
BIG DATA

BASES DE DATOS EN AWS AWS GLUE Y ATHENA

EDUARD LARA

INDICE

1. AWS Glue
2. AWS Athena
3. Laboratorio S3-Glue-Athena

1. AWS GLUE

- AWS Glue es un servicio de integración de datos sin servidor que facilita el descubrimiento, la preparación y la combinación de datos para análisis, aprendizaje automático y desarrollo de aplicaciones.
- AWS Glue proporciona todas las capacidades necesarias para la integración de datos, de modo que pueda comenzar a analizar sus datos y ponerlos en uso en minutos en lugar de meses.
- AWS Glue proporciona interfaces visuales y basadas en código para facilitar la integración de datos.
- Los usuarios pueden encontrar y acceder fácilmente a los datos mediante el catálogo de datos de AWS Glue.

2. AWS ATHENA

- AWS Athena Amazon Athena es un servicio de consulta interactivo que facilita el análisis de datos en Amazon S3 mediante SQL estándar.
- Athena no tiene servidor, por lo que no hay infraestructura que administrar y solo paga por las consultas que ejecuta.
- Athena es fácil de usar. Simplemente, apunte a sus datos en Amazon S3, defina el esquema y comience a realizar consultas mediante SQL estándar. La mayoría de los resultados se entregan en segundos.
- Con Athena, no es necesario realizar trabajos ETL complejos para preparar los datos para el análisis.

2. AWS ATHENA

- Permite que cualquier persona con conocimientos de SQL analice rápidamente conjuntos de datos a gran escala.
- Athena está integrado de fábrica con AWS Glue Data Catalog, lo que le permite crear un repositorio de metadatos unificado en varios servicios, rastrear fuentes de datos para descubrir esquemas y completar su catálogo con definiciones de tablas y particiones nuevas y modificadas, y mantener el control de versiones del esquema.

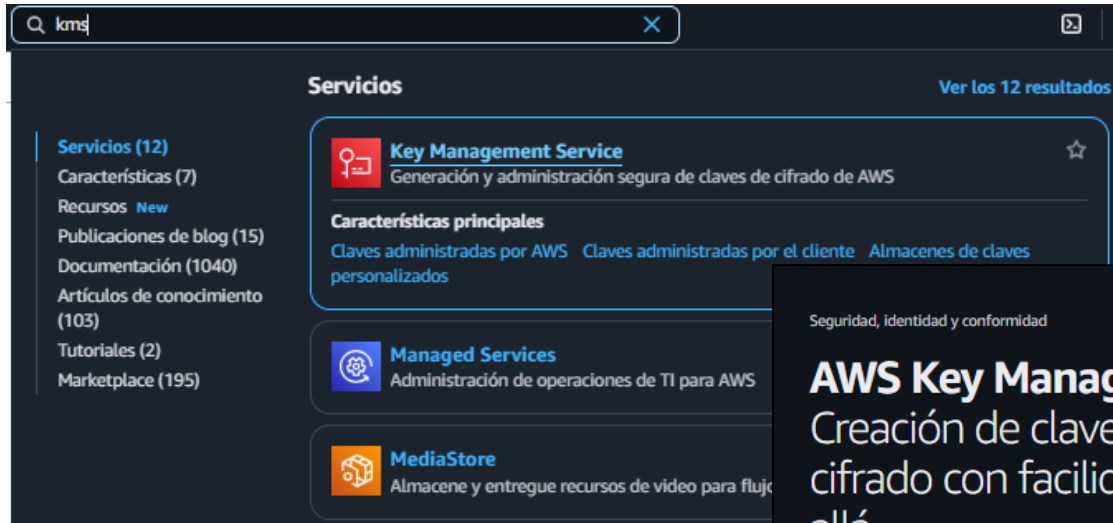
3. LABORATORIO S3 - GLUE - ATHENA

Paso 0. Realizaremos los siguientes pasos:

- 1) Utilizaremos *AWS Key Management Service (KMS)* para encriptar una instantanea de base de datos.
- 2) Crearemos una copia de una base de datos *MySQL*.
- 3) Crearemos un *Bucket* de *S3* a donde exportar snapshot
- 4) Exportaremos la copia de seguridad o snapshot a un *Bucket S3*.
- 5) Aprovecharemos el rastreador de *AWS Glue* para descubrir y almacenar metadatos en el Catálogo de datos.
- 6) *Athena* utilizará estos metadatos para ejecutar consultas *SQL* directamente sobre nuestros datos almacenados en *S3*.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 1. Primero crearemos una llave KMS para cifrado. Ubicaremos el servicio KMS dentro de la consola de AWS. Estando allí pulsaremos "Crear clave".



A screenshot of the AWS Key Management Service (KMS) landing page. The page has a dark blue header with the text 'Seguridad, identidad y conformidad'. The main heading is 'AWS Key Management Service' followed by the subtitle 'Creación de claves y control del cifrado con facilidad en AWS y más allá'. Below this, a paragraph describes KMS as a managed service for key creation and control. On the right side, there is a white box with the text 'Empezar ahora' (Get started now) and a subtext 'Para crear una clave, haga clic en el botón a continuación.' (To create a key, click the button below.) with a yellow 'Crear una clave' (Create a key) button. Below this, there is a 'Precios' (Pricing) section with a link 'Más información' (More information). At the bottom, there is a 'Cómo funciona' (How it works) section with a paragraph explaining KMS functionality. On the right side of the bottom section, there is an 'Introducción' (Introduction) section with a link 'Qué es AWS Key Management Service' and a link 'Introducción a KMS'.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 2. Elegimos el tipo de clave como "Simétrico"; esto implica que la clave servirá tanto para cifrar como para descifrar.

KMS > Claves administradas por el cliente > Crear clave

Presentamos la nueva experiencia Crear clave
Hemos mejorado la experiencia de creación de claves con un editor de políticas mejorado. [Háganos saber lo que piensa](#) o [puede usar la experiencia anterior](#).

Paso 1
Configurar clave

Paso 2
Añadir etiquetas

Paso 3 - opcional
Definir permisos de administración de claves

Paso 4 - opcional
Definir permisos de uso de claves

Paso 5 - opcional
Editar política de claves

Paso 6
Revisar

Configurar clave

Tipo de clave [Ayuda para elegir](#)

☒ **Simétrico**
Una única clave que se utiliza para cifrar y descifrar datos o generar y verificar códigos HMAC

☐ **Asimétrico**
Un par de claves públicas y privadas que se utilizan para cifrar y descifrar datos, firmar y verificar mensajes o derivar secretos compartidos

Uso de claves [Ayuda para elegir](#)

☒ **Cifrado y descifrado**
Utilice la clave solo para cifrar y descifrar datos.

☐ **Generar y verificar MAC**
Utilice la clave solo para generar y verificar códigos de autenticación de mensajes basados en hash (HMAC).

► Opciones avanzadas

Cancelar **Siguiente**

3. LABORATORIO S3 - GLUE - ATHENA

Paso 3. Asignaremos un nombre a la clave y una descripción que facilite la identificación de su propósito.

KMS > Claves administradas por el cliente > Crear clave

Paso 1

Configurar clave

Paso 2

Añadir etiquetas

Paso 3 - opcional

Definir permisos de administración de claves

Paso 4 - opcional

Definir permisos de uso de claves

Paso 5 - opcional

Editar política de claves

Paso 6

Revisar

Añadir etiquetas

Alias
Puede cambiar el alias en cualquier momento. [Más información](#)

Alias
clave_KMS

Descripción - Opcional
Puede cambiar la descripción en cualquier momento.
Descripción
Clave KMS para encriptar instantanea de Base de Datos MySQL

Etiquetas - Opcional
Puede usar etiquetas para clasificar e identificar sus claves de KMS y hacer un seguimiento de los costos de AWS. Cuando agrega etiquetas a los recursos de AWS, se genera un informe de asignación de costos para cada etiqueta. [Más información](#)
Esta clave no tiene etiquetas.
[Agregar etiqueta](#)
Puede agregar hasta 50 etiquetas más.

Cancelar

Saltar a Revisar

Anterior

Siguiente

3. LABORATORIO S3 - GLUE - ATHENA

Paso 4. Concedemos permisos de administración de claves al rol labrole para que pueda administrar esta clave.

KMS > Claves administradas por el cliente > Crear clave

Presentamos la nueva experiencia Crear clave
Hemos mejorado la experiencia de creación de claves con un editor de políticas mejorado. [Háganos saber lo que piensa](#) o puede [usa la experiencia anterior](#).

Paso 1
Configurar clave

Paso 2
Añadir etiquetas

Paso 3 - opcional
Definir permisos de administración de claves

Paso 4 - opcional
Definir permisos de uso de claves

Paso 5 - opcional
Editar política de claves

Paso 6
Revisar

Definir permisos de administración de claves - *opcional*

Administradores de claves (1/21)
Seleccione los roles y usuarios de IAM autorizados para administrar esta clave mediante la API de KMS. Estos administradores se añadirán a la política de claves con el identificador de sentencia (Sid) «Permitir la administración de la clave». La modificación de este Sid podría afectar a la capacidad de la consola para actualizar la sentencia del administrador incluida en la política de claves. [Más información](#)

Q lab X 3 coincidencias < 1 >

| | Nombre | Ruta | Tipo |
|-------------------------------------|-------------------------|------|------|
| <input checked="" type="checkbox"/> | LabRole | / | Role |
| <input type="checkbox"/> | RoleForLambdaModLabRole | / | Role |
| <input type="checkbox"/> | voclabs | / | Role |

Eliminación de claves

☒ Permita que los administradores de claves eliminen esta clave.

[Cancelar](#) [Saltar a Revisar](#) [Anterior](#) [Siguiente](#)

3. LABORATORIO S3 - GLUE - ATHENA

Paso 5. Concedemos permisos de uso de claves al rol labrole para que pueda usar esta clave.

KMS > Claves administradas por el cliente > Crear clave

Presentamos la nueva experiencia Crear clave
Hemos mejorado la experiencia de creación de claves con un editor de políticas mejorado. [Háganos saber lo que piensa](#) o puede [usar la experiencia anterior](#).

Paso 1
Configurar clave

Paso 2
Añadir etiquetas

Paso 3 - opcional
Definir permisos de administración de claves

Paso 4 - opcional
Definir permisos de uso de claves

Paso 5 - opcional
Editar política de claves

Paso 6
Revisar

Definir permisos de uso de claves - *opcional*

Usuarios de claves (1/21)

Seleccione los roles y usuarios de IAM autorizados a usar esta clave en las operaciones criptográficas. Estos usuarios se añadirán a la política de claves con los identificadores de sentencia (Sid) «Permitir el uso de la clave» y «Permitir la conexión de recursos persistentes». La modificación de estos Sids podría afectar a la capacidad de la consola para actualizar las sentencias de usuario de la política de claves. [Más información](#)

3 coincidencias

| <input type="checkbox"/> | Nombre | Ruta | Tipo |
|-------------------------------------|-------------------------|------|------|
| <input checked="" type="checkbox"/> | LabRole | / | Role |
| <input type="checkbox"/> | RoleForLambdaModLabRole | / | Role |
| <input type="checkbox"/> | vociabs | / | Role |

Otras cuentas de AWS

Especifique las cuentas de AWS que pueden usar esta clave. Los administradores de las cuentas que especifique son responsables de la administración de los permisos que autorizan a los usuarios y los roles de IAM a usar esta clave. [Más información](#)

Agregar otra cuenta de AWS

Cancelar

Saltar a Revisar

Anterior

Siguiente

Paso 6. Dejamos por defecto la política de claves

[administradas por el cliente](#) > [Crear clave](#)

Presentamos la nueva experiencia Crear clave

Hemos mejorado la experiencia de creación de claves con un editor de políticas mejorado. Háganos saber lo que piensa o puede usar la experiencia anterior.

12

3. LABORATORIO S3 - GLUE - ATHENA

Paso 7. En Revisar le damos al botón Finalizar y creamos la clave

KMS > Claves administradas por el cliente > Crear clave

Presentamos la nueva experiencia Crear clave
Hemos mejorado la experiencia de creación de claves con un editor de políticas mejorado. [Háganos saber lo que piensa](#) o puede [usa la experiencia anterior](#).

Paso 1
Configurar clave

Paso 2
Añadir etiquetas

Paso 3 - opcional
Definir permisos de administración de claves

Paso 4 - opcional
Definir permisos de uso de claves

Paso 5 - opcional
Editar política de claves

Paso 6
Revisar

Revisar

Configuración de la clave

Tipo de clave
Simétrico

Especificación de la clave
SYMMETRIC_DEFAULT

Uso de claves
Cifrado y descifrado

Origen
AWS KMS

Regionalidad
Clave de una sola región

No puede cambiar la configuración de la clave una vez creada la clave.

Editar

Alias y descripción

Alias
clave_KMS

Descripción
Clave KMS para encriptar instantanea de Base de Datos MySQL

Editar

Etiquetas

| Clave | Valor |
|-------------------------------|-------|
| No hay datos | |
| No hay etiquetas para mostrar | |

Editar

Claves administradas por el cliente (1)

Q Filtrar claves por propiedades o etiquetas

Acciones de clave ▾

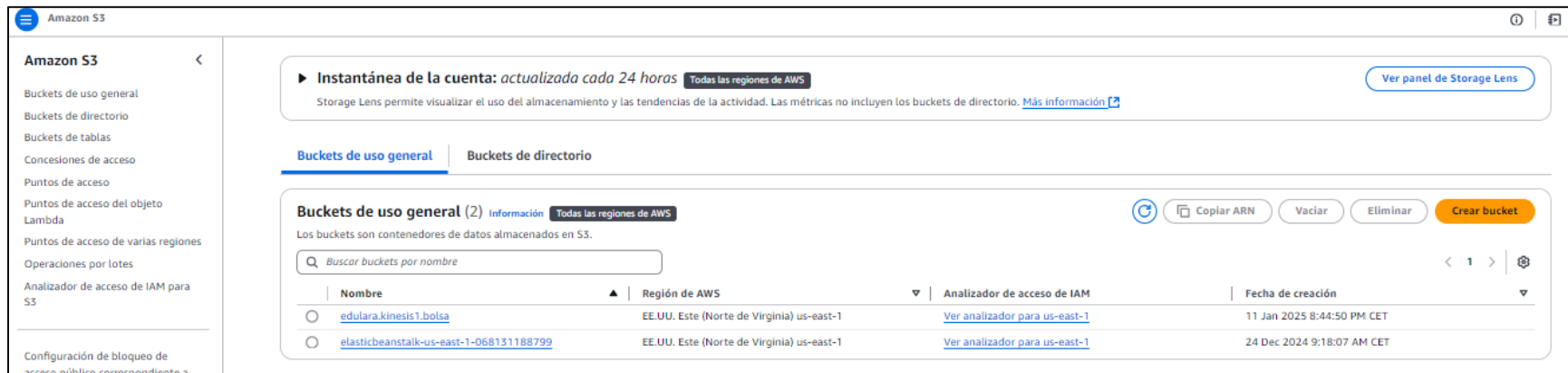
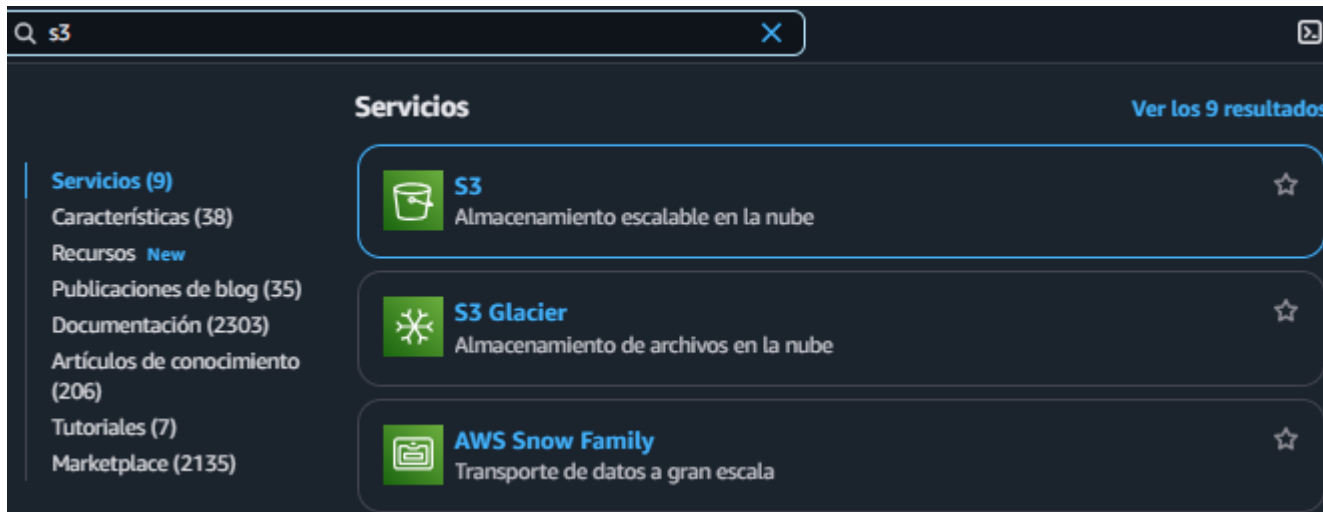
Crear clave

< 1 > ⚙

| <input type="checkbox"/> | Alias | ID de clave | Estado | Tipo de clave | Especificación de la clave ⓘ | Uso de claves |
|--------------------------|-----------|-----------------------------------|------------|---------------|------------------------------|----------------------|
| <input type="checkbox"/> | clave_KMS | b3c3d599-af4f-42c3-9467-0f9028... | Habilitada | Simétrico | SYMMETRIC_DEFAULT | Cifrado y descifrado |

3. LABORATORIO S3 - GLUE - ATHENA

Paso 8. Vamos al servicio de S3 en la consola de AWS y elegimos crear Bucket.



3. LABORATORIO S3 - GLUE - ATHENA

Paso 9. Dejamos la región Norte de Virginia (us-east-1). El nombre del bucket debe de ser único "bucket-mysql-nombreAlumno". Indicamos que sea de uso general

Crear bucket [Información](#)

Los buckets son contenedores de datos almacenados en S3.

Configuración general

Región de AWS

EE.UU. Este (Norte de Virginia) us-east-1

Tipo de bucket [Información](#)

☒ Uso general

Recomendado para la mayoría de los casos de uso y patrones de acceso. Los buckets de uso general son del tipo de bucket de S3 original. Permiten una combinación de clases de almacenamiento que almacenan objetos de forma redundante en múltiples zonas de disponibilidad.

☐ Directorio

Recomendado para casos de uso de baja latencia. Estos buckets utilizan únicamente la clase de almacenamiento S3 Express One Zone, que proporciona un procesamiento más rápido de los datos dentro de una única zona de disponibilidad.

Nombre del bucket [Información](#)

bucket-mysql-elara

El nombre del bucket debe ser único dentro del espacio de nombres global y seguir las reglas de nomenclatura del bucket. [Consulte las reglas para la asignación de nombres de buckets](#)

Copiar la configuración del bucket existente: opcional

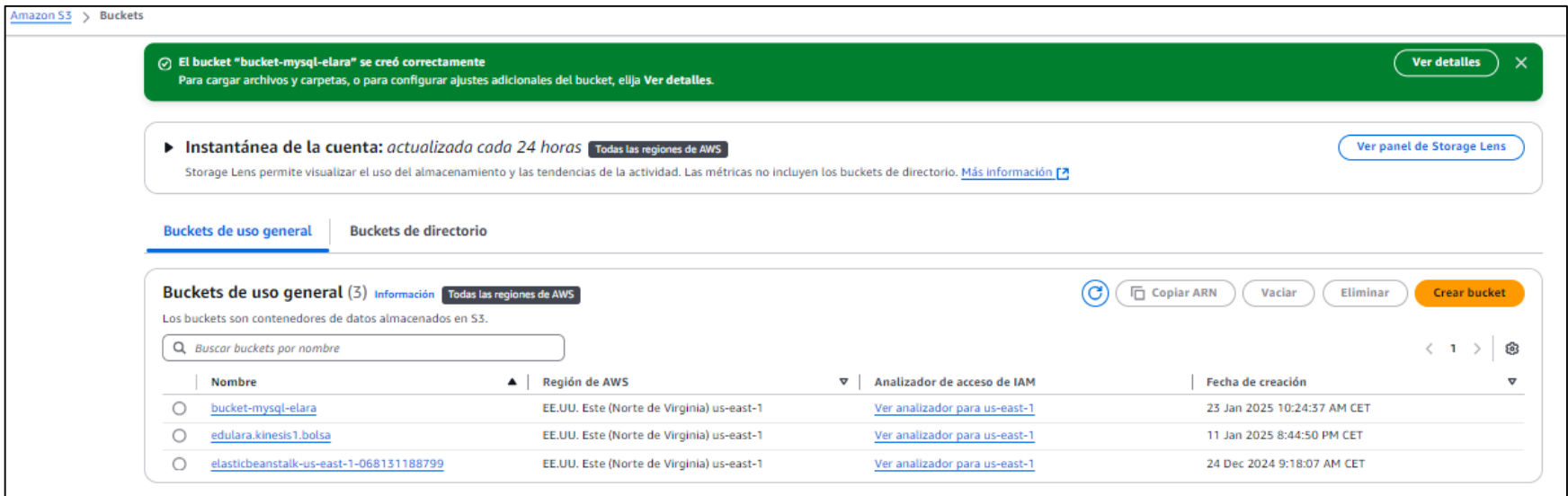
Solo se copia la configuración del bucket en los siguientes ajustes.

[Elegir el bucket](#)

Formato: s3://bucket/prefijo

3. LABORATORIO S3 - GLUE - ATHENA

Paso 10. Dejamos el resto de configuraciones predeterminadas. No necesitamos que nuestro bucket sea de acceso público, tampoco requerimos versionado ni ninguna configuración avanzada. Le damos al botón de crear bucket



The screenshot shows the Amazon S3 Buckets console. At the top, a green notification bar states: "El bucket 'bucket-mysql-elara' se creó correctamente. Para cargar archivos y carpetas, o para configurar ajustes adicionales del bucket, elija Ver detalles." Below this, a section for "Instantánea de la cuenta" indicates it is updated every 24 hours. The main area shows "Buckets de uso general (3)" with a search bar and a table of buckets. The table has columns for Nombre, Región de AWS, Analizador de acceso de IAM, and Fecha de creación. Three buckets are listed: bucket-mysql-elara, edulara.kinesis1.bolsa, and elasticbeanstalk-us-east-1-068131188799.

| Nombre | Región de AWS | Analizador de acceso de IAM | Fecha de creación |
|---|---|---|-----------------------------|
| bucket-mysql-elara | EE.UU. Este (Norte de Virginia) us-east-1 | Ver analizador para us-east-1 | 23 Jan 2025 10:24:37 AM CET |
| edulara.kinesis1.bolsa | EE.UU. Este (Norte de Virginia) us-east-1 | Ver analizador para us-east-1 | 11 Jan 2025 8:44:50 PM CET |
| elasticbeanstalk-us-east-1-068131188799 | EE.UU. Este (Norte de Virginia) us-east-1 | Ver analizador para us-east-1 | 24 Dec 2024 9:18:07 AM CET |

3. LABORATORIO S3 - GLUE - ATHENA

Paso 11. Dado que este bucket será el destino de copias de seguridad, como buena práctica, podríamos acceder al bucket y crear un directorio específico para nuestra base de datos. Entramos en el bucket recién creado y vamos a Crear carpeta.

The screenshot shows the Amazon S3 console interface for a bucket named 'bucket-mysql-elara'. The 'Objetos' tab is active, displaying a list of objects. The interface includes a search bar, a table of objects, and a 'Crear carpeta' button.

bucket-mysql-elara Información

Objetos | Metadatos Vista previa | Propiedades | Permisos | Métricas | Administración | Puntos de acceso

Objetos (0) Información

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

Buscar objetos por prefijo

Acciones: Copiar URI de S3, Copiar URL, Descargar, Abrir, Eliminar, Acciones, Crear carpeta, Cargar

| Nombre | Tipo | Última modificación | Tamaño | Clase de almacenamiento |
|----------------------------------|------|---------------------|--------|-------------------------|
| No hay objetos | | | | |
| No tiene objetos en este bucket. | | | | |

Cargar

3. LABORATORIO S3 - GLUE - ATHENA

Paso 12. A la carpeta le ponemos el nombre "copy-MyDB", el nombre de la base de datos de RDS MySQL

Crear carpeta [Información](#)

Utilice carpetas para agrupar los objetos en buckets. Al crear una carpeta, S3 creará un objeto con el nombre que usted especifique seguido de una barra inclinada (/). Este objeto luego aparecerá como una carpeta en la consola. [Más información](#)

ⓘ Su política de bucket podría bloquear la creación de carpetas
Si su política de bucket impide cargar objetos sin etiquetas, metadatos o beneficiarios específicos de la lista de control de acceso (ACL), no podrá crear una carpeta con esta configuración. En su lugar, puede utilizar la [configuración de carga](#) para cargar una carpeta vacía y especificar la configuración adecuada.

Carpeta

Nombre de la carpeta

copy-MyDB /

Los nombres de las carpetas no pueden contener "/" [Consulte las reglas de nomenclatura](#)

Cifrado del lado del servidor [Información](#)

El cifrado del lado del servidor protege los datos en reposo.

ⓘ La siguiente configuración de cifrado se aplica únicamente al objeto de carpeta y no a los objetos de subcarpeta.

Cifrado del lado del servidor

☒ No especificar una clave de cifrado
La configuración del bucket para el cifrado predeterminado se utiliza para cifrar el objeto de carpeta al almacenarlo en Amazon S3.

☐ Especificar una clave de cifrado
La clave de cifrado especificada se utiliza para cifrar el objeto de carpeta antes de almacenarlo en Amazon S3.

⚠ Si la política del bucket requiere que los objetos se cifren con una clave de cifrado específica, deberá especificar la misma clave de cifrado al crear una carpeta. De lo contrario, se producirá un error al crear la carpeta.

[Cancelar](#) [Crear carpeta](#)

bucket-mysql-elara [Información](#)

[Objetos](#) | [Metadatos Vista previa](#) | [Propiedades](#) | [Permisos](#) | [Métricas](#) | [Administración](#) | [Puntos de acceso](#)

Objetos (1) [Información](#)

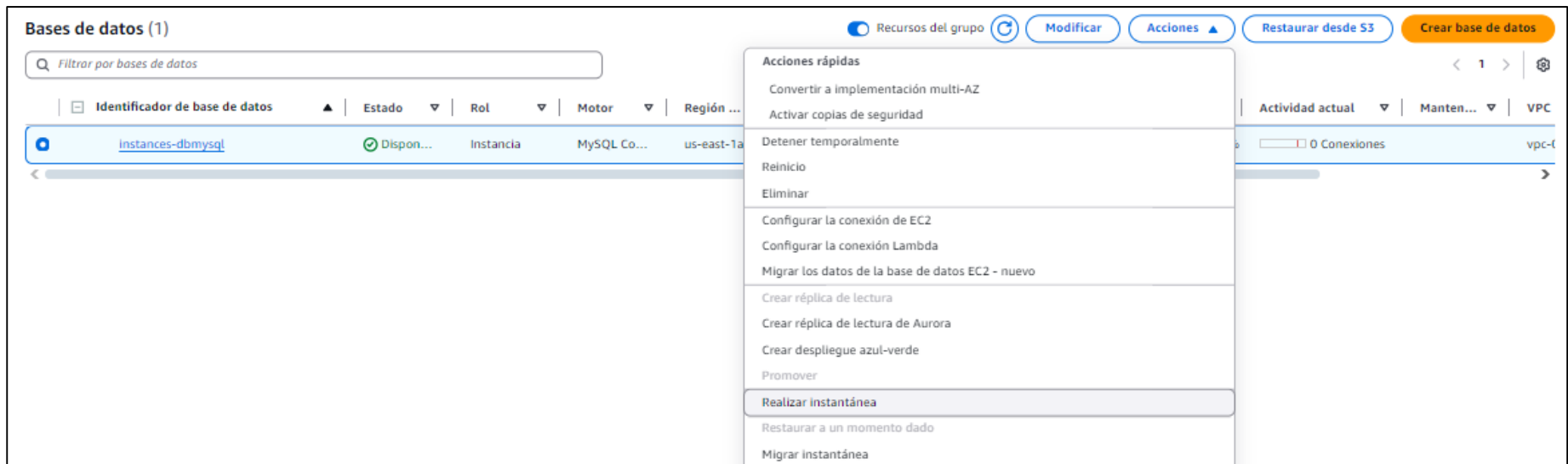
[🔄](#) [Copiar URI de S3](#) [Copiar URL](#) [Descargar](#) [Abrir](#) [Eliminar](#) [Acciones](#) [Crear carpeta](#) [Cargar](#)

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

| <input type="checkbox"/> | Nombre | Tipo | Última modificación | Tamaño | Clase de almacenamiento |
|--------------------------|------------|---------|---------------------|--------|-------------------------|
| <input type="checkbox"/> | copy-MyDB/ | Carpeta | - | - | - |

3. LABORATORIO S3 - GLUE - ATHENA

Paso 13. Crearemos una copia de la base de datos MyDB. Vamos al servicio RDS y seleccionamos nuestra base de datos. Hacemos clic en el botón de acciones y elegimos "Realizar instantánea".



3. LABORATORIO S3 - GLUE - ATHENA

Paso 14. Asignamos un nombre a la instantánea 'snap-MyDB'. Hacemos clic en "Realizar instantánea". Esta acción suele demorar unos minutos

Tomar instantánea de la base de datos

Preferencias

Para tomar una instantánea de base de datos, elija una base de datos y asigne un nombre a la instantánea de base de datos.

Tipo de instantánea

- ☒ Instancia de base de datos
☐ Clúster de base de datos

Instancia de base de datos

Identificador de instancia de base de datos. Es la clave única que identifica una instancia de base de datos.

instances-dbmysql

Nombre de instantánea

Identificador de la instantánea de la base de datos.

snap-MyDB

El identificador de la instantánea no distingue entre mayúsculas y minúsculas, pero se almacena todo en minúsculas, como en "mysnapshot". No puede ser nulo, vacío o estar en blanco. Debe contener de 1 a 255 caracteres alfanuméricos o guiones. El primer carácter debe ser una letra. No puede terminar con un guión ni contener dos guiones consecutivos.

Cancelar

Realizar instantánea

3. LABORATORIO S3 - GLUE - ATHENA

Paso 15. La creación de la copia de seguridad de la base de datos tarda unos minutos

Creando la Instantánea snap-MyDB.

Notificaciones 0 0 1 0 1

Instantáneas de

Manual Sistema Compartido conmigo Pública Servicio de copias de seguridad Exportaciones en Amazon S3

Instantáneas manuales (3)

Acciones Realizar instantánea

Filtrar por instantáneas manuales

| <input type="checkbox"/> | Nombre de la instantánea | Versión del motor | Instancia o clúster de base de datos | Hora de creación de la instantánea | Hora de creación de la instancia de base de datos | Estado |
|--------------------------|---|-------------------------|--------------------------------------|-------------------------------------|---|------------|
| <input type="checkbox"/> | elara-aurora-final-snapshot | 8.0.mysql_aurora.3.08.0 | elara-aurora | January 22, 2025, 18:29 (UTC+01:00) | January 13, 2025, 12:02 (UTC+01:00) | Disponible |
| <input type="checkbox"/> | instances-dbmysql-snapshot | 8.0.39 | instances-dbmysql | January 22, 2025, 17:34 (UTC+01:00) | January 13, 2025, 23:38 (UTC+01:00) | Disponible |
| <input type="checkbox"/> | snap-mydb | 8.0.39 | instances-dbmysql | January 23, 2025, 10:50 (UTC+01:00) | January 22, 2025, 19:14 (UTC+01:00) | Creando |

Se ha creado Instantánea snap-MyDB. Ver detalles

Notificaciones 0 0 2 0 0

Instantáneas de

Manual Sistema Compartido conmigo Pública Servicio de copias de seguridad Exportaciones en Amazon S3

Instantáneas manuales (3)

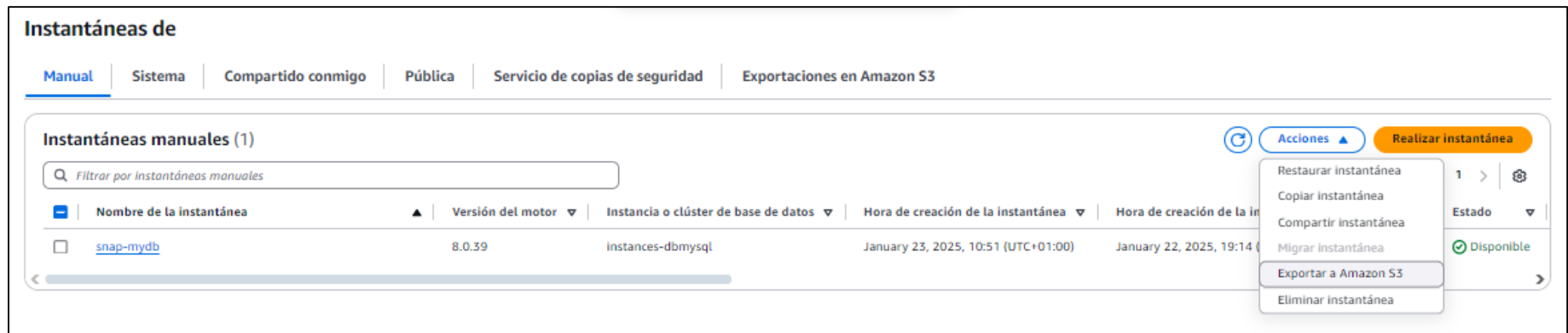
Acciones Realizar instantánea

Filtrar por instantáneas manuales

| <input type="checkbox"/> | Nombre de la instantánea | Versión del motor | Instancia o clúster de base de datos | Hora de creación de la instantánea | Hora de creación de la instancia de base de datos | Estado |
|--------------------------|---|-------------------------|--------------------------------------|-------------------------------------|---|------------|
| <input type="checkbox"/> | elara-aurora-final-snapshot | 8.0.mysql_aurora.3.08.0 | elara-aurora | January 22, 2025, 18:29 (UTC+01:00) | January 13, 2025, 12:02 (UTC+01:00) | Disponible |
| <input type="checkbox"/> | instances-dbmysql-snapshot | 8.0.39 | instances-dbmysql | January 22, 2025, 17:34 (UTC+01:00) | January 13, 2025, 23:38 (UTC+01:00) | Disponible |
| <input type="checkbox"/> | snap-mydb | 8.0.39 | instances-dbmysql | January 23, 2025, 10:50 (UTC+01:00) | January 22, 2025, 19:14 (UTC+01:00) | Creando |

3. LABORATORIO S3 - GLUE - ATHENA

Paso 16. Exportaremos la Instantánea al Bucket S3. Seleccionamos la instantánea, y en el botón "Acciones" elegimos la opción "Exportar a Amazon S3"



The screenshot shows the Amazon RDS console interface for managing snapshots. The top navigation bar includes links for Manual, Sistema, Compartido conmigo, Pública, Servicio de copias de seguridad, and Exportaciones en Amazon S3. The main section is titled "Instantáneas de" and shows a list of manual snapshots under the heading "Instantáneas manuales (1)". A search bar is present with the placeholder text "Filtrar por instantáneas manuales". The table lists the following snapshot:

| | Nombre de la instantánea | Versión del motor | Instancia o clúster de base de datos | Hora de creación de la instantánea | Hora de creación de la instantánea |
|--------------------------|---------------------------|-------------------|--------------------------------------|-------------------------------------|-------------------------------------|
| <input type="checkbox"/> | snap-mydb | 8.0.39 | instances-dbmysql | January 23, 2025, 10:51 (UTC+01:00) | January 22, 2025, 19:14 (UTC+01:00) |

On the right side of the table, there is a dropdown menu labeled "Acciones" with a blue arrow icon. The menu is open, showing the following options: Restaurar instantánea, Copiar instantánea, Compartir instantánea, Migrar instantánea, Exportar a Amazon S3 (highlighted), and Eliminar instantánea. To the right of the menu, there is a status indicator showing "1" and a gear icon, and a status label "Estado" with a dropdown arrow. Below the status label, it says "Disponibles" with a green checkmark icon.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 17. Ingresamos un identificador para la exportación, no es mas que un nombre.

En la sección "datos exportados" queremos exportar la totalidad de nuestros datos, ya que también existe la opción de exportaciones parciales.

[RDS](#) > [Instantáneas de](#) > [snap-mydb](#) > Exportar a Amazon S3

Exportar a Amazon S3 Información

Utilice la exportación de instantáneas de la base de datos de RDS a Amazon S3 para extraer los datos de las instantáneas y almacenarlos en un formato comprimido y consultable en un bucket de S3 en su cuenta de AWS.

Configuración

Identificador de la exportación

Escriba un nombre para identificar la exportación. El nombre debe ser único en todas las exportaciones de instantáneas de base de datos que pertenezcan a su cuenta de AWS en la región de AWS actual.

El identificador de la exportación no distingue entre mayúsculas y minúsculas, pero se almacena todo en minúsculas (por ejemplo, "miexportación"). Restricciones: de 1 a 60 caracteres alfanuméricos o guiones. El primer carácter debe ser una letra. No puede contener dos guiones consecutivos. No puede terminar con un guion.

Datos exportados

Formato de los datos exportados

Cantidad de datos que se van a exportar

Los datos que se exportan son los que están disponibles al inicio de la exportación.

☒ Todos (20 GB)
Se exportan todos los datos de la base de datos.

☐ Parcial
Se exporta una parte de la base de datos. Se puede definir qué parte mediante el uso de identificadores.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 18. En la sección "Destino de S3" elegimos el bucket creado anteriormente "bucket-mysql-nombreAlumno", y en el campo "prefijo" va el nombre de la carpeta donde insertaremos esta exportación "copy-MyDB"

Destino de S3

Bucket de S3

bucket-mysql-elara

Prefijo de S3 -opcional | [Información](#)

copy-MyDB/

Para agrupar objetos en un bucket, S3 utiliza un prefijo antes de los nombres de los objetos. La barra inclinada (/) en el prefijo representa una carpeta. Por ejemplo, utilice el prefijo "exportaciones/2019/" para una carpeta de 2019 en una carpeta de exportaciones. RDS añadirá el prefijo con el carácter "/"

3. LABORATORIO S3 - GLUE - ATHENA

Paso 19. En la sección de "Rol IAM" crearemos un nuevo rol. Por defecto se crea con políticas que permiten insertar datos en S3.

Rol de IAM

Rol de IAM
Elige o crea un rol de IAM para conceder permisos de escritura al bucket de S3.

LabRole

Cifrado

Clave de AWS KMS [Información](#)

claveKMS

Cuenta
068131188799

ID de clave de KMS
988cf961-ba11-45f0-b77d-7d299850298e

Información sobre precios
El costo de exportación de datos de instantáneas a Amazon S3 se basa en el tamaño de la instantánea. [Más información](#)
Se aplican cargos adicionales por almacenar datos exportados en Amazon S3. [Más información](#)

The principal export.rds.amazonaws.com isn't allowed to assume the IAM role arn:aws:iam::068131188799:role/LabRole or the IAM role arn:aws:iam::068131188799:role/LabRole

1. If you don't have permissions to do the following changes, contact your AWS Administrator
2. Navigate to the IAM console
3. In the left navigation pane, click on 'Roles'
4. Search for the role 'LabRole' and click on it
5. If the role doesn't exist, create a new role:

- Click 'Create role'
- Choose 'AWS service' as the trusted entity type
- Select 'RDS' as the use case
- Click 'Next'
- Attach necessary permissions (e.g., AmazonS3FullAccess for S3 export)
- Name the role 'LabRole'
- Click 'Create role'

6. If the role exists, modify its trust relationship:

- Click on the 'Trust relationships' tab
- Click 'Edit trust relationship'
- Add the following statement to the policy document:

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "export.rds.amazonaws.com"
  },
  "Action": "sts:AssumeRole"
}
```

- Click 'Update Trust Policy'

7. Ensure the role has necessary permissions for S3 access:

- Click on the 'Permissions' tab
- If needed, click 'Add permissions' and then 'Attach policies'
- Search for and attach 'AmazonS3FullAccess' or a more specific custom policy

8. Return to the RDS console and retry the export operation

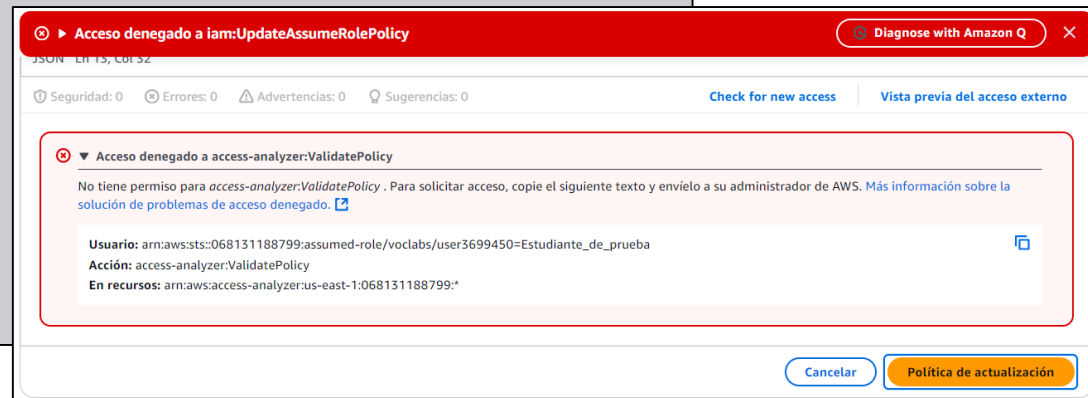
Use of Amazon Q is subject to the [AWS Responsible AI Policy](#). Outputs should be evaluated for accuracy and appropriateness for your use case.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 19. En la capa gratuita no hay ningún rol que tenga la propiedad `export.rds.amazonaws.com`. No se puede crear ningún rol, ni tampoco se puede modificar las propiedades de los roles para realizar esta practica

Editar la política de confianza

```
1 {  
2   "Version": "2012-10-17",  
3   "Statement": [  
4     {  
5       "Effect": "Allow",  
6       "Principal": {  
7         "AWS": "arn:aws:iam::068131188799:role/LabRole",  
8       "Service": [  
9         "events.amazonaws.com",  
10        "servicecatalog.amazonaws.com",  
11        "dynamodb.amazonaws.com",  
12        "kinesisanalytics.amazonaws.com",  
13        "export.rds.amazonaws.com",  
14        "rds.amazonaws.com",  
15        "lambda.amazonaws.com",  
16        "rekognition.amazonaws.com",  
17        "forecast.amazonaws.com",  
18        "iotevents.amazonaws.com",
```



3. LABORATORIO S3 - GLUE - ATHENA

Paso 20. En la seccion "Cifrado" vamos a elegir la clave KMS que creamos al inicio. La clave de cifrado se utiliza para cifrar los datos antes de ser almacenados en el bucket de S3. Esto proporciona una capa adicional de seguridad, especialmente cuando los datos son sensibles o confidenciales.

Procede a realizar la exportacion, esto puede demorar entre 20 y 25 minutos, dependiendo de la cantidad de datos.

Cifrado

Clave de AWS KMS [Información](#)

clave_KMS ▼

Cuenta
068131188799

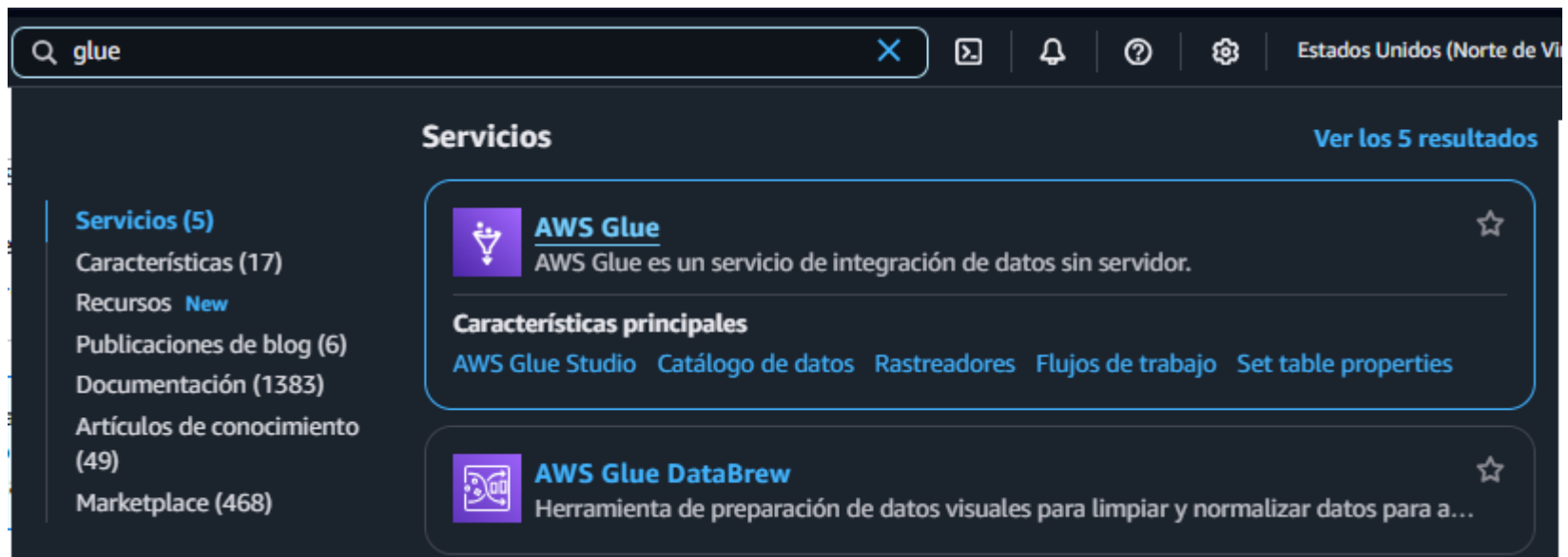
ID de clave de KMS
b3c3d599-af4f-42c3-9467-0f902863186f

Información sobre precios
El costo de exportación de datos de instantáneas a Amazon S3 se basa en el tamaño de la instantánea. [Más información](#)
Se aplican cargos adicionales por almacenar datos exportados en Amazon S3. [Más información](#)

Cancelar Exportar a Amazon S3

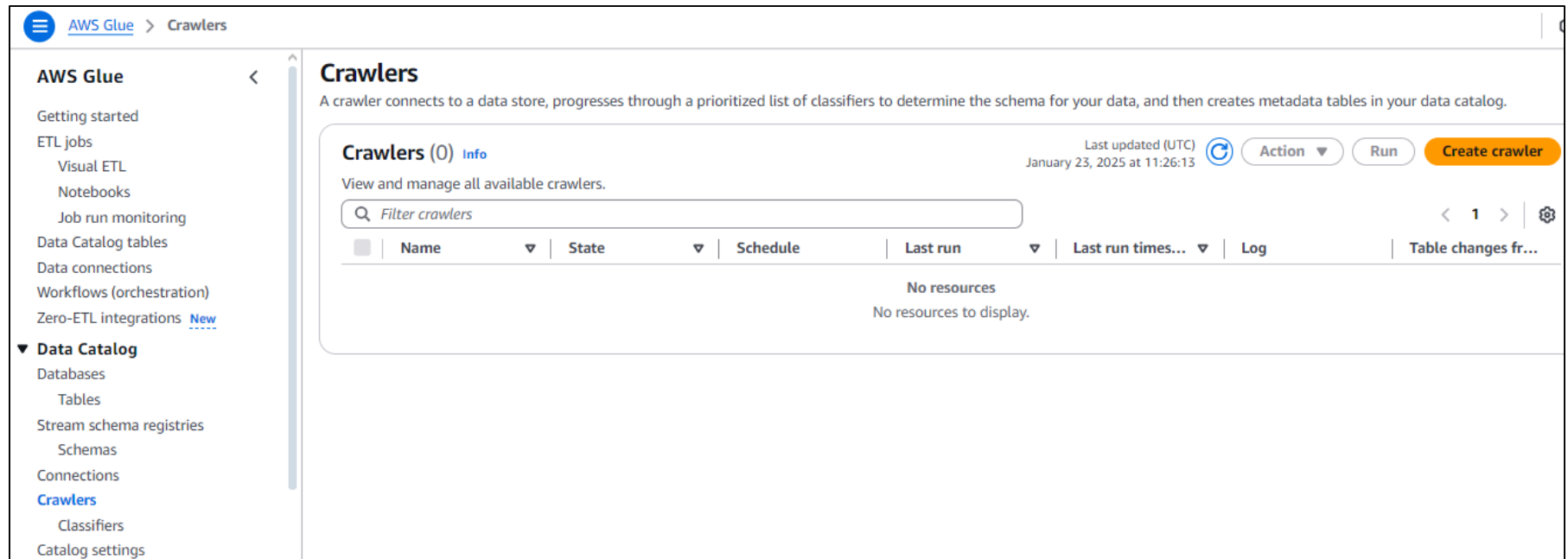
3. LABORATORIO S3 - GLUE - ATHENA

Paso 21. Crearemos un rastreador (crawler) de AWS Glue. Este rastreador permite extraer metadatos esenciales de una copia de nuestra base de datos, que reside en un bucket S3. Estos metadatos facilitaran consultas analíticas desde AWS Athena. Buscamos el servicio AWS Glue desde la consola.



3. LABORATORIO S3 - GLUE - ATHENA

Paso 22. Una vez allí, en la columna izquierda, sección "Data catalog" busca la opción "Crawlers"



The screenshot shows the AWS Glue console interface. On the left, the navigation menu is expanded to the 'Data Catalog' section, where 'Crawlers' is selected. The main content area is titled 'Crawlers' and includes a description: 'A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.' Below this, it says 'Crawlers (0) Info' and 'View and manage all available crawlers.' There is a search bar labeled 'Filter crawlers' and a table with columns: Name, State, Schedule, Last run, Last run times..., Log, and Table changes fr... The table is currently empty, displaying 'No resources' and 'No resources to display.' At the top right of the main area, there are buttons for 'Action', 'Run', and 'Create crawler', along with a timestamp 'Last updated (UTC) January 23, 2025 at 11:26:13'.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 23. Crear crawlers, asigne un nombre y una descripcion, click a "Siguiente".

The screenshot shows the 'Set crawler properties' step in the AWS Glue console. On the left, a vertical list of steps is shown: Step 1 (Set crawler properties, selected), Step 2 (Choose data sources and classifiers), Step 3 (Configure security settings), Step 4 (Set output and scheduling), and Step 5 (Review and create). The main area is titled 'Set crawler properties' and contains two sections. The first section, 'Crawler details', includes a 'Name' field with the value 'Mi_crawler' and a description field with the value 'Crawler para metadatos'. The second section, 'Tags - optional', is currently collapsed. At the bottom right, there are 'Cancel' and 'Next' buttons.

Step 1
● **Set crawler properties**

Step 2
○ Choose data sources and classifiers

Step 3
○ Configure security settings

Step 4
○ Set output and scheduling

Step 5
○ Review and create

Set crawler properties

Crawler details [Info](#)

Name

Mi_crawler

Name can be up to 255 characters long. Some character set including control characters are prohibited.

Description - optional

Crawler para metadatos

Descriptions can be up to 2048 characters long.

► **Tags - optional**

Use tags to organize and identify your resources.

[Cancel](#) [Next](#)

3. LABORATORIO S3 - GLUE - ATHENA

Paso 24. En la sección de "Data Sources" agregaremos la fuente de los datos de donde el crawler obtendrá la metadata, en nuestro caso es un bucket de S3.

The screenshot shows the 'Choose data sources and classifiers' step in the AWS Glue console. On the left, a vertical navigation pane lists five steps: 'Step 1: Set crawler properties', 'Step 2: Choose data sources and classifiers' (which is selected with a blue circle), 'Step 3: Configure security settings', 'Step 4: Set output and scheduling', and 'Step 5: Review and create'.

The main content area is titled 'Choose data sources and classifiers'. It contains a section for 'Data source configuration' with the question 'Is your data already mapped to Glue tables?'. There are two radio button options: 'Not yet' (selected) with the subtext 'Select one or more data sources to be crawled.', and 'Yes' with the subtext 'Select existing tables from your Glue Data Catalog.'.

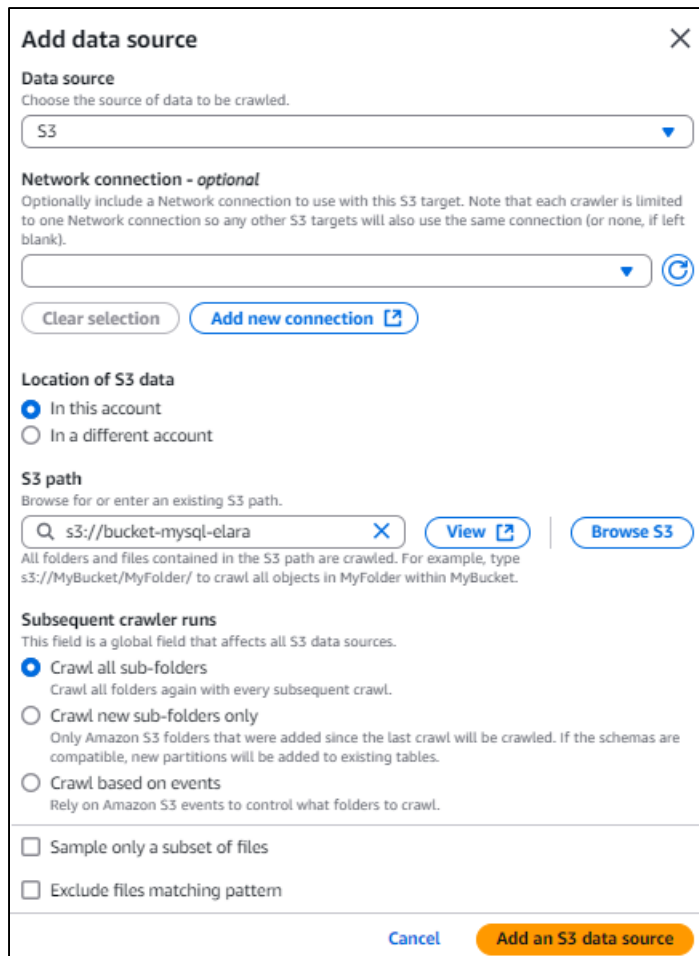
Below this is a section for 'Data sources (1)' with an 'Info' link. It states 'The list of data sources to be scanned by the crawler.' and includes 'Edit', 'Remove', and 'Add a data source' buttons. A table lists the current data source:

| Type | Data source | Parameters |
|------|-------------------------|-------------|
| S3 | s3://bucket-mysql-elara | Recrawl all |

At the bottom, there is a section for 'Custom classifiers - optional' with a description: 'A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.' At the bottom right, there are 'Cancel', 'Previous', and 'Next' buttons.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 25. Al pulsar "Add sources", indicaremos "S3" como origen de los datos, y seleccionaremos nuestro Bucket



The screenshot shows the 'Add data source' dialog box in the AWS Glue console. The 'Data source' dropdown is set to 'S3'. Under 'Network connection - optional', there is a dropdown menu and a 'Clear selection' button. The 'Location of S3 data' section has two radio buttons: 'In this account' (selected) and 'In a different account'. The 'S3 path' section has a text input field containing 's3://bucket-mysql-elara', a 'View' button, and a 'Browse S3' button. Below this, there are three radio buttons for 'Subsequent crawler runs': 'Crawl all sub-folders' (selected), 'Crawl new sub-folders only', and 'Crawl based on events'. At the bottom, there are two checkboxes: 'Sample only a subset of files' and 'Exclude files matching pattern'. The dialog has a 'Cancel' button and an 'Add an S3 data source' button.

Add data source

Data source
Choose the source of data to be crawled.

S3

Network connection - optional
Optionally include a Network connection to use with this S3 target. Note that each crawler is limited to one Network connection so any other S3 targets will also use the same connection (or none, if left blank).

Clear selection Add new connection

Location of S3 data

☒ In this account
☐ In a different account

S3 path
Browse for or enter an existing S3 path.

s3://bucket-mysql-elara View Browse S3

Subsequent crawler runs
This field is a global field that affects all S3 data sources.

☒ Crawl all sub-folders
Crawl all folders again with every subsequent crawl.

☐ Crawl new sub-folders only
Only Amazon S3 folders that were added since the last crawl will be crawled. If the schemas are compatible, new partitions will be added to existing tables.

☐ Crawl based on events
Rely on Amazon S3 events to control what folders to crawl.

☐ Sample only a subset of files

☐ Exclude files matching pattern

Cancel Add an S3 data source

Elegimos la opción de rastrear todas las subcarpetas que existan dentro del bucket
Hacemos click en "Agregar fuente de datos S3" para finalizar esta tarea.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 26. Agregada la fuente, pasamos al siguiente paso,

The screenshot shows the 'Choose data sources and classifiers' step in the AWS Glue console. On the left, a vertical list of steps is shown: Step 1 (Set crawler properties), Step 2 (Choose data sources and classifiers - selected), Step 3 (Configure security settings), Step 4 (Set output and scheduling), and Step 5 (Review and create). The main content area is titled 'Choose data sources and classifiers'. It features a 'Data source configuration' section with a question 'Is your data already mapped to Glue tables?'. Two radio buttons are present: 'Not yet' (selected) with the subtext 'Select one or more data sources to be crawled.', and 'Yes' with the subtext 'Select existing tables from your Glue Data Catalog.'. Below this is a 'Data sources (1)' section with an 'Info' link. It includes a table with columns 'Type', 'Data source', and 'Parameters'. The table contains one entry: 'S3', 's3://bucket-mysql-elara', and 'Recrawl all'. To the right of the table are buttons for 'Edit', 'Remove', and 'Add a data source'. At the bottom, there is a section for 'Custom classifiers - optional' with a descriptive text. At the very bottom right, there are 'Cancel', 'Previous', and 'Next' buttons.

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Choose data sources and classifiers

Data source configuration

Is your data already mapped to Glue tables?

☒ Not yet
Select one or more data sources to be crawled.

☐ Yes
Select existing tables from your Glue Data Catalog.

Data sources (1) [Info](#)

The list of data sources to be scanned by the crawler.

| Type | Data source | Parameters |
|--------------------------|-------------------------|-------------|
| <input type="radio"/> S3 | s3://bucket-mysql-elara | Recrawl all |

[Edit](#) [Remove](#) [Add a data source](#)

► Custom classifiers - optional

A classifier checks whether a given file is in a format the crawler can handle. If it is, the classifier creates a schema in the form of a StructType object that matches that data format.

[Cancel](#) [Previous](#) [Next](#)

3. LABORATORIO S3 - GLUE - ATHENA

Paso 27. En el paso Security Setting, crearemos un rol, solo debemos asignarle un nombre y este se creara con los permisos necesarios para que el crawler acceda a S3, colocar un nombre descriptivo puede ser buena practica para identificar o reutilizar, click en "Siguiente"

The screenshot shows the 'Configure security settings' step in the AWS Glue console. On the left, a vertical navigation pane lists five steps: 'Set crawler properties', 'Choose data sources and classifiers', 'Configure security settings' (which is highlighted with a blue circle), 'Set output and scheduling', and 'Review and create'. The main content area is titled 'Configure security settings' and contains three sections. The first section, 'IAM role', includes an 'Info' link, a dropdown menu for 'Existing IAM role' with 'LabRole' selected, a 'View' button with an external link icon, and two buttons: 'Create new IAM role' and 'Update chosen IAM role'. A note below states: 'Only IAM roles created by the AWS Glue console and have the prefix "AWSGlueServiceRole-" can be updated.' The second section, 'Lake Formation configuration - optional', includes a 'Learn more' link and a checkbox labeled 'Use Lake Formation credentials for crawling S3 data source'. The checkbox is currently unchecked. Below the checkbox, a note explains: 'Checking this box will allow the crawler to use Lake Formation credentials for crawling the data source. If the data source is registered in another account, you must provide the registered account ID. Otherwise, the crawler will crawl only those data sources associated to the account. Only applicable to S3, Glue Catalog, Iceberg, and Hudi data sources.' The third section, 'Security configuration - optional', has a right-pointing triangle icon and the text 'Enable at-rest encryption with a security configuration.' At the bottom right of the console, there are three buttons: 'Cancel', 'Previous' (highlighted with a blue border), and 'Next' (highlighted with an orange background).

Step 1
Set crawler properties

Step 2
Choose data sources and classifiers

Step 3
Configure security settings

Step 4
Set output and scheduling

Step 5
Review and create

Configure security settings

IAM role [Info](#)

Existing IAM role

LabRole [View](#)

[Create new IAM role](#) [Update chosen IAM role](#)

Only IAM roles created by the AWS Glue console and have the prefix "AWSGlueServiceRole-" can be updated.

Lake Formation configuration - optional

Allow the crawler to use Lake Formation credentials for crawling the data source. [Learn more.](#)

☐ Use Lake Formation credentials for crawling S3 data source

Checking this box will allow the crawler to use Lake Formation credentials for crawling the data source. If the data source is registered in another account, you must provide the registered account ID. Otherwise, the crawler will crawl only those data sources associated to the account. Only applicable to S3, Glue Catalog, Iceberg, and Hudi data sources.

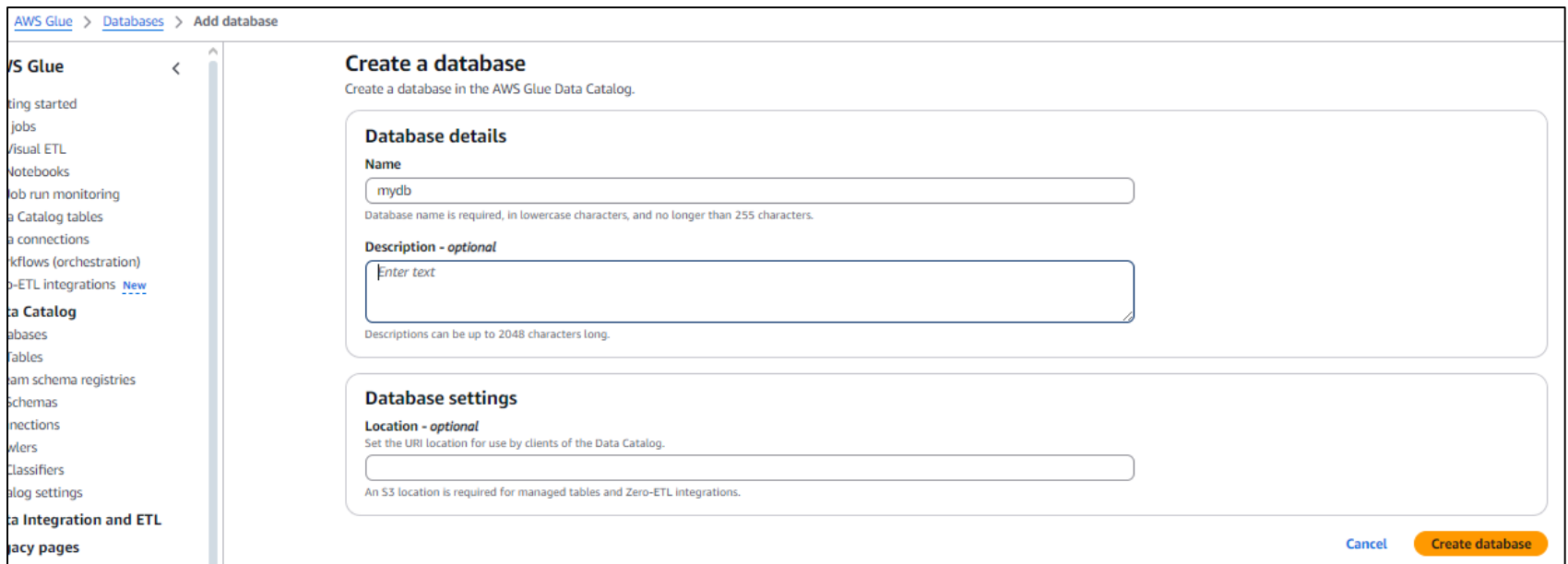
► Security configuration - optional

Enable at-rest encryption with a security configuration.

[Cancel](#) [Previous](#) [Next](#)

3. LABORATORIO S3 - GLUE - ATHENA

Paso 28. Ahora debemos agregar una base de datos, click en "Agregar base de datos".



The screenshot shows the AWS Glue console interface for creating a new database. The breadcrumb navigation at the top reads 'AWS Glue > Databases > Add database'. On the left, a sidebar menu lists various AWS Glue services, with 'Databases' highlighted under the 'Data Catalog' section. The main content area is titled 'Create a database' and includes the instruction 'Create a database in the AWS Glue Data Catalog.' Below this, there are two main sections: 'Database details' and 'Database settings'. In the 'Database details' section, the 'Name' field contains 'mydb' and has a note stating 'Database name is required, in lowercase characters, and no longer than 255 characters.' The 'Description - optional' field is empty with a placeholder 'Enter text' and a note stating 'Descriptions can be up to 2048 characters long.' In the 'Database settings' section, the 'Location - optional' field is empty with a note stating 'Set the URI location for use by clients of the Data Catalog.' and 'An S3 location is required for managed tables and Zero-ETL integrations.' At the bottom right of the form, there are two buttons: 'Cancel' and 'Create database'.

[AWS Glue](#) > [Databases](#) > Add database

Create a database

Create a database in the AWS Glue Data Catalog.

Database details

Name

Database name is required, in lowercase characters, and no longer than 255 characters.

Description - optional

Descriptions can be up to 2048 characters long.

Database settings

Location - optional

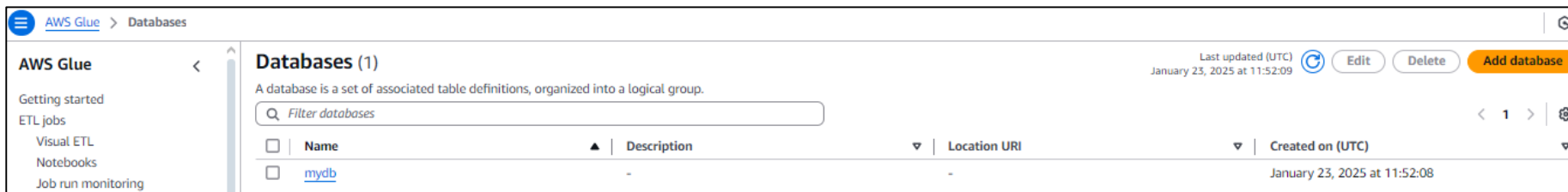
Set the URI location for use by clients of the Data Catalog.

An S3 location is required for managed tables and Zero-ETL integrations.

[Cancel](#) [Create database](#)

3. LABORATORIO S3 - GLUE - ATHENA

Paso 29. Esta base de datos actúa como un repositorio organizado para los metadatos extraídos por el rastreador, incluyendo información sobre tablas y columnas. La creación de esta base de datos no solo estructura de manera ordenada la información, sino que también facilita la realización de consultas analíticas más efectivas en Athena en fases posteriores del proceso.



The screenshot displays the AWS Glue Databases console. On the left, a navigation menu includes 'AWS Glue', 'Getting started', 'ETL jobs', 'Visual ETL', 'Notebooks', and 'Job run monitoring'. The main content area is titled 'Databases (1)' and includes a description: 'A database is a set of associated table definitions, organized into a logical group.' Below this is a search bar labeled 'Filter databases'. A table lists the databases with columns for 'Name', 'Description', 'Location URI', and 'Created on (UTC)'. One database, 'mydb', is listed with a description of '-' and a creation time of 'January 23, 2025 at 11:52:08'. In the top right corner, there are buttons for 'Edit', 'Delete', and 'Add database', along with a 'Last updated (UTC)' timestamp of 'January 23, 2025 at 11:52:09'.

| Name | Description | Location URI | Created on (UTC) |
|------|-------------|--------------|------------------------------|
| mydb | - | - | January 23, 2025 at 11:52:08 |

3. LABORATORIO S3 - GLUE - ATHENA

Paso 30. Después de crear la base de datos, la seleccionamos como nuestra base de datos objetivo, y dejamos el resto de configuraciones predeterminadas

The screenshot shows the 'Set output and scheduling' step in the AWS Glue console. On the left, a vertical list of steps is shown: Step 1 (Set crawler properties), Step 2 (Choose data sources and classifiers), Step 3 (Configure security settings), Step 4 (Set output and scheduling, which is the current step and highlighted with a blue circle), and Step 5 (Review and create). The main content area is titled 'Set output and scheduling' and contains two sections: 'Output configuration' and 'Crawler schedule'. The 'Output configuration' section includes a 'Target database' dropdown menu set to 'mydb', with 'Clear selection' and 'Add database' buttons. Below this is a 'Table name prefix - optional' text input field with the placeholder 'Type a prefix added to table names'. Further down is a 'Maximum table threshold - optional' section with a text input field and a description: 'This field sets the maximum number of tables the crawler is allowed to generate. In the event that this number is surpassed, the crawl will fail with an error. If not set, the crawler will automatically generate the number of tables depending on the data schema.' Below this is a collapsed 'Advanced options' section. The 'Crawler schedule' section includes a description: 'You can define a time-based schedule for your crawlers and jobs in AWS Glue. The definition of these schedules uses the Unix-like cron syntax. Learn more.' and a 'Frequency' dropdown menu set to 'On demand'. At the bottom right, there are three buttons: 'Cancel', 'Previous', and 'Next'.

El rastreador de Glue se dirigirá a S3, recuperará la información, y la almacenará en la recién creada base de datos en el Catálogo de AWS Glue.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 31. En el último paso, hacemos una revisión de todos los parámetros establecidos y le damos a "Crear crawler"

Step 1

Set crawler properties

Step 2

Choose data sources and classifiers

Step 3

Configure security settings

Step 4

Set output and scheduling

Step 5

Review and create

Review and create

Step 1: Set crawler properties

Edit

Set crawler properties

| Name | Description | Tags |
|------------|------------------------|------|
| MI_crawler | Crawler para metadatos | - |

Step 2: Choose data sources and classifiers

Edit

Data sources (1) Info

The list of data sources to be scanned by the crawler.

| Type | Data source | Parameters |
|------|-------------------------|-------------|
| S3 | s3://bucket-mysql-elara | Recrawl all |

Step 3: Configure security settings

Edit

Configure security settings

| IAM role | Security configuration | Lake Formation configuration |
|----------|------------------------|------------------------------|
| LabRole | - | - |

Step 4: Set output and scheduling

Edit

Set output and scheduling

| Database | Table prefix - optional | Maximum table threshold - optional | Schedule |
|----------|-------------------------|------------------------------------|-----------|
| myddb | - | - | On demand |

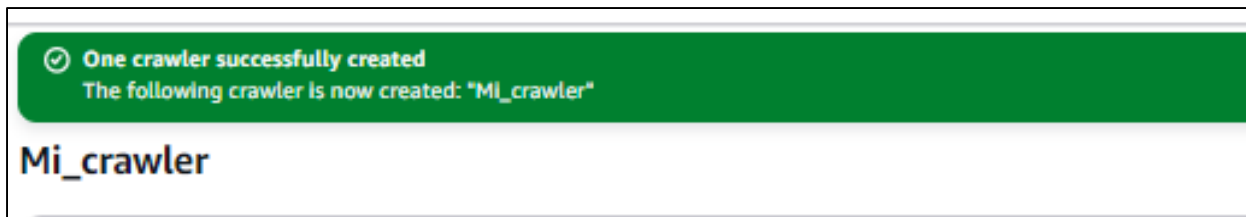
Cancel

Previous

Create crawler

3. LABORATORIO S3 - GLUE - ATHENA

Paso 32. Cuando se ha creado el crawler, lo ejecutamos en la opción "Run crawler" para que haga su función



✓ Crawler successfully starting

The following crawler is now starting: "Mi_crawler"

Run crawler

Edit

Delete

Mi_crawler

Last updated (UTC)
January 23, 2025 at 11:57:22

Crawler properties

| | | | |
|-------------------------|------------------------|------------------------------|--------------|
| Name | IAM role | Database | State |
| Mi_crawler | LabRole | mydb | READY |
| Description | Security configuration | Lake Formation configuration | Table prefix |
| Crawler para metadatos | - | - | - |
| Maximum table threshold | | | |
| - | | | |

► Advanced settings

Crawler runs

Schedule

Data sources

Classifiers

Tags

Crawler runs (1)

The list of crawler runs for this crawler.

Filter data

Filter by a date and time range

< 1 >

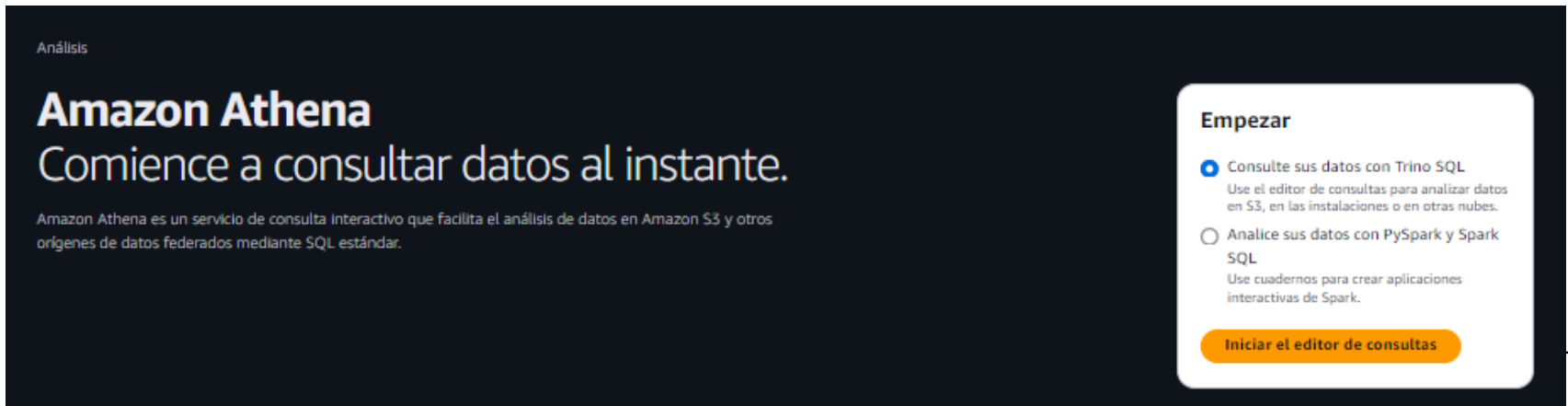
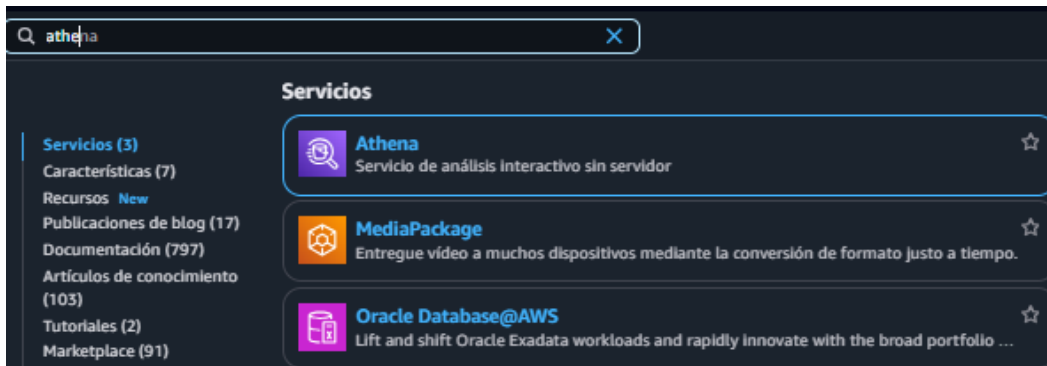
| Start time (UTC) | End time (UTC) | Current/last duration | Status | DPU hours | Table changes |
|------------------------------|------------------------------|-----------------------|-----------|-----------|---------------|
| January 23, 2025 at 11:58:29 | January 23, 2025 at 11:59:43 | 01 min 13 s | Completed | 0.037 | - |

3. LABORATORIO S3 - GLUE - ATHENA

Paso 33. Habilitamos el Crawler para que pueda utilizar nuestra Clave KMS (Key Management Service) que hemos configurado. Accede a AWS Key Management Service (KMS) y busca la clave correspondiente haciendo clic sobre ella. Desplázate hasta la sección de 'Usuarios y Claves' dentro de la configuración de la clave KMS. Agrega el rol asociado a nuestro rastreador a la lista de usuarios autorizados para utilizar esta clave. Esto garantiza que el rastreador tenga los permisos necesarios para interactuar con KMS y usar la clave para descriptar la información almacenada durante el proceso de rastreo en nuestro bucket S3.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 34. Una vez completada la configuración, ya podemos realizar consultas SQL directamente desde Athena. Vamos a la consola de AWS y buscamos el servicio Athena, para comenzar a explorar y analizar nuestros datos



3. LABORATORIO S3 - GLUE - ATHENA

Paso 35. Si ya ejecutamos el crawler, al ingresar en AWS Athena ya podremos ver nuestra base de datos en la sección de Datos. En Origen de los datos se encuentra el AWSDataCatalog y en Base de datos mydb.

The screenshot displays the Amazon Athena console's query editor. The top navigation bar includes 'Amazon Athena' and 'Editor de consultas'. Below this, a tabbed interface shows 'Editor' as the active tab, with other tabs for 'Consultas recientes', 'Consultas guardadas', and 'Configuración'. A 'Grupo de trabajo' dropdown is set to 'primary'. Two informational banners are present: one about configuring S3 results and another about SQL code suggestions. On the left, the 'Datos' sidebar shows 'Origen de datos' as 'AwsDataCatalog', 'Catálogo' as 'None', and 'Base de datos' as 'mydb'. It also includes a 'Tablas y vistas' section with a 'Crear' button and a search bar. The main area, titled 'Consulta 1', contains a large text editor for SQL queries. At the bottom, a status bar shows 'SQL Ln 1, Col 1' and buttons for 'Ejecutar', 'Explicar', 'Cancelar', 'Borrar', and 'Crear'. A toggle at the bottom right allows users to 'Volver a utilizar los resultados de la consulta hasta hace 60 minutos'.

3. LABORATORIO S3 - GLUE - ATHENA

Paso 36. Antes de realizar consultas es necesario ir a la pagina de Administrar la configuración, haciendo click en el botón Editar Ajustes. Indicamos un destino para almacenar los resultados de las consultas que hagamos, que puede ser el mismo bucket creado. Opcionalmente podemos crear una carpeta dentro del bucket, indicándolo después del nombre del bucket

Amazon Athena > Editor de consultas > Administrar la configuración

Administrar la configuración

Ubicación y codificación de los resultados de la consulta

Location of query result - optional
Enter an S3 prefix in the current region where the query result will be saved as an object.

Q s3://bucket-mysql-elara X View Browse S3

You can create and manage lifecycle rules for this bucket
Use Amazon S3 lifecycle rules to store your query results and metadata cost effectively or to delete them after a period of time. [Learn more](#) Lifecycle configuration

Expected bucket owner - optional
Specify the AWS account ID that you expect to be the owner of your query results output location bucket.

Enter AWS account ID

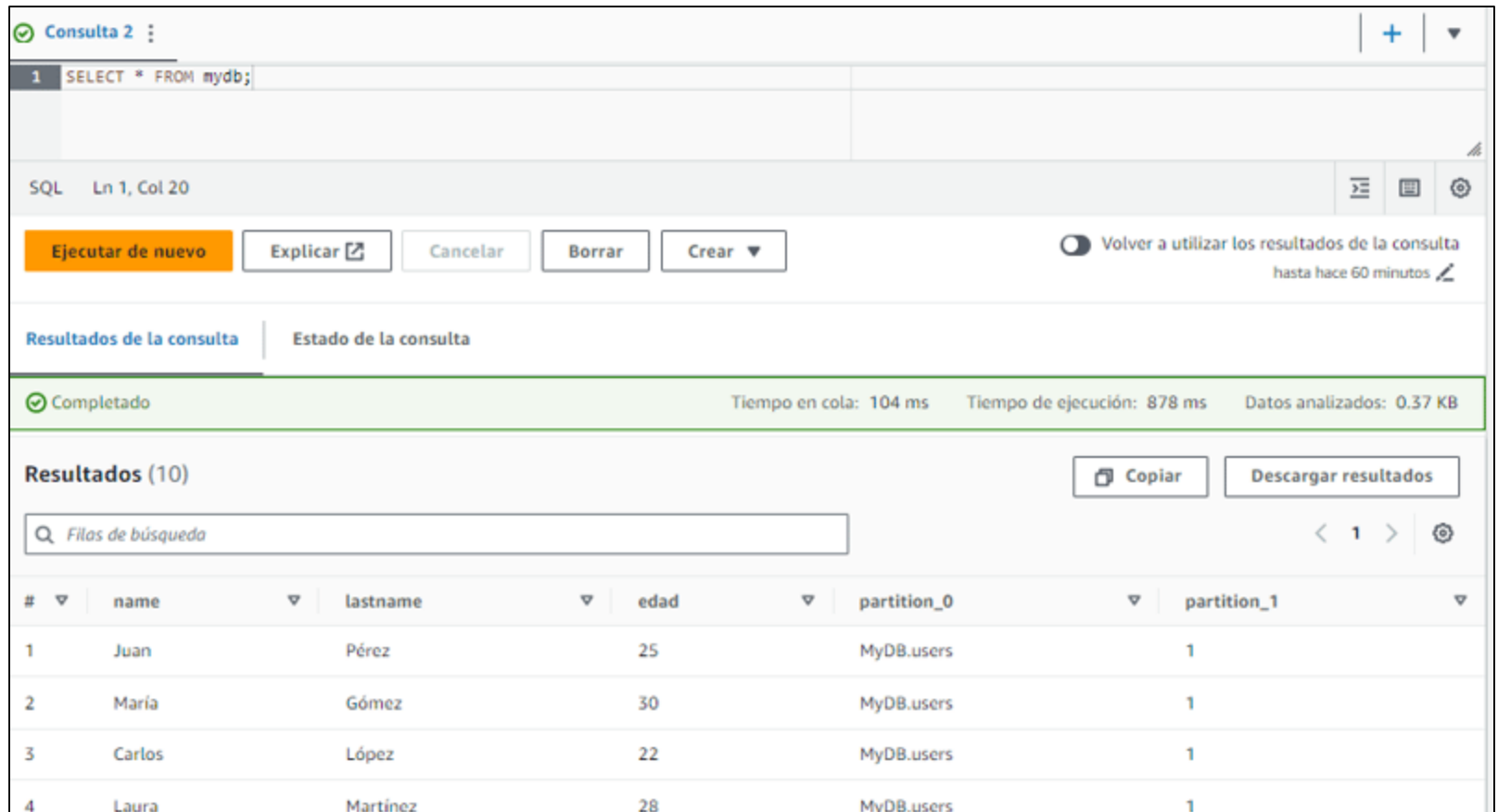
☐ Assign bucket owner full control over query results
Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

☐ Encrypt query results

Cancelar Guardar

3. LABORATORIO S3 - GLUE - ATHENA

Paso 37. Ahora ya podemos realizar consultas sobre nuestra base de datos desde Athena



Consulta 2

```
1 SELECT * FROM mydb;
```

SQL Ln 1, Col 20

Ejecutar de nuevo Explicar Cancelar Borrar Crear

☐ Volver a utilizar los resultados de la consulta hasta hace 60 minutos

Resultados de la consulta Estado de la consulta

Completado Tiempo en cola: 104 ms Tiempo de ejecución: 878 ms Datos analizados: 0.37 KB

Resultados (10) Copiar Descargar resultados

Filas de búsqueda

| # | name | lastname | edad | partition_0 | partition_1 |
|---|--------|----------|------|-------------|-------------|
| 1 | Juan | Pérez | 25 | MyDB.users | 1 |
| 2 | María | Gómez | 30 | MyDB.users | 1 |
| 3 | Carlos | López | 22 | MyDB.users | 1 |
| 4 | Laura | Martínez | 28 | MyDB.users | 1 |