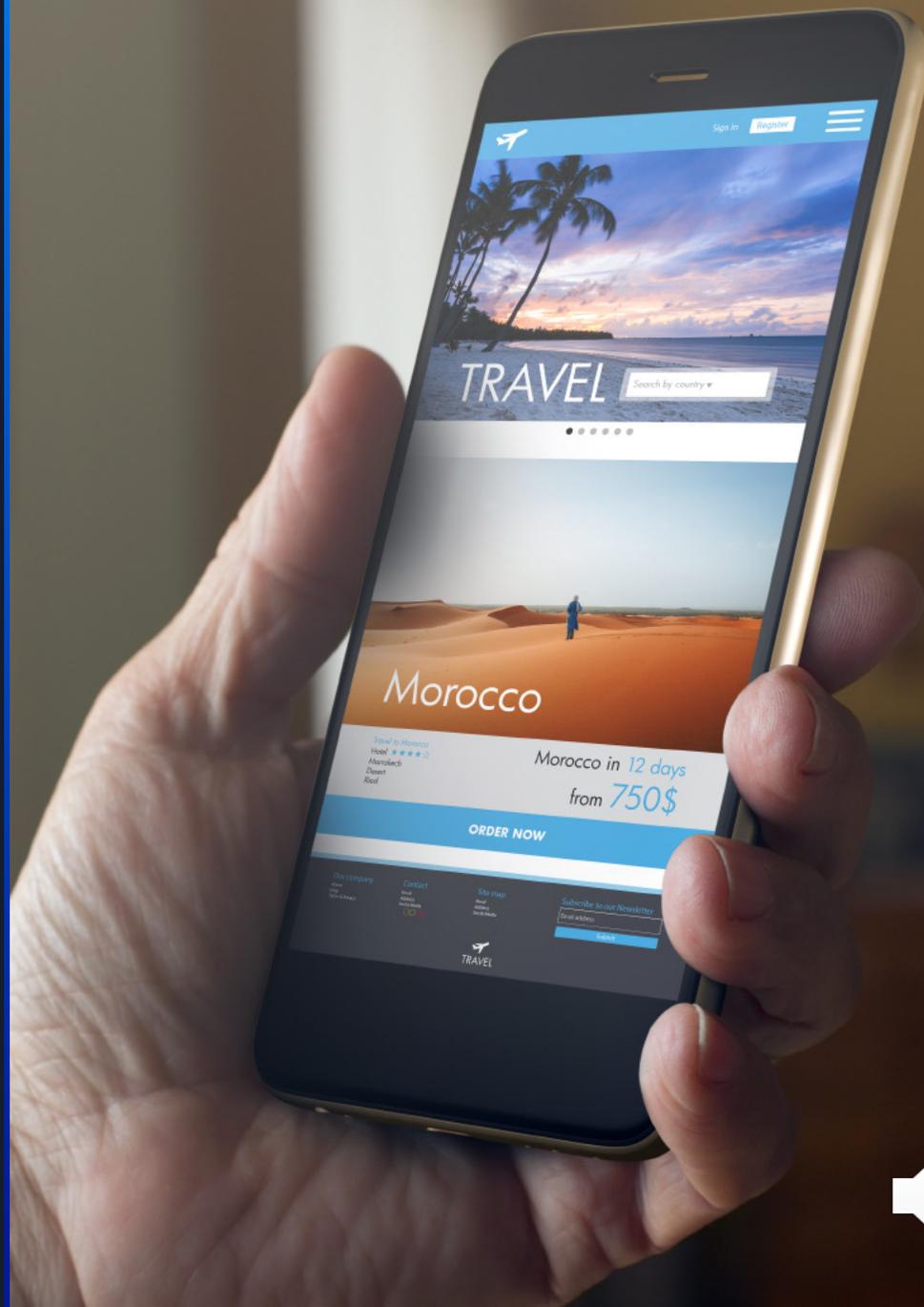
 旅の好み

企業データ

 飛行機とホテルの空き状況

 プロモーション

 似た旅行者の情報



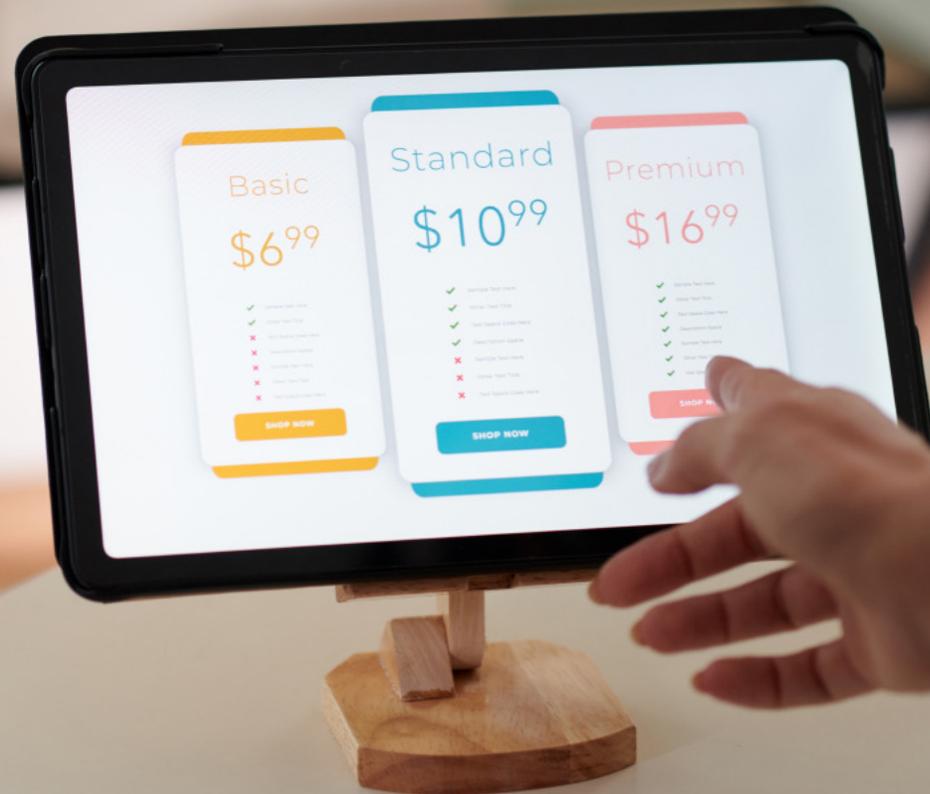
ブランドに関するコンテンツ

目的

会社のブランドで定められたサウンド、見た目や操作感に合わせて、変化する在庫状況に応じたマーケティングコンテンツと広告を作成する

データ

-  リアルタイムの在庫状況
-  ブランドガイドライン
-  過去の広告やコンテンツ



より関連性の高い 対話型検索

目的

企業データに基づいて
顧客と従業員に回答を提供する

データ

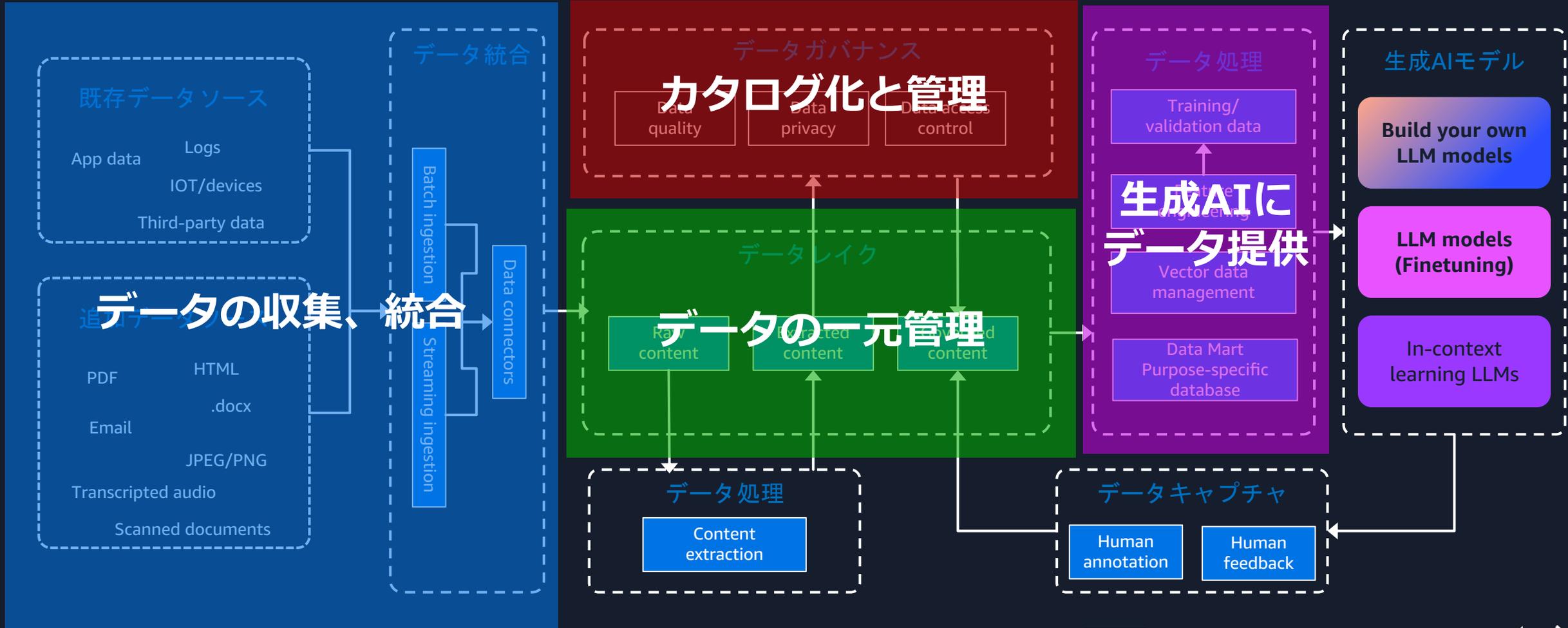
 Slack

 内部ガイド

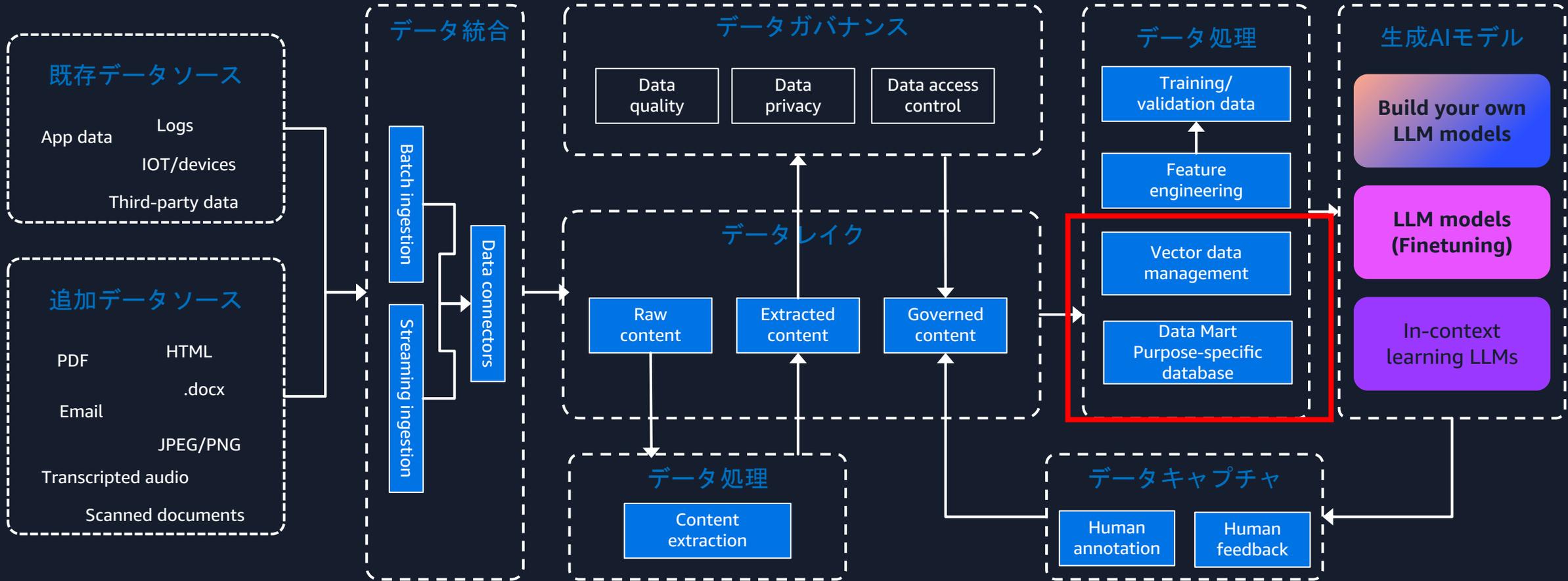
 SharePoint



生成 AI アプリケーションにおける AWS データサービスの役割



本セッションの範囲



Net new capabilities for generative AI



新たな生成 AI のデータパターン

1



検索拡張生成 (RAG)

事前にトレーニングされたモデルに、自社のドメイン固有のコンテキストデータを加えた結果を生成

例: ドメイン固有の、要件が限定された仮想エージェント

2



既存モデルの ファインチューニング

既存のモデルをベースに、ドメイン固有のデータによる追加学習を実施

例: ドメイン特化型の知識エージェント

3



独自モデルの作成

ドメイン固有データによるモデルのトレーニングを実施

例: 特定のデータで学習を行った、高度にドメインに特化したアプリケーション

新たな生成 AI のデータパターン

1



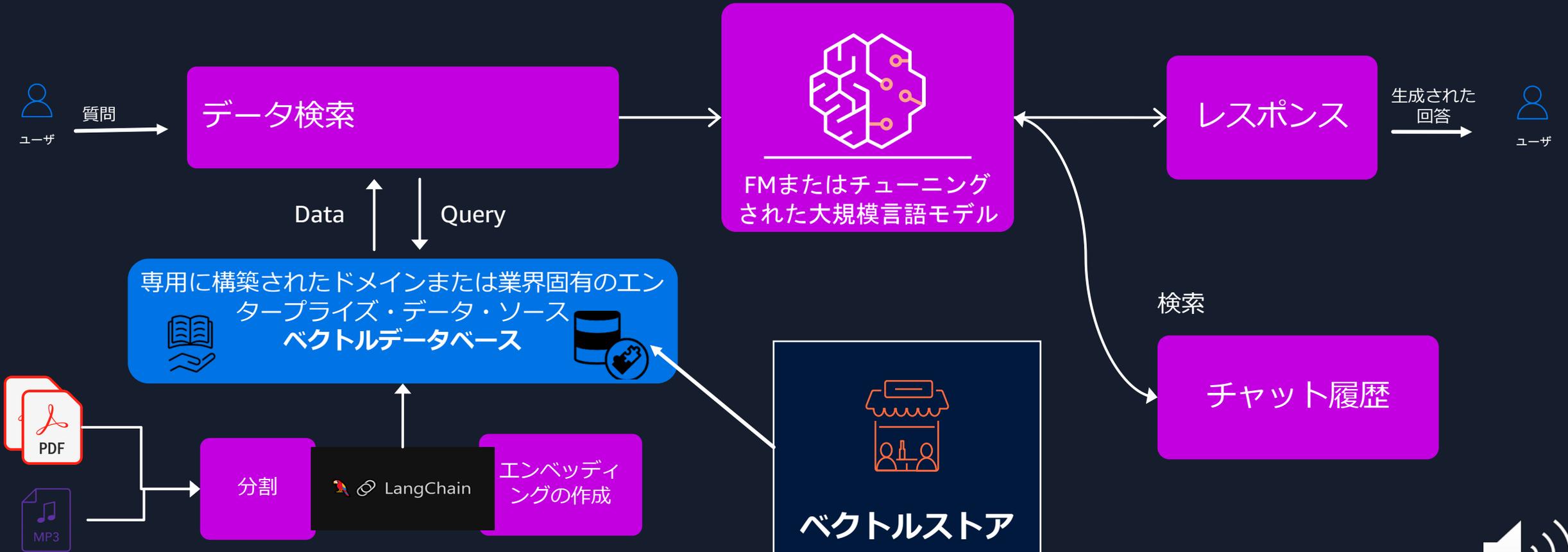
検索拡張生成 (RAG)

事前にトレーニングされたモデルに、自社のドメイン固有のコンテキストデータを加えた結果を生成

例: ドメイン固有の、要件が限定された仮想エージェント

Retrieval Augmented Generation (RAG)

RAGアプローチは、ドメイン固有の企業知識ベースから、ユーザーの要求に最も関連する情報を検索する



ベクトル埋め込みとは

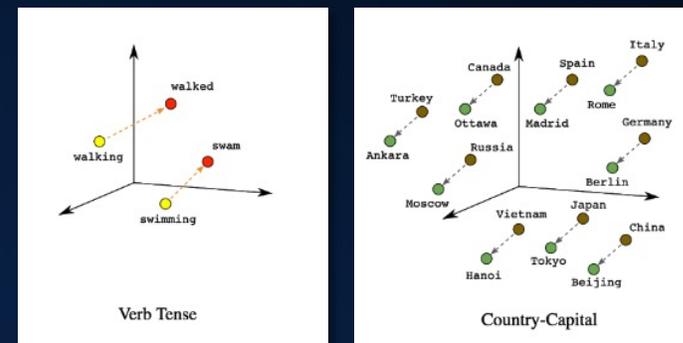


セマンティック要素:

- 単語、フレーズ
- 段落、文書
- シーン、曲セクション
- 顔、検出された画像内の要素
- その他...



0.35 0.1 0 0.9 001.0 00 0001.0 0 0 0...
0.35 0.1 0 0.8 001.0 00 0001.0 0 0 0...
0.15 0.1 0 0.7 001.0 00 0001.0 0 0 0...

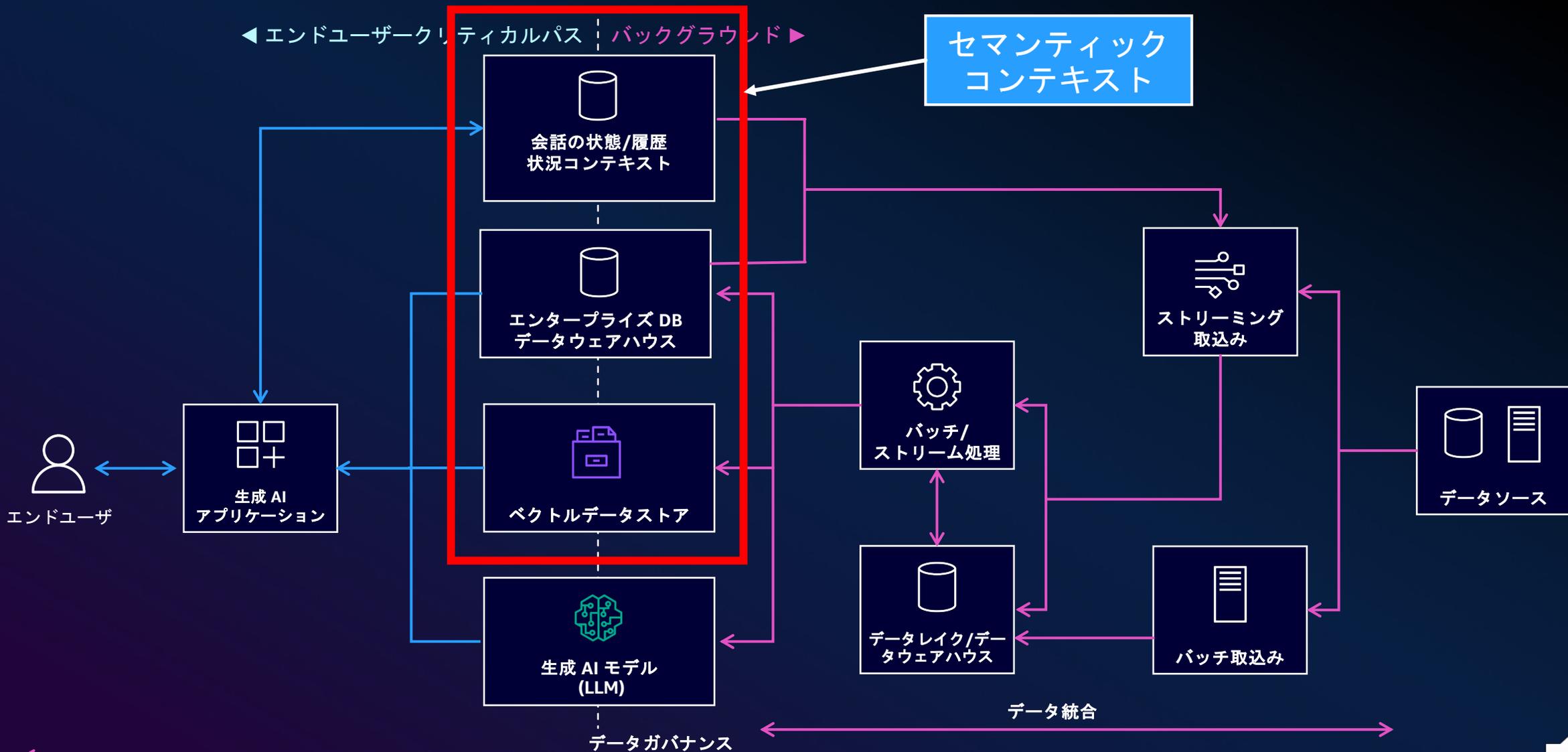


3D 簡略表示。埋め込みには数千のディメンションがあります。
ソース:<https://daleonai.com/embeddings-explained>

埋め込み:ベクトル要素が意味を持った表現であることにより、生成 AI で使用可能に

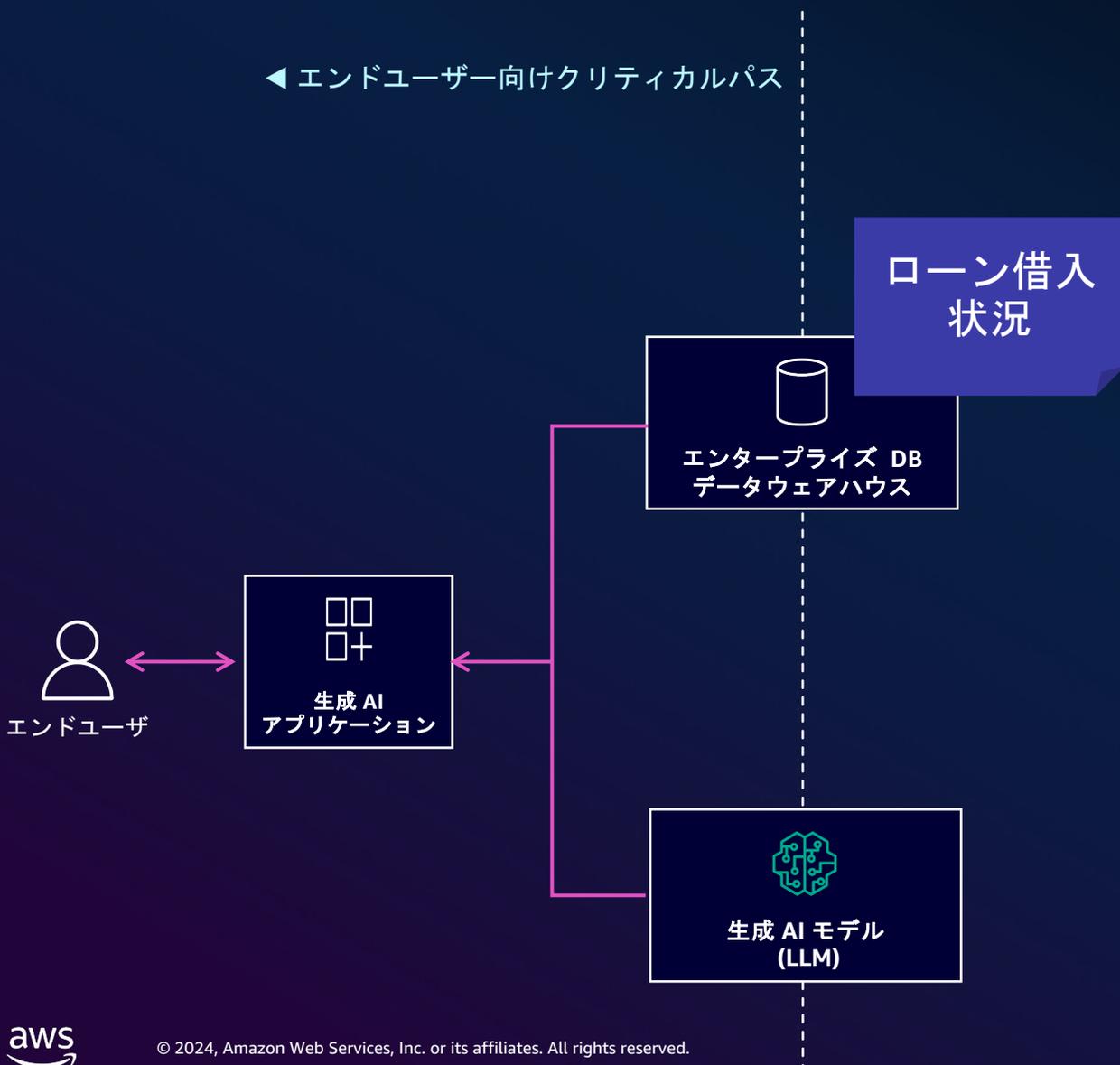


RAG リファレンスアーキテクチャ



エンドユーザー側のデータの流れ

◀ エンドユーザー向けクリティカルパス



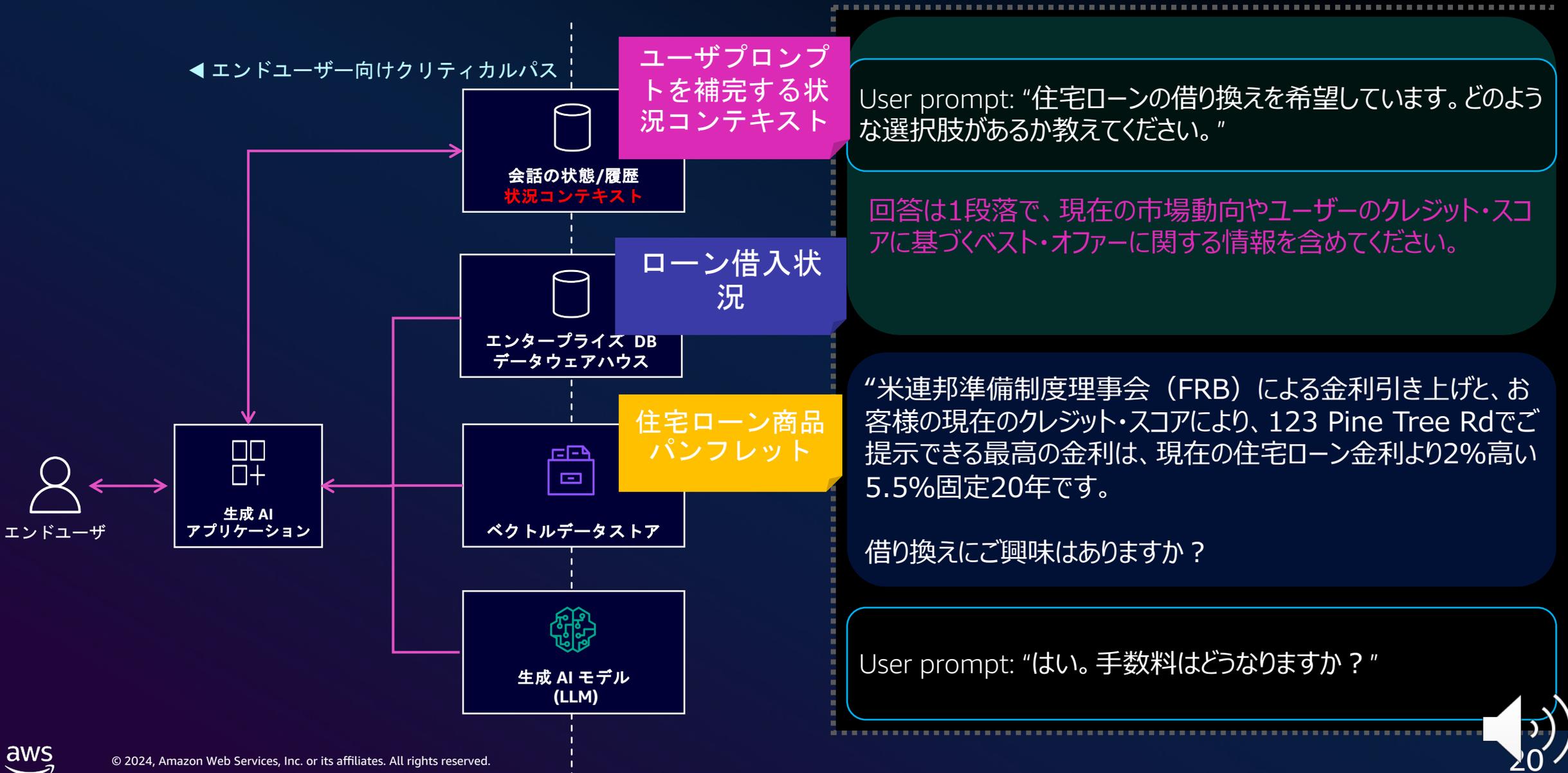
User prompt: "住宅ローンの借り換えを希望しています。どのような選択肢があるか教えてください。"

"現在お住まいの123 Pine Tree Rdの住宅ローンは、年利3.5%の固定金利で、2024年8月1日現在、元本残高220,000ドルを今後6年4ヶ月で支払うことになっています。

この住宅ローンを借り換えたいのですか？"

User prompt: "はい、その通りです。"

エンドユーザー側のデータの流れ



新たな生成 AI のデータパターン

2



既存モデルの ファインチューニング

既存のモデルをベースに、
ドメイン固有のデータによる
追加学習を実施

例: ドメイン特化型の知識エージェント

3



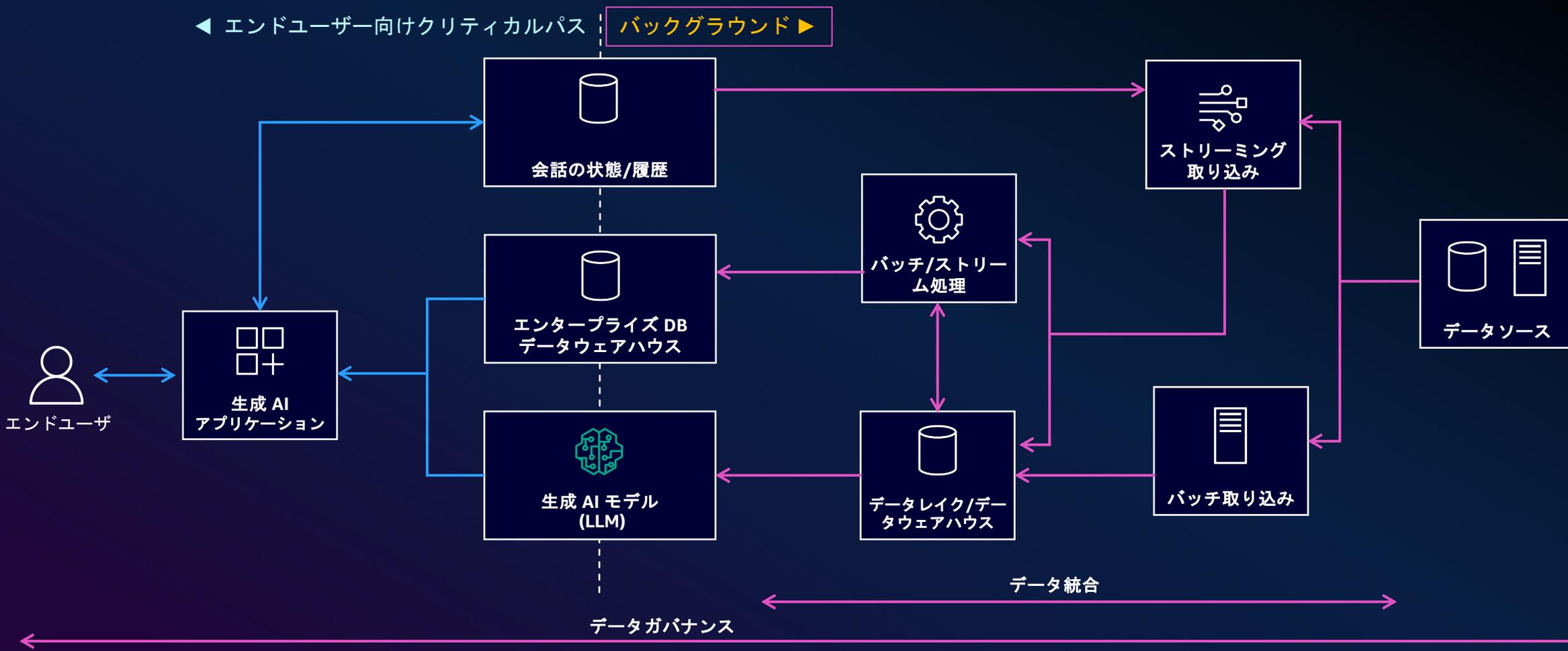
独自モデルの作成

ドメイン固有データによる
モデルの関するモデルのト
レーニングを実施

例: 特定のデータで学習を行った、高
度にドメインに特化したアプリケー
ション



LLM のファインチューニングと独自モデルの学習



生成AIデータ基盤としてのAWSマネージドデータベースサービス

マネージド型で提供されるデータベースと分析サービス



Aurora



RDS



DynamoDB



DocumentDB



ElastiCache



MemoryDB



Keyspaces



EMR



Elasticsearch Service



Managed Streaming for Kafka



Relational databases



Non-relational databases

Hadoop and Spark



Operational analytics



Real-time analytics

ORACLE

MySQL

Microsoft SQL Server

PostgreSQL

MariaDB

mongoDB

Apache CASSANDRA

redis

M

hadoop

APACHE SPARK

elasticsearch

logstash

kibana

kafka

マネージドサービスの有用性

生産性

下記のようなお客様のビジネスに直結しない作業を AWS のインフラストラクチャに任せることができます。

- ハードウェアの設置
- ミドルウェアの導入と設定
- パッチ適用
- スケーリング

AWS は 24時間365日体制でモニタリング含めてインフラストラクチャを維持管理し、お客様のエンジニアがビジネスに直結する作業に集中する時間を産み出します。

俊敏性

AWS ではデータの利活用を進めるために下記のようなサービスを提供しています。

- データ統合
- データレプリケーション
- ダウンタイムを最小化したアップグレード

AWS 上でお客様のデータベースが稼働していれば、上記のサービスを活用してスムーズにお客様のデータを統合し、利活用を進める事ができます。

コスト効率

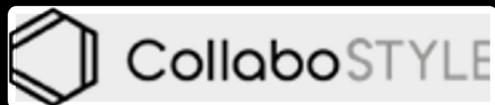
マネージドサービスを利用するインスタンスとして Graviton インスタンスを使用することで最大で 30% のコスト削減が可能になります。

利用するサービスによっては、Amazon OpenSearch Service における UltraWarm ストレージのように大量データを低コストで格納することが可能です。
(90% 程度のコスト削減)

コラボスタイル、Aurora Serverless v2 の採用 でスケールアップ時のダウンタイムをゼロに。 負荷の高かったスケールアップ作業からも解放



お客様プロフィール



INDUSTRY
ISV

COUNTRY
日本

「ワークスタイルの未来を切り拓く」を理念に、より良い仕事環境を築くためのワークフローツール「コラボフロー」などのSaaS製品を提供しています。



ビジネスの課題

- ユーザー数の増加に伴いデータベース負荷が高くなり、Aurora PostgreSQLの頻繁なスケールアップとそれに伴うダウンタイムが発生。監視や検証作業などの運用工数が増大
- スケールアップを行うとリザーブドインスタンス (RI) によるコスト最適化効果が下がってしまうため、RI をどのように購入するか精度の高いキャパシティ予想が必要に。



ソリューション

- Aurora Serverless v2を採用で、ダウンタイム削減と運用負荷削減へ
- Aurora Serverless v2 へはリーダーインスタンスでAurora Serverless v2の効果を確認しながら段階的に移行することで、移行に伴う検証負荷も最小限に



導入効果

- スケールアップ時のダウンタイムがゼロに。スケールアップに伴う検証作業や監視なども不要になり、負荷の高かったスケールアップ作業からも解放
- Aurora Serverless v2への変更による性能影響や、スケールアップ時のパフォーマンス劣化は発生せず
- Aurora Serverless v2への移行もインスタンスタイプの変更のみ

“ Aurora Serverless v2導入で、スケールアップ時の運用負荷が削減でき、SREチーム本来の業務に集中できるようになりました。 ”

成島 丈実 氏

株式会社コラボスタイル 開発部 チームマネージャー



ヤオコー、システム内製化を推進し、AWS 上に展開したデータ基盤で店舗運営の効率化と顧客満足度向上を推進



お客様プロフィール



INDUSTRY

小売

COUNTRY

日本

埼玉県を中心に関東圏に 200 店舗近く展開するスーパーマーケットチェーン。



ビジネスの課題

- IT 環境の効率向上とシステム構築の外部依存からの脱却



ソリューション

- 自社でハンドリングできる AWS 環境を準備：各種システムと接続し、Amazon S3 や Amazon EC2、AWS Lambda、Amazon RDS などを利用して、他システムへ連携できるプラットフォームを構築



導入効果

- 全社横断のデータ連携基盤の構築
- 内製開発組織の組成と育成
- インフラコストの最適化
- ソフトウェアプロダクトの開発スピード向上

“ データ連携基盤を内製化したことにより、業務アプリのマイクロサービス化、デジタル実験店舗での検証、他店舗への展開といったサイクルがスピーディに回るようになりました。 ”

小笠原 暁史 氏

株式会社ヤオコー デジタル統括部 部長



ワンダープラネット株式会社、 人気スマートフォンゲームのデータベースに AWS Graviton プロセッサを採用 アクセス集中時の処理能力向上とコスト削減を両立



お客様プロフィール



INDUSTRY
ゲーム

COUNTRY
日本

『アリスフィクション』『クラッシュフィーバー』『ジャンプチヒーローズ』等、世界に向けてスマートフォン向けアプリ・ゲームの企画・開発・運営を行う、名古屋に本社をおくエンターテインメントサービス企業。



ビジネスの課題

- アクセス集中時の機会損失
- 動作担保のためのインフラコスト
- インフラ配備・監視などの運用負担



ソリューション

- マネージドデータベースによる生産性向上
- インフラコストの削減



導入効果

- アクセス集中時のパフォーマンス向上
- 年間 **13%** のデータベースサーバーコスト削減
- 新規タイトルにも AWS Graviton を採用

“ AWS Graviton プロセッサを採用した結果、クラッシュフィーバーでは明確に負荷対応の軽減とコスト削減ができました。移行のハードルもないため、安心してほかのプロダクトへの導入を考えられました。

楠善一 氏
ワンダープラネット株式会社 執行役員 VPoE 兼 EDMO 室長



マネージドサービスへの移行 - NoSQL/インメモリデータベース

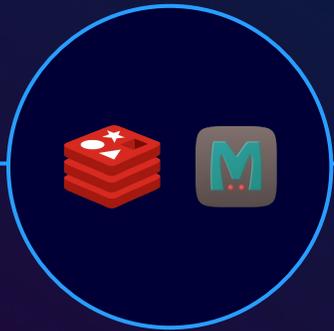
オンプレミス/お客様が管理する NoSQL/インメモリデータベースのマネージドサービスへの移行

- 既存のアプリケーションを変更する事なく移行
- データベースの運用管理負荷を軽減
- パフォーマンス、可用性、スケーラビリティ、セキュリティを向上



Amazon ElastiCache

Redis &
Memcached compatible



オープンソースの
Redis/Memcached と
完全互換

Extreme
Performance



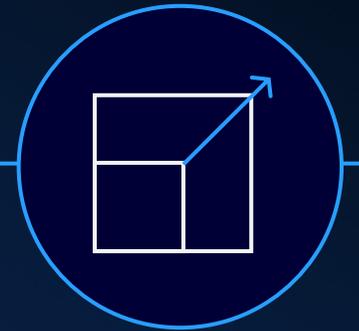
ミリ秒の
レスポンスを実現する
インメモリデータストア/
キャッシュ

Fully
Managed



AWS がハードウェア、ソフ
トウェアの導入設定及び構
成、モニタリングを実施

Easily
Scalable



書き込み及び読み取り
がスケール可能

Amazon MemoryDB for Redis

Redis
compatible



オープンソースの
Redis と完全互換

Ultra-fast
Performance



マイクロ秒の読み込みと
ミリ秒の書き込みレイテン
シを数百万 TPS と両立

Durability and
High Availability



耐障害性及び高可用性を
実現するため
Multi-AZ 構成が可能

High
Scalability



クラスタあたり 100TB 以上
のストレージを提供



ベクトル検索
に対応

Amazon OpenSearch Serverless

クラスターの管理なしに全文検索
とログ検索を実施



管理が容易

クラスターのサイジング、スケーリング、チューニング、シャードとインデックスのライフサイクル管理が不要に



速度

リソースを自動的にスケールし、高速なデータ取り込みレートとクエリ応答時間を一貫して維持



エコシステム

既存の OpenSearch クライアント、パイプライン、API を使用して数秒で利用を開始できる

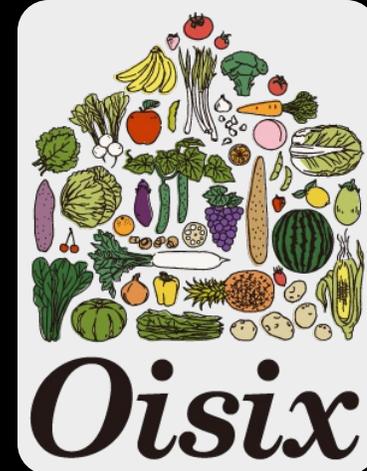


費用対効果

事前のリソースプロビジョニングは不要



Oisix ra daichi、Amazon RDS for Oracle の負荷削減のため Amazon ElastiCache を採用。検討からアプリケーションの改修、導入まで約1ヶ月で実施し、Amazon RDS for Oracle の CPU使用率を 20% 以上削減



お客様プロフィール

Oisix ra daichi

デジタルネイティブビジネス

日本

有機野菜やミールキットなど、安心安全でおいしい食材を宅配するネットスーパー。ミールキットの先駆けである Kit Oisix は累計1億4000万食を突破。近年はフードロス削減等に取り組み、サステナブルな食卓の実現を目指している。



ビジネスの課題

- オンプレミスのシステムを2021年に AWS に移行し、データベースも RDS for Oracle に移行。コロナ禍による需要増により、移行後の CPU 使用率が高い状態で推移していた。ピーク時にはビジネスへの影響も出ている状況で、SQL やデータベースチューニングで対応していたが、更なる改善が必要な状況だった。



ソリューション

- RDS for Oracle に対するチューニングを実施しつつ、負荷の高い処理のデータをキャッシュするために ElastiCache を採用。
- ElastiCache の採用検討からアプリケーションの解析、修正、実装までの期間は約1ヶ月だった。



導入効果

- CPU 使用率は、プライマリで約20%、Read Replica は約50%削減。Read Replica をスケールダウンすることで、コスト削減を実現。
- ElastiCache 導入前はピーク時間帯に Oracle への負荷が上昇していたが、導入後は Oracle のピーク時間帯がなくなり、サービスとしても安定させることができた。
- 今後、プライマリのスケールダウンを実施し更なるコストダウンを見込む。

“

キャッシュ効果が高い箇所とリアルタイムデータの必要性を見極めて ElastiCache でキャッシュする、というアプリケーションとして当たり前のことをしただけです。

”

川上 徹氏

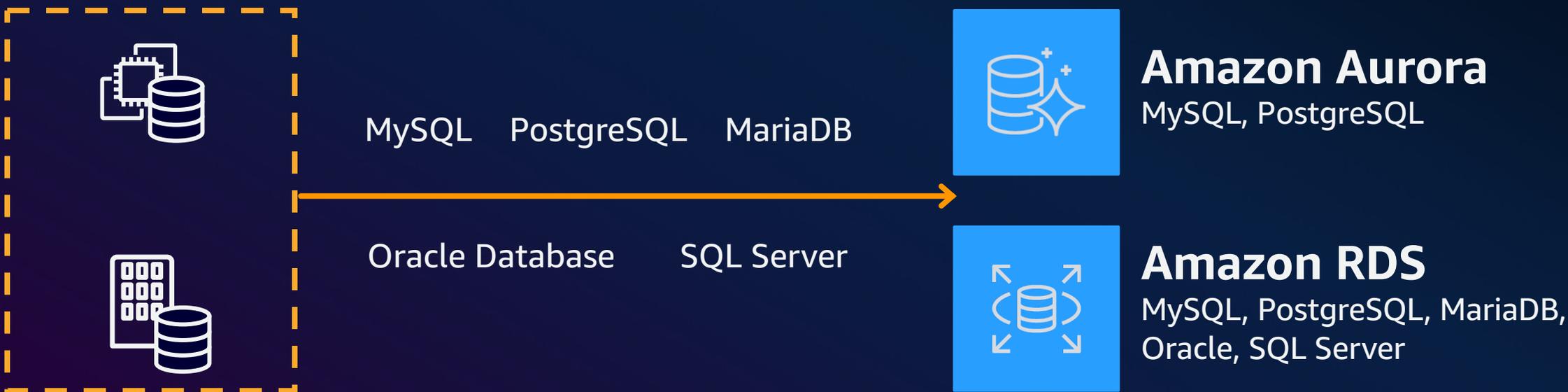
オイシックス・ラ・大地株式会社 システム本部 リードエンジニア



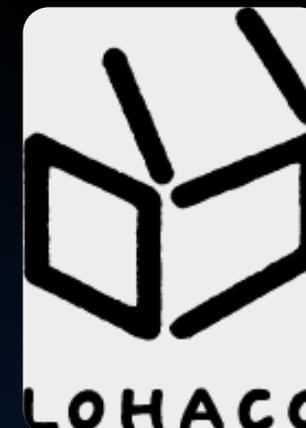
マネージドサービスへの移行 - リレーショナルデータベース

オンプレミス/お客様が管理するリレーショナルデータベースのマネージドサービスへの移行

- 既存のアプリケーションを変更する事なく移行
- データベースの運用管理負荷を軽減
- パフォーマンス、可用性、スケーラビリティ、セキュリティを向上



アスクル、ECサイトの LOHACO を Oracle RAC から RDS for Oracle に移行することで運用コストを 8% 削減し、オンプレミスと比較してパフォーマンスも最大約 10% 改善



お客様プロフィール



INDUSTRY
DNB

COUNTRY
日本

アスクルは「お客様のために進化する」というDNAのもと「仕事場とくらしと地球の明日に「うれしい」を届け続ける」ことをパーパス（存在意義）として、事業所向け通信販売サービス“ASKUL”や個人向け通信販売サービス“LOHACO”を展開している会社です。



ビジネスの課題

- LOHACO では、フロント側を AWS、Oracle RAC をオンプレミスで運用していたため、ディスク追加作業や半年に一度の定期アップグレード作業など、サーバー運用担当者の作業負荷が高かった
- AWS へ移行することによる Oracle Database のライセンスコスト増加や、可用性、負荷分散、移行にかかる時間が問題となった



ソリューション

- 移行先は、マネージドサービスによる運用負荷削減を考慮して RDS for Oracle を前提に検討
- ライセンスの最適化を行うために AWS が提供する Database Freedom Workshop のパフォーマンス分析で、現状の約 1/6 に削減できることを確認。可用性はマルチ AZ 構成、負荷分散は一部の機能を Aurora PostgreSQL に移行することで実現。
- データ移行は、Database Link 経由の Data pump とマテリアライズドビューの採用で 3TB のデータ移行時間を削減



導入効果

- 2ヶ月の検証と4ヶ月の移行作業後、2022年9月に移行を実施。移行は6時間で実施し、移行後のトラブル等は発生せず
- AWS 移行により、インフラ保守などの運用コストを約 8% 削減
- RDS for Oracle への移行と全てのサービスが AWS 環境に集約された結果、パフォーマンスがオンプレミスと比較して最大約 10% 改善

“ AWS 移行で、インフラのコード化が促進されたことなどにより、拡張性の低さやアップグレードの保守工数が改善、開発の俊敏性向上など、大きなメリットを感じています。また、困った時も Sales や SA に迅速に対応して頂くことができ、大変助かっています。

川合 陵平 氏

アスクル株式会社 EC プロダクト本部 エンジニアリング 5 部 Dept



マクロミル、SQL Server を Aurora PostgreSQL に 移行することでデータベースのライセンスコストをゼロに。 メンテナンスに要していた時間も30%削減。

Build your Data Culture

お客様プロフィール



INDUSTRY
ISV/SaaS

COUNTRY
日本

国内オンラインリサーチ業界のリーディングカンパニー。豊富なリサーチ実績とノウハウ、90以上の国と地域でリサーチ可能なグローバルネットワーク、独自に構築した消費者パネルから得られる多様なデータを活用し、お客様のマーケティング課題の解決に向けて最適なソリューションを提供。



ビジネスの課題

- 処理量の増加に伴い、意図しないロック等により、処理が正常に完了しない障害が発生しており、繁忙期の深夜待機など非生産的な業務に工数が発生していた。
- データ容量増加に伴いインスタンスを拡張する必要があったが、商用ライセンス数の制限によりインスタンスの拡張ができなかった。
- EC2上でSQL Serverをインストールし運用しており、パッチ適用やストレージ拡張などのメンテナンスにも工数を要していた。



ソリューション

- 可用性や柔軟性、性能、コストなどの観点から、Aurora PostgreSQLの採用を検討。PoCの結果、繁忙期の負荷で現環境より安定的に稼働できることを確認し、移行を決断。
- 海外の子会社（2022年当時）のエンジニア3名と共に300本以上のプロシージャを3ヶ月で移行。
- AWSからデータベースに強いスペシャリストのアサインなど、課題に対する適切な支援により移行作業の効率化が図れた。



導入効果

- Aurora PostgreSQLへの移行によりシステムの可用性の向上、ライセンスコストの削減を実現。
- パッチ適用工数や、繁忙期での1ヶ月の夜間待機など、非生産的な作業工数がほぼゼロに。
- Database Freedom Workshopで、Aurora PostgreSQLや移行について深い知識を得ることができ、運用後も安定した運用を実現。

“**メンテナンス作業や、繁忙期の夜間待機など非生産的な作業に多くの工数が発生していました。Auroraへの移行でこのような作業がなくなり、本来フォーカスすべき取り組みに集中できるようになりました。**”

浦濱暢介氏

株式会社マクロミル グローバルテクノロジー本部



農林中央金庫、JAバンクの大規模基幹システムをAWSに移行。Amazon Aurora への更改などでTCOを13年間で100億円以上削減見込み



お客様プロフィール



業種
金融

国名
日本

農業協同組合（JA）、漁業協同組合（JF）、森林組（JForest）を基盤とする全国金融機関。会員、農林水産業者、農林水産業に関連する企業等への貸出を行うとともに、国内外で多様な投融資を行い、会員への安定的な収益還元に貢献する。



ビジネスの課題

- 情報系をクラウド移行し、経営環境の変化に伴うコスト削減要求の高まりに対応
- ビジネスの加速に対応する拡張性・柔軟性の向上、他システムとの接続容易性の実現



ソリューション

- コスト競争力と可用性を両立するAmazon Auroraを採用
- AWS プロフェッショナルサービスを活用「セキュリティアーキテクチャ実装ガイドライン」を策定



導入効果

- 13年間のTCO削減見込みが100億円以上
- DBのライセンス費用を3割削減
- AWSに移行したデータ量が25TB

“

大規模システムのクラウド移行を成功させたことは、IT部門全体にとってエポックメイキングな出来事で、カルチャーを大きく変えるきっかけになりました。

”

半場 雄二 氏

農林中央金庫 理事 兼 常務執行役員 IT 統括責任者



データ基盤のための 包括的なサービス群

変革と統合

保存、クエリ、分析

アクション

データソース

アプリケーション

分析 & ML

IoT/デバイス

ベクトルデータ型を
サポートした RDBMS

データウェアハウス

機械学習

ユーザー

アプリ/ログ

ベクトルデータ型を
サポートした NoSQL DB

データレイク

生成 AI

アプリ

第三者データ

ストリーミングサービス

ビッグデータ

ビジネスインテリジェンス

デバイス

カタログ化と管理

Thank you!

