

アーキテクチャ図

GenerativeAI での取得拡張生成のためのリアルタイムストーリーミングの探索



GenerativeAI での取得拡張生成のためのリアルタイムストーリーミングの探索: アーキテクチャ図

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon の商標およびトレードドレスは Amazon 以外の製品およびサービスに使用することはできません。また、お客様に誤解を与える可能性がある形式で、または Amazon の信用を損なう形式で使用することもできません。Amazon が所有していないその他のすべての商標は Amazon との提携、関連、支援関係の有無にかかわらず、それら該当する所有者の資産です。

Table of Contents

GenerativeAI での取得拡張生成のためのリアルタイムストリーミングの探索 i

GenerativeAI での取得拡張生成のためのリアルタイムストリーミングの探索 1

詳細情報 2

寄稿者 2

図の履歴 3

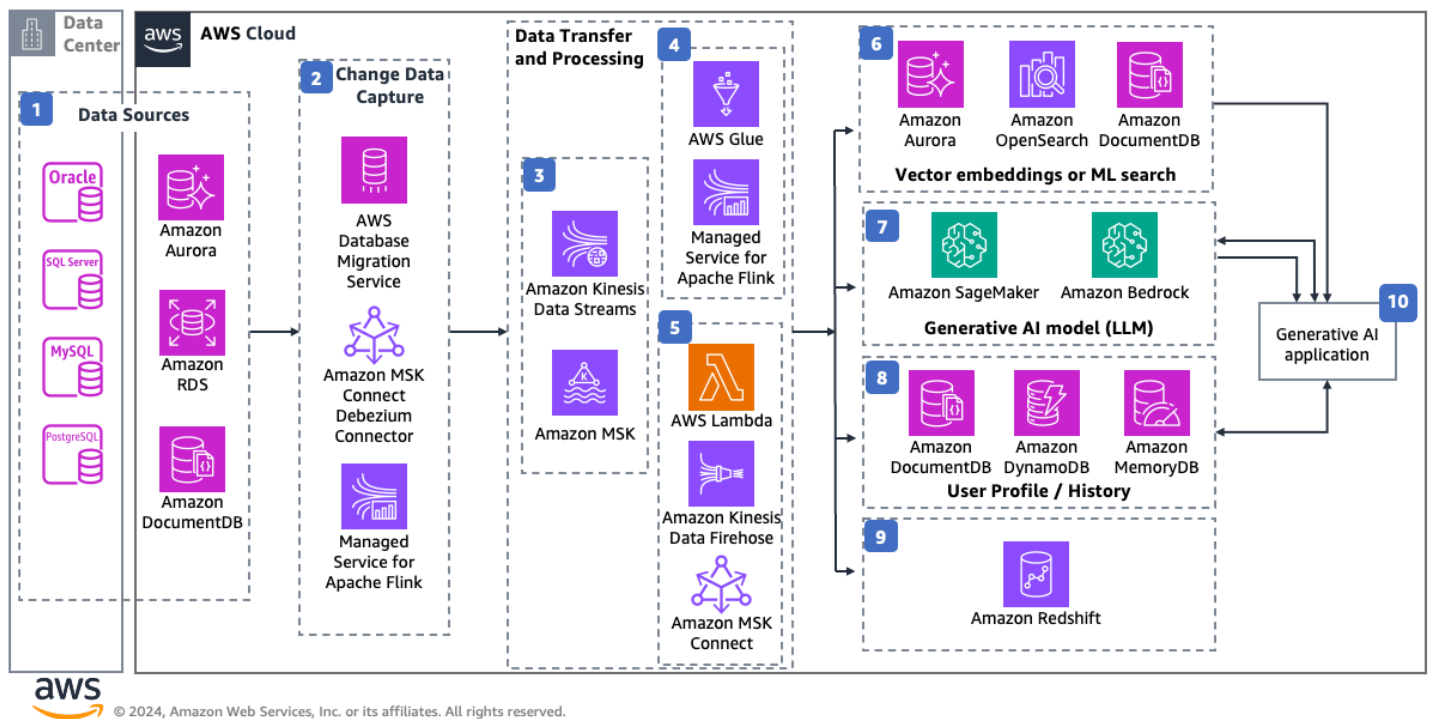
..... iv

GenerativeAI での取得拡張生成のためのリアルタイムストリーミングの探索

公開日: 2024 年 8 月 12 日 ([図の履歴](#))

このアーキテクチャは、生成 AI アプリケーションの AWS でのストリーミングデータサービスと検索拡張生成 (RAG) の統合を示しています。

GenerativeAI での取得拡張生成のためのリアルタイムストリーミングの探索



1. 変更データキャプチャ (CDC) のデータソースには、Oracle、SQL Server、MySQL、PostgreSQL、Amazon Aurora、Amazon RDS などのオンプレミスまたは AWS データベースが含まれ、すべてのデータが取得拡張生成 (RAG) モデルにファネルされます。
2. AWS Database Migration Service および Amazon MSK Connect with Debezium コネクタは、データベースの 1 回限りのデータ移行と継続的なデータレプリケーションに役立ちます。AWS Database Migration Service および Amazon MSK Connect with Debezium コネクタは、ソースデータベースから変更をキャプチャしてストリーミングし、キャプチャされたのと同じ順序でターゲットに適用します。

3. AWS Database Migration Service と Amazon MSK Connect with Debezium コネクタを使用すると、Amazon Kinesis Data Streams または Amazon Managed Streaming for Apache Kafka (Amazon MSK) へのデータのストリーミングが可能になり、大規模なストリームの収集と処理が容易になります。
4. AWS Glue Spark Streaming と Amazon Managed Service for Apache Flink を利用することで、データ消費要件を満たすための特殊なデータ処理パイプラインを構築できます。
5. データシンクサービスとして機能する AWS Lambda、Amazon Kinesis Data Firehose、Amazon MSK Connect を使用すると、Data Lake や Amazon Redshift などの宛先にソースデータを直接転送できます。
6. pgvector、Amazon Opensearch、Amazon DocumentDB で Amazon Aurora PostgreSQL を利用すると、ベクトル埋め込みを生成して、効率的なデータ取得、ベクトル表現管理、スケーラビリティ、リアルタイム推論機能を実現できます。 Amazon DocumentDB
7. Amazon SageMaker と Amazon Bedrock は、コーパス内の関連情報を検出し、ベクトル化されたドメイン固有のデータセットに対して類似度検索を実行し、このデータを生成モデルの入力として使用する手段を提供します。
8. ユーザープロファイルと会話履歴を維持するために、Amazon DocumentDB、Amazon DynamoDB、Amazon MemoryDB には適切なオプションが用意されています。
9. Amazon Redshift サービスを活用してデータの永続性を確保し、生成 AI RAG モデルのデータ入力を強化します。
10. RAG アプリケーションの仕組みの詳細については、[「検索拡張生成 \(RAG\)」](#)を参照してください。

詳細情報

詳細については、「」を参照してください。

- [AWS アーキテクチャアイコン](#)
- [AWS アーキテクチャセンター](#)
- [AWS Well-Architected](#)

寄稿者

このリファレンスアーキテクチャ図の寄与要因は次のとおりです。

- Jatinder Singh (jsinghtq@)、シニアテクニカルアカウントマネージャー

- Manpreet Kour (mkour@)、シニアテクニカルアカウントマネージャー
- Ali Alemi (alialem@)、シニア WW SSA ストーリーミング

図の履歴

このリファレンスアーキテクチャ図の更新について通知を受け取るには、RSS フィードをサブスクライブします。

変更	説明	日付
初版発行	リファレンスアーキテクチャ図が最初に公開されました。	2024 年 8 月 22 日

Note

RSS 更新をサブスクライブするには、使用しているブラウザで RSS プラグインが有効になっている必要があります。

翻訳は機械翻訳により提供されています。提供された翻訳内容と英語版の間で齟齬、不一致または矛盾がある場合、英語版が優先します。