

Amazon Bedrock Crash Course



Ankit Khare
DevRel @Rockset

Agenda



Part 1

Bedrock - All you need to know

- Key Features
- FMs
- Customizing FMs
- Architecture
- Agents

35 min

Part 2

Building Apps using APIs

Hands-on part on Google Colab

- Simple text responses from FMs
- Simple RAG implementation
- Finding movies using Vector Similarity search

45 min

Part 3

Q&A

Ask away all your questions !

15 min and beyond

Customization Focused



Key Features



Accelerate development of generative AI applications using FMs through an API, without managing infrastructure



Choose FMs from AI21 Labs, Anthropic, Stability AI, and Amazon to find the right FM for your use case



Privately customize FMs using your organization's data



Enhance your data protection using comprehensive AWS security capabilities



Responsible AI provided by supported model providers; Amazon Titan supports AI best practices

Leading FMs Supported



Amazon Titan

Amazon Titan FMs are a family of models built by Amazon that are pretrained on large datasets, which makes them powerful, general-purpose models



Jurassic-2

Multilingual LLMs for text generation in Spanish, French, German, Portuguese, Italian, and Dutch



Claude

LLM for conversations, question answering, and workflow automation based on research into training honest and responsible AI systems

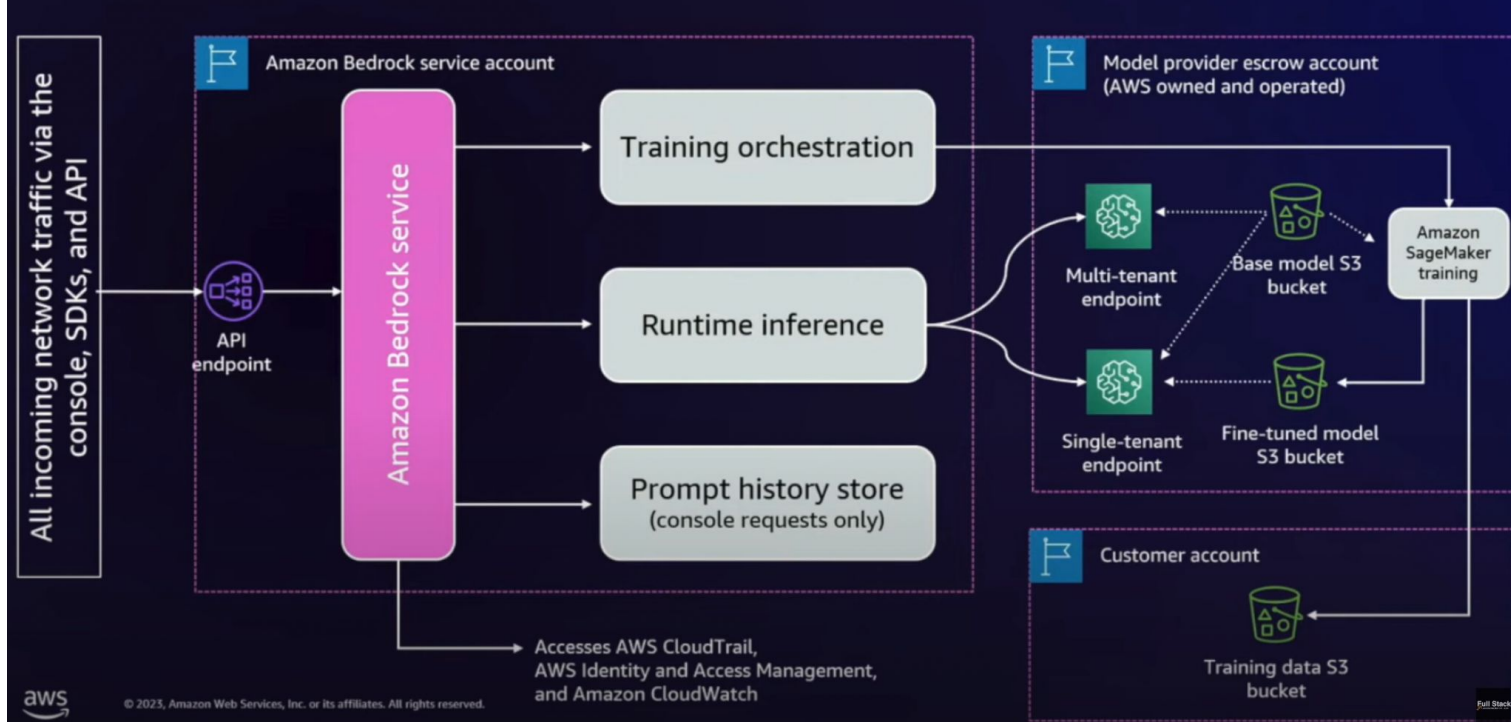


Stable Diffusion

Generation of unique, realistic, high-quality images, art, logos, and designs

Visibility into the Black box

Complete architecture overview



Code Completion in Bedrock

1

Start driving
productivity
with Amazon
CodeWhisperer
today

2

Explore FMs
through Amazon
Bedrock and other
FMs on JumpStart

3

Get started on a
PoC for your top
use cases

Code Completion in Bedrock

1

Start driving
productivity
with Amazon
CodeWhisperer
today

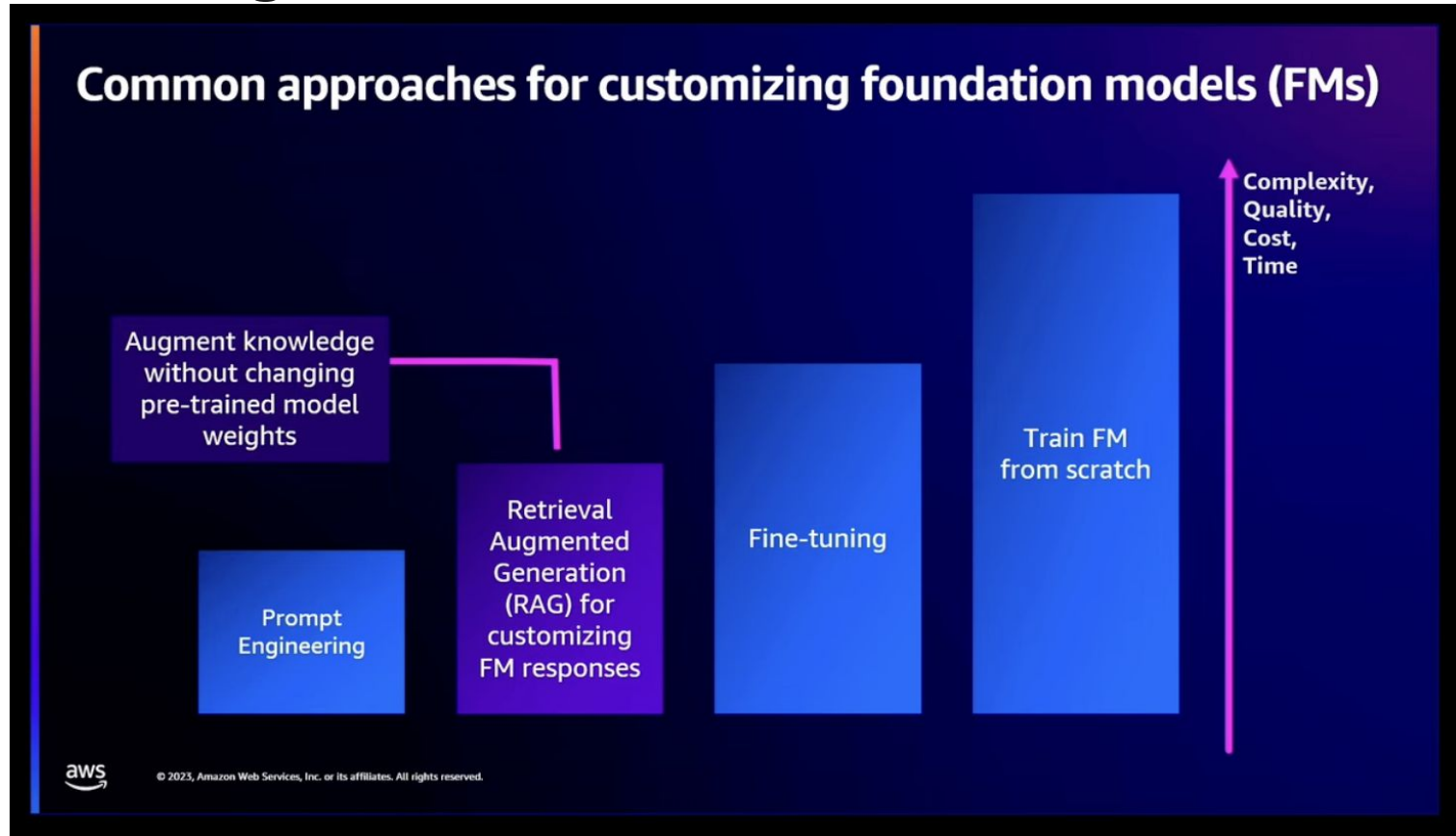
2

Explore FMs
through Amazon
Bedrock and other
FMs on JumpStart

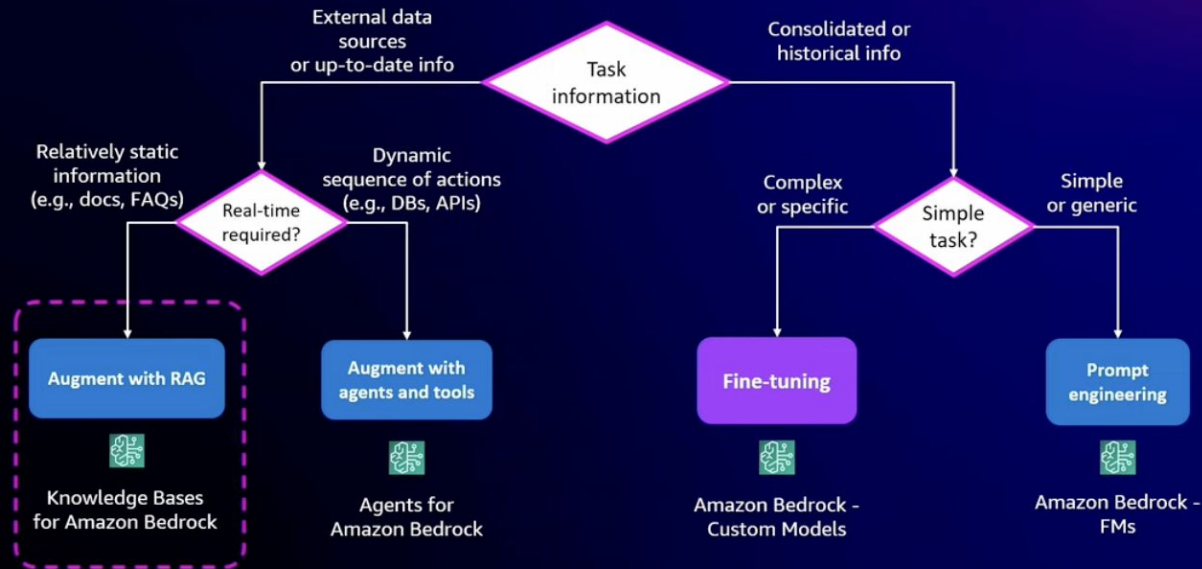
3

Get started on a
PoC for your top
use cases

Customizing FMs



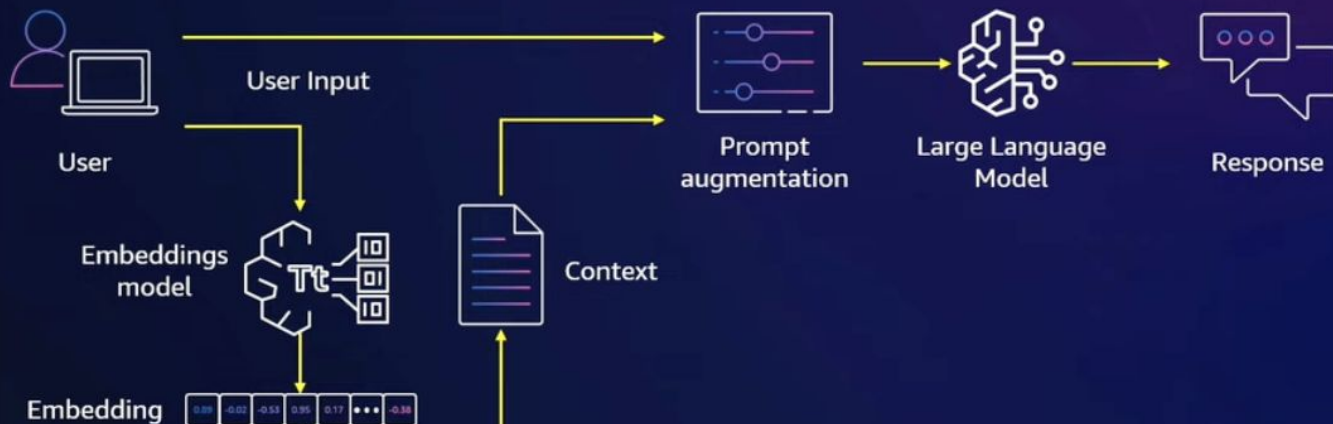
Customize vs. augment



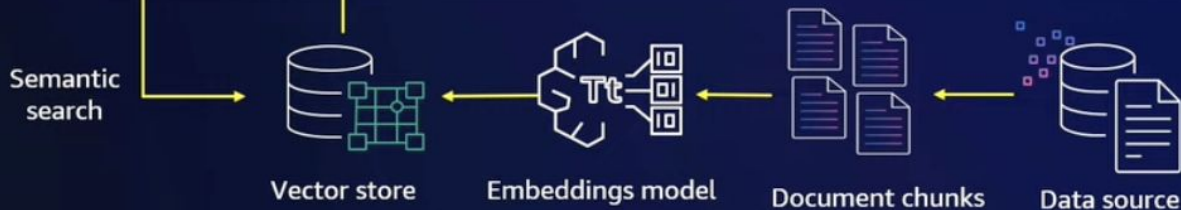
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

RAG in Action

Text Generation Workflow



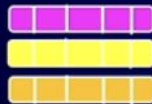
Data Ingestion Workflow



However, when it comes to implementing RAG, there are challenges...



Managing
multiple data
sources



Creating vector
embeddings for large
volumes of data



Incremental
updates to vector
store



Coding effort



Scaling retrieval
mechanism

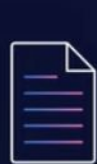


Orchestration

Data Ingestion Workflow

KNOWLEDGE BASES FOR AMAZON BEDROCK

Fully
managed
data
ingestion
workflow



New data



Data source



Document chunks



Embeddings
model



Vector store

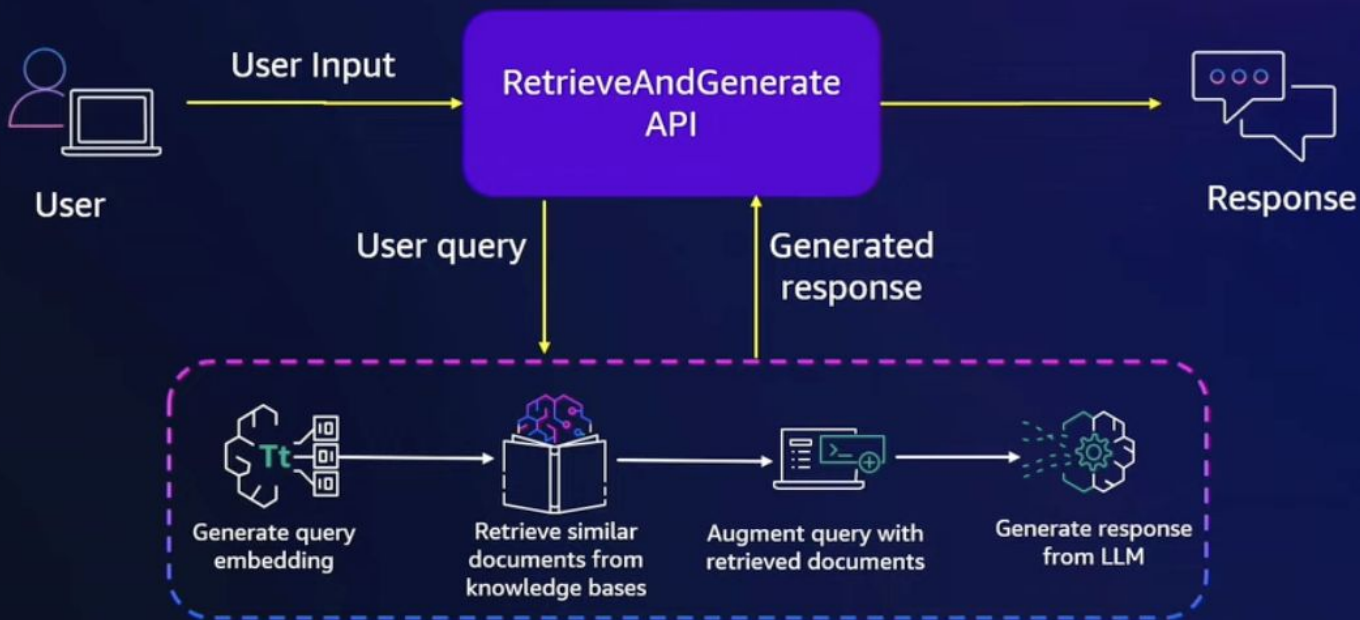
- Choose your data source (Amazon S3)
- Support for incremental updates
- Multiple data file formats supported
- Choose your chunking strategy
 - Fixed chunks
 - No chunking
 - Default (200 tokens)
- Choose your embedding model
 - Amazon Titan
- Choose your vector store
 - Open search serverless
 - Pinecone
 - Redis



RetrieveAndGenerate API

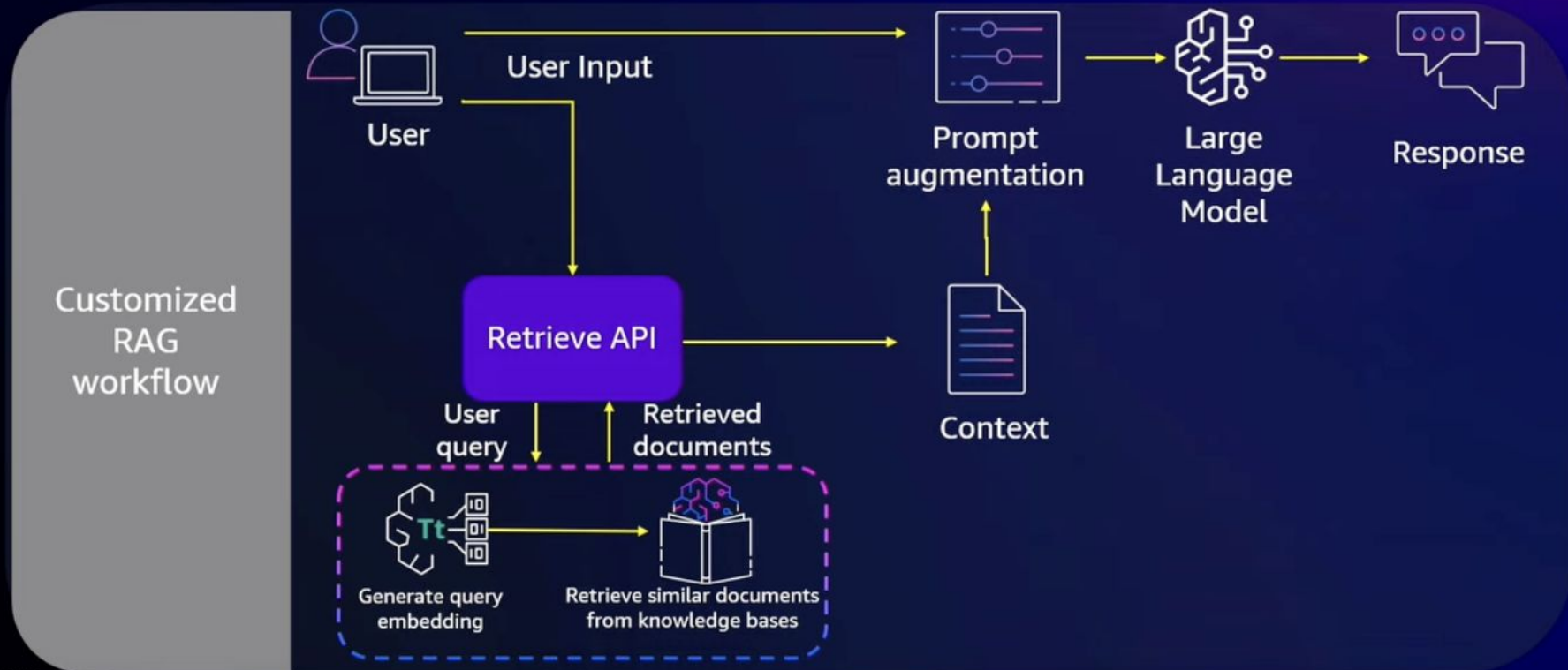
KNOWLEDGE BASES FOR AMAZON BEDROCK

Fully
managed
RAG

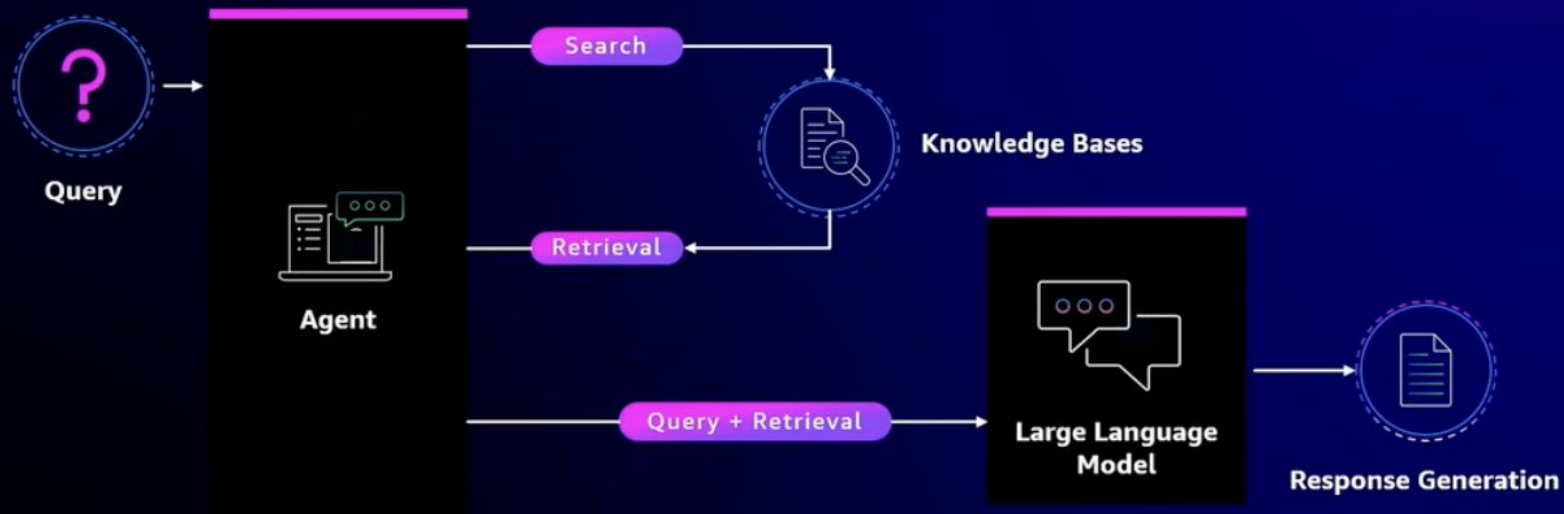


Customize RAG workflows using Retrieve API

KNOWLEDGE BASES FOR AMAZON BEDROCK



Knowledge Bases integration with Agents



Follow us for upcoming workshops and more:

- **Workshop resources -**

- https://drive.google.com/drive/folders/1GXGmBKeQzzWMS_8Yat7XSVeCoH5cMa4i?usp=sharing

- **Subscribe on YouTube**, turn the notifications on, and get an alert whenever workshop recording is available - https://www.youtube.com/channel/UCy4qLzJ7yuEmsIN2Mm5Pn-w?sub_confirmation=1

- Rockset LinkedIn - <https://www.linkedin.com/company/rocksetcloud>

- Ankit Khare LinkedIn - <https://www.linkedin.com/in/deeplearnerak/>

- Interested in building search, analytics, and AI applications, get your free slot for the **Index conference**, May 16 2024 - <https://rockset.com/index-conf/>

- Questions! Feel free to reach out to me - ankitkhare@rockset.com

- Rockset Product support - support@rockset.com