BIG DATA

BASES DE DATOS EN AWS LABORATORIO RDS

EDUARD LARA

INDICE

- 1. Introducción bases de datos en AWS
- 2. Base de datos relacionales en AWS
- 3. Base de datos de pares clave valor
- 4. Base de datos documentales
- 5. Base de datos en memoria
- 6. Base de datos de gráficos
- 7. Base de datos de series temporales
- 8. Laboratorio RDS-MySQL- Workbench
- 9. Laboratorio conexión RDS desde EC2

1. INTRODUCCION BASE DE DATOS AWS

- Las aplicaciones necesitan:
 - bases de datos capaces de almacenar desde terabytes hasta petabytes de nuevos tipos de datos.
 - Proporcionar acceso a los datos con latencia de mseg
 - Procesar millones de solicitudes por segundo
 - Escalar para admitir millones de usuarios en cualquier parte del mundo.
- □ Por tanto se necesitan bases de datos relacionales y no relacionales, diseñadas especialmente para gestionar las necesidades específicas de las aplicaciones.
- Amazon WS ofrece una gran variedad de bases de datos personalizadas para todos los casos de uso específicos de sus aplicaciones.

2. BASE DE DATOS RELACIONALES AWS

- □ Las bases de datos relacionales almacenan datos cuyas relaciones y esquema están predefinidos, diseñadas para admitir transacciones y conservar la integridad referencial, así como la coherencia de datos.
- Se utiliza en aplicaciones tradicionales ERP, CRM y comercio electrónico.
- ☐ Herramientas de bases de datos relacionales en AWS:
 - Amazon Aurora
 - Amazon RDS
 - Amazon Redshift

2. BASE DE DATOS RELACIONALES AWS

- □ Amazon Aurora. Es una base de datos relacional compatible con MySQL y PostgreSQL creada para la nube. Combina el rendimiento y la disponibilidad de las bases de datos tradicionales con la simplicidad y la rentabilidad de las bases de datos de código abierto.
- □ Amazon RDS. Es un servicio para bases de datos relacionales en la nube, sencillo de configurar, completamente administrado y que proporciona los principales motores de búsqueda de bases de datos conocidos como Amazon Aurora, MySQL, PostgreSQL, MariaDB, Oracle y SQL Server.

2. BASE DE DATOS RELACIONALES AWS

- Amazon Redshift. Es un servicio de almacenamiento de datos ágil y escalable que permite analizar todos los datos de su almacenamiento de datos de manera simple y rentable. Ofrece:
 - Un rendimiento 10 veces superior a cualquier otro almacén de datos gracias al uso del aprendizaje automático.
 - La ejecución masiva de consultas en paralelo
 - Almacenamiento en columnas en un disco de alto rendimiento.

3. BASE DE DATOS DE PARES CLAVE VALOR

- □ Las bases de datos de pares clave valor están optimizadas para almacenar y recuperar pares de clave valor en grandes volúmenes de datos en milisegundos, sin la sobrecarga en el rendimiento y las limitaciones de escala propias de las bases de datos relacionales
- □ Se utilizan en aplicaciones a escala de Internet, pujas en tiempo real, carros de la compra y en preferencias de clientes.
- □ El mejor ejemplo es Amazon DinamoDB

3. BASE DE DATOS DE PARES CLAVE VALOR

Amazon DinamoDB

- □ Es un servicio de base de datos no SQL rápido y flexible para cualquier escala
- Admite las aplicaciones a escala más grande del mundo y proporciona tiempos de respuesta de milisegundos.
- □ Replican sus datos en varias regiones AWS para darle acceso rápido y local a los datos
- □ Con DynamoDB no hay servidores que aprovisionar, ya que aumenta o reduce automáticamente las tablas para ajustar la capacidad y mantener el rendimiento.
- □ Casos de uso: Aplicaciones web sin servidor, almacén de datos para microservicios, aplicaciones móviles, tecnología publicitaria, videojuegos e IOT (internet de las cosas)

4. BASE DE DATOS DOCUMENTALES

- Las bases de datos documentales están diseñadas para almacenar datos en forma de documentos.
- □ Se utiliza en la administración de contenido, personalización y aplicaciones móviles.
- Amazon Document DB es un servicio de bases de datos de documentos ágil, escalable y de alta disponibilidad y completamente administrado, que es compatible con cargas de trabajo de MongoDB

5. BASE DE DATOS EN MEMORIA

- Las bases de datos en memoria se utilizan en aplicaciones que requieren acceso en tiempo real a los datos.
- □ Al almacenar los datos directamente en memoria, estas bases de datos proporcionan una latencia de microsegundos que es muy útil cuando la latencia de milisegundos no es suficiente.
- □ Se utiliza en el almacenamiento en caché, marcadores de videojuegos y análisis en tiempo real
- Tenemos dos herramientas
 - Amazon ElastiCache para Redis.
 - Amazon ElastiCache para Memcached

5. BASE DE DATOS EN MEMORIA

- Amazon ElastiCache para Redis, que es un almacén de datos en memoria increíblemente rápido en la nube. Compatible con redes que ofrece una latencia inferior a un milisegundo para aplicaciones en tiempo real.
- □ Amazon ElastiCache para Memcached es un servicio de almacén de clave valor en memoria, compatible con Memcached, que se puede utilizar tanto como caché como como almacén de datos. Es ideal para casos de uso en los que los datos de acceso frecuente deben permanecer en memoria.

6. BASE DE DATOS DE GRAFICOS

- Bases de datos de gráficos Las bases de datos de gráficos se utilizan en aplicaciones que deben permitir a millones de usuarios consultar las relaciones entre conjuntos de datos altamente conectados, así como navegar por ellas con una latencia de milisegundos.
- Se utilizan en la detección de fraudes, redes sociales y motores de recomendaciones
- Amazon Neptune es un servicio de base de datos de gráficos de confianza y completamente administrado que permite crear y ejecutar fácilmente aplicaciones que funcionan con conjuntos de datos altamente conectados.

7. BASE DE DATOS DE SERIES TEMPORALES

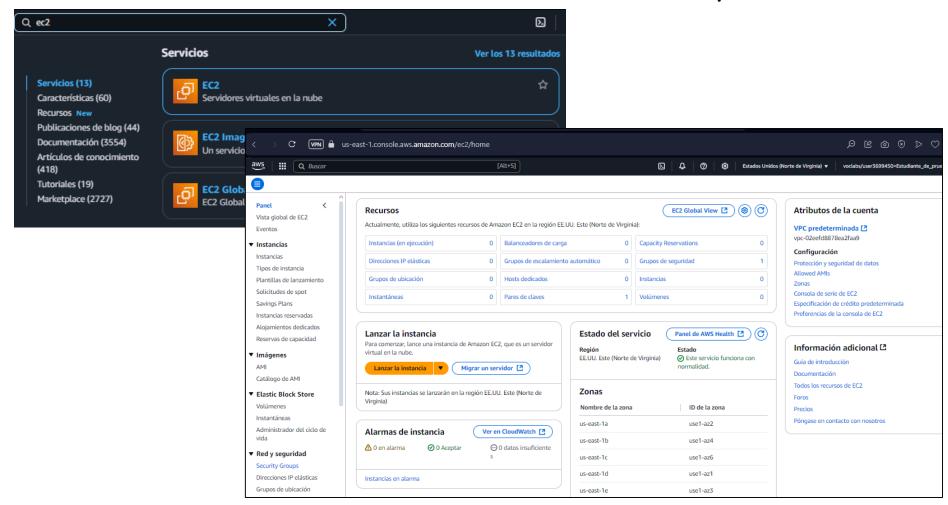
- Las bases de datos de series temporales se utilizan para recopilar, sintetizar y sacar conclusiones de forma eficaz de enormes cantidades de datos que cambian con el paso del tiempo, denominados datos de series temporales.
- □ Se utilizan en aplicaciones de IoT (internet de las cosas), DevOps y telemetría industrial
- □ Amazon TimeStream: Es un servicio de bases de datos de serie temporal, ágil, escalable y completamente administrado para aplicaciones operativas y compatibles con IoT, que facilita el almacenamiento y análisis de billones de eventos diarios

Paso O. Amazon Relational Database Service o Amazon RDS) es un conjunto de servicios administrados que simplifican las tareas de configuración, operación y escalado de bases de datos en la nube de manera flexible Permite elegir entre 7 motores populares, como:

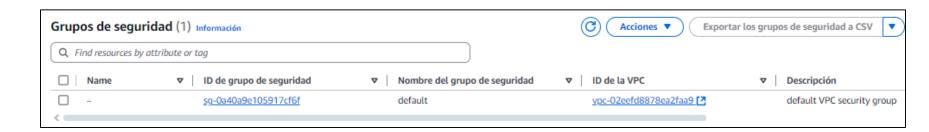
- Amazon Aurora (compatible con MySQL y PostgreSQL),
- MySQL
- MariaDB
- PostgreSQL
- Oracle
- SQL Server

Se puede implementar en instalaciones locales con Amazon RDS en AWS Outposts

Paso 1. Una vez iniciada sesión en la Consola de AWS, buscamos el servicio EC2 en la barra de búsqueda.



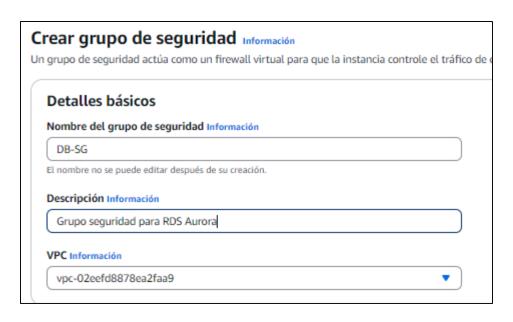
Paso 2. Dentro de la consola de EC2 vamos al menú de la izquierda a la sección de Redes y Seguridad. Seleccionamos la opción Security Groups.



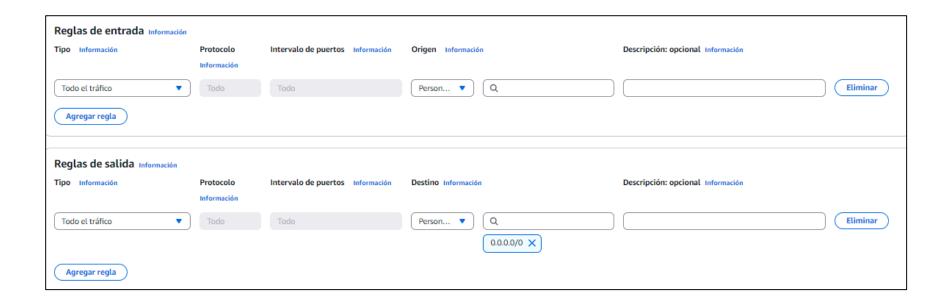
Vamos a crear un grupo de seguridad para nuestra base de datos, el cual definirá las reglas de tráfico entrante y saliente.

Paso 3. Le damos a crear un nuevo grupo de seguridad. Asignamos un nombre al grupo de seguridad (DB-Security Group). Se debe agregar una descripción por ejemplo: Grupo seguridad RDS MySQL.

Usaremos la VPC por defecto.



Paso 4. En la sección de reglas de entrada, en la parte inferior seleccionamos el botón Agregar Regla. Este es un paso crucial al momento de crear un grupo de seguridad.



Paso 5. En Tipo de Trafico seleccionamos MySQL/Aurora y automáticamente se selecciona el puerto 3306 de MySQL. En Origen elegimos cualquier fuente, con la anotación 0.0.0.0/0. Nuestro grupo de seguridad permite trafico de entrada desde cualquier IP.

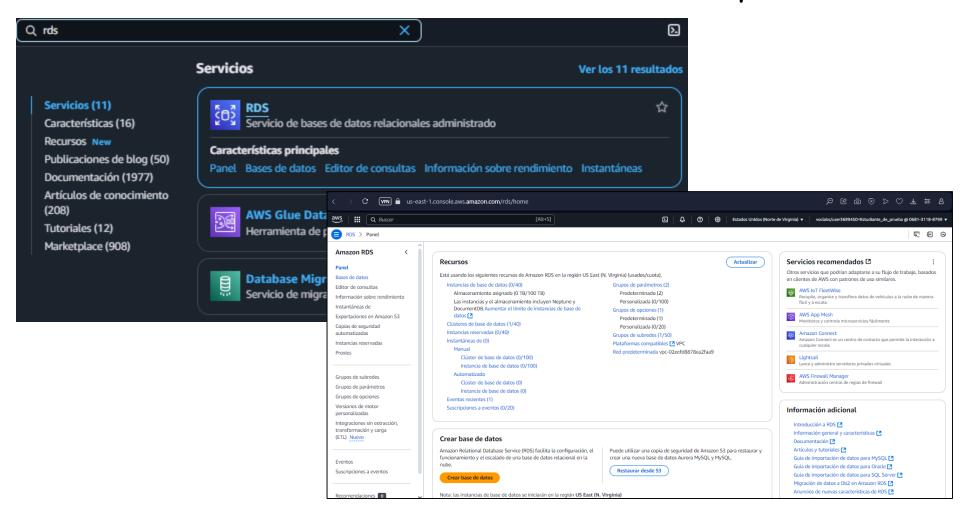
Reglas de entrada Información							
Tipo Información	Protocolo Información	Intervalo de puertos Información	Origen Información	Descripción: opcional Información			
	Informacion						
MYSQL/Aurora ▼	TCP	3306	Anywh • Q. 0.0.0.0/0	Eliminar			
			0.0.0.0/0 🗙				
Agregar regla							
△ Las reglas cuyo origen es 0.0.0.0/0 o ::/0 permiten a todas las direcciones IP acceder a la instancia. Recomendamos configurar reglas de grupo de seguridad para permitir el acceso únicamente desde direcciones IP conocidas.							
Reglas de salida Información							
Tipo Información	Protocolo	Intervalo de puertos Información	Destino Información	Descripción: opcional Información			
	Información						
Todo el tráfico ▼	Todo	Todo	Person ▼ Q	Eliminar			
			0.0.0.0/0 X				
Agregar regla							

Paso 6. Una vez hemos acabado, le damos al boton "Crear grupo de seguridad"



20

Paso 7. Vamos a crear nuestra base de datos. Buscamos el servicio de RDS utilizando la barra de búsqueda



Paso 8. Estando en la consola de RDS, en el menú de la izquierda seleccionamos "Bases de datos" y luego elegimos "Crear base de datos".

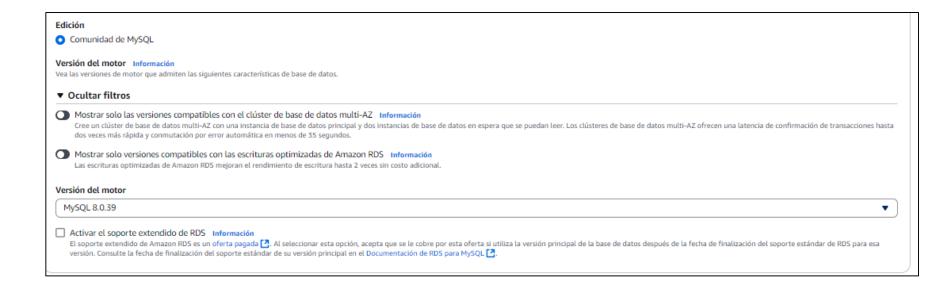


22

Paso 9. Elegimos el modo de creación Estándar, ya que permite personalizar configuraciones como seguridad y alta disponibilidad, etc. El modo sencillo usa configuraciones recomendadas por defecto, algunas modificables después de la creación. Usaremos una base de datos MySQL

rear base de datos Información	
Elegir un método de creación de base de datos	
 Creación estándar Puede definir todas las opciones de configuración, incluidas las de disponibilidad, seguridad, copias de seguridad mantenimiento. 	Creación sencilla Utilice las configuraciones recomendadas. Algunas opciones de configuración se pueden cambiar después de crear la base de datos.
Opciones del motor Tipo de motor Información	
O Aurora (MySQL Compatible)	Aurora (PostgreSQL Compatible)
• MySQL	O PostgreSQL
○ MariaDB	Oracle
Microsoft SOL Server	O IBM Db2

Paso 10. Dejamos la versión del motor seleccionada por defecto, en este caso la versión MySQL 8.0.39.



Paso 11. En la sección de plantillas, elegimos la opción "Capa gratuita" para evitar costes innecesarios.

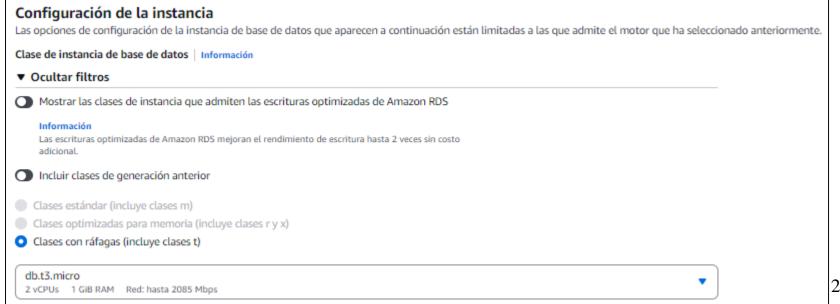


Si hubiéramos elegido "Producción" o "Desarrollo y Pruebas", se activa la sección de "Disponibilidad y Durabilidad" que en nuestro caso no es necesario.

Paso 12. En la sección de configuración, indicamos un identificador único de instancia de la base de datos. Este identificador no es el nombre de la base de datos en sí. Le ponemos instances-dbmysql. Podemos poner admin como nombre de usuario y admin2025 como password pej

Configuración					
Identificador de instancias de bases de datos Información Escriba un nombre para la instancia de base de datos. El nombre debe ser único en relación con todas las instancias de base de datos pertenecientes a su cuenta de AWS en la región de AWS actual.					
instances-dbmysql					
El identificador de la instancia de base de datos no distingue entre mayúsculas y minúsculas, pero se almacena con todas las letras en minúsculas (como en "mydbinstance"). Restricciones: de 1 a 63 caracteres alfanuméricos o guiones. El primer carácter debe ser una letra. No puede contener dos guiones consecutivos. No puede terminar con un guion.					
▼ Configuración de credenciales					
Nombre de usuario maestro Información Escriba un ID de inicio de sesión para el usuario maestro de la instancia de base de datos.					
admin					
1 a 16 caracteres alfanuméricos. El primer carácter debe ser una letra.					
Administración de credenciales Puede usar AWS Secrets Manager o administrar sus credenciales de usuario maestro.					
Administrado en AWS Secrets Manager - más seguro RDS genera una contraseña y la administra durante todo su ciclo de vida mediante AWS Secrets Manager. O Autoadministrado Cree su propia contraseña o pida a RDS que cree una contraseña para que pueda administrarla.					
Generar contraseña automáticamente Amazon RDS puede generar una contraseña en su nombre, o bien puede especificar su propia contraseña.					
Contraseña maestra Información					
Password strength Weak					
Restricciones mínimas: al menos 8 caracteres ASCII imprimibles. No puede contener ninguno de los siguientes símbolos: / ' " @					
Confirmar la contraseña maestra Información					

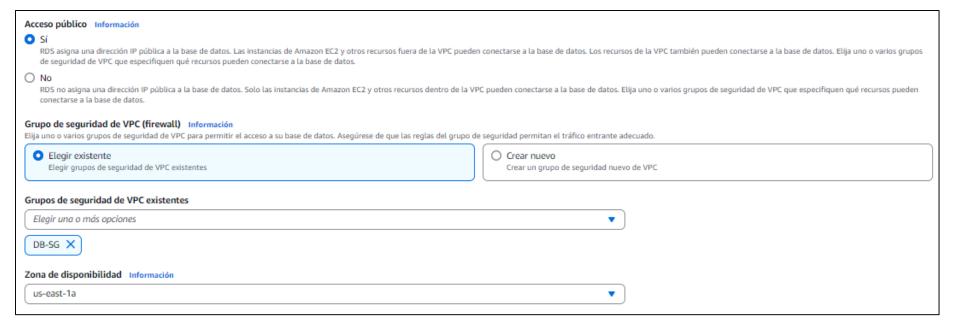
Paso 13. En la sección de "Configuración de la instancia", seleccionaremos "Clases con ráfagas" y en concreto la instancia de tipo t3.micro, que están dentro de la capa gratuita. Recordar que del tipo de instancia depende la capacidad de procesamiento, rendimiento y transaccionalidad de la base de datos.



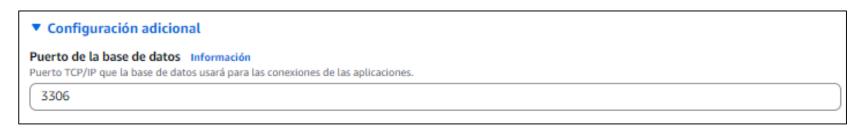
Paso 14. En la sección Almacenamiento, dejamos la configuración por defecto. Esto implica un tipo de almacenamiento SSD de uso general (Gp2) con un valor mínimo de almacenamiento de 20GB. En la sección de 'Escalado automático', deshabilitamos esta opción, no es necesario un escalado automático de la base de datos



Paso 15. En la sección de conectividad, habilitamos el acceso publico y elegimos el grupo de seguridad que creamos al inicio DB-SG. En la zona de disponibilidad, podemos usar us-east-1a para permitir que otros recursos en la misma zona de disponibilidad puedan interactuar con nuestra base de datos.



Paso 16. En configuracion adicional dejaremos el puerto 3306 (Puerto TCP/IP que la base de datos usará para las conexiones de las aplicaciones.)



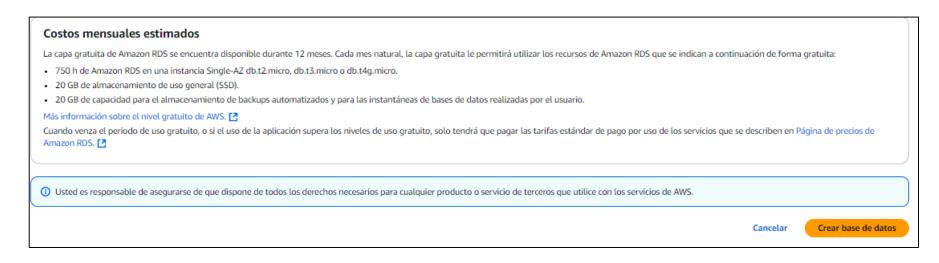
Paso 17. En la sección de autenticación de bases de datos, seleccionaremos Autentificacion con contraseña. Ya asignamos previamente un usuario y contraseña.



Paso 18. Vamos a la sección de configuración adicional. Importante: Debemos establecer un nombre inicial para nuestra base de datos. Si no, solo se crearía la instancia de base de datos (no la base de datos) y tendríamos que acceder a ella posteriormente para crearla. Deshabilitamos las copias de seguridad automatizadas

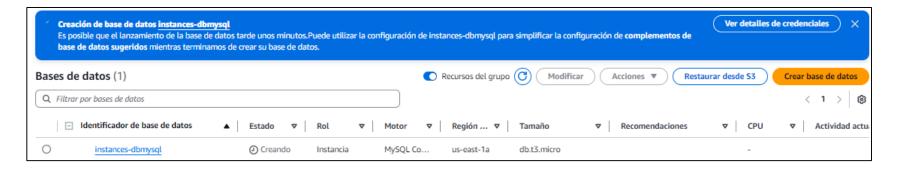
▼ Configuración adicional Opciones de base de datos, cifrado activado, copia de seguridad desactivado, retroceder desactivado, mantenimiento, Registros de CloudWatch, eliminar protección desactivado.
Opciones de base de datos
Nombre de base de datos inicial Información
MyDB
Si no especifica un nombre de base de datos, Amazon RDS no crea una base de datos.
Grupo de parámetros de base de datos Información
default.mysql8.0 ▼
Grupo de opciones Información
default:mysql-8-0 ▼
Copia de seguridad
Habilitar las copias de seguridad automatizadas. Crea una instantánea de un momento dado de su base de datos

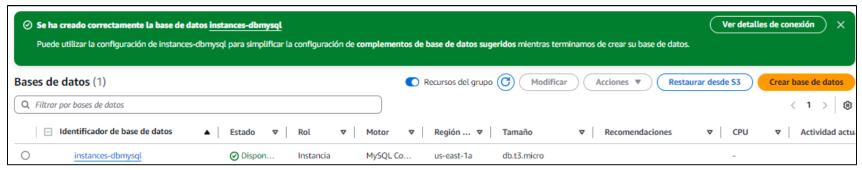
Paso 19. El resto de puntos los dejamos con la información por defecto. Al final, veremos un resumen del coste de la base de datos.



Nota: Una vez realizada la practica se recomienda borrar la base de datos para que ese coste no se genere

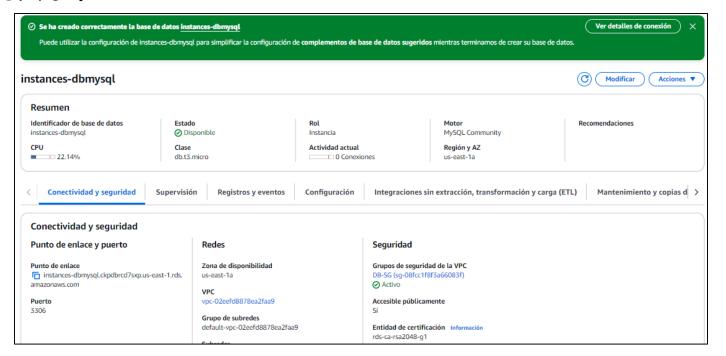
Paso 20. Le damos al botón 'Crear base de datos'. Esto tomara algunos minutos. En breve tenemos nuestra base de datos lista para cargarle datos y lanzar consultas.



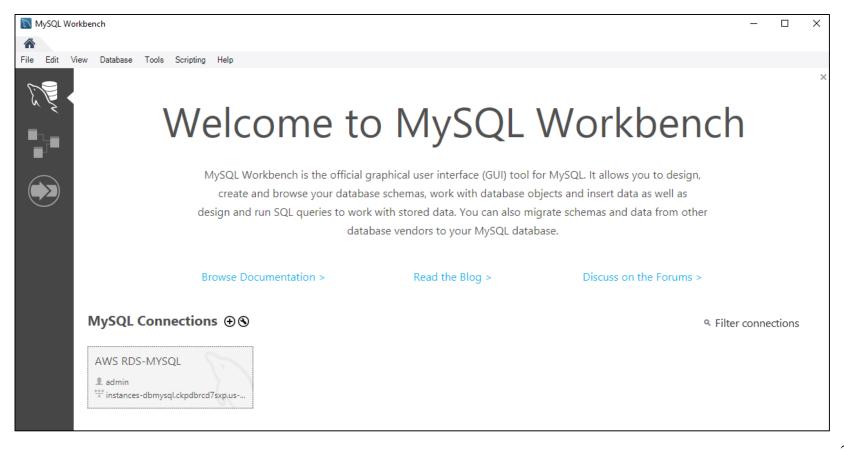


_

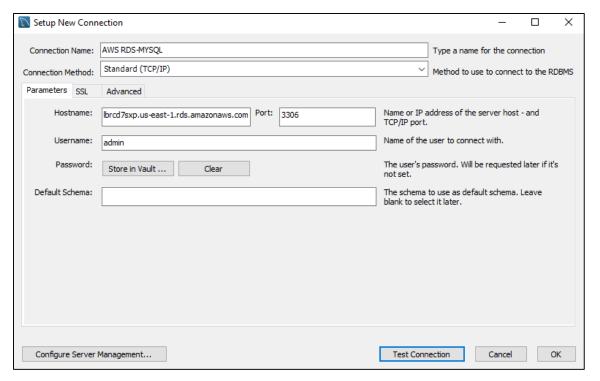
Paso 21. Hacemos clic en nuestra base de datos y vamos a la sección 'Conectividad y seguridad'. Aquí encontramos un resumen e información que necesitaremos para establecer la conexión con los programas Mysql Workbench o SqlElectron, como el 'Punto de enlace' y el 'Puerto'.



Paso 22. Abrimos el cliente MySql Workbench. Agregamos una nueva conexión:

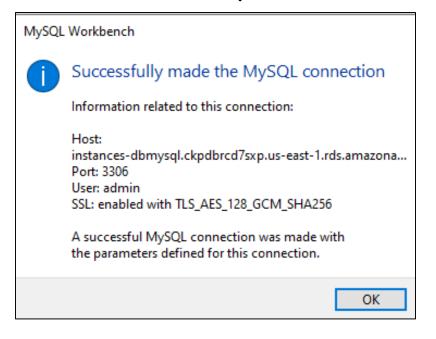


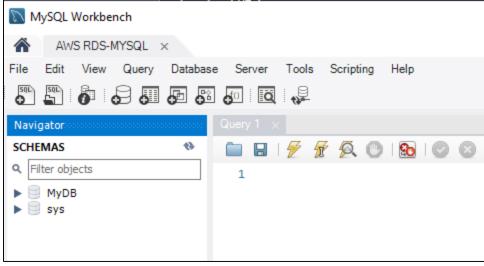
Paso 23. En el campo 'Conection Name', asignamos un nombre para la conexión. En Hostname colocamos el punto de enlace copiado de nuestra base de datos en AWS. Indicamos el usuario (admin) y la contraseña (admin2025) que establecimos cuando creamos la base de datos



8. LABORATORIO RDS-MYSQL-WORKBENCH

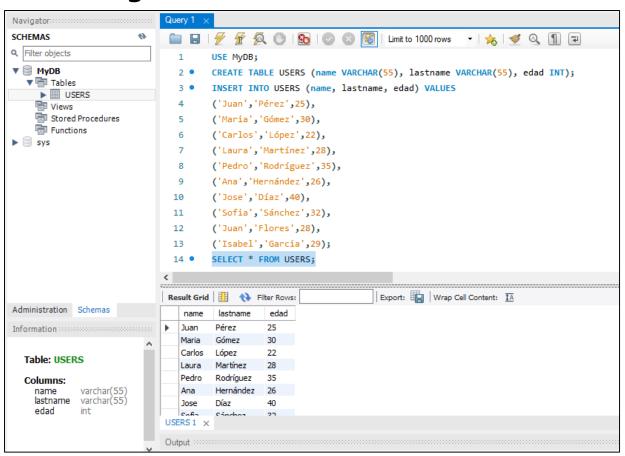
Paso 24. Podemos hacer click en el botón 'Test Connection' para asegurarnos de que hemos introducido bien todos los parámetros y la conexión ha sido exitosa. En este caso podemos entrar y visualizar nuestra database MyDB



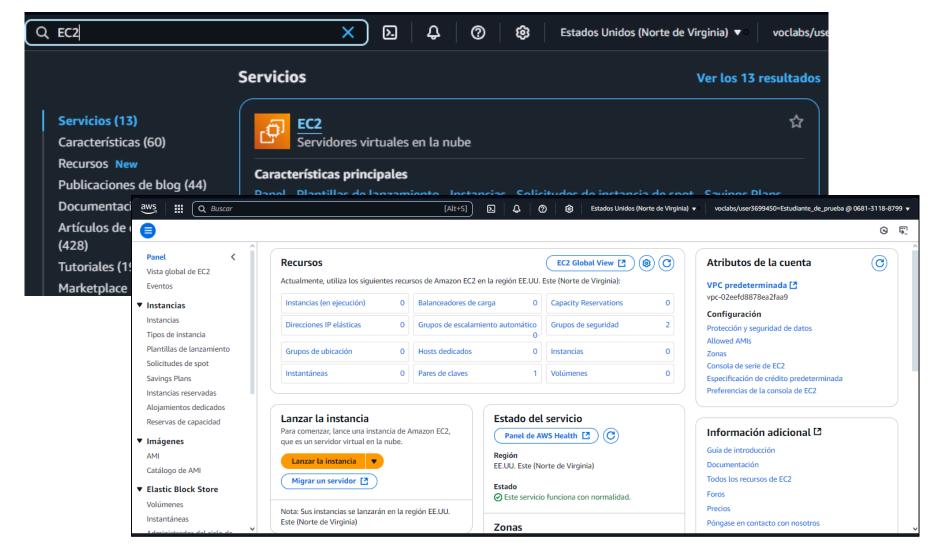


8. LABORATORIO RDS-MYSQL-WORKBENCH

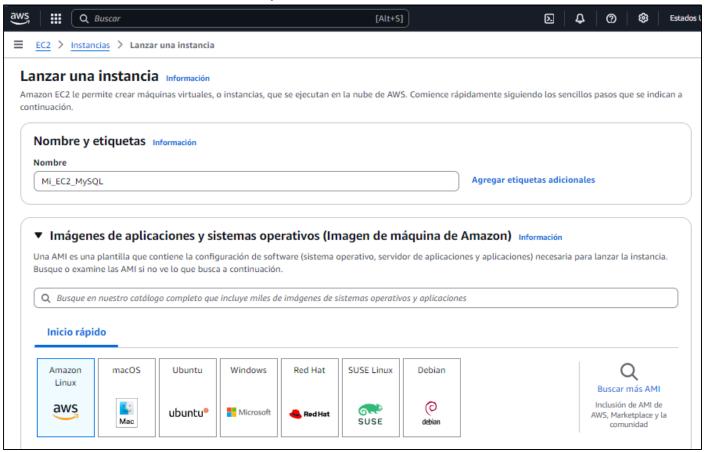
Paso 25. Podemos ejecutar el siguiente script, que accede a la base de datos MyDB, crea una tabla user, e introduce 13 registros en la tabla user.



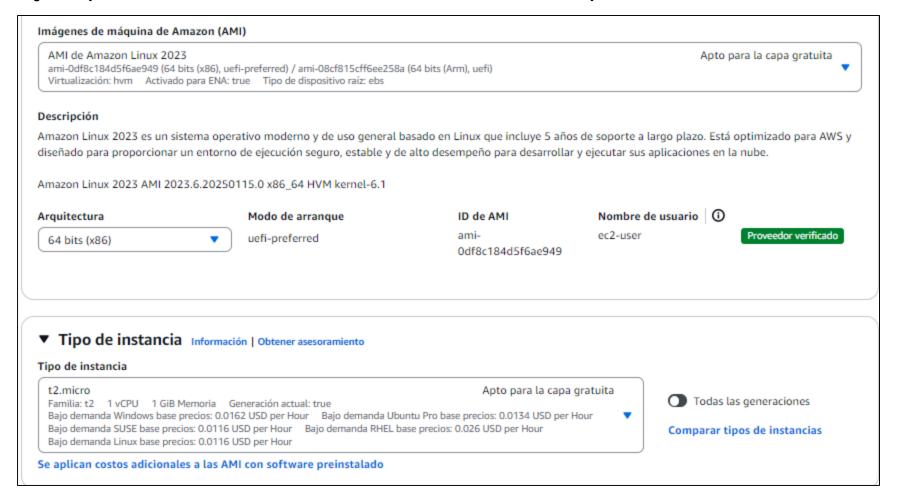
Paso 1. Abrimos la consola de Amazon EC2



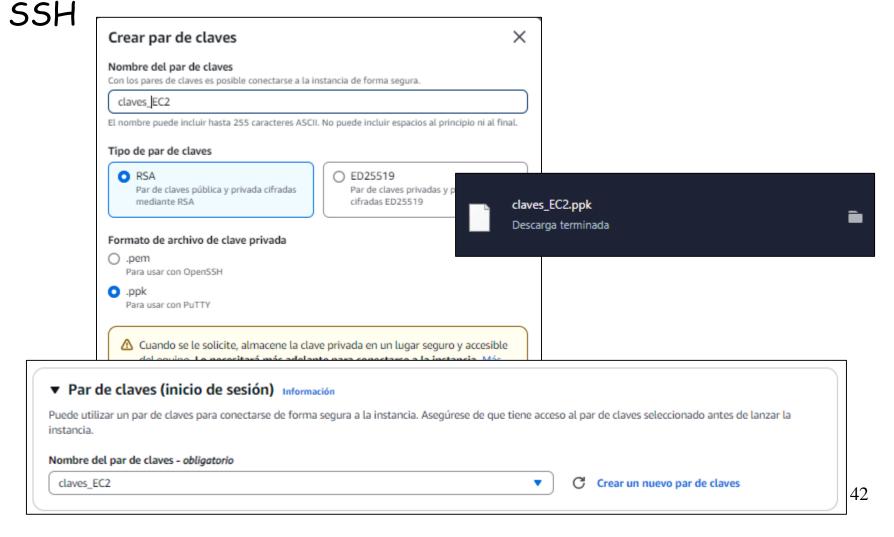
Paso 2. Le damos a Lanzar instancia. Le ponemos un nombre (pej: Mi_EC2_MySQL) y seleccionamos Amazon Linux como Sistema Operativo.



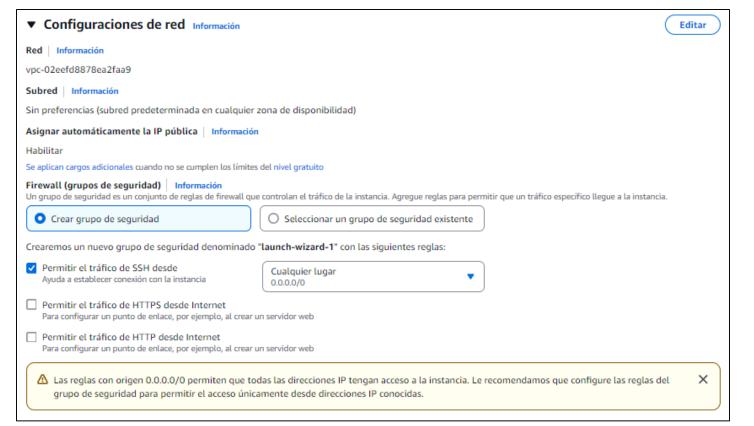
Paso 3. Elegimos una instancia de la capa gratuita por ejemplo AMI Amazon Linux 2023 de tipo t2.micro



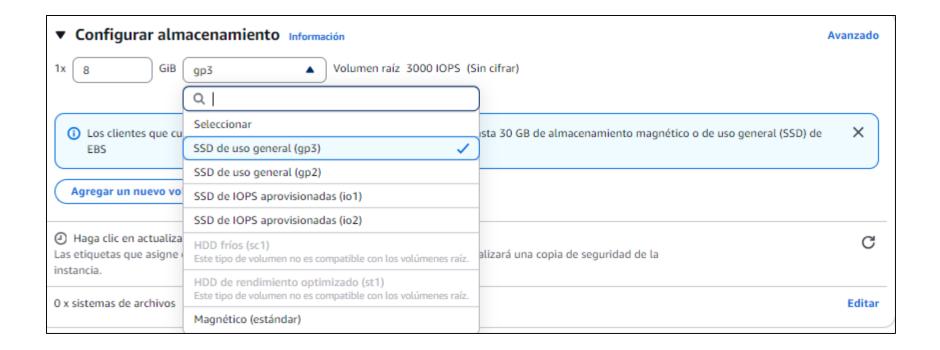
Paso 4. Creamos el par de claves RSA para el acceso con



Paso 5. En configuración de red dejamos todo por defecto: VPC es la misma que utiliza la instancia de bbdd de RDS, habilitada la asignación automáticamente la IP publica, Nuevo grupo de seguridad que permita SSH, etc



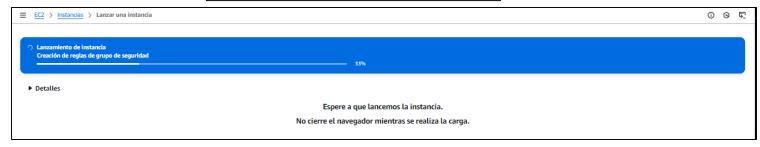
Paso 6. En configurar almacenamiento dejamos las opciones por defecto 8 GB en SSD de uso general (gp3)



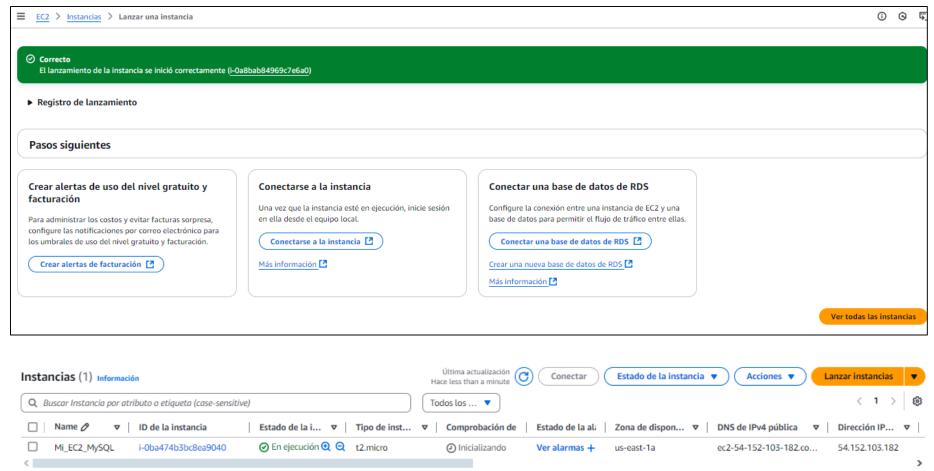
44

Paso 7. Dejamos las opciones de Detalles avanzados tal y como esta y le damos a Lanzar Instancia

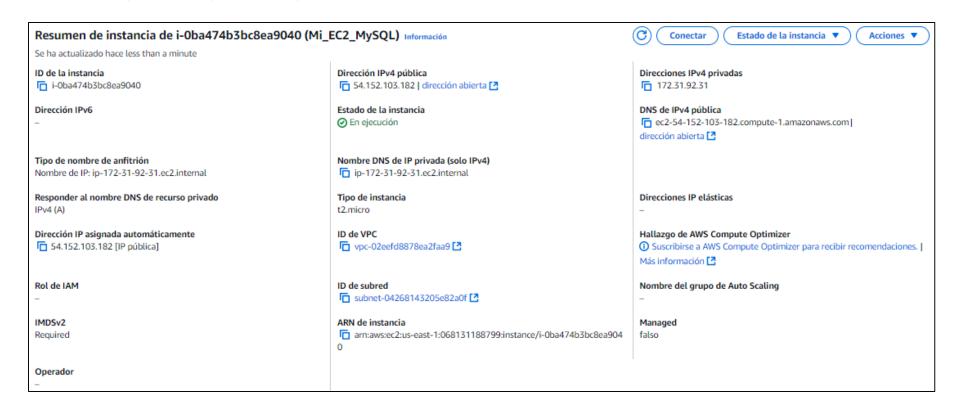
▼ Resumen		
Número de instancias Información		
1		
Imagen de software (AMI)		^
Amazon Linux 2023 AMI 2023.6.2má ami-0df8c184d5f6ae949	s información	
Tipo de servidor virtual (tipo de instatz.micro	nncia)	
Firewall (grupo de seguridad) Nuevo grupo de seguridad		
Almacenamiento (volúmenes) Volúmenes: 1 (8 GiB)		
Nivel gratuito: El primer año incinstancias t2.micro (o t3.micro e t2.micro no esté disponible) en lal mes, 750 horas de uso de dire mes, 30 millones de E/S, 2 millo	n las regiones en las que as AMI del nivel gratuito cciones IPv4 públicas al	Y
Cancelar	Lanzar instancia	
1	Código de versión preliminar	



Paso 8. Vemos que hemos creado una instancia y esta activa

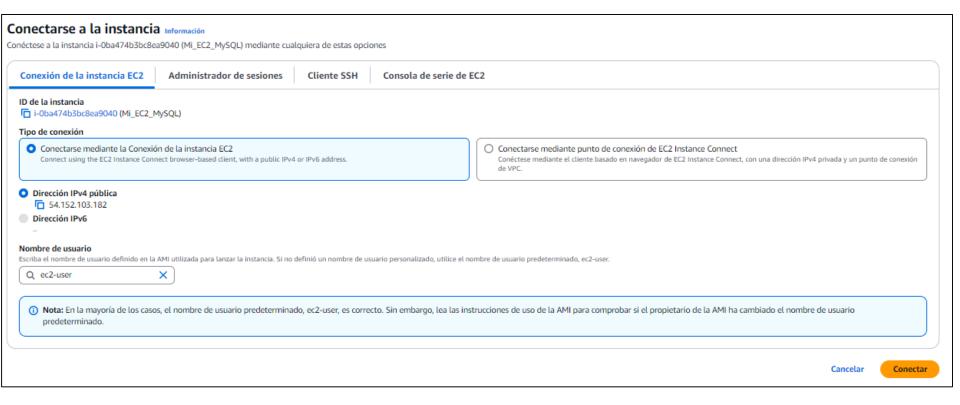


Paso 9. Si entramos en la instancia, vemos un resumen de sus principales parámetros



47

Paso 10. Si hacemos click en conectar podemos ver los parámetros de conexión que usaremos desde el cliente ssh



Paso 11. Descargamos el clicente ssh putty

Download PuTTY: latest release (0.82)

Home | FAQ | Feedback | Licence | Updates | Mirrors | Keys | Links | Team Download: Stable · Pre-release · Snapshot | Docs | Privacy | Changes | Wishlist

This page contains download links for the latest released version of PuTTY. Currently this is 0.82, released on 2024-11-27.

When new releases come out, this page will update to contain the latest, so this is a good page to bookmark or link to. Alternatively, here is a permanent link to the 0.82 release.

Release versions of PuTTY are versions we think are reasonably likely to work well. However, they are often not the most up-to-date version of the code available. If you have a problem with this release, then it might be worth trying out the <u>pre-release builds of 0.83</u>, or the <u>development snapshots</u>, to see if the problem has already been fixed in those versions.

Package files

You probably want one of these. They include versions of all the PuTTY utilities (except the new and slightly experimental Windows pterm).

(Not sure whether you want the 32-bit or the 64-bit version? Read the FAQ entry.)

We also publish the latest PuTTY installers for all Windows architectures as a free-of-charge download at the Microsoft Store; they usually take a few days to appear there after we release them.

MSI ('Windows Installer')

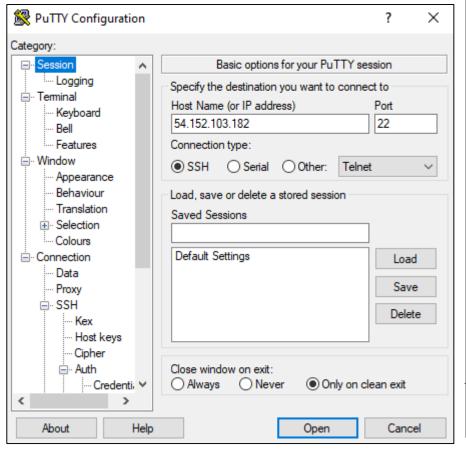
 64-bit x86:
 putty-64bit-0.82-installer.msi
 (signature)

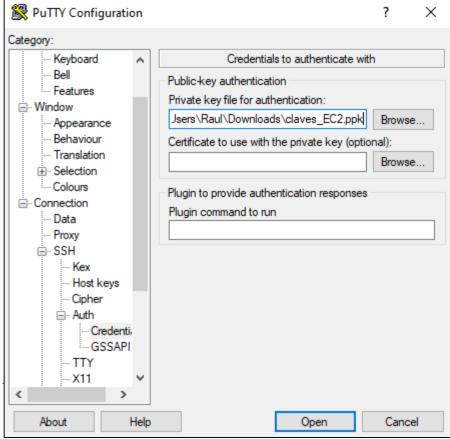
 64-bit Arm:
 putty-arm64-0.82-installer.msi
 (signature)

 32-bit x86:
 putty-0.82-installer.msi
 (signature)

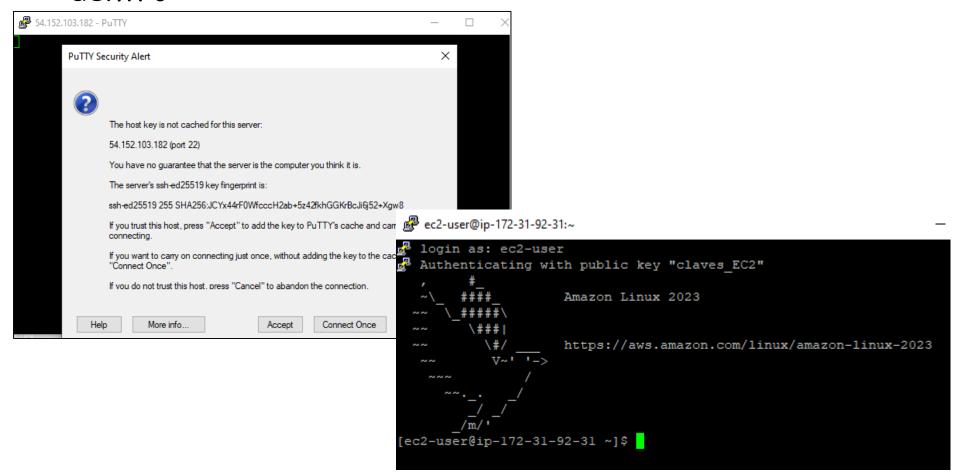


Paso 12. Configuramos en putty la dirección IP publica de la maquina virtual EC2. También indicamos el fichero de claves para las credenciales





Paso 13. Le damos al botón Open para realizar la conexión. Aceptamos para confiar en el host y ya estamos dentro



Paso 14. Debemos instalar el cliente mysql en una maquina Amazon Linux 2023. Primero actualizamos el sistema

sudo dnf update -y

```
[ec2-user@ip-172-31-92-31 ~]$ sudo dnf update -y
Last metadata expiration check: 0:25:38 ago on Wed Jan 22 19:03:39 2025.
Dependencies resolved.
Nothing to do.
Complete!
[ec2-user@ip-172-31-92-31 ~]$
```

Paso 15. Habilitamos el repositorio MySQL. Amazon Linux 2023 no incluye MySQL en sus repositorios predeterminados, por lo que es necesario habilitar el repositorio oficial de MySQL. Descargamos el repositorio MySQL para EL9 (compatible con Amazon Linux 2023): sudo dnf install -y https://dev.mysql.com/get/mysql80community-release-el9-1.noarch.rpm

Paso 16. Verificamos que el repositorio se haya agregado correctamente

sudo dnf repolist enabled | grep mysql

Paso 17. Importamos manualmente la clave GPG de MySQL que valida los paquetes del repositorio:

sudo rpm --import https://repo.mysql.com/RPM-GPG-KEY-mysql-2022

```
[ec2-user@ip-172-31-92-31 ~]$ sudo rpm --import https://repo.mysq1.com/RPM-GPG-KEY
-mysq1-2022
[ec2-user@ip-172-31-92-31 ~]$
```

Paso 18. Instalamos sólo el cliente MySQL: sudo dnf install -y mysql-community-client

Nota: Si hay problemas relacionados con la clave GPG, se puede realizar la instalación del cliente mysql desactivando temporalmente la verificación: sudo dnf install -y mysql-community-client --nogpgcheck

sade and morally mysqr community chem magpgeneer

Paso 19. Una vez instalado, verificamos la versión del cliente MySQL:

```
mysql --version
```

```
[ec2-user@ip-172-31-92-31 ~]$ mysql --version
mysql Ver 8.0.41 for Linux on x86_64 (MySQL Community Server - GPL)
[ec2-user@ip-172-31-92-31 ~]$
```

Paso 20. Nos conectamos la base de datos MySQL utilizando el cliente recién instalado:

mysql -h <host-del-servidor> -P 3306 -u <usuario> -p

```
@ ec2-user@ip-172-31-92-31:~
                                                                                      ×
[ec2-user@ip-172-31-92-31 ~]$ mysql -h instances-dbmysql.ckpdbrcd7sxp.us-east-1.rds.a
mazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 49
Server version: 8.0.39 Source distribution
Copyright (c) 2000, 2025, Oracle and/or its affiliates.
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
mysql> show databases;
 Database
  information schema
 mysql
  performance schema
 rows in set (0.00 sec)
```

Paso 21. Ejecutamos los comando habituales y visualizamos la tabla users

```
mysql> use MyDB;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -\mathtt{A}
Database changed
mysql> show tables;
 Tables in MyDB
 row in set (0.00 sec)
nysql> SELECT * FROM USERS;
         lastname
                      edad
  Juan | Pérez
                         25
  Maria | Gómez
                         30
  Carlos | López
                         22
  Laura | Martinez
                         28
  Pedro | Rodríguez |
                         35
         | Hernández
                         26
  Ana
  Jose
         | Díaz
                         40
  Sofia | Sánchez
                         32
  Juan
         | Flores
                         28
  Isabel | Garcia
                         29
10 rows in set (0.00 sec)
mysql>
```