



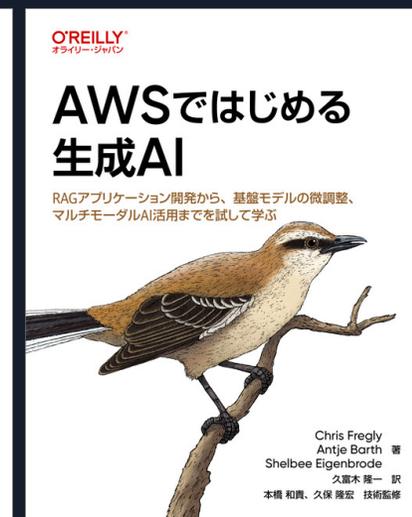
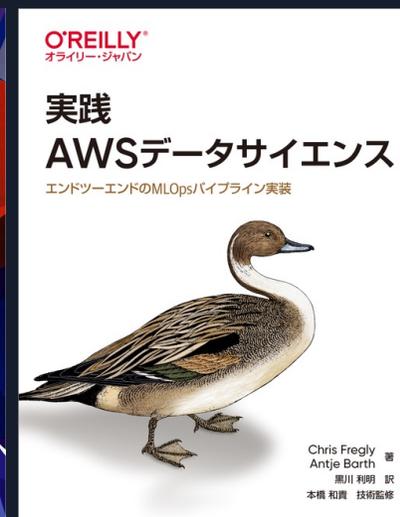
AWS と DataStax で加速する 生成 AI × データ活用

Kazuki Motohashi, Ph.D.

GTM Specialist SA – AI/ML
AWS Japan

本橋 和貴, Ph.D.

アマゾン ウェブ サービス ジャパン 合同会社
機械学習スペシャリストソリューションアーキテクト



AWS の AI/ML サービスのお客様に対する技術的な
支援を行いながら市場開拓を推進しています。

オライリージャパン社「実践 AWS データサイエンス」
「AWS ではじめる生成 AI」 翻訳・技術監修



アジェンダ

AWS の生成 AI スタック

Amazon Bedrock

生成 AI に社内データを取り込む方法

DataStax Astra DB on AWS

Langflow と Amazon Bedrock の統合

まとめ



生成 AI による変革



GENERATIVE AI

can **transform industries**



AWS の生成 AI スタック

アプリケーション利用 モデルを活用するアプリケーション層

 Amazon Q  AWS App Studio

アプリケーション開発 モデルを活用しアプリケーションを開発するためのツール

 Amazon Bedrock

Guardrails | Agents | Studio | Customization | Custom Model Import | Amazon Models

モデルの学習と推論のためのインフラストラクチャ

 SageMaker  Trainium  Inferentia  GPUs

 UltraClusters  EFA  EC2 Capacity Blocks  Nitro  Neuron

Amazon Bedrock

基盤モデルを利用して
生成 AI アプリケーションを
スケーラブルに構築する
最も簡単な方法



インフラを管理することなく
基盤モデルを API 経由で利用して
生成 AI アプリケーションの開発を加速



Amazon, AI21 Labs, Anthropic, Cohere, Meta,
Mistral, Stability AI の基盤モデルを選択可能

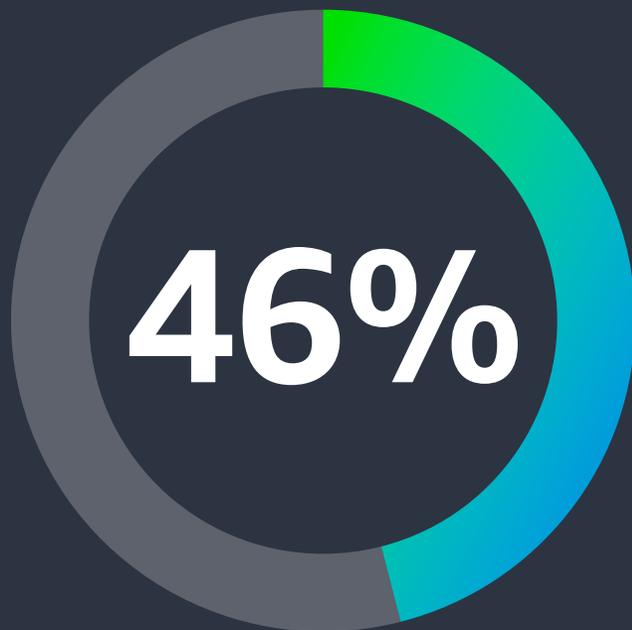


組織内のデータを利用してセキュアに
基盤モデルをカスタマイズ

Amazon Bedrock で利用可能なモデル

AI21 labs	amazon	ANTHROPIC	cohere	∞ Meta	MISTRAL AI	stability.ai
文脈に基づいた回答 要約、パラフレーズ	テキスト要約、 生成、Q&A、検索、 画像生成	要約、 複雑な論理的判断、 執筆、コード生成	テキスト生成、 検索、分類	Q&A、文書理解	テキスト要約、分類、 補完、コード生成、 Q&A	高品質な画像、 アート生成
Jamba 1.5 Large Jamba 1.5 Mini Jamba-Instruct Jurassic-2 Ultra Jurassic-2 Mid	Amazon Titan Image Generator V2 Amazon Titan Image Generator Amazon Titan Text Premier Amazon Titan Text Lite Amazon Titan Text Express Amazon Titan Text Embeddings V2 Amazon Titan Text Embeddings Amazon Titan Multimodal Embeddings	Claude 3.5 Sonnet v2 Claude 3.5 Sonnet Claude 3.5 Haiku Claude 3 Opus Claude 3 Sonnet Claude 3 Haiku Claude Claude Instant	Command R+ Command R Embed English Embed Multilingual Command Command Light	Llama 3.2 90B Llama 3.2 11B Llama 3.2 3B Llama 3.2 1B Llama 3.1 405B Llama 3.1 70B Llama 3.1 8B Llama 3 70B Llama 3 8B Llama 2 70B Llama 2 13B	Mistral Large 2 Mistral Large Mixtral 8x7B Mistral 7B	Stable Image Ultra Stable Diffusion 3 Large Stable Image Core Stable Diffusion XL

企業のデータこそが 企業の差別化要素



CDO (Chief Data Officer) は生成 AI を実装する上で**データ品質が最大の課題**と認識している



CDO は生成 AI イニシアチブにとって**エンドツーエンドのデータ戦略**がクリティカルであると認識している

～ 生成 AI 活用シナリオ ① ～

よりパーソナライズされた 仮想エージェント

目的

旅行予約サイトでそれぞれのユーザーの好みに
合わせた旅程プランを生成したい

必要なデータ

顧客プロフィール

 旅行履歴

 ウェブ履歴

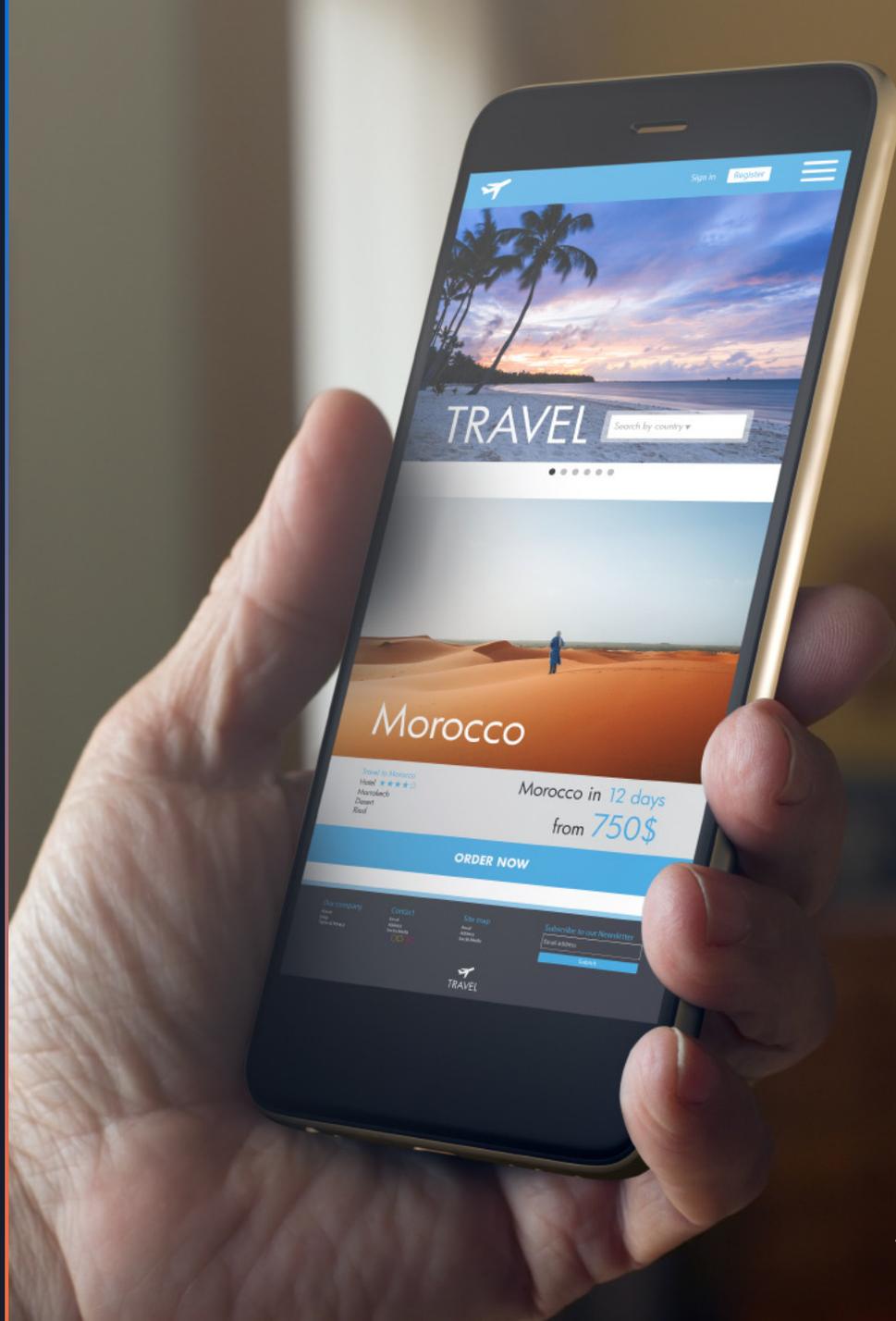
 旅行の好み

会社のデータ

 フライトとホテルの
リストと空き状況

 プロモーションの有無

 似た属性のユーザー情報



生成 AI に社内データを取り込む方法



プロンプト エンジニアリング

生成 AI の
プロンプトに
直接知識を取り込む



RAG (検索拡張生成)

ナレッジベースの
検索結果を
知識として与える



ファイン チューニング

特定のタスクに
対して特化した
能力を得る



独自モデル の開発

大量の学習データ
による一般知識や
専門知識の獲得

EASIER



難易度



COMPLEX

Amazon Bedrock Knowledge Bases

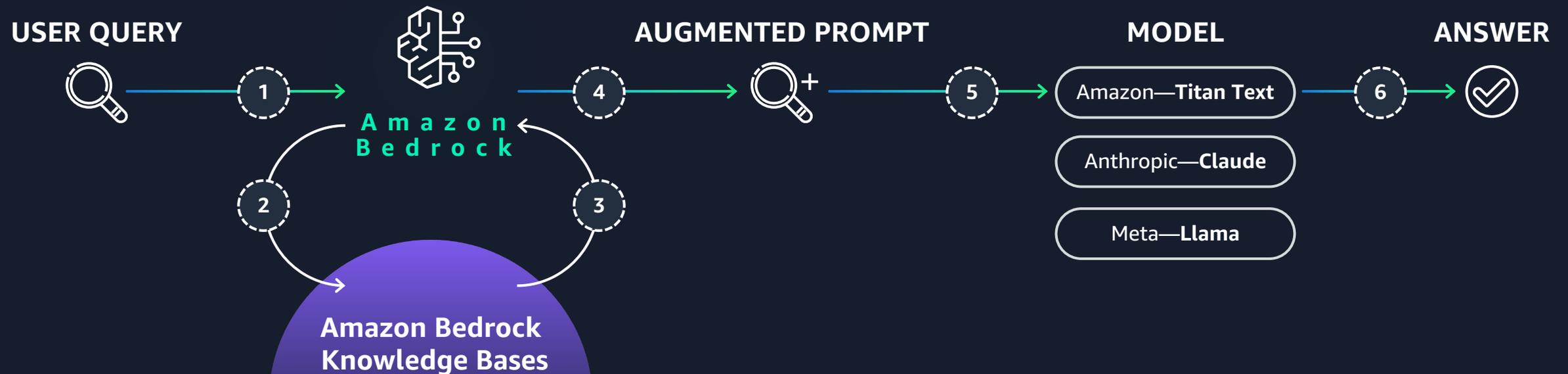
RAG を簡単に構築できるマネージドサービス

基盤モデルをデータソースに安全に接続し、RAG でより適切な回答を提供

取り込み、検索、生成を含むフルマネージド RAG ワークフロー

複数ターンにわたる会話のための組み込みセッションコンテキスト管理

検索機能付きの自動引用による透明性の向上



DataStax Astra DB on AWS

aws
PARTNER
Generative AI
Competency

- accenture
- BOSCH
- LUMIQ
- AllCloud
- ADASTRA
- Deloitte.
- MongoDB.
- DEDICATED
- MYSCALE
- IBM
- CAYLENT
- Capgemini
- REPLY
- firemind.
- mission cloud
- flexa
- INNOVATIVE
- cognizant
- ops@uru
- DATASTAX
- doit
- tcs TATA CONSULTANCY SERVICES
- nlx ANTHROPIC
- intserve
- 神州数码 Digital China
- MLG
- LOKA
- 八斗智能
- Booz Allen.
- wipro
- elastic
- ECCOii 华研网络
- Shellcode
- devoteam
- slalom
- compass.uol
- stability.ai
- fractal
- AI SERA
- goML
- rackspace technology
- Crayon
- eCloudrover 伊克罗德
- NeuronsLab
- ICS compute Re-Invent Go.Beyond
- Ultrapower 神州泰岳
- Provectus
- Kua.ai
- cohere
- zilliz
- weaviate
- HCLTech
- PRESIDIO
- CloudHesive
- ZERO & ONE
- PREDICTi
- 聚云科技
- quantiphi
- TensorIoT
- nVIDIA
- SingleStore
- HuggingFace
- KG DATA 南京柯基数据科技有限公司
- minFy
- cloudthat
- NETRON 網創資訊
- NeuraFlah
- NRI
- ScaleCapacity
- rapyder solutions, simplified.
- zeb
- AI21 labs
- LG CNS
- SIAPARTNERS
- Weights & Biases
- C3.ai



AWS Marketplace で DataStax Astra DB を購入

[AWS Marketplace](#) > [Generative AI](#) > [Software as a Service \(SaaS\)](#) > [DataStax Astra DB](#)



DataStax Astra DB Info

Sold by: [DataStax, Inc.](#)

[View purchase options](#)

[Try for free](#)

[Free Trial](#)

DataStax delivers the real-time vector data that Gen AI apps need, with seamless integration to AWS.

5 ★★★★★ (1) [1 AWS reviews](#) | [31 external reviews](#)

<https://aws.amazon.com/marketplace/pp/prodview-myrrjh2h77gofe>

AWS Marketplace はサードパーティのソフトウェア、データ、サービスを検索、購入、デプロイ、管理できる厳選されたデジタルカタログ

DataStax Astra DB も SaaS 製品として出品されており、フリートライアル込みでサブスクライブすることが可能

Langflow と Amazon Bedrock の統合

RAG やマルチエージェント AI アプリケーションを
ノーコード・ローコードで構築するプラットフォーム

コードを書かずに Amazon Bedrock の豊富な基盤モデル
を活用した処理のフローを構築

利用方法

1. AWS のアクセスキー / シークレットキーを入力
2. 基盤モデルを選択
3. 入力と出力のコンポーネントに接続

aws Amazon Bedrock

Generate text using Amazon Bedrock LLMs.

Input

Type something...

Model ID

anthropic.claude-3-haiku-20240307-v1:0

Access Key

Type something...

Secret Key

Type something...

Region Name

us-east-1

Text

Language Model

Langflow と Amazon Bedrock を用いた RAG システム構築 ①

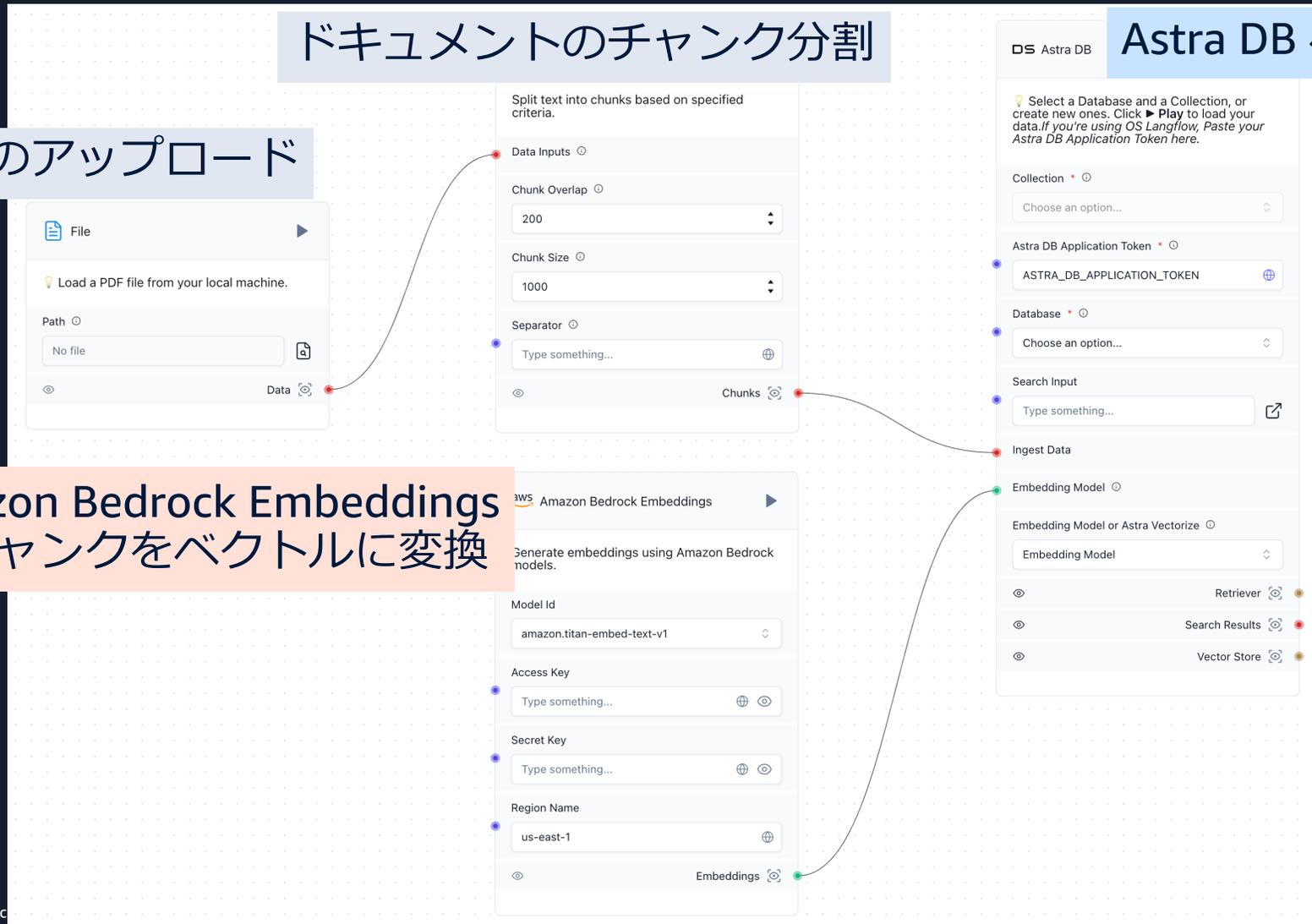
Astra DB へのドキュメントの登録

ファイルのアップロード

ドキュメントのチャンク分割

Astra DB への登録

Amazon Bedrock Embeddings
でチャンクをベクトルに変換



Langflow と Amazon Bedrock を用いた RAG システム構築 ②

RAG の質問回答を実行

ユーザークエリの入力

Astra DB で類似チャンクの検索

プロンプトに
クエリと
検索結果を
埋め込む

Amazon Bedrock
で検索結果に基づいた
回答を生成

回答を出力

検索結果を取得

Amazon Bedrock Embeddings
でクエリをベクトルに変換

Langflow のいいところ

- Datastax のマネージドサービスの SaaS として、ライブラリのインストールなど不要ですぐに始められる
- AWS との統合を含めてビルトインのコンポーネントが豊富 ⇨
 - RAG はいかにうまくドキュメントをベクトル DB に保管するかが質を左右するが、Unstructured との統合により非構造データの読み込みも簡単にできる
- 作成したアプリケーションを実行するための API もすぐに発行でき、AWS の他のサービスとの連携も容易

Components	
↓	Inputs >
↑	Outputs >
[-]	Prompts >
☰	Data >
🔗	Models >
🔧	Helpers >
☰	Vector Stores >
📄	Embeddings >
Experimental BETA ▾	
👤	Agents >
🌟	Astra Assistants >
🔄	Chains >
📎	Loaders >
🔗	Utilities >
🔗	Link Extractors >
📄	Memories >
📄	Output Parsers >
🧪	Prototypes >
📄	Retrievers >
✂️	Text Splitters >
📁	Toolkits >
🔗	Tools >

まとめ

まとめ

- AWS では幅広いお客様のニーズに応えるため3階層の生成 AI サービススタックを提供している
- 生成 AI アプリケーション開発を支援する Amazon Bedrock では、多数の生成 AI モデルをサーバーレス API として提供している
- 生成 AI 単体としてだけでなく RAG などの仕組みを通じて社内外のデータを生成 AI と組み合わせることで会社独自の価値を出すことができる
- Langflow を用いて DataStax Astra DB と Amazon Bedrock をノーコードで連携して RAG などの生成 AI システムを構築できる
- AWS Marketplace を介して DataStax Astra DB をサブスクライブ可能

Thank you!

Kazuki Motohashi, Ph.D.

X: kmotohas@