Amazon Bedrock Crash Course

Ankit Khare
DevRel @Rockset

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

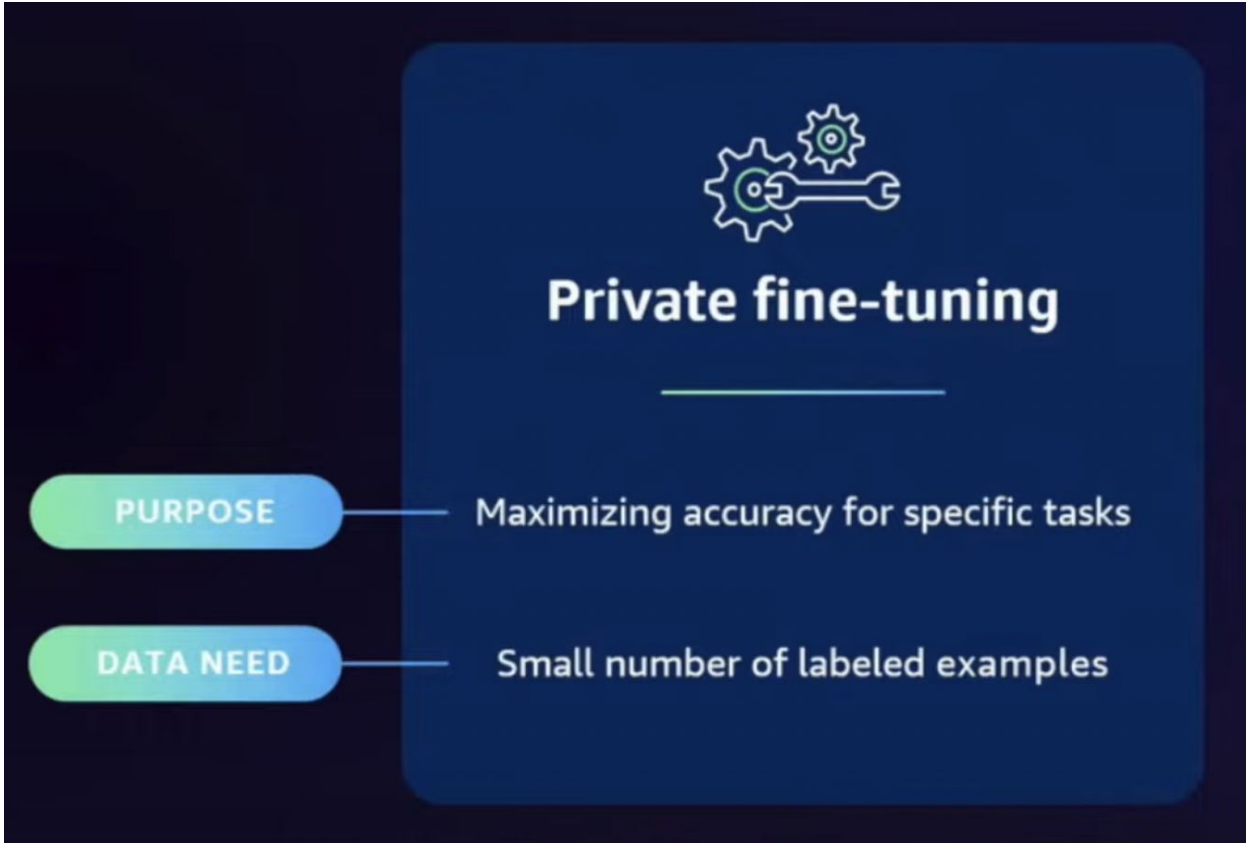
Part 3

Q&A

Ask away all your questions !

15 min and beyond

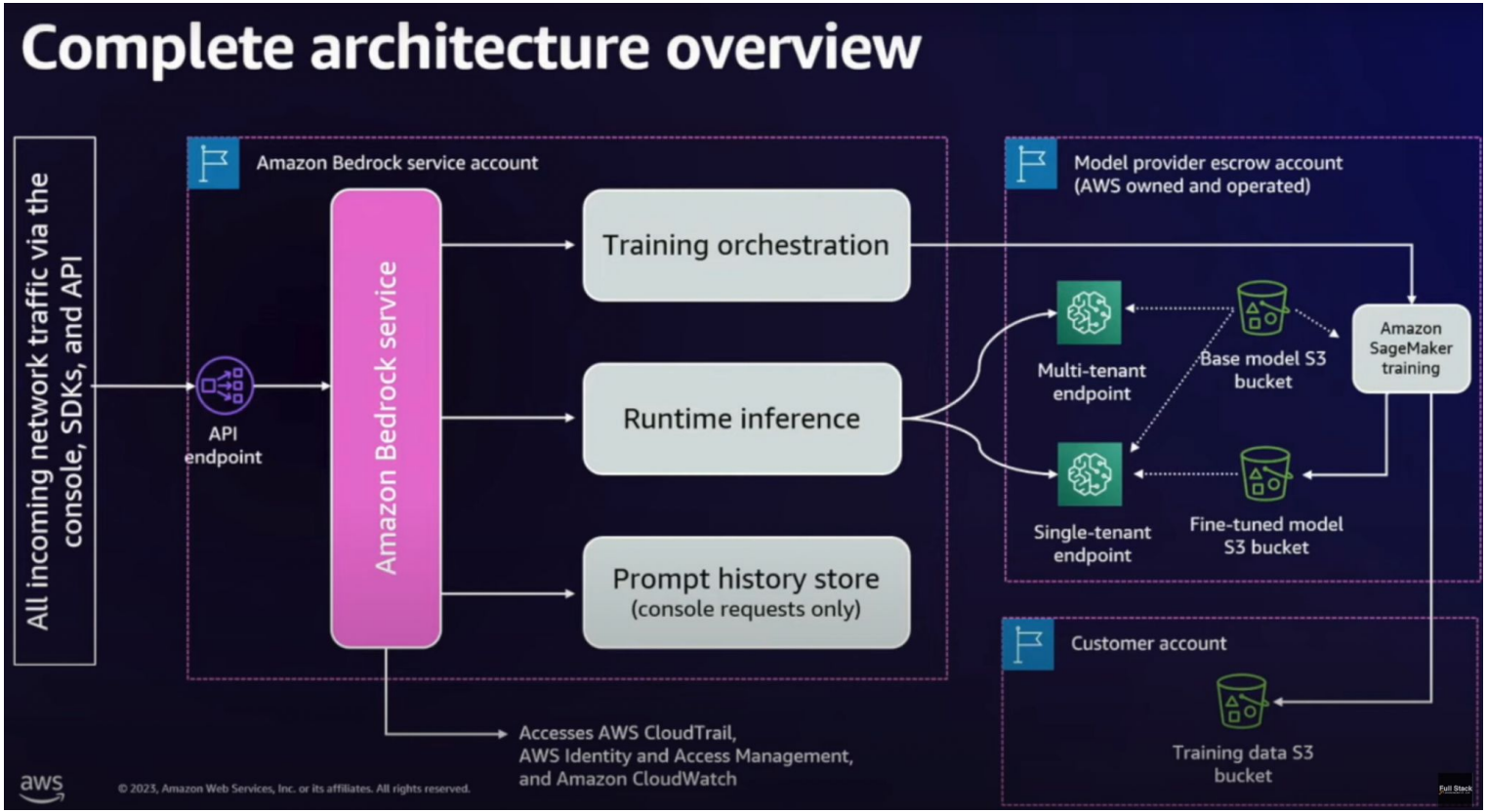
Customization Focused



Key Features

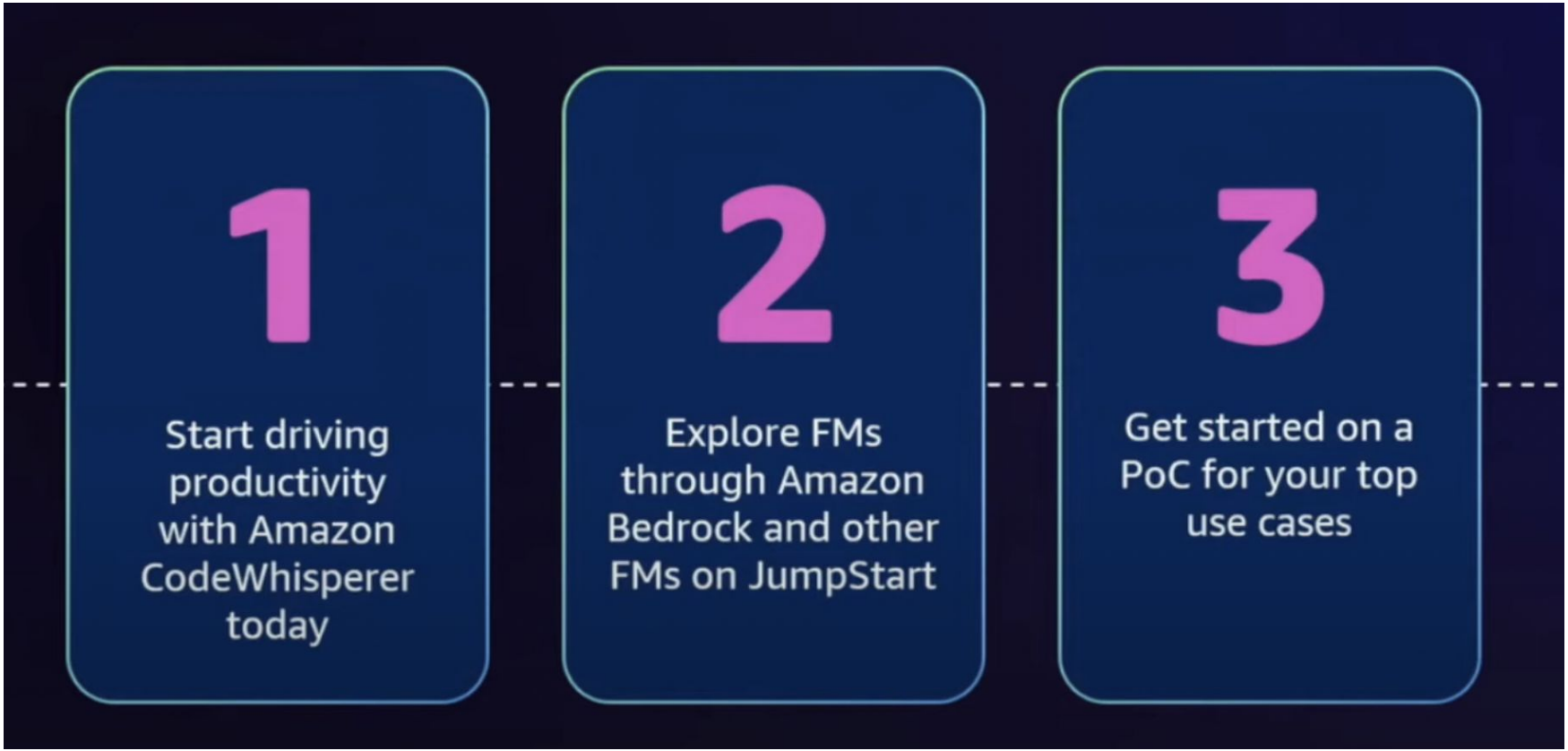
Leading FMs Supported

Visibility into the Black box

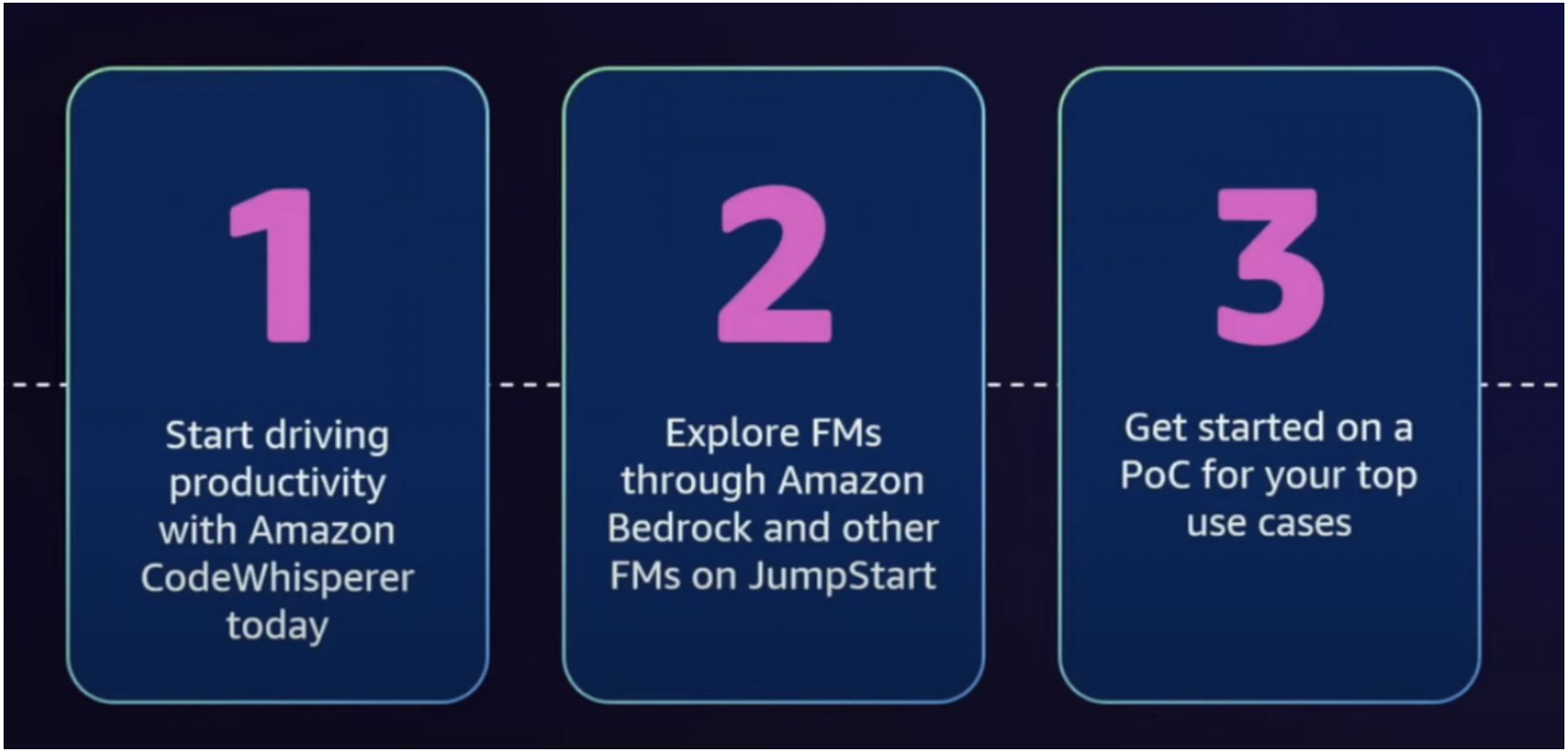


ROCKSET

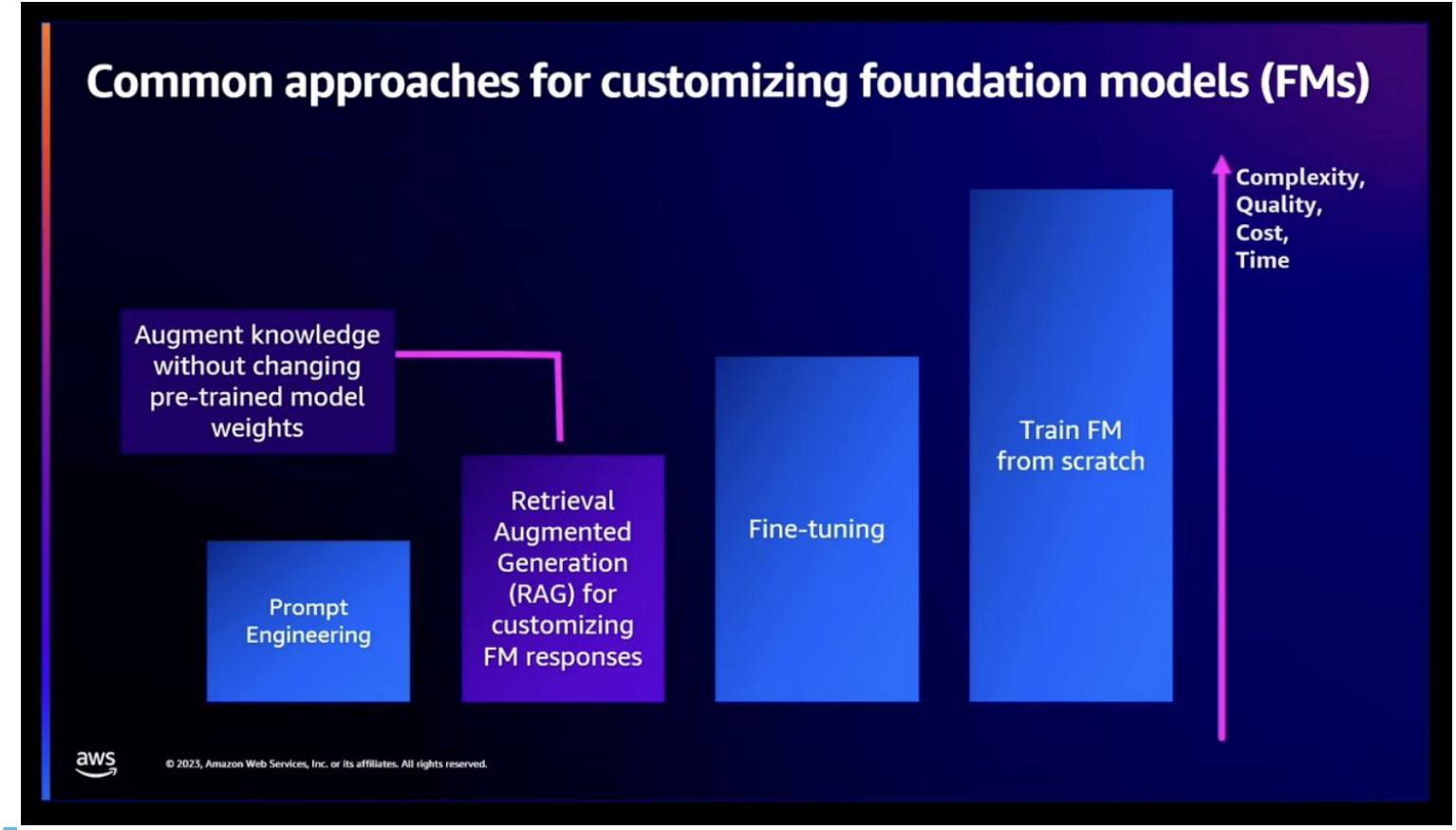
Code Completion in Bedrock



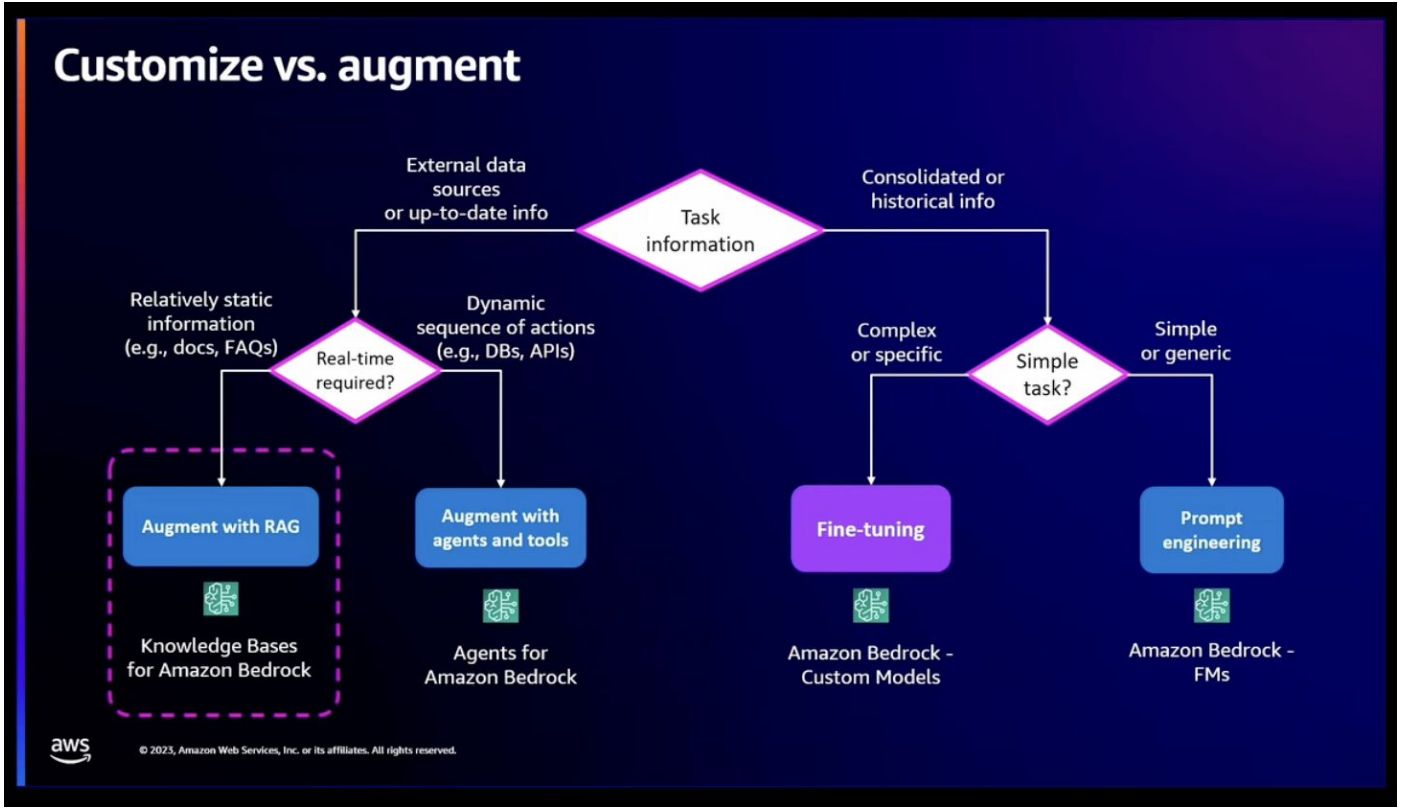
Code Completion in Bedrock



Customizing FMs

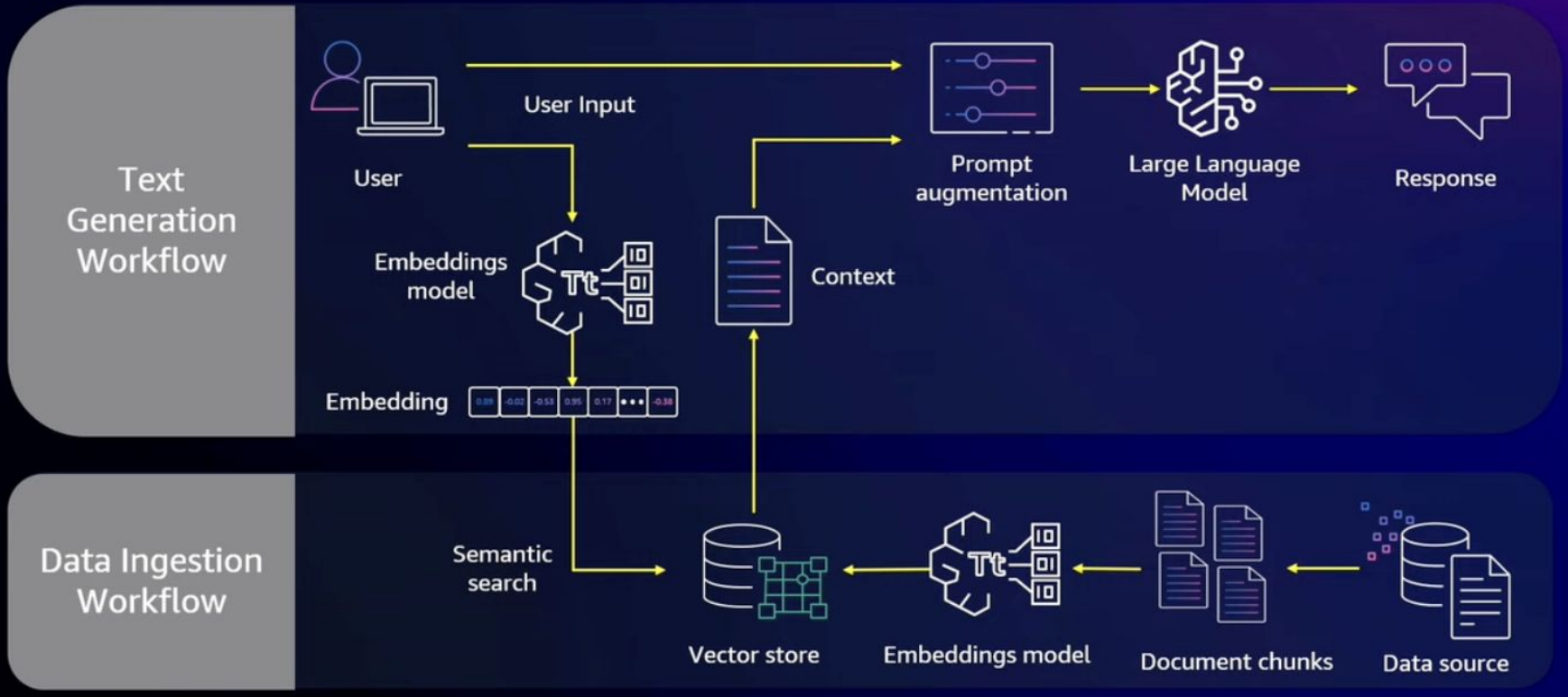


ROCKSET


Customize vs. augment
External data sources or up-to-date info
Task information
Consolidated or historical info
Relatively static information (e.g., docs, FAQs)
Real-time required?
Dynamic sequence of actions (e.g., DBs, APIs)
Complex or specific
Simple task?
Simple or generic
Augment with RAG
Augment with agents and tools
Fine-tuning
Prompt engineering
Knowledge Bases for Amazon Bedrock
Agents for Amazon Bedrock
Amazon Bedrock - Custom Models
Amazon Bedrock - FMs
aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ROCKSET

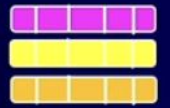
RAG in Action



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

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

aws

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

Data Ingestion Workflow

KNOWLEDGE BASES FOR AMAZON BEDROCK

Fully managed data ingestion workflow

- 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

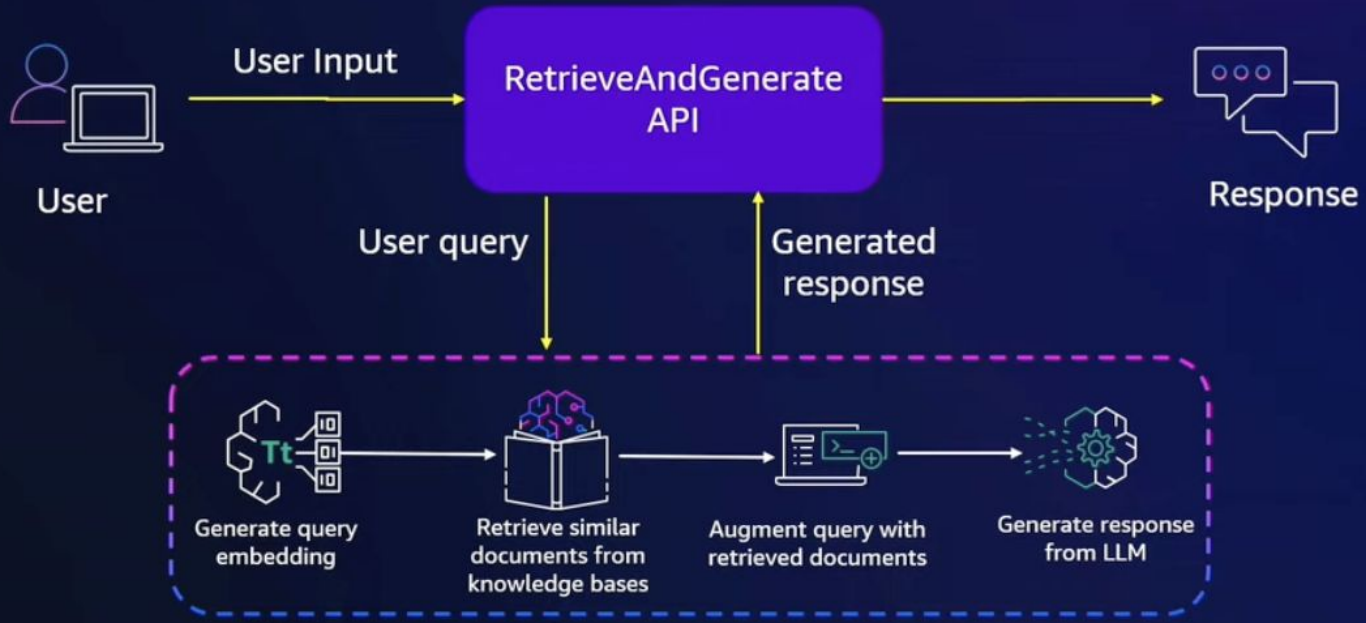
aws

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

RetrieveAndGenerate API

KNOWLEDGE BASES FOR AMAZON BEDROCK

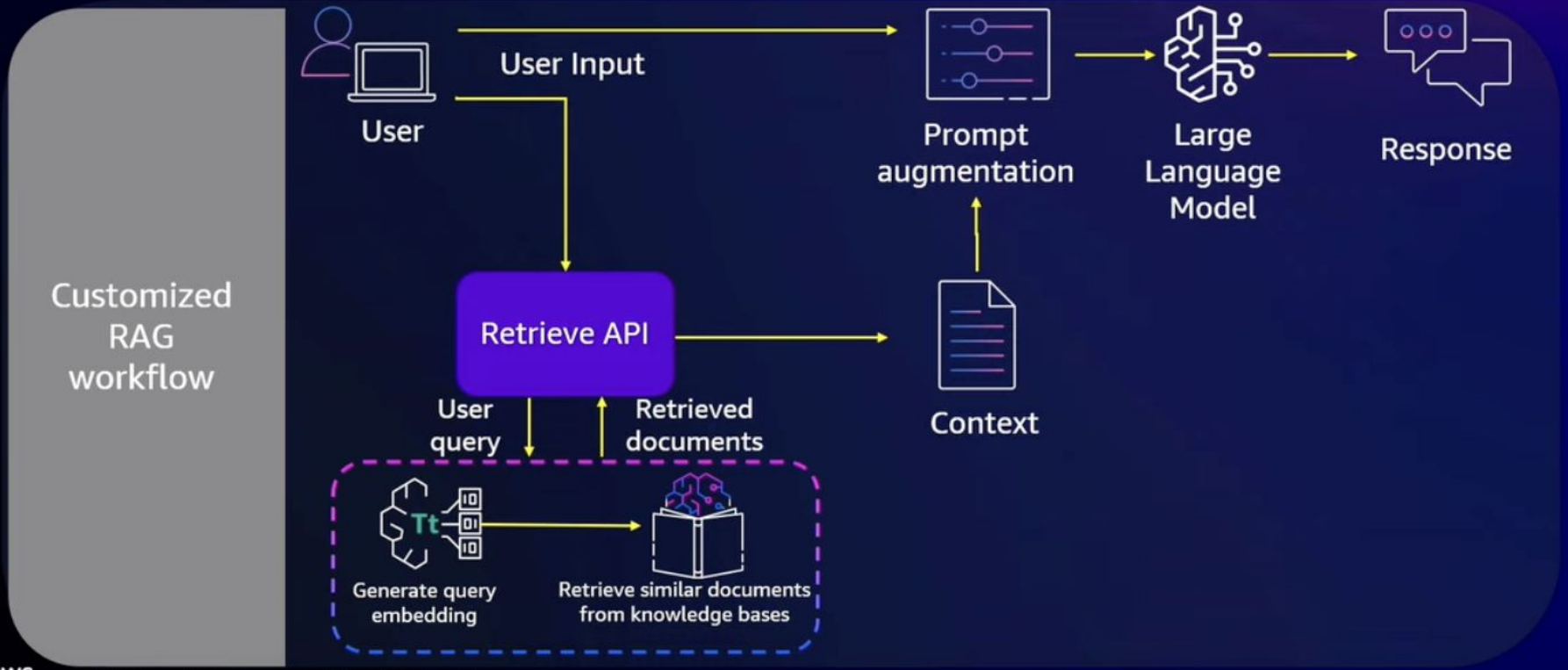
Fully managed RAG



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

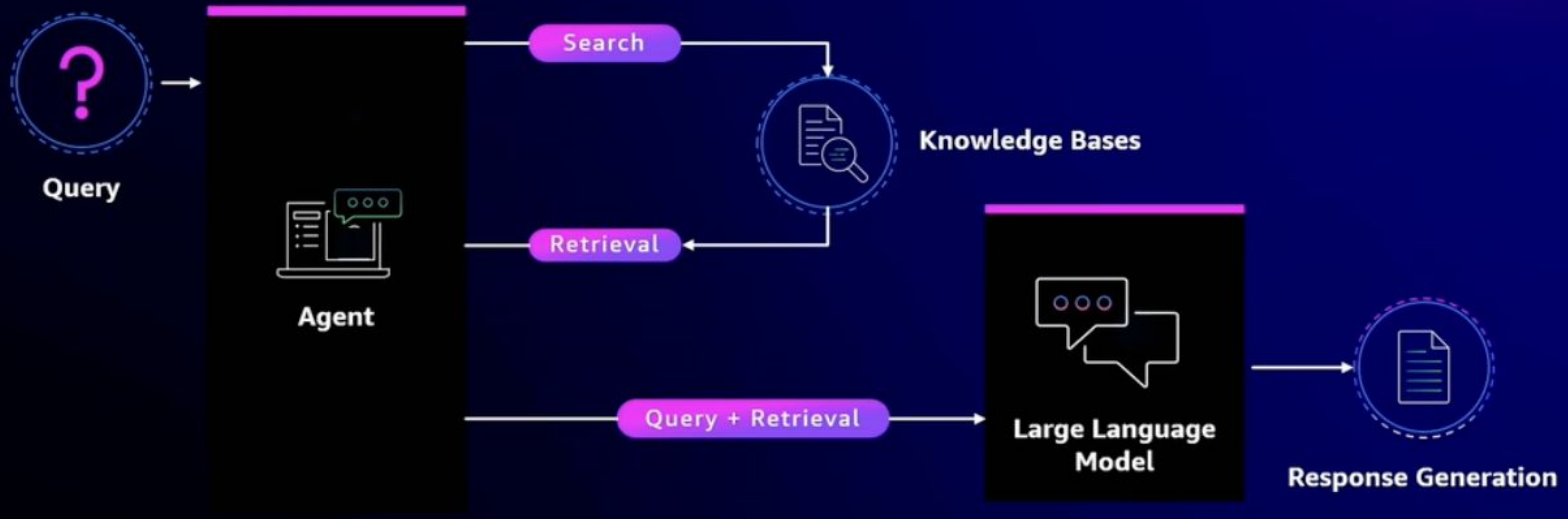
Customize RAG workflows using Retrieve API

KNOWLEDGE BASES FOR AMAZON BEDROCK



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

Knowledge Bases integration with Agents



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

Follow us for upcoming workshops and more:

- **Workshop resources** - https://drive.google.com/drive/folders/1GXGmBKeQzzWMS_8Yat7XSVeCoH5cMa4j?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