AWS
re:Invent
DECEMBER 2 – 6, 2024 | LAS VEGAS, NV

AIM396-NEW

Unlocking power of structured data with Amazon Bedrock Knowledge Bases

Saket Saurabh
(he/him)
Principal Software Engineer
Amazon Bedrock Knowledge Bases
AWS

Ron Widha
(he/him)
Senior Manager, Software Development
Amazon Bedrock Knowledge Bases
AWS

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

Agenda

01 AI for consumers vs. businesses

02 Different approaches to customizing AI

03 What is Retrieval Augmented Generation (RAG)

04 RAG for structured data

05 Demo

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

What was the **last prompt** you gave to an LLM?

Can you improve this paragraph
for style and grammar?

Can you summarize this article?

Can you draft an email about this topic?

Can you solve this math problem?

Can you explain this complex topic?

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

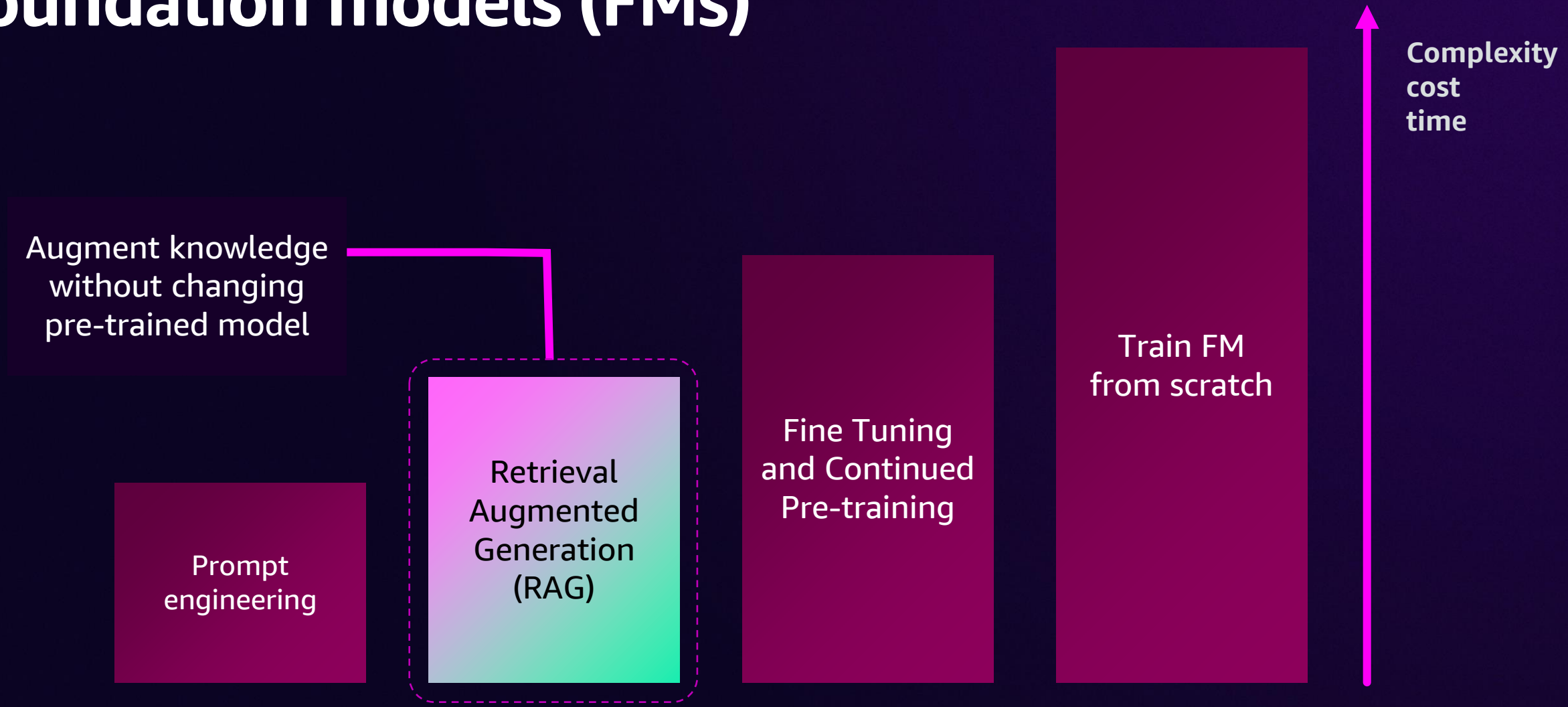
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What **prompts** should **AnyCompany** support?

What’s your return policy?

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

Common approaches for customizing foundation models (FMs)

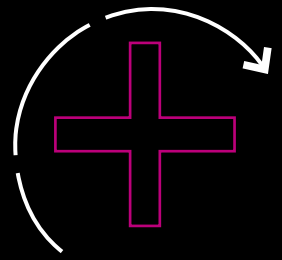


© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Retrieval Augmented Generation (RAG)?

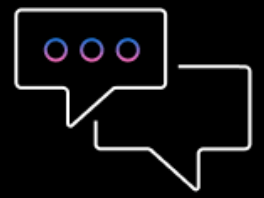
Retrieval

Fetches the relevant content from the external knowledge base or data sources based on a user query



Augmentation

Adding the retrieved relevant context to the user prompt, which goes as an input to the foundation model

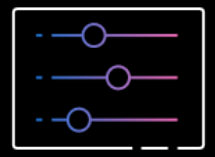


Generation

Response from the foundation model based on the augmented prompt

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

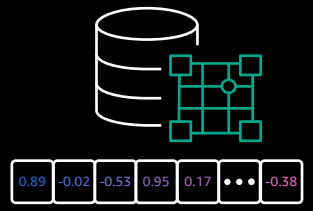
Types of retrieval



Rule-based

Fetches unstructured data like documents

e.g., Keyword searches



Semantic search

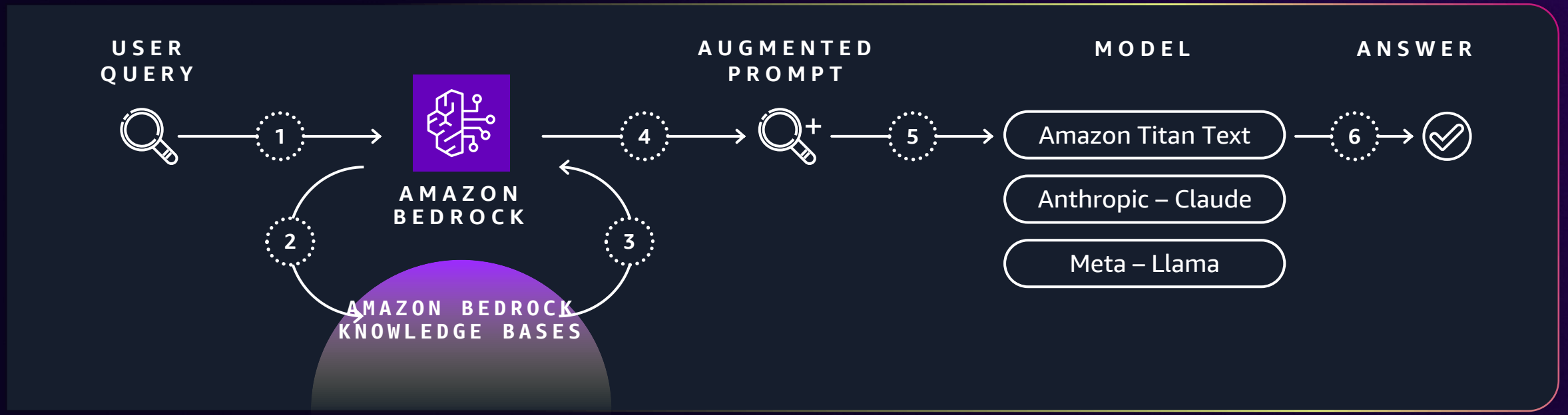
Get relevant documents based on text embeddings

New York → Subway, Statue of Liberty, Tall buildings

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

Amazon Bedrock Knowledge Bases

FOR UNSTRUCTURED DATA



"What's your return policy for this item?"

"Return policy
- Merchandise, including beauty, must be returned within 30 days
- Merchandise must not be worn, opened, used, altered or washed
- Merchandise must have all tags attached, including seals, and be returned in original packaging"

```
"You are a customer support
agent. Given the context
information and not prior
knowledge, answer the query.

Context: Return policy . . .

Query: What's your return
policy?"
```

"Hi, rest assured that if for some reason the item isn't to your satisfaction, you can always return the item within 30 days, as long as . . ."

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

What **prompts** should **AnyCompany** support?

What's your return policy?

Do you have size 9 available?

What was the top selling product in November 2024?

What's my order status?

Who was the "Prestige Partner" in November 2024?

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

Example question: What was the top selling product in November 2024?

order_item_id	product_id	order_id	quantity	price
4319	4044	1437	1	89.99
4320	4015	1437	1	24.99
4321	4034	1437	1	19.99
…	…	…	…	…

User_id	username	first_name	Last_name
7001	alejandro_rosalez	Alejandro	Rosalez
7002	akua_mansa	Akua	Mansa
7003	anacarolina_silva	Ana Carolina	Silva
…	…	…	…

id	order_status	user_id	total_amount	ship_mthd	ts1
1437	Delivered	7002	214.96	EXP	2024-11-19 03:45:36
1438	Delivered	7116	629.92	STD	2024-11-04 06:05:15
…	…		…	…	

product_id	name	stock_quantity	price	seller_id
4001	Smartphone X	50	699.99	501
4002	Laptop Pro	20	1299.99	501
…	…	…	…	…

seller_id	Store_name	User_id	description
501	Tech Store	7001	A cutting-edge electronics store ..
502	Kitchen Store	7005	go-to destination for top-quality home appliances ..
…	…	…	…

aws

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

Accessing structured data requires more than semantic search

Example question: What was the top selling product in November 2024?

order_item_id	product_id	order_id	quantity	price
4319	4044	1437	1	89.99
4320	4015	1437	1	24.99
4321	4034	1437	1	19.99
…	…	…	…	…

User_id	username	first_name	Last_name
7001	alejandro_rosalez	Alejandro	Rosalez
7002	akua_mansa	Akua	Mansa
7003	anacarolina_silva	Ana Carolina	Silva
…	…	…	…

id	order_status	user_id	total_amount	ship_mthd	ts1
1437	Delivered	7002	214.96	EXP	2024-11-19 03:45:36
1438	Delivered	7116	629.92	STD	2024-11-04 06:05:15
…	…		…	…	

product_id	name	stock_quantity	price	seller_id
4001	Smartphone X	50	699.99	501
4002	Laptop Pro	20	1299.99	501
…	…	…	…	…

seller_id	Store_name	User_id	description
501	Tech Store	7001	A cutting-edge electronics store ..
502	Kitchen Store	7005	go-to destination for top-quality home appliances ..
…	…	…	…

aws

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

Accessing structured data requires more than semantic search

Example question: What was the top selling product in November 2024?

order_item_id	product_id	order_id	quantity	price
4319				
4320				
4321				
…				

User_id	username	first_name	Last_name
			Rosalez
			Mansa
			Silva
			…

To answer questions like

"What was the top selling product in November 2024?"

We need to translate the natural language query into a (SQL) query

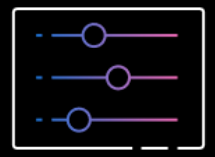
product_id	name			
4001	Smartp			
4002	Laptop Pro			
…	…	…	…	…

			dge electronics
502	Kitchen Store	7005	go-to destination for top-quality home appliances ..
…	…	…	…

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

Types of retrieval

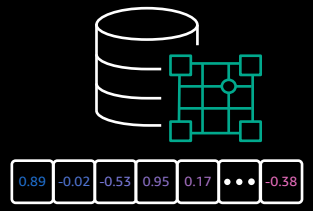
Natural Language to SQL (NL2SQL) is the RAG for structured data



Rule-based

Fetches unstructured data like documents

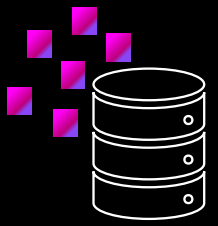
e.g., Keyword searches



Semantic search

Get relevant documents based on text embeddings

New York → Subway, Statue of Liberty, Tall buildings



Structured data

Transactional retrieval from database or API

e.g., Select customers from All_orders where order == ‘XYZ’

aws

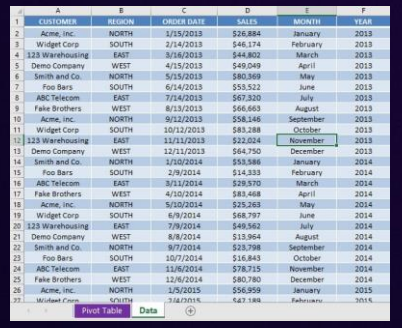
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Bedrock Knowledge Bases

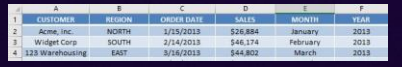
INTRODUCING STRUCTURED DATA RETRIEVAL IN BEDROCK KNOWLEDGE BASES

"What was the top selling product in November?"

```
SELECT customer_id, SUM(order_total) AS
total_revenue

FROM customer_orders

WHERE order_date BETWEEN '2024-01-01'
AND '2024-01-31'

GROUP BY customer_id ORDER BY
total_revenue DESC;
```



"In November, the top selling product is YYYY"



aws

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

Why is NL2SQL hard?

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

Why is NL2SQL hard?

Count the number of schools in Alameda County that have less than 100 test takers

LLM output

```
SELECT COUNT(*)
FROM schools
WHERE district = 'Alameda'
AND test_takers < 100
```

Expected output

```
SELECT COUNT(*)
FROM schools s
JOIN satscores ss
ON s.CDSCode=ss.cds
WHERE s.County = 'Alameda'
AND ss.NumTstTakr < 100
```

Personalized to your database, tables, and schema

aws

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

Why is NL2SQL hard?

List the product descriptions of the transactions in gas stations in the Czech Republic

LLM output

```
SELECT DISTINCT T3.Description
FROM transactions_1k AS T1
INNER JOIN gasstations AS T2
ON T1.GasStationID = T2.GasStationID
INNER JOIN products AS T3
ON T1.ProductID = T3.ProductID
WHERE T2.Country = 'Czech Republic';
```

Expected output

```
SELECT DISTINCT T3.Description
FROM transactions_1k AS T1
INNER JOIN gasstations AS T2
ON T1.GasStationID = T2.GasStationID
INNER JOIN products AS T3
ON T1.ProductID = T3.ProductID
WHERE T2.Country = 'CZE';
```

Personalized to your data

aws

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

Why is NL2SQL hard?

Of all the contestants, what is the contestant number and name of the contestant who got the least votes?

LLM output

```
SELECT contestant_number, contestant_name
FROM contestants inner join votes on
contestants.contestant_number =
votes.contestant_number
GROUP BY contestant_number, contestant_name
ORDER BY count(*) asc
LIMIT 1;
```

OperationalError ambiguous column name: contestant_number

Expected output

```
SELECT c.contestant_number, c.contestant_name
FROM contestants c inner join votes v on
c.contestant_number = v.contestant_number
GROUP BY c.contestant_number,
c.contestant_name
ORDER BY count(*) asc
LIMIT 1;
```

Personalized to your SQL query engine

aws

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

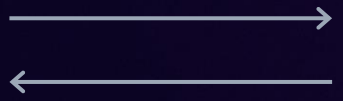
How it works

STRUCTURED DATA RETRIEVAL USING GENERATIVE AI

Personalized

Context aware

Secure



Amazon Bedrock Knowledge Bases

Provides structured data retrieval over Amazon Redshift and Amazon SageMaker Lakehouse

Database metadata

Query configurations

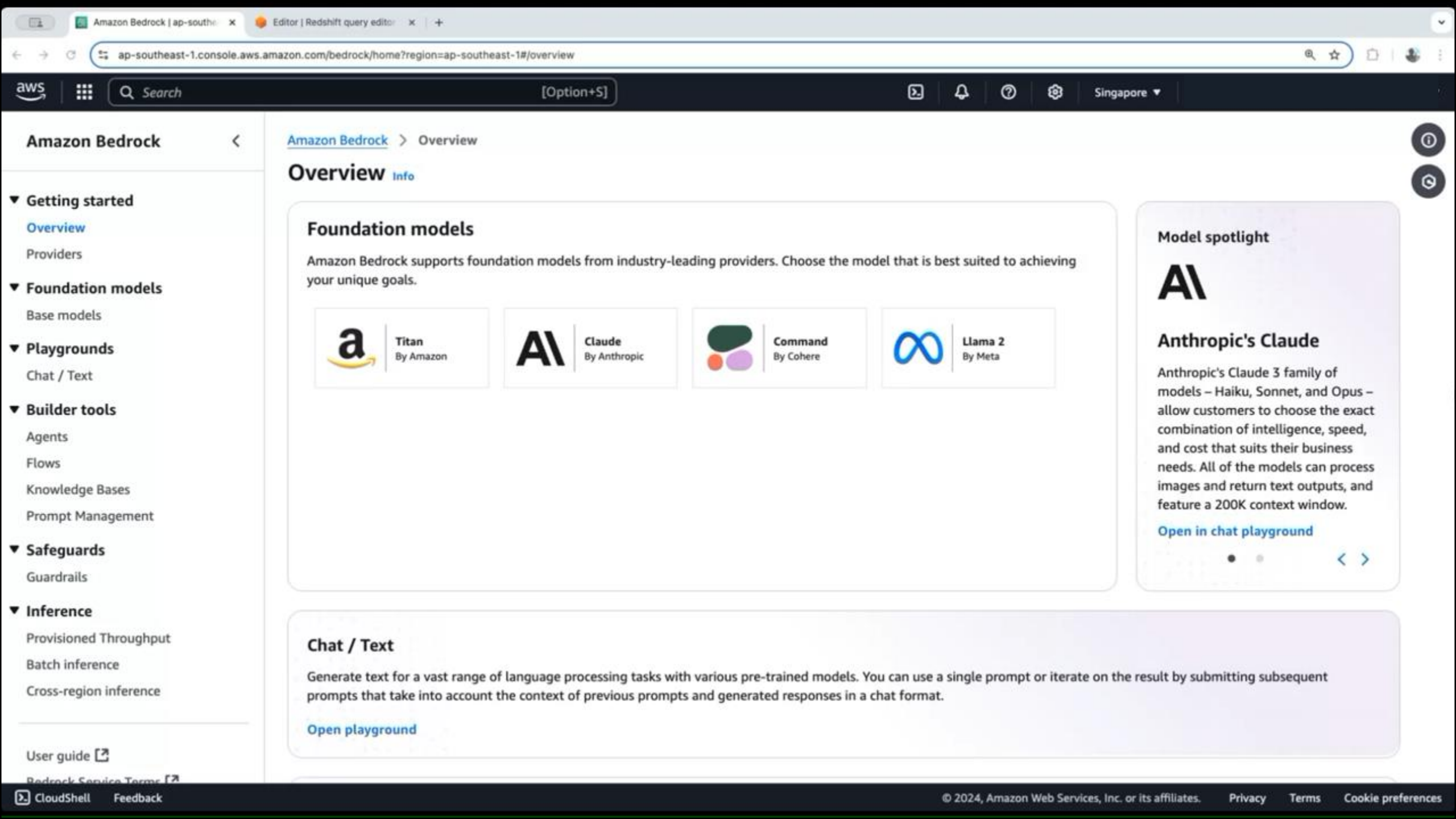
Conversation history

Content processed by Amazon Bedrock Knowledge Bases is not stored or used for service improvement

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

Demo

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


Amazon Bedrock | ap-southe
Editor | Redshift query editor
ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/overview
aws
Search
[Option+S]
Singapore
Amazon Bedrock
Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference
User guide
Amazon Bedrock > Overview
Overview Info
Foundation models
Amazon Bedrock supports foundation models from industry-leading providers. Choose the model that is best suited to achieving your unique goals.
Titan
By Amazon
Claude
By Anthropic
Command
By Cohere
Llama 2
By Meta
Model spotlight
Anthropic's Claude
Anthropic's Claude 3 family of models – Haiku, Sonnet, and Opus – allow customers to choose the exact combination of intelligence, speed, and cost that suits their business needs. All of the models can process images and return text outputs, and feature a 200K context window.
Open in chat playground
Chat / Text
Generate text for a vast range of language processing tasks with various pre-trained models. You can use a single prompt or iterate on the result by submitting subsequent prompts that take into account the context of previous prompts and generated responses in a chat format.
Open playground
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Amazon Bedrock

Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference

User guide
Bedrock Service Terms

Step 1
Provide Knowledge Base details
Step 2
Configure query engine
Step 3
Review and create

Provide Knowledge Base details

Knowledge Base details

Knowledge Base name

knowledge-base-quick-start-8hzqt

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 50 characters.

Knowledge Base description - *optional*

Enter description

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 200 characters.

Data source details

Query engine

Select the query engine to use for retrieving data from your data stores.

Amazon Redshift

A fully managed, petabyte-scale data warehouse service in the cloud that lets you access and analyze data without needing to configure provisioning and capacity.

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Amazon Bedrock

Getting started
Overview
Providers

Foundation models
Base models

Playgrounds
Chat / Text

Builder tools
Agents
Flows
Knowledge Bases
Prompt Management

Safeguards
Guardrails

Inference
Provisioned Throughput
Batch inference
Cross-region inference

User guide
Bedrock Service Terms

Knowledge Base description - *optional*

Enter description

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 200 characters.

Data source details

Query engine

Select the query engine to use for retrieving data from your data stores.

Amazon Redshift

A fully managed, petabyte-scale data warehouse service in the cloud that lets you access and analyze data without needing to configure provisioning and capacity.

IAM permissions

Certain permissions are necessary to access other services or perform actions in order to create this resource. For more information, see service role for Amazon Bedrock

Runtime role

- (●) Create and use a new service role
- ( ) Use an existing service role

Service role name

AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt

© 2024, Amazon Web Services, Inc. or its affiliates.

Amazon Bedrock

Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference

User guide
Bedrock Service Terms

Step 1
Provide Knowledge Base details
Step 2
Configure query engine
Step 3
Review and create

Configure query engine

Configure for the chosen data source

Amazon Redshift

You can connect to Amazon Redshift so that your web experience users can query for information in databases and tables using natural language.

Query engine details

Connection options

To create new clusters or workgroups, visit Amazon Redshift.

- (•) Redshift serverless
  Select if you are using Redshift Serverless offering.
- ( ) Redshift provisioned
  Select if your databases are provisioned on dedicated compute nodes.

Workgroup

Select a workgroup

Authentication

- (•) IAM Role
- ( ) Temporary credentials user name
- ( ) Secrets manager

The Service role named **AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt** was configured in Step 1.

© 2024, Amazon Web Services, Inc. or its affiliates.

Amazon Bedrock

- Getting started
  - Overview
  - Providers
- Foundation models
  - Base models
- Playgrounds
  - Chat / Text
- Builder tools
  - Agents
  - Flows
  - Knowledge Bases
  - Prompt Management
- Safeguards
  - Guardrails
- Inference
  - Provisioned Throughput
  - Batch inference
  - Cross-region inference

User guide
Bedrock Service Terms

Amazon Bedrock > Knowledge Bases > Create knowledge base with structured data store

Step 1
Provide Knowledge Base details
Step 2
Configure query engine
Step 3
Review and create

Configure query engine

Configure for the chosen data source

Amazon Redshift

You can connect to Amazon Redshift so that your web experience users can query for information in databases and tables using natural language.

Query engine details

Connection options

To create new clusters or workgroups, visit Amazon Redshift.

- ○ Redshift serverless
  Select if you are using Redshift Serverless offering.
- ◉ Redshift provisioned
  Select if your databases are provisioned on dedicated compute nodes.

Cluster

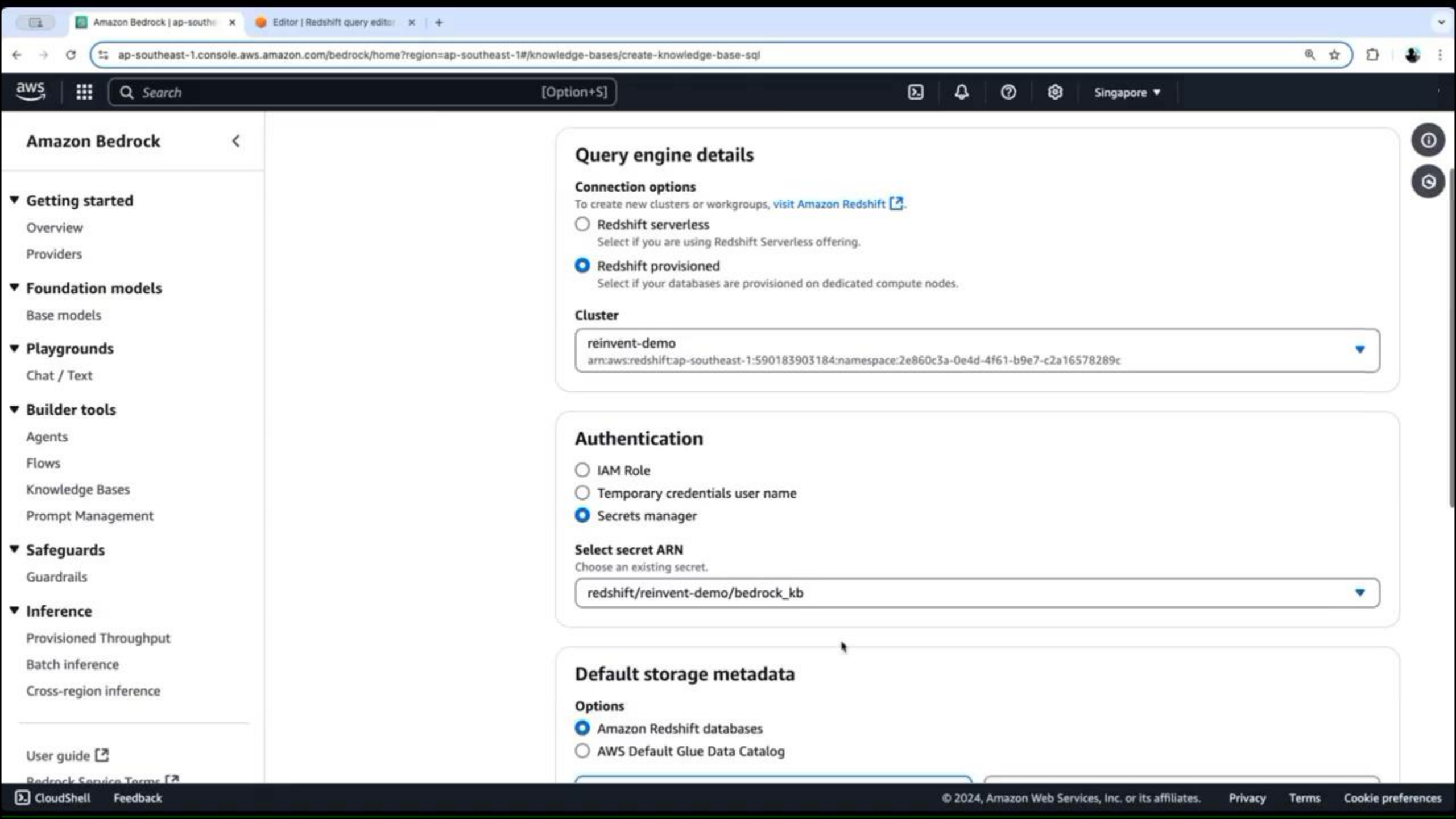
reinvent-demo
arn:aws:redshift:ap-southeast-1:590183903184:namespace:2e860c3a-0e4d-4f61-b9e7-c2a16578289c

Authentication

- ◉ IAM Role
- ○ Temporary credentials user name
- ○ Secrets manager

ⓘ The Service role named

© 2024, Amazon Web Services, Inc. or its affiliates.


Amazon Bedrock | ap-southe
Editor | Redshift query editor
ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/knowledge-bases/create-knowledge-base-sql
aws
Search
[Option+S]
Singapore
Amazon Bedrock
Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference
User guide
Query engine details
Connection options
To create new clusters or workgroups, visit Amazon Redshift.
Redshift serverless
Select if you are using Redshift Serverless offering.
Redshift provisioned
Select if your databases are provisioned on dedicated compute nodes.
Cluster
reinvent-demo
arn:aws:redshift:ap-southeast-1:590183903184:namespace:2e860c3a-0e4d-4f61-b9e7-c2a16578289c
Authentication
IAM Role
Temporary credentials user name
Secrets manager
Select secret ARN
Choose an existing secret.
redshift/reinvent-demo/bedrock_kb
Default storage metadata
Options
Amazon Redshift databases
AWS Default Glue Data Catalog
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Amazon Bedrock

Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference

User guide
Bedrock Service Terms

Temporary credentials user name
Secrets manager

Select secret ARN
Choose an existing secret.
redshift/reinvent-demo/bedrock_kb

Default storage metadata

Options
Amazon Redshift databases
AWS Default Glue Data Catalog

Redshift database list
Choose database

Enter database name
Enter Redshift database name

Database
Select a database for your knowledge base to be able to query.
mp_demo

Query configurations - *optional*
You can add different configurations to your queries.

Maximum query time
If the query takes longer than the duration you set, it will time out.
200 seconds
Specify a time between 1 and 200 seconds.

▸ **Descriptions** (0)

© 2024, Amazon Web Services, Inc. or its affiliates.

Amazon Bedrock

Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference

User guide
Bedrock Service Terms

Step 1
Provide Knowledge Base details
Step 2
Configure query engine
Step 3
Review and create

Review and create

Step 1: Provide details

Edit

Knowledge Base details

Knowledge Base name
reinvent-demo

Knowledge Base description
—

Service role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt

Knowledge base type
Knowledge base use SQL database

Query engine
Amazon Redshift

KMS key for transient data storage
AWS default KMS key

Log Deliveries
—

Step 2: Configure query engine

Redshift details

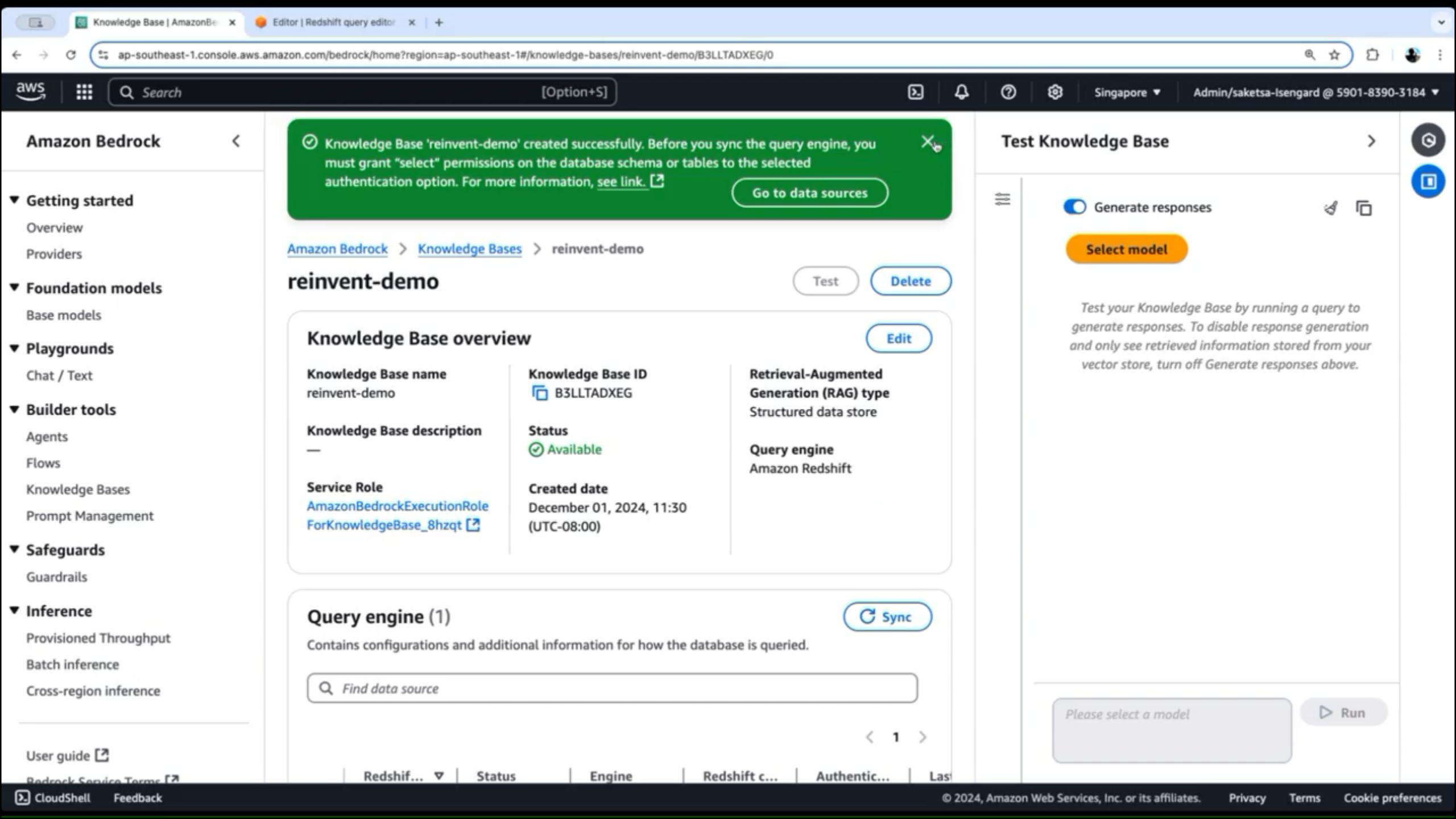
Resource details

Connection options
Redshift cluster

Redshift cluster
reinvent-demo

Redshift database
mp_demo

© 2024, Amazon Web Services, Inc. or its affiliates.


Knowledge Base | AmazonBe
Editor | Redshift query editor
ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/knowledge-bases/reinvent-demo/B3LLTADXEG/0
aws
Search
[Option+S]
Singapore
Admin/saketsa-Isengard @ 5901-8390-3184
Amazon Bedrock
Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference
User guide
Knowledge Base 'reinvent-demo' created successfully. Before you sync the query engine, you must grant "select" permissions on the database schema or tables to the selected authentication option. For more information, see link.
Go to data sources
Amazon Bedrock > Knowledge Bases > reinvent-demo
reinvent-demo
Test
Delete
Knowledge Base overview
Edit
Knowledge Base name
reinvent-demo
Knowledge Base description
—
Service Role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt
Knowledge Base ID
B3LLTADXEG
Status
Available
Created date
December 01, 2024, 11:30 (UTC-08:00)
Retrieval-Augmented Generation (RAG) type
Structured data store
Query engine
Amazon Redshift
Query engine (1)
Sync
Contains configurations and additional information for how the database is queried.
Find data source
1
Redshif...
Status
Engine
Redshift c...
Authentic...
Last
Test Knowledge Base
Generate responses
Select model
Test your Knowledge Base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.
Please select a model
Run
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Amazon Bedrock | ap-southe | Editor | Redshift query editor

ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/knowledge-bases/reinvent-demo/B3LLTADXEG/0

aws | Search [Option+S] | Singapore

Amazon Bedrock > Knowledge Bases > reinvent-demo

reinvent-demo

Test | Delete

Knowledge Base overview

Edit

Knowledge Base name
reinvent-demo

Knowledge Base description
—

Service Role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt

Knowledge Base ID
B3LLTADXEG

Status
Available

Created date
December 01, 2024, 11:30 (UTC-08:00)

Retrieval-Augmented Generation (RAG) type
Structured data store

Query engine
Amazon Redshift

Query engine (1)

Sync

Contains configurations and additional information for how the database is queried.

Find data source

1

	Redshif...	Status	Engine	Redshift c...	Authentic...	Last sync date
○	mp_demo	COMPLETE	PROVISIO...	reinvent-d...	USERNAM...	Sun Dec 01 2024 11:30:54 GMT-...

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Test Knowledge Base

Generate responses

Select model

Test your Knowledge Base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model

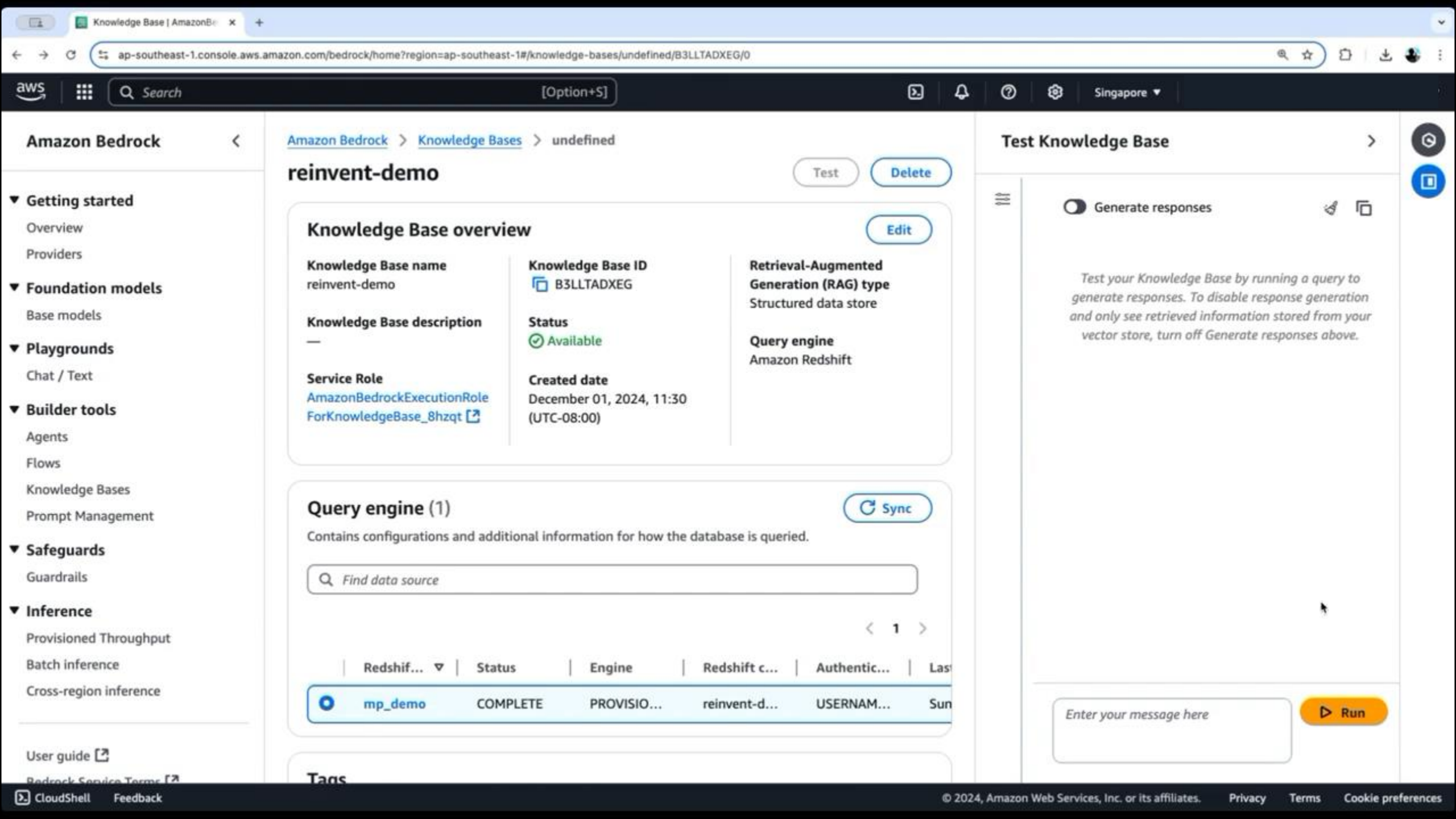
Run

CloudShell | Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy | Terms | Cookie preferences

Improving SQL generation

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


Knowledge Base | AmazonBe
ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/knowledge-bases/undefined/B3LLTADXEG/0
aws
Search
[Option+S]
Singapore
Amazon Bedrock
Getting started
Overview
Providers
Foundation models
Base models
Playgrounds
Chat / Text
Builder tools
Agents
Flows
Knowledge Bases
Prompt Management
Safeguards
Guardrails
Inference
Provisioned Throughput
Batch inference
Cross-region inference
User guide
Amazon Bedrock > Knowledge Bases > undefined
reinvent-demo
Test
Delete
Knowledge Base overview
Edit
Knowledge Base name
reinvent-demo
Knowledge Base description
—
Service Role
AmazonBedrockExecutionRole ForKnowledgeBase_8hzqt
Knowledge Base ID
B3LLTADXEG
Status
Available
Created date
December 01, 2024, 11:30 (UTC-08:00)
Retrieval-Augmented Generation (RAG) type
Structured data store
Query engine
Amazon Redshift
Query engine (1)
Sync
Contains configurations and additional information for how the database is queried.
Find data source
1
Redshif...
Status
Engine
Redshift c...
Authentic...
Last
mp_demo
COMPLETE
PROVISIO...
reinvent-d...
USERNAM...
Sun
Tags
Test Knowledge Base
Generate responses
Test your Knowledge Base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.
Enter your message here
Run
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Knowledge Base | AmazonBe

ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/knowledge-bases/undefined/B3LLTADXEG/0

aws | Search [Option+S] | Singapore

Amazon Bedrock

- **Getting started**
  - Overview
  - Providers
- **Foundation models**
  - Base models
- **Playgrounds**
  - Chat / Text
- **Builder tools**
  - Agents
  - Flows
  - Knowledge Bases
  - Prompt Management
- **Safeguards**
  - Guardrails
- **Inference**
  - Provisioned Throughput
  - Batch inference
  - Cross-region inference

User guide

Bedrock Service Terms

Amazon Bedrock > Knowledge Bases > undefined

reinvent-demo

Test | Delete

Knowledge Base overview

Edit

Knowledge Base name
reinvent-demo

Knowledge Base description
—

Service Role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt

Knowledge Base ID
B3LLTADXEG

Status
Available

Created date
December 01, 2024, 11:30 (UTC-08:00)

Retrieval-Augmented Generation (RAG) type
Structured data store

Query engine
Amazon Redshift

Query engine (1)

Sync

Contains configurations and additional information for how the database is queried.

Find data source

1

Redshif...	Status	Engine	Redshift c...	Authentic...	Last
mp_demo	COMPLETE	PROVISIO...	reinvent-d...	USERNAM...	Sun

Tags

Test Knowledge Base

Generate responses

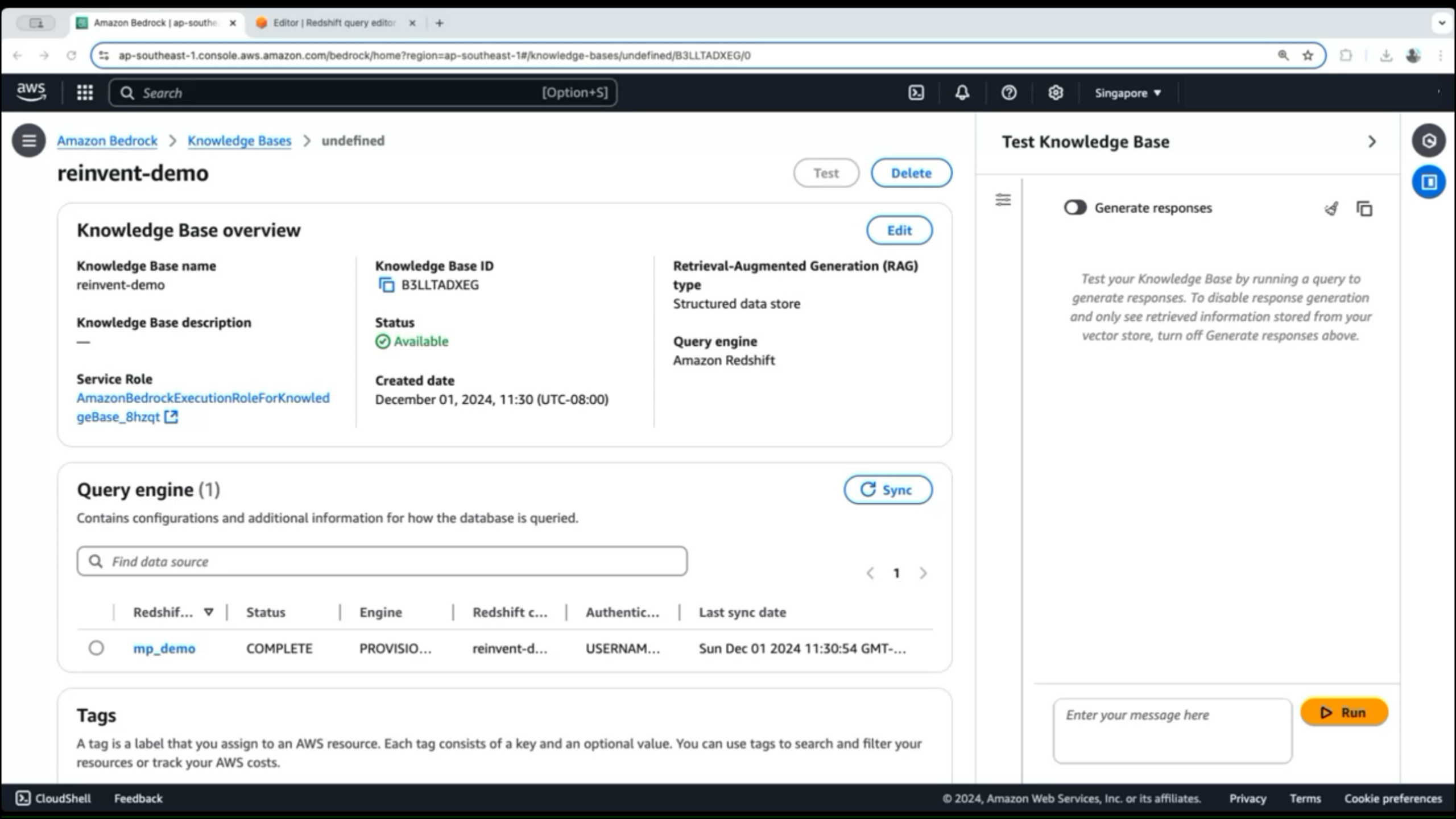
average time from order creation to shipping in november 2024

1. SELECT AVG(EXTRACT(DAY FROM (o.ts3 - o.ts1))) AS avg_delay_days FROM public.orders o WHERE DATE_TRUNC('month', o.ts1) = '2024-11-01';

Enter your message here

Run

CloudShell | Feedback | © 2024, Amazon Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences


Amazon Bedrock | ap-southe
Editor | Redshift query editor
ap-southeast-1.console.aws.amazon.com/bedrock/home?region=ap-southeast-1#/knowledge-bases/undefined/B3LLTADXEG/0
Search
[Option+S]
Singapore
Amazon Bedrock > Knowledge Bases > undefined
reinvent-demo
Test
Delete
Knowledge Base overview
Edit
Knowledge Base name
reinvent-demo
Knowledge Base description
—
Service Role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt
Knowledge Base ID
B3LLTADXEG
Status
Available
Created date
December 01, 2024, 11:30 (UTC-08:00)
Retrieval-Augmented Generation (RAG) type
Structured data store
Query engine
Amazon Redshift
Query engine (1)
Sync
Contains configurations and additional information for how the database is queried.
Find data source
1
Redshif...
Status
Engine
Redshift c...
Authentic...
Last sync date
mp_demo
COMPLETE
PROVISIO...
reinvent-d...
USERNAM...
Sun Dec 01 2024 11:30:54 GMT-...
Tags
A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.
Test Knowledge Base
Generate responses
Test your Knowledge Base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.
Enter your message here
Run
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

Safety features

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

Amazon Bedrock

- Getting started
  - Overview
  - Providers
- Foundation models
  - Base models
- Playgrounds
  - Chat / Text
- Builder tools
  - Agents
  - Flows
  - Knowledge Bases
  - Prompt Management
- Safeguards
  - Guardrails
- Inference
  - Provisioned Throughput
  - Batch inference
  - Cross-region inference

User guide
Bedrock Service Terms

reinvent-demo

Test Delete

Knowledge Base overview

Edit

Knowledge Base name
reinvent-demo

Knowledge Base description
—

Service Role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt

Knowledge Base ID
B3LLTADXEG

Status
Available

Created date
December 01, 2024, 11:30 (UTC-08:00)

Retrieval-Augmented Generation (RAG) type
Structured data store

Query engine
Amazon Redshift

Query engine (1)

Sync

Contains configurations and additional information for how the database is queried.

Find data source

1

Redshif...	Status	Engine	Redshift c...	Authentic...	Last
mp_demo	COMPLETE	PROVISIO...	reinvent-d...	USERNAM...	Sun

Tags

Test Knowledge Base

Generate responses

Test your Knowledge Base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Enter your message here

Run

© 2024, Amazon Web Services, Inc. or its affiliates.

Amazon Bedrock

- Getting started
  - Overview
  - Providers
- Foundation models
  - Base models
- Playgrounds
  - Chat / Text
- Builder tools
  - Agents
  - Flows
  - Knowledge Bases
  - Prompt Management
- Safeguards
  - Guardrails
- Inference
  - Provisioned Throughput
  - Batch inference
  - Cross-region inference

User guide
Bedrock Service Terms

reinvent-demo

Test Delete

Knowledge Base overview

Edit

Knowledge Base name
reinvent-demo

Knowledge Base description
—

Service Role
AmazonBedrockExecutionRoleForKnowledgeBase_8hzqt

Knowledge Base ID
B3LLTADXEG

Status
Available

Created date
December 01, 2024, 11:30 (UTC-08:00)

Retrieval-Augmented Generation (RAG) type
Structured data store

Query engine
Amazon Redshift

Query engine (1)

Sync

Contains configurations and additional information for how the database is queried.

Find data source

1

Redshif...	Status	Engine	Redshift c...	Authentic...	Last
mp_demo	COMPLETE	PROVISIO...	reinvent-d...	USERNAM...	Sun

Tags

Test Knowledge Base

Generate responses

Test your Knowledge Base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Enter your message here

Run

© 2024, Amazon Web Services, Inc. or its affiliates.

Did you catch all that?

Choices to fit your needs

Query engine

Amazon Redshift Serverless and Amazon Redshift provisioned clusters

Data storage

Amazon Redshift Managed Storage and AWS Glue Data Catalog

Auth methods

AWS Identity and Access Management (IAM) and Amazon Redshift DB user

Advanced query configurations

Table and column descriptions

Table and column inclusion/exclusion

Curated queries

Safety features

Permission boundary

Update protection

Delete protection

aws

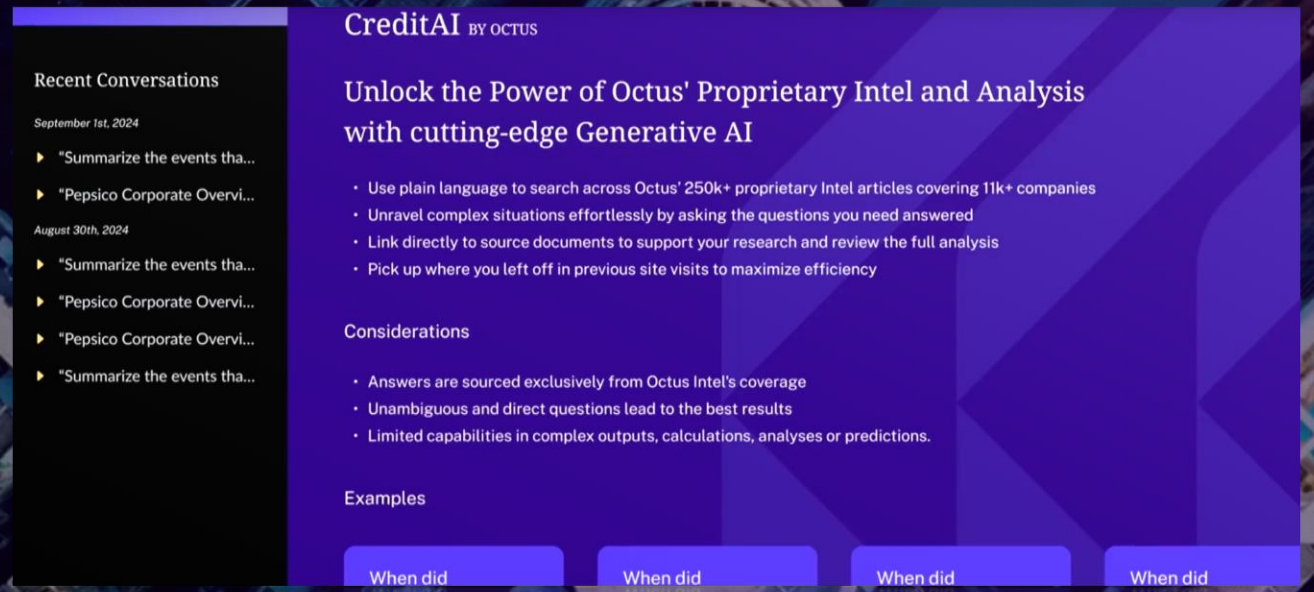
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

OCTUS | Formerly Reorg

Credit intelligence and data resource for investment banks, buy side firms, law firms, and advisory firms.

Octus, a credit intelligence company, plans to leverage the new structured data retrieval capability in Knowledge Bases to allow end users to query structured data using natural language.

By connecting Knowledge Bases to Octus' existing Master Data Management (MDM) system, end-user prompts can be translated into SQL queries that Amazon Bedrock uses to retrieve the relevant information and return it to the user as part of the application's response. This will help Octus' chatbots deliver precise, data-driven insights to its users and enhance the users' interactions with the company's array of data products.



aws

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

Recap

- 4 ways to customize AI experiences
- Retrieval Augmented Generation (RAG) gives FM access to latest information from traditional data sources
- Unstructured data is best handled by semantic search
- Structured data is handled by natural language-to-SQL conversion
- Bedrock Knowledge Bases is a fully managed service to orchestrate your RAG workflow, capable of handling both structured and unstructured data

aws

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

Next steps

- Try Amazon Bedrock Knowledge Bases today!
- Complete the survey
- Come and say hi – we'd love to talk with you

Try Amazon Bedrock Knowledge Bases today!

Bedrock Knowledge Bases Code Samples

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

Thank you!

Please complete the session survey in the mobile app

Try Amazon Bedrock Knowledge Bases today!

Bedrock Knowledge Bases Code Samples

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