

Guía Práctica: AWS Bedrock con RAG

Introducción

Esta guía te llevará paso a paso en la implementación de **AWS Bedrock** con **Retrieval-Augmented Generation (RAG)**, un enfoque que combina modelos de lenguaje con bases de conocimiento personalizadas para generar respuestas más precisas y contextuales.

¿Qué es RAG?

RAG (Retrieval-Augmented Generation) es una técnica que:

- **Recupera** información relevante de una base de conocimiento
- **Aumenta** el prompt del modelo con esta información
- **Genera** respuestas más precisas y fundamentadas

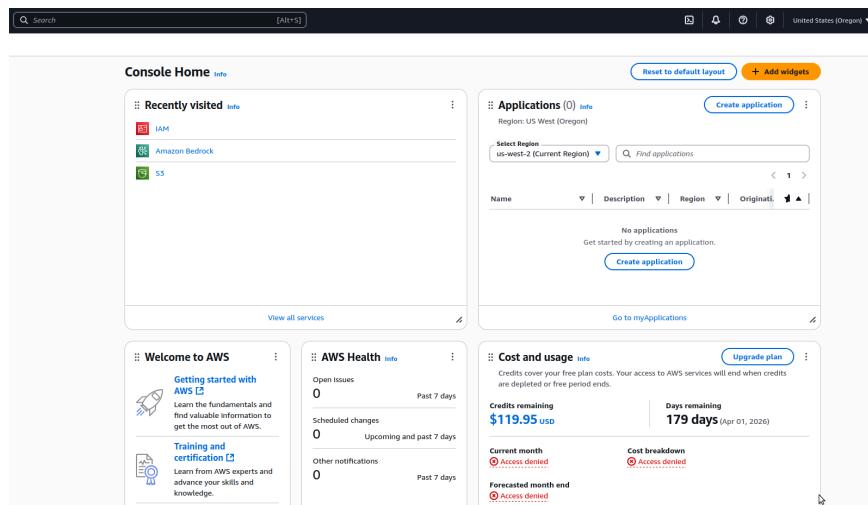
Paso 1: Configuración Inicial de AWS

1.1 Crear Cuenta AWS

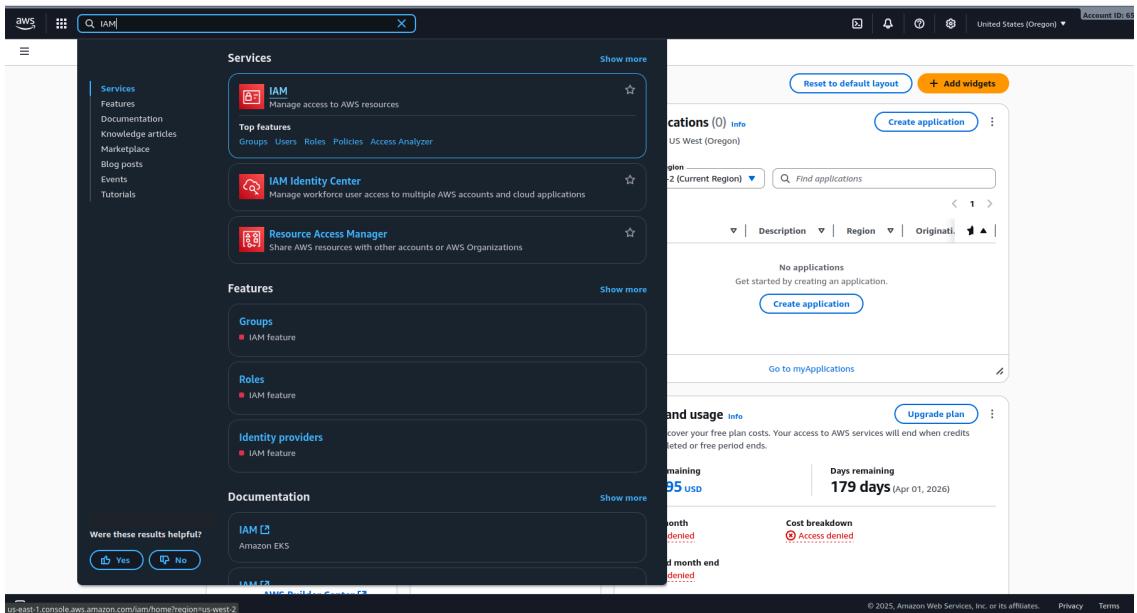
- Ve a <https://aws.amazon.com>
- Haz clic en "Create an AWS Account"
- Completa el proceso de registro (requiere tarjeta de crédito, pero hay tier gratuito)

1.2 Configurar IAM User

Una vez completado el anterior paso, seremos redirigidos aquí:



1. Buscar IAM en la barra de navegación superior



una vez hecho click serás redirigido a:

A screenshot of the AWS IAM Dashboard. The top navigation bar includes the AWS logo, a search bar with 'Search' and 'Alt+S', and a breadcrumb trail 'IAM > Dashboard'. The left sidebar shows 'Identity and Access Management (IAM)' and 'Access management' (User groups, Users, Roles, Policies). It also lists 'Access reports' (Access Analyzer, Resource analysis, Unused access, Analyzer settings, Credential report, Organization activity, Service control policies, Resource control policies), 'IAM Identity Center', and 'AWS Organizations'. The main dashboard area features a blue banner for 'New access analyzers available' (Access Analyzer now analyzes internal access patterns to your critical resources within a single account or across your entire organization). Below this is the 'IAM Dashboard' with 'Security recommendations' (Root user has MFA, Root user has no active access keys), 'IAM resources' (Resources in this AWS Account: User groups 0, Users 0, Roles 4, Policies 3, Identity providers 0), and a 'What's new' section with a 'View all' link and a list of recent changes (Amazon Bedrock introduces API keys for streamlined development, AWS Service Reference Information now supports annotations for service actions, AWS expands resource control policies (RCPs) support to two additional services, AWS IAM now enforces MFA for root users across all account types).

Ir a "users" en el menú lateral

2. Crear usuario:

- User name

3. Permisos:

- Attach existing policies directly
- Seleccionar: por simpleza le asignaremos el **AdministratorAccess** y ya

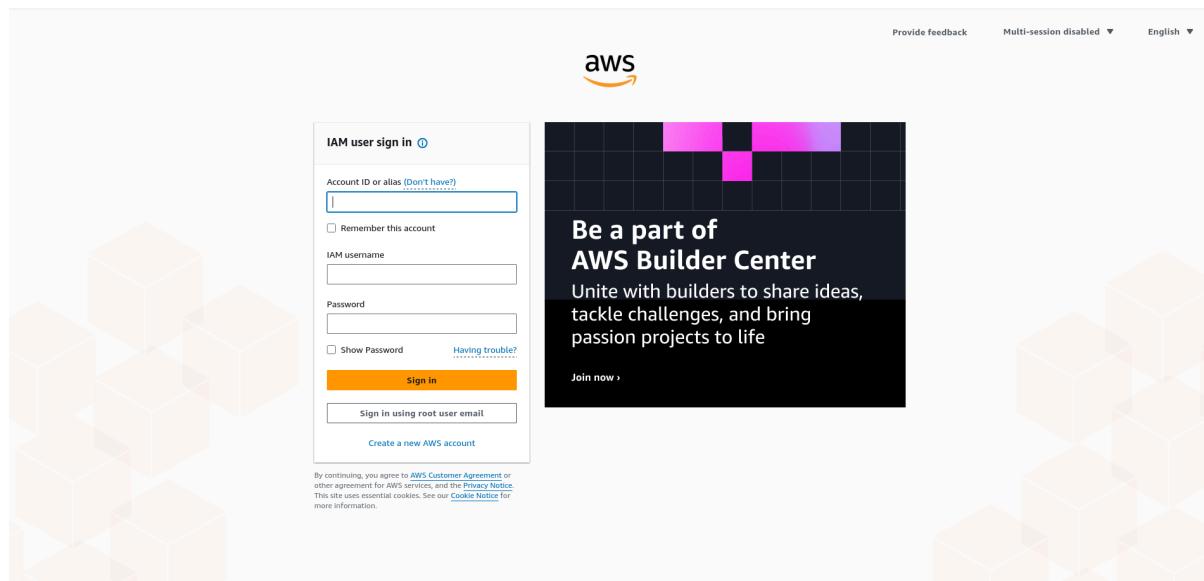
4. Crear usuario y guardar las credenciales

- Hacer click en crear usuario y ser redirigido a:

The screenshot shows the AWS IAM Users page. At the top, a green banner says "User created successfully". Below it, a table lists one user: "pablox1". The table has columns for User name, Path, Groups, Last activity, MFA, Password age, Console last sign-in, Access key ID, Active key age, Access key last use, and ARN. On the right side of the table, there are "Delete" and "Create user" buttons. The left sidebar shows navigation options like Dashboard, Access management (selected), and Access reports.

5. Con el usuario creado, copiamos nuestro account-id que sale en la parte superior derecha que se puede observar en la anterior imagen, para posteriormente hacer sign out.

6. Iniciar sesión con el usuario creado. Se realiza el inicio de sesión



Nota: Inicialmente la password es la misma que la cuenta principal, pero cuando entres con el usuario que acabas de crear te pedirá cambiarla para ese usuario en específico.

Paso 2: Preparar Datos en Amazon S3

2.1 ¿Por qué usar S3?

Amazon S3 nos permite almacenar los documentos que nuestra base de conocimiento utilizará para el RAG.

1. Ubicarse en el home

The screenshot shows the AWS Home page. On the left, there's a sidebar with links like Amazon Bedrock, Billing and Cost Management, and Amazon Polly. The main area has three sections: "Welcome to AWS" (with Getting started with AWS, Training and certification, and AWS Builder Center), "AWS Health" (with Open issues, Scheduled changes, and Other notifications), and "Cost and usage" (showing credits remaining of \$99.95 USD, days remaining of 181 days, and a chart of forecasted monthly costs from May 25 to Oct 25).

2. Buscar "S3" en la barra de navegación superior

The screenshot shows the search results for "S3". The top navigation bar has a search bar with "S3" typed in. The results show the "Services" section with cards for S3 (Scalable Storage in the Cloud), S3 Glacier (Archive Storage in the Cloud), and AWS Snow Family (Large Scale Data Transport). Below that is the "Features" section with cards for Imports from S3 (DynamoDB feature) and Feature spotlight (S3 feature). At the bottom is the "Resources" section with a card for Introducing resource search (enabling cross-region resources for account search). To the right of the search results is the same "Cost and usage" dashboard as in the previous screenshot.

3. Crear bucket:

- dejamos todos los campos por defecto, en este caso lo único que se cambió fue el nombre, se le colocó “ia2demo”
- Después de crear el bucket, serás redireccionado a la lista de buckets.

The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with options like 'General purpose buckets', 'Storage Lens', and 'AWS Marketplace for S3'. The main area displays a table for 'General purpose buckets'. One row is selected, showing the bucket name 'ia2demo', the 'AWS Region' as 'US West (Oregon) us-west-2', and the 'Creation date' as 'October 3, 2025, 23:10:59 (UTC-05:00)'. To the right of the table, there are three cards: 'Account snapshot', 'External access summary - new', and 'Block Public Access settings for this account'.

2.2 Subir Documentos

- Seleccionar el bucket recién creado**
- Hacer clic en "Upload" para ser redireccionado aquí:**

The screenshot shows the 'Upload objects' dialog box. At the top, it says 'Upload objects - S3 bucket L...'. Below that, it shows the URL 'us-west-2.console.aws.amazon.com/s3/upload/ia2demo?region=us-west-2&bucketType=general'. The main area has a large button labeled 'Upload' with a dashed blue outline. Below it is a section titled 'Upload' with the sub-section 'Info'. It says 'Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDKs or Amazon S3 REST API. Learn more'. There's a large dashed blue box for dragging files. Below that is a table titled 'Files and folders (0)' with columns for 'Name', 'Folder', 'Type', and 'Size'. A note says 'All files and folders in this table will be uploaded.' and 'No files or folders'. At the bottom of the dialog are sections for 'Destination', 'Permissions', and 'Properties', each with a 'Cancel' and 'Upload' button.

- Añadir archivos:** Seleccionar documentos PDF, TXT o DOC, en este caso añadiremos el pdf que se encuentra en el repositorio llamado “BATMAN”:

4. Hacer clic en "Upload" y debería de mostrarse esto:

The screenshot shows the AWS Lambda console interface. At the top, there's a green success message: "Upload succeeded. For more information, see the Files and folders table." Below it, a banner says "Upload: status" with a "Close" button. A note says "After you navigate away from this page, the following information is no longer available." The main area has a "Summary" section with "Destination" set to "s3://iazedemo". It shows two rows: "Succeeded" (1 file, 210.9 KB (100.0%)) and "Failed" (0 files, 0 B (0%)). Below this is a "Files and folders" table with one entry: "BATMAN.pdf" (application/pdf, 210.9 KB, Succeeded). The table has columns for Name, Folder, Type, Size, Status, and Error.

Paso 3: Configurar AWS Bedrock

3.1 Acceder a Bedrock

- Volver al home
- Buscar "Amazon Bedrock" en la barra de navegación

The screenshot shows the AWS search results for "Amazon bedrock". The search bar at the top contains the query. On the left, a sidebar lists services like Features, Resources, Documentation, and Marketplace. The main content area shows three service cards: "Amazon Bedrock" (selected), "Amazon Bedrock AgentCore", and "AWS Private Certificate Authority". Below these are sections for "Features" (Prompt Management, Imported Models, Marketplace Deployments) and "Resources" (Introducing resource search). On the right, there's a "Billing and usage" section showing a chart of remaining credits (181 days) and a "Create application" button. The top navigation bar includes account information (Account ID: 6584-8522-1516, United States (Oregon), Pablo).

- Seleccionar el servicio

3.2 Habilitar Model Access

¿Por qué? AWS requiere que habilites explícitamente los modelos que quieras usar.

1. Ir a "Model access" en el menú lateral

The screenshot shows the Amazon Bedrock console interface. On the left, there is a sidebar with several sections: Image / Video playground, Watermark detection, Infer (Cross-region inference, Batch inference, Provisioned Throughput, Custom model on-demand), Tune (Custom models, Prompt router models, Imported models, Marketplace model deployment), Build (Agents, Flows, Knowledge Bases, Automated Reasoning, Guardrails, Prompt Management, Data Automation, AgentCore), Assess (Evaluations), Configure and learn (Settings, Model access, User guide, Bedrock Service Terms). The 'Model access' option is highlighted with a red rectangle.

The main content area has three main sections: Overview (Info), Model catalog, and Model spotlight (DeepSeek V3.1). The Overview section includes a 'Get started by using API Keys' button and a 'Test' section with a 'Chat / Text playground'. The Model catalog section has a 'View Model catalog' button. The Model spotlight section features a 'DeepSeek V3.1' card with a hybrid reasoning model and an 'Open in chat playground' button. There is also a watermark detection section with a code snippet and a 'View watermark detection' button.

2. Hacer clic en "Modify model access"

The screenshot shows the 'Modify model access' page. At the top, there is a yellow banner with the message: '⚠️ Model access page retiring Oct 8 2025. Starting Oct 8 2025, Amazon Bedrock will simplify access to all serverless foundation models, and any new models, by automatically enabling them for every AWS account, eliminating the need to manually activate access through the Bedrock console. Account administrators retain full control over model access through IAM policies [?] and Service Control Policies (SCPs) [?] to restrict model access as needed.' Below the banner, there is a note: 'Note: Selected new models launched before Oct 8 2025 will be auto-enabled, so they won't appear on this page; you can still manage their access via IAM [?]/SCPs [?]. For a complete list of supported models, refer to our documentation [?].'

The main content area has a 'What is Model access?' section with a 'Modify model access' button. Below that is a 'Base models (63)' section with a search bar and a 'Find model' button. A 'Group by provider' dropdown is also present. The base models table lists 12 models under the 'Amazon' provider:

Models	Access status	Modality	EULA
Titan Text G1 - Lite	Access granted	Text	EULA
Titan Text G1 - Express	Available to request	Text	EULA
Titan Embeddings G1 - Text	Available to request	Embedding	EULA
Titan Text Embeddings V2	Available to request	Embedding	EULA
Titan Image Generator G1	Available to request	Image	EULA
Titan Image Generator G1 v2	Available to request	Image	EULA

3. Seleccionar modelos:

- Seleccionamos toda la familia de modelos de amazon:

The screenshot shows the 'Base models' section of the Amazon Bedrock interface. It lists various AI models grouped by provider. The columns include 'Models', 'Access status', 'Modality', and 'EULA'. The 'Access status' column indicates whether access is granted or available to request. The 'Modality' column specifies the type of data the model can process (Text, Embedding, Image, Text & Vision). The 'EULA' column shows the terms of use for each model.

4. Hacer clic en "Next" y posteriormente en "Save changes"

Nota: Hay que esperar unos minutos para que nos habiliten el acceso a estos modelos.

Paso 4: Crear Base de Conocimiento (Knowledge Base)

4.1 ¿Qué es una Knowledge Base?

Es el componente central del RAG que:

- **Almacena** tus documentos procesados
- **Crea embeddings** (representaciones vectoriales del texto)
- **Permite búsquedas semánticas** para encontrar información relevante

4.2 Proceso de Creación

1. Ir a "Knowledge bases" en el menú lateral
2. Seleccionar "Create knowledge base" → "With vector store"

The screenshot shows the 'Knowledge Bases' creation interface in the Amazon Bedrock console. On the left, there's a sidebar with navigation links like 'Discover', 'Infer', 'Tune', and 'Build'. The main area has a 'How it works' section with three steps: 'Create a Knowledge Base with', 'Test the Knowledge Base', and 'Use the Knowledge Base'. Below this is a 'Create Knowledge Base' form with fields for 'Name', 'Status', 'Type', 'Data sources', 'Description', 'Creation time', and 'Last sync date'. A dropdown menu for 'Type' includes options like 'Unstructured data', 'Knowledge Base with vector store', 'Kendra GenAI Index', 'Structured data', and 'Structured data store'. At the bottom, there's a 'Create' button.

Paso 4.2.1: Configure knowledge base details

- dejaremos todos los campos por defecto y hacemos click en “Next”:

The screenshot shows the 'Provide Knowledge Base details' step in the Amazon Bedrock console. On the left, there's a sidebar with navigation links for Discover, Test, Infer, Tune, Build, and Knowledge Bases. The 'Knowledge Bases' section is expanded, showing Automated Reasoning, Guardrails, Prompt Management, and Data Automation. The main panel shows the 'Provide Knowledge Base details' step selected. It has four tabs: Step 1 (selected), Step 2, Step 3, and Step 4. The Step 1 tab contains fields for 'Knowledge Base name' (set to 'knowledge-base-quick-start-f35dv') and 'Knowledge Base description - optional'. Below these are sections for 'IAM permissions' (with 'Create and use a new service role' selected) and 'Service role name' (set to 'AmazonBedrockExecutionRoleForKnowledgeBase_f35dv'). The 'Choose data source type' section shows three options: 'Amazon S3' (selected), 'Web Crawler - Preview', and 'Custom'. The 'Amazon S3' option is described as an object storage service that stores data in general purpose buckets up to 5 buckets as data sources.

Paso 4.2.2: Configure data source

The screenshot shows the 'Configure data source' step in the Amazon Bedrock console. The sidebar and navigation are the same as in the previous step. The main panel shows the 'Configure data source' step selected. It has four tabs: Step 1, Step 2 (selected), Step 3, and Step 4. The Step 2 tab shows the 'Amazon S3' data source configuration. It includes fields for 'Data source name' (set to 'knowledge-base-quick-start-fdrp2-data-source'), 'Data source location' (set to 'This AWS account'), and an 'S3 URI' input field containing 's3://<bucket-name>/<prefix>/<object>'. There are also sections for 'Parsing strategy' (set to 'Amazon Bedrock default parser'), 'Foundation models as a parser' (unchecked), and 'Chunking strategy' (set to 'Default chunking'). A note at the bottom says 'Automatically splits text into chunks of about 300 tokens in size, by default. If a document is less than or already 300 tokens, it's not split an...'. A 'Delete' button is visible in the top right corner of the configuration area.

Los campos los dejamos por defecto a excepción del S3 URI, ahí le haremos click en “Browse S3” y seleccionaremos el bucket que creamos anteriormente para posteriormente clickear en “Next”.

Paso 4.2.3: Configure embeddings and vector store

Step 1
Provide Knowledge Base details

Step 2
Configure data source

**Step 3
Configure data storage and processing**

Step 4
Review and create

Configure data storage and processing

Choose an embeddings model to convert the data that you will provide in the next step, and provide details for a vector data store in which Bedrock can store, manage, and update your embeddings. The embeddings model and vector store cannot be changed after creation of Knowledge Base.

Embeddings model

Select an embeddings model to convert your data into an embedding. Your selection may limit vector stores and embedding types that are available. Pricing depends on the model. [Learn more](#)

Select model

Additional configurations

Vector store Info

Let Amazon create a vector store on your behalf or select a previously created store to allow Bedrock to store, update and manage embeddings. You will be billed directly from the vector store provider. [Learn more](#)

Vector store creation method

Quick create a new vector store - Recommended
A vector store will be created on your behalf in this AWS account during Knowledge Base creation.

Use an existing vector store
Connect to an existing vector store to store, update, and manage embeddings.

Vector store type - new Info

A vector store holds the vector embeddings representation of your data. This choice cannot be changed later. Not all vector stores support binary vector embeddings.

Select a vector store

Cancel Previous Next

- **Embeddings model:** seleccionamos el modelo "Amazon Titan Embeddings V2"
 - **¿Qué son embeddings?** Son representaciones numéricas del texto que permiten comparar similitudes semánticas
- **Vector store:** "Quick create a new vector store"
- **Vector store type:** Amazon S3 vectores preview

Así quedaría la configuración completa para posteriormente clickear en "Next":

Step 1
Provide Knowledge Base details

Step 2
Configure data source

**Step 3
Configure data storage and processing**

Step 4
Review and create

Configure data storage and processing

Choose an embeddings model to convert the data that you will provide in the next step, and provide details for a vector data store in which Bedrock can store, manage, and update your embeddings. The embeddings model and vector store cannot be changed after creation of Knowledge Base.

Embeddings model

Select an embeddings model to convert your data into an embedding. Your selection may limit vector stores and embedding types that are available. Pricing depends on the model. [Learn more](#)

Titan Text Emb... On-demand

Additional configurations

Vector store Info

Let Amazon create a vector store on your behalf or select a previously created store to allow Bedrock to store, update and manage embeddings. You will be billed directly from the vector store provider. [Learn more](#)

Vector store creation method

Quick create a new vector store - Recommended
A vector store will be created on your behalf in this AWS account during Knowledge Base creation.

Use an existing vector store
Connect to an existing vector store to store, update, and manage embeddings.

Vector store type - new Info

A vector store holds the vector embeddings representation of your data. This choice cannot be changed later. Not all vector stores support binary vector embeddings.

Amazon S3 Vectors - Preview
Select to store and index vector embeddings optimized for the cost-effective and durable storage of large, long-term vector data sets while maintaining sub-second query performance.

Additional configurations

Cancel Previous Next

Paso 4.2.4: Review and create

Bases > Create knowledge base with vector store

S3 URI s3://ia2demo	Chunking strategy Default	S3 bucket for Lambda function -
		Data deletion policy DELETE

Step 3: Configure data storage and processing

Embeddings model		
Model Titan Text Embeddings v2	Embedding type Float vector embeddings	Vector dimensions 1024
Vector store		
Quick create vector store - recommended Amazon S3 Vectors	Encryption key ARN -	
Multimodal storage destination		
S3 URI -		

Cancel **Previous** **Create Knowledge Base**

- Revisar el resumen de la configuración
- Hacer clic en "Create knowledge base"
- Si se ha creado correctamente la base de conocimiento serás redireccionado aquí:

Amazon Bedrock > Knowledge Bases > knowledge-base-quick-start-4sg22

Discover Overview Model catalog API keys Infer Chat / Text playground Image / Video playground Watermark detection Tune Cross-region inference Batch inference Provisioned Throughput Custom model on-demand New Build Agents Flows Knowledge Bases Automated Reasoning Guardrails Prompt Management Data Automation	knowledge-base-quick-start-4sg22 <p>(+) Knowledge Base 'knowledge-base-quick-start-4sg22' created successfully. Sync one or more data sources to index your content for searching. Syncing can take from a few minutes to a few hours.</p> <p>Knowledge Base overview</p> <table border="1"> <tr> <td>Knowledge Base name knowledge-base-quick-start-4sg22</td> <td>Knowledge Base ID Z7OKLR9GGL</td> <td>Log Deliveries Configure log deliveries and event logs in the Edit page.</td> </tr> <tr> <td>Knowledge Base description —</td> <td>Status Available</td> <td>Retrieval-Augmented Generation (RAG) type Vector store</td> </tr> <tr> <td>Service Role AmazonBedrockExecutionRoleForKnowledgeBase_4sg22</td> <td>Created date October 04, 2025, 16:54 (UTC-05:00)</td> <td></td> </tr> </table> <p>Data source (1) Data sources contain information returned when querying a Knowledge Base.</p> <table border="1"> <thead> <tr> <th>Find data source</th> <th>Sync</th> <th>Stop sync</th> <th>Add</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/> knowledge... Available S3 6584522... s3://ia2de...</td> <td>Default</td> <td>Default</td> <td>Delete</td> </tr> </tbody> </table> <p>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.</p> <table border="1"> <thead> <tr> <th>Key</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td colspan="2">No tags</td> </tr> </tbody> </table>			Knowledge Base name knowledge-base-quick-start-4sg22	Knowledge Base ID Z7OKLR9GGL	Log Deliveries Configure log deliveries and event logs in the Edit page.	Knowledge Base description —	Status Available	Retrieval-Augmented Generation (RAG) type Vector store	Service Role AmazonBedrockExecutionRoleForKnowledgeBase_4sg22	Created date October 04, 2025, 16:54 (UTC-05:00)		Find data source	Sync	Stop sync	Add	<input type="checkbox"/> knowledge... Available S3 6584522... s3://ia2de...	Default	Default	Delete	Key	Value	No tags	
	Knowledge Base name knowledge-base-quick-start-4sg22	Knowledge Base ID Z7OKLR9GGL	Log Deliveries Configure log deliveries and event logs in the Edit page.																					
	Knowledge Base description —	Status Available	Retrieval-Augmented Generation (RAG) type Vector store																					
	Service Role AmazonBedrockExecutionRoleForKnowledgeBase_4sg22	Created date October 04, 2025, 16:54 (UTC-05:00)																						
	Find data source	Sync	Stop sync	Add																				
<input type="checkbox"/> knowledge... Available S3 6584522... s3://ia2de...	Default	Default	Delete																					
Key	Value																							
No tags																								

4.3 Sincronizar Data Source

¿Por qué? Para procesar los documentos y crear los embeddings.

1. Seleccionar la knowledge base creada
2. Ir a la pestaña "Data sources"
3. Seleccionar el data source

Data source overview

- Name:** knowledge-base-quick-start-jk3q5-data-source
- Data source type:** Amazon S3
- Serverside KMS key:** -
- Account ID:** -
- Chunking strategy:** Default
- Lambda function:** -

Data source ID: YS8P7DNXDS
Created date: October 04, 2025, 16:54 (UTC-05:00)
Status: Available
Data deletion policy: DELETE
Parsing strategy: Default
S3 bucket for Lambda function: -

Sync history (0)

Start time	End time	Status	Source files	Metadata files	Failed files	Added	Deleted	Modified	Metadata fil..
No sync history									

Documents (0)

Find document	Delete document	Add documents
< 1 >		

4. Hacer clic en "Sync"

Nota: Si aparece error de permisos, verificar que el modelo de embedding que colocaste esté habilitado en "Model access".

Paso 5: Probar la Knowledge Base

5.1 Test en Consola

1. Ir a "Knowledge bases" en el menú lateral
2. Ir a tu knowledge base creada
3. Hacer clic en "Test knowledge base"

Knowledge Base overview

- Knowledge Base name:** knowledge-base-quick-start-4sg22
- Knowledge Base description:** -
- Service Role:** AmazonBedrockExecutionRoleForKnowledgeBase_4sg22

Knowledge Base ID: Z7OKL95GGU
Status: Available
Created date: October 04, 2025, 16:54 (UTC-05:00)

Log Deliveries: Configure log deliveries and event logs in the [Edit](#) page.
Retrieval-Augmented Generation (RAG) type: Vector store

Data source (1)

Find data source	Data source ...	Status	Data source type	Account ID	Source Link	Last sync time	Last sync warn...	Chunking stra...	Parsing strategy	Data deletion ...
	<input type="checkbox"/> knowledge-ba...	Available	S3	65848522151...	s3://la2demo	October 04, 20...	-	Default	Default	Delete

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.

Key	Value
No tags	

Embedding model

4. **Seleccionar modelo:** Puedes usar cualquiera, sin embargo para esta guía se usará el "Nova Micro 1.0" ya que habilitamos el uso de todos estos modelos de amazon anteriormente.
5. **Hacer preguntas** sobre el contenido de tus documentos

The screenshot shows the AWS Bedrock interface. On the left, under 'Configurations', there's a section for 'Retrieval and response generation' where 'Retrieval and response generation: data sources and model' is selected. A 'Model' section shows 'Nova Micro 1.0' is chosen. Below that are sections for 'Source' (with 'Source chunks' and 'Search Type'), 'Data manipulation' (with 'Filters' and 'Guardrails'), and a 'Preview' window. The 'Preview' window displays a question 'Quién es Batman?' followed by a detailed response about Batman's origin and history. At the bottom of the preview window is a text input field with placeholder text: 'Write a prompt (Shift + ENTER to start a new line, and ENTER to generate a response)'.

5.2 ¿Qué observar?

- **Respuestas contextuales** basadas en tus documentos
- **Citas numeradas** que referencian las fuentes
- **Hacer clic en las citas** para ver el texto original

Paso 6: Demo Práctica con Python

6.1 Código de la Demo

descarga el archivo llamado “`demo_bedrock.py`” del repositorio

6.2 Instalación de Librerías Requeridas

Antes de ejecutar el código, instala la librería necesaria:

```
pip install boto3
```

¿Qué es boto3?

- Es el **SDK oficial de AWS para Python**
- Permite interactuar con servicios de AWS como Bedrock, S3, IAM, etc.

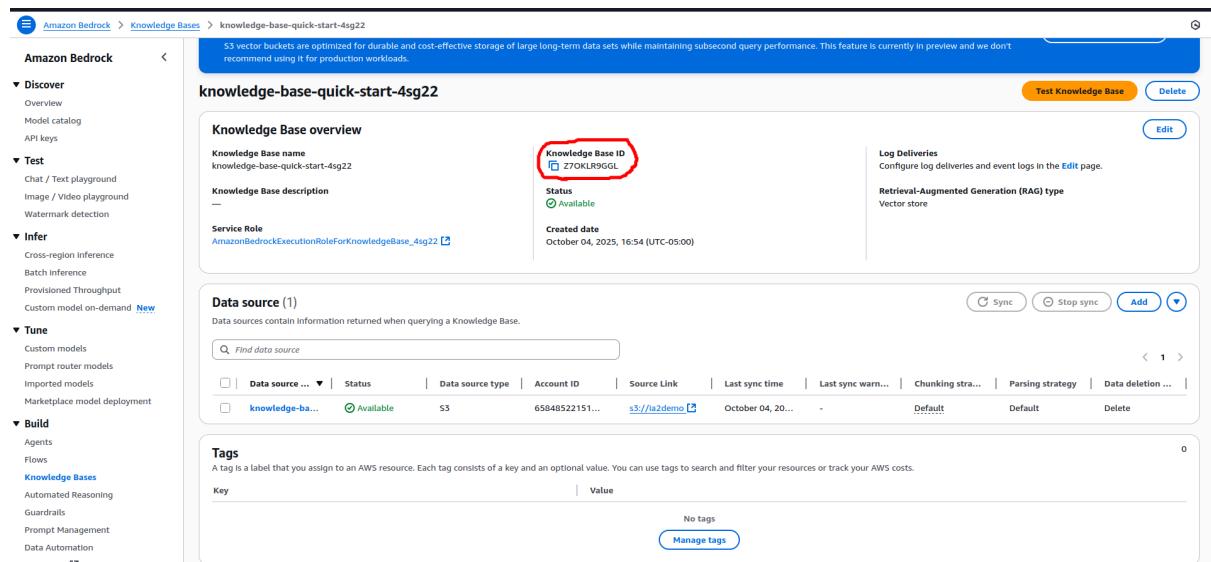
6.3 Configuración de Valores

Reemplaza estos 4 valores en el código:

1. **KNOWLEDGE_BASE_ID**: Tu Knowledge Base ID de Bedrock
2. **AWS_ACCESS_KEY**: Tu AWS Access Key ID
3. **AWS_SECRET_KEY**: Tu AWS Secret Access Key
4. **AWS_REGION**: **us-west-2** (ya configurada, pero debes cambiarla según tu región)
5. **MODEL_NAME**: **nova-micro-1-0** (ya configurado)

¿Dónde encontrar el **KNOWLEDGE_BASE_ID**?

se encuentra dirigiéndose a la base de conocimiento que creamos anteriormente:



The screenshot shows the Amazon Bedrock Knowledge Bases console. On the left, there's a sidebar with navigation links for Discover, Test, Infer, Tune, and Build. The main area displays the 'knowledge-base-quick-start-4sg22' overview. It includes sections for Knowledge Base overview (with a note about vector buckets), Knowledge Base description, Service Role (AmazonBedrockExecutionRoleForKnowledgeBase_4sg22), and Log Deliveries (Retrieval-Augmented Generation (RAG) type). Below this, there's a Data source table with one entry ('knowledge-base-quick-start-4sg22') and a Tags section where a tag named 'knowledge-base-quick-start-4sg22' is listed.

¿Dónde encontrar el **AWS_ACCESS_KEY** y **AWS_SECRET_KEY**?

Haremos click al nuestro nombre de usuario que aparece en la parte superior derecha de la pantalla y le haremos click a security credentials

The screenshot shows the AWS Console Home page. At the top right, it displays the account ID (6584-0522-1516), United States (Oregon) region, and a message about free access ending on April 1, 2026. The 'Security credentials' section is highlighted.

Free plan status

- Credits remaining: \$119.95 USD
- Days remaining: 179 days

Applications (0)

No applications. Get started by creating an application.

Cost and usage

Credits cover your free plan costs. Your access to AWS services will end when credits are depleted or free period ends.

Security credentials

Account ID: 6584-0522-1516
 Account color: Unset
 IAM user: pablox1
 Account: Organization: Service Quotas: Billing and Cost Management: Security credentials:

Turn on multi-session support | Switch role | Sign out

Serás redirigido a la siguiente página, donde deberás hacer click en “create access key para después seleccionar la opción de CLI”:

The screenshot shows the 'My security credentials' page. The 'AWS IAM credentials' tab is selected.

Account details

- User name: pablox1
- AWS account ID: [REDACTED]
- User ARN: [REDACTED]
- Canonical user ID: [REDACTED]

AWS IAM credentials

Console sign-in

- Console sign-in link: https://658485221516.sigin.aws.amazon.com/console
- Console password: Updated 20 hours ago (2025-10-04 16:50 GMT-5)
- Last console sign-in: 3 hours ago (2025-10-05 10:11 GMT-5)

Multi-factor authentication (MFA) (0)

Use MFA to increase the security of your AWS environment. Signing in with MFA requires an authentication code from an MFA device. Each user can have a maximum of 8 MFA devices assigned. Learn more

Type	Identifier	Certifications	Created on
No MFA devices. Assign an MFA device to improve the security of your AWS environment			

Create access key

Después de los pasos anteriores nos arrojará las credenciales las cuales deberemos reemplazar en el código:

The screenshot shows the AWS IAM Access Key creation wizard. The current step is 'Step 3: Retrieve access keys'. A green header bar at the top says 'Access key created' with the note: 'This is the only time that the secret access key can be viewed or downloaded. You cannot recover it later. However, you can create a new access key any time.' Below this, there's a table with two columns: 'Access key' and 'Secret access key'. The 'Access key' column contains a redacted value, and the 'Secret access key' column contains a redacted value followed by '*****' and a 'Show' link. To the right of the table is a 'Download .csv file' button and an orange 'Done' button. On the left, a sidebar lists three steps: 'Access key best practices & alternatives' (selected), 'Set description tag' (disabled), and 'Retrieve access keys' (selected).

Nota: cuando tenga las credenciales inmediatamente pasalas al código, una vez hecho click en "Done" ya no puedes verlas.

6.3 Ejecutar la Demo

```
(env) pablo@pablo-VivoBook-ASUSLaptop-X415DA-M415DA:~/Desktop/IA2/practica aws$ python demo_bedrock.py
💡 Pregunta lo que quieras sobre tus documentos:
Tú: quien es batman?
Bot: Batman es un personaje ficticio que ha trascendido las páginas del cómic para convertirse en un símbolo universal de disciplina, inteligencia y justicia implacable. Su origen se remonta a 1939, cuando Bob Kane y Bill Finger crearon un nuevo héroe para Detective Comics. A diferencia de otros superhéroes, Batman no posee superpoderes, sino que su fuerza proviene de su voluntad inquebrantable, de una mente brillante, y del dolor de un trauma que transformó la tragedia en propósito. El asesinato de sus padres no solo lo marcó emocionalmente, sino que lo convirtió en el reflejo más humano del mito heroico. Desde su creación, Batman ha sido reinventado y reinterpretado por escritores, artistas y cineastas. Cada generación ha visto una nueva versión de este arquetipo moderno, una representación de la capacidad humana para enfrentarse a la oscuridad interior y exterior sin perder la cordura ni la ética.
```

Paso 7: Limpieza de Recursos (Opcional pero Recomendado si ya no harás uso de estas herramientas para que no consuma creditos)

7.1 Eliminar Knowledge Base

1. Ve a AWS Bedrock → Knowledge bases
2. Selecciona tu knowledge base
3. Haz clic en "Delete"
4. Confirma la eliminación

7.2 Eliminar Bucket S3

1. Ve a Amazon S3 → Buckets
2. Selecciona tu bucket
3. Haz clic en "Empty" (vaciar bucket primero)

4. Luego haz clic en "Delete"

Conclusiones

Esta práctica demostró la implementación exitosa de un sistema de IA generativa empresarial usando AWS Bedrock con RAG, donde se creó una base de conocimiento que combina modelos fundacionales como Nova Micro con documentos personalizados almacenados en S3, permitiendo respuestas precisas basadas en información específica. El proceso incluyó la configuración completa desde la creación de usuarios IAM con políticas adecuadas, implementación de buckets S3 para almacenamiento documental, habilitación de modelos en Bedrock, construcción de knowledge bases con vector stores para búsqueda semántica, y desarrollo de una aplicación Python funcional que consume las APIs de Bedrock, validando así la capacidad de AWS para democratizar el acceso a IA avanzada y permitir la creación de asistentes inteligentes contextuales que responden basándose en documentación corporativa específica, todo ello implementable en cuestión de horas sin requerir expertise especializado en machine learning.