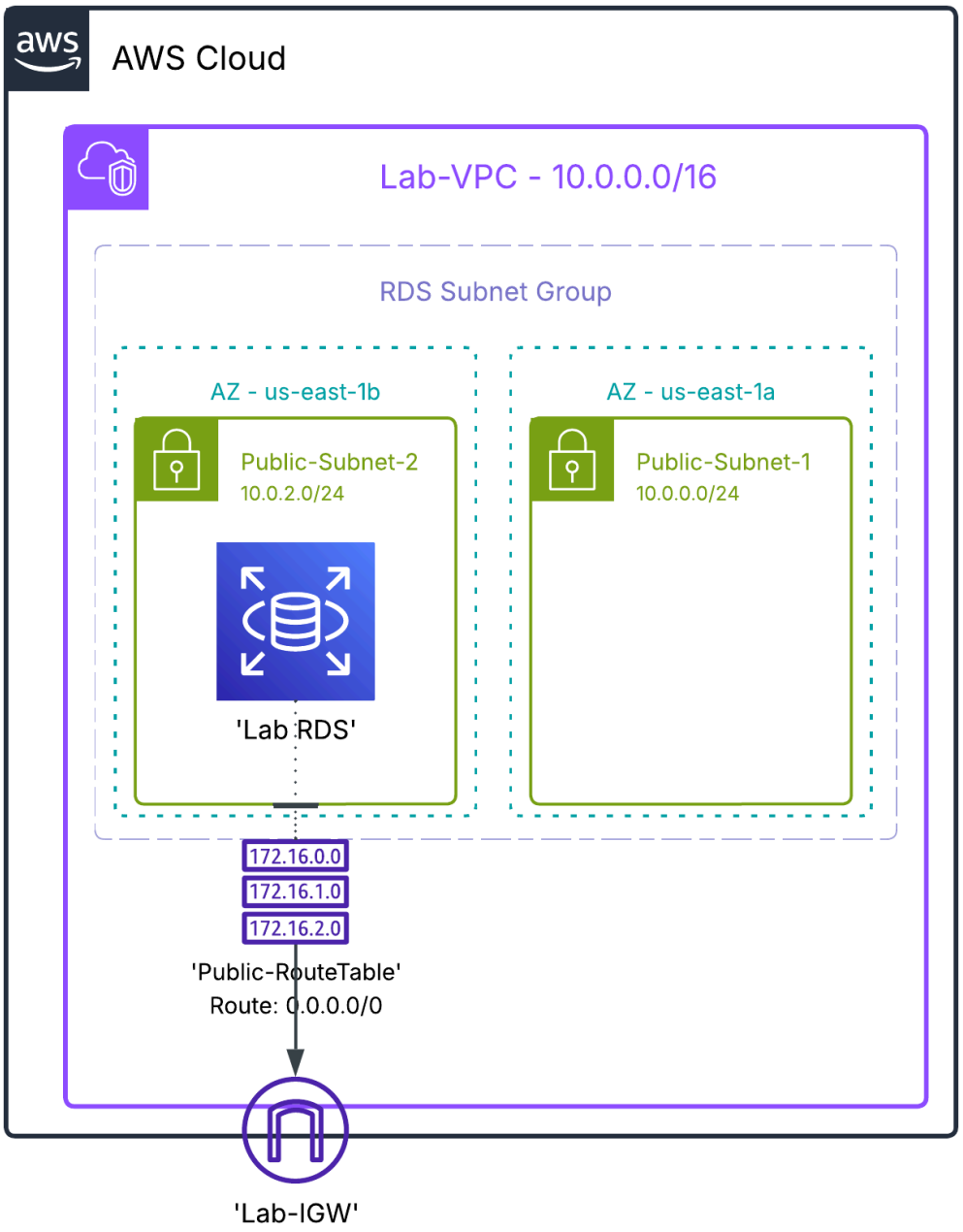
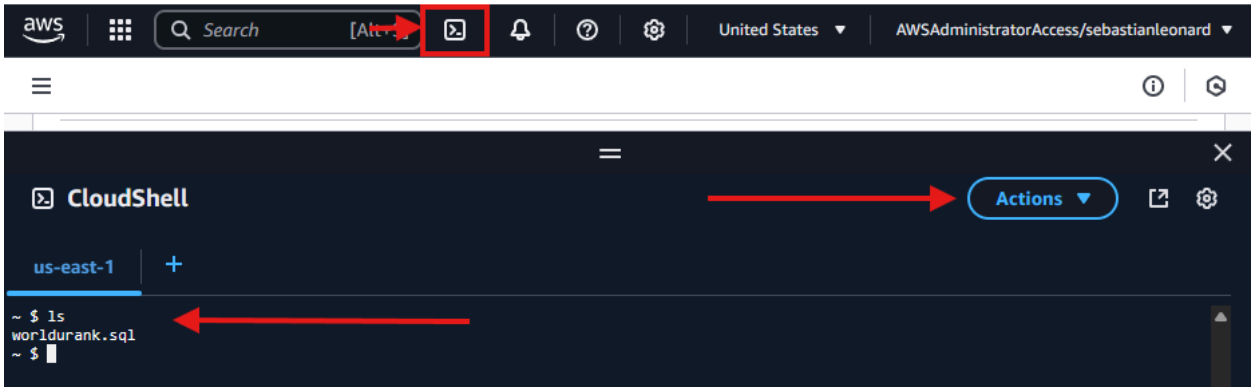
Guía Laboratorio de Migración de Datos AWS DMS

Nota Inicial: La VPC con RDS se levanta automáticamente desde el template de CloudFormation. Se proporcionará a los alumnos un enlace a un S3 donde se encuentra el archivo worldurank.sql. Este script SQL deberá ser descargado y luego subido al entorno de CloudShell en la consola de AWS.



1. Instalación de sqlcmd en CloudShell



(recuerda subir el archivo .sql al CloudShell y verificas con $ ls que tenemos el archivo dentro)

En la consola de AWS, arriba a la derecha habrá un logo de código, damos click y se abre el CloudShell. En el desplegable 'Actions' elegimos la opción "Upload File" y elegimos el archivo worldurank.sql que descargamos previamente.

Estando en el ambiente, será necesario instalar sqlcmd. Sigue estos pasos:

1.1. Añadir el repositorio de Microsoft

sudo curl https://packages.microsoft.com/config/rhel/7/prod.repo | sudo tee /etc/yum.repos.d/msprod.repo

1.2. Instalar las herramientas y el driver ODBC

sudo yum install -y mssql-tools msodbcsql17

si te pide aceptar license terms, escribe 'yes' en el input.

1.3. Hacer que el comando sqlcmd sea permanente para futuras sesiones

echo 'export PATH="$PATH:/opt/mssql-tools/bin"' >> ~/.bashrc

1.4. Aplicar la configuración a la sesión actual

source ~/.bashrc

1.5. Verificar la instalación (opcional)

sqlcmd -?

Output esperado:
~ $ sqlcmd -?
Microsoft (R) SQL Server Command Line Tool Version 17.10.0001.1 Linux Copyright (C) 2017 Microsoft...

2. Creación de la Base de Datos en RDS

Una vez instalado sqlcmd, ejecuta el siguiente comando para crear una base de datos dentro de la instancia RDS:

sqlcmd -S RDS_ENDPOINT -U admin -P "#LabDBase3!" -Q "CREATE DATABASE UniversityDB;"

IMPORTANTE: Reemplaza RDS_ENDPOINT por el endpoint de la RDS del laboratorio.

3. Ejecución del Script SQL

Usa el siguiente comando para ejecutar el script worldurank.sql dentro de la base de datos recién creada:

sqlcmd -S RDS_ENDPOINT -U admin -P "#LabDBase3!" -d UniversityDB -i worldurank.sql

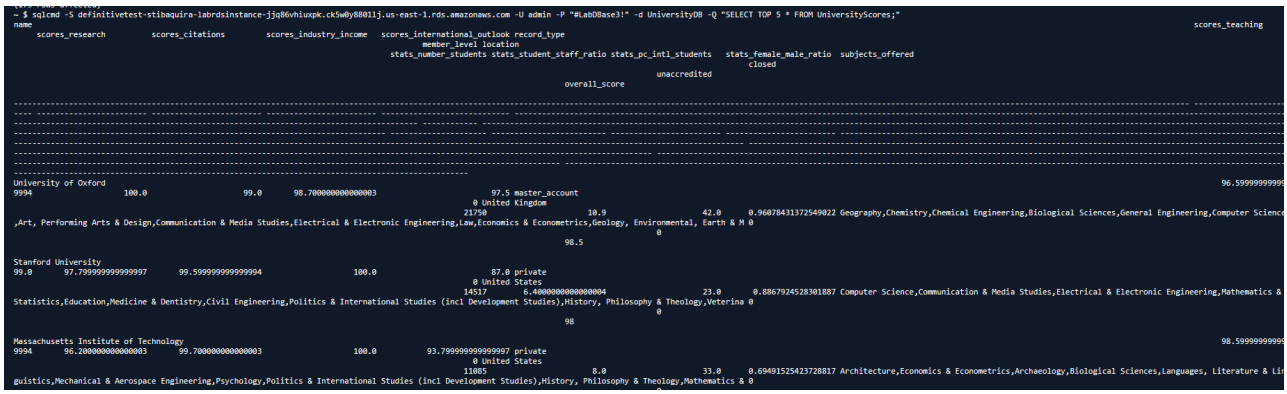
Si el output es algo como "(500 rows affected)" y mensajes similares, significa que funcionó correctamente.

4. Verificación de la Importación de Datos

Para comprobar que los datos se importaron correctamente, consulta las primeras 5 filas de la tabla:

sqlcmd -S RDS_ENDPOINT -U admin -P "#LabDBase3!" -d UniversityDB -Q "SELECT TOP 5 * FROM UniversityScores;"

El output debería mostrar todas las columnas del archivo worldurank.sql. (no importa si no se visualiza como tabla)



5. Preparación para la Instancia DMS

Una vez verificado que la RDS tiene la base de datos con los datos insertados, proseguimos con el laboratorio. (Ya puedes cerrar el CloudShell)

6. Creación de Recursos Previos a DMS

Antes de iniciar con la creación de la instancia DMS, es necesario crear:

1. Un **Security Group** para permitir las conexiones a los diferentes puntos de migración.
2. Un **Subnet Group** de replicación para configurar la instancia DMS.

6.1. Creación del Security Group

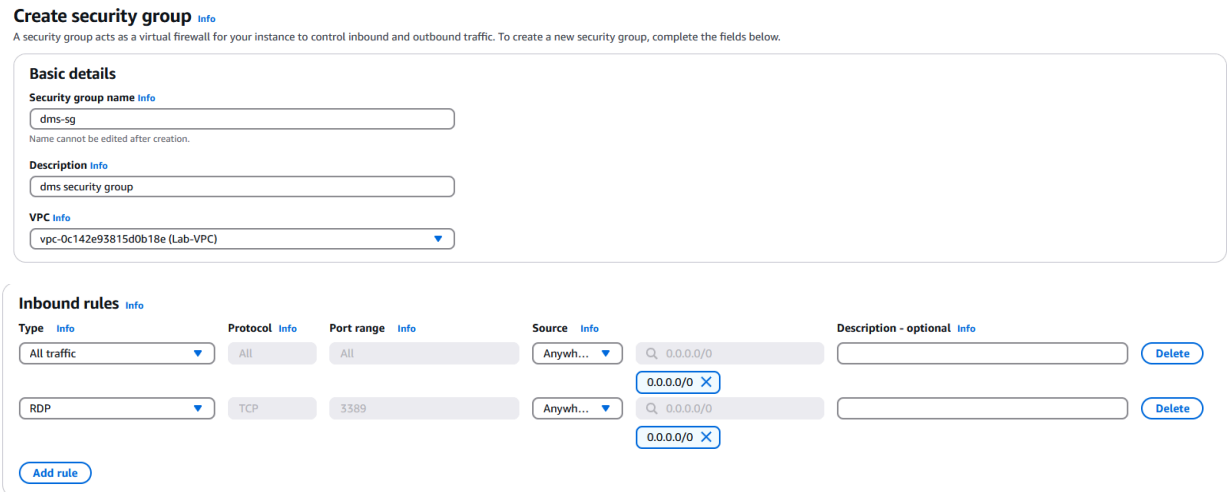
En la AWS Management Console:

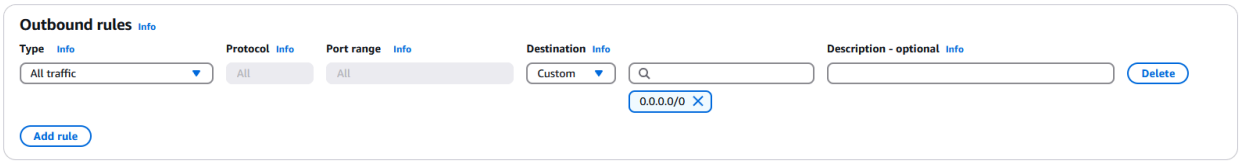
1. Ve a **EC2**.
2. En el panel de navegación, selecciona **Network & Security > Security Groups**.
3. Haz clic en **Create security group**.

Configuración del Security Group:

- **VPC:** Selecciona la Lab-VPC.
- **Inbound rules:**
  - Regla 1 (por defecto):
    - **Type:** All traffic
    - **Source:** 0.0.0.0/0
  - Regla 2:
    - **Type:** RDP (Protocolo TCP, Puerto 3389 se autocompletarán)
    - **Source:** 0.0.0.0/0
- **Outbound rules:**
  - Regla 1:
    - **Type:** All traffic
    - Source: 0.0.0.0/0

      Finalmente, crea el Grupo de Seguridad.



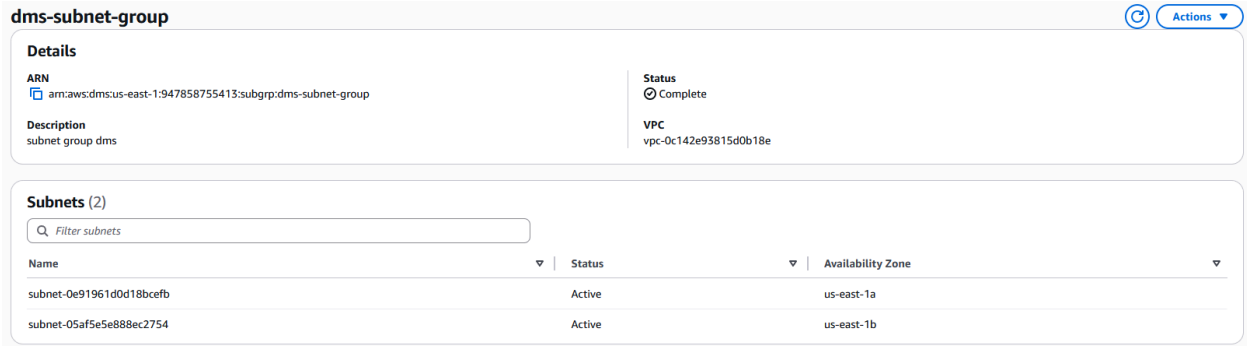


7. Configuración en DMS (Database Migration Service)

Busca **DMS** dentro de la AWS Management Console.

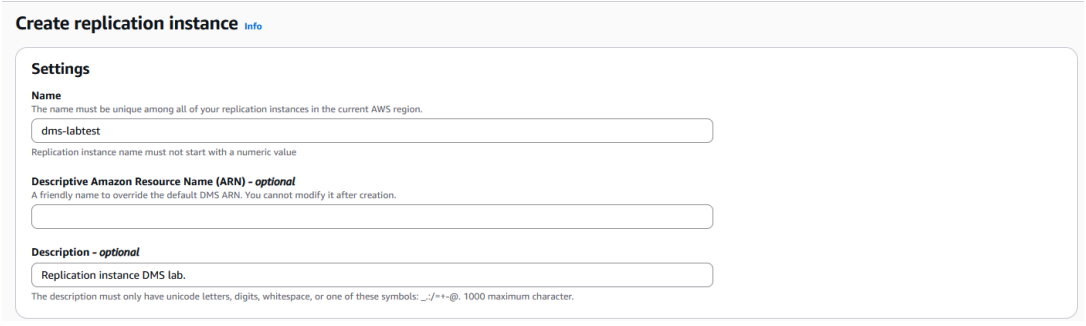
7.1. Creación del Subnet Group en DMS

1. En la página de DMS, ve a **Subnet groups**.
2. Haz clic en **Create subnet group**.
3. **Name:** dms-subnetgroup
4. **VPC:** Elige la Lab-VPC.
5. **Add Subnets:** Selecciona las 2 subnets de la Lab-VPC.
6. Agrega etiquetas (altamente recomendable) y haz clic en **Create**.

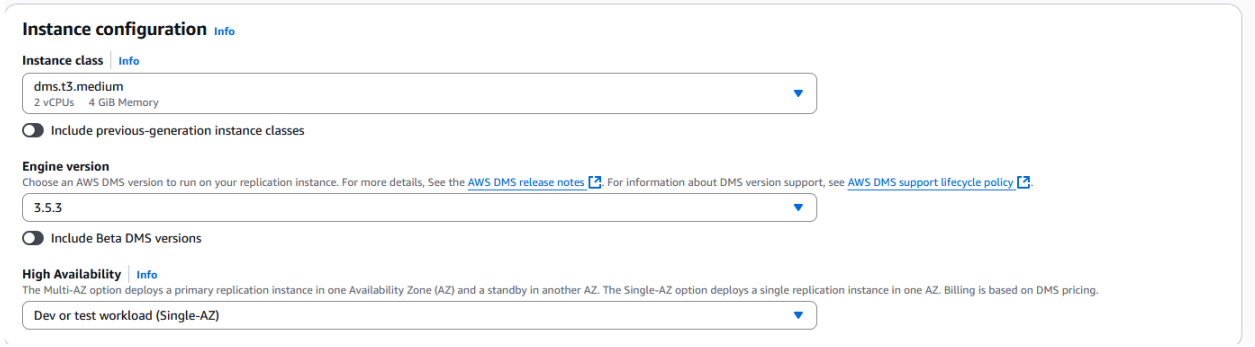


7.2. Creación de la Instancia de Replicación DMS

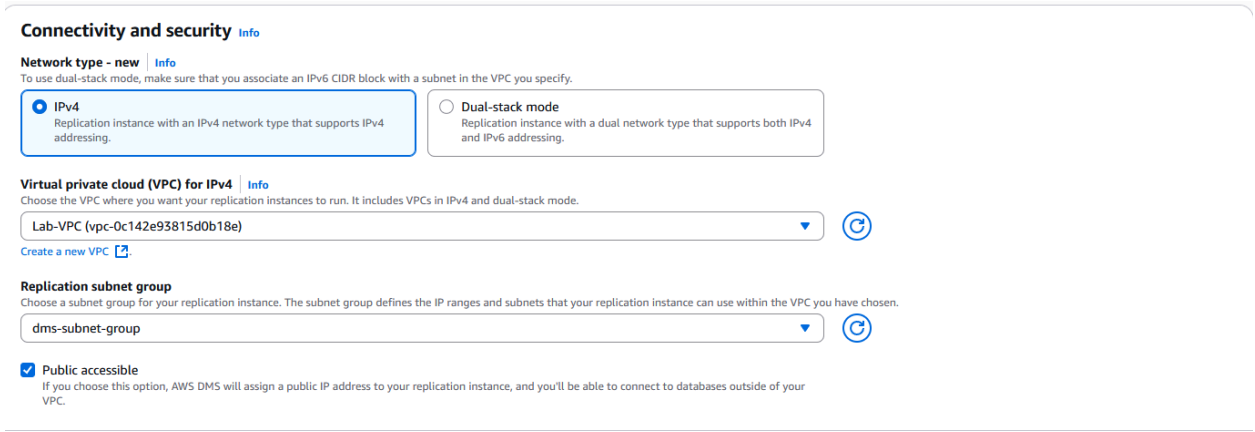
1. En la página de DMS, selecciona **Replication instances** y luego **Create replication instance**.
2. **Name y Description:** Especifica un nombre y descripción.



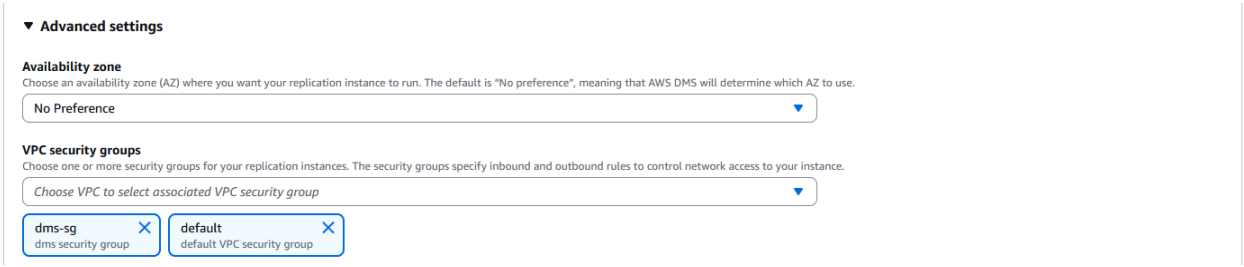
3. **Instance class:** Elige la clase de instancia adecuada.
4. **Engine version:** Selecciona la versión **3.5.3** (para evitar errores con los motores de BD)



5. **VPC:** Elige Lab-VPC.
6. **Publicly accessible:** Habilita esta opción.



7. **High Availability:** Si es obligatorio, selecciona Dev or Test Single AZ.
8. **Replication subnet group:** Selecciona el grupo de subredes creado (dms-subnetgroup).
9. **Availability zone:** No preference, si es obligatorio elige cualquiera.
10. **VPC security group(s):** Selecciona el Security Group creado anteriormente y quita el default.



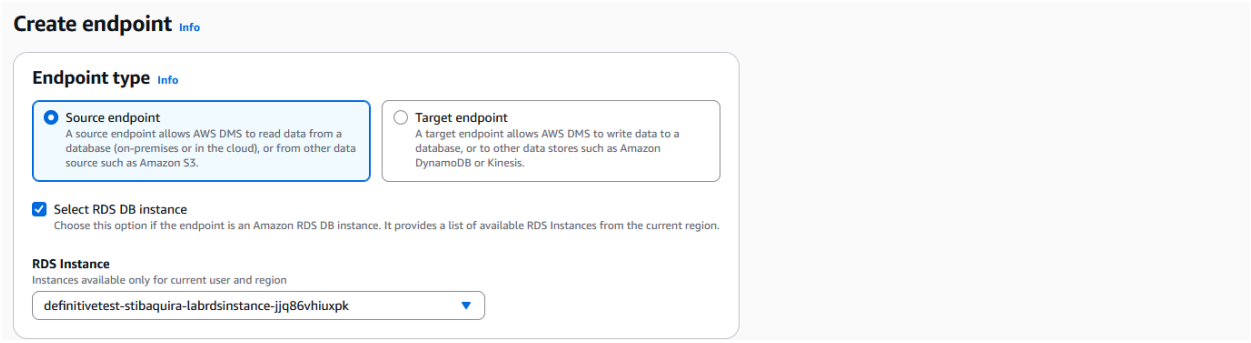
11. Indica una etiqueta (altamente recomendable) y crea la instancia DMS.

8. Creación de Endpoints en DMS

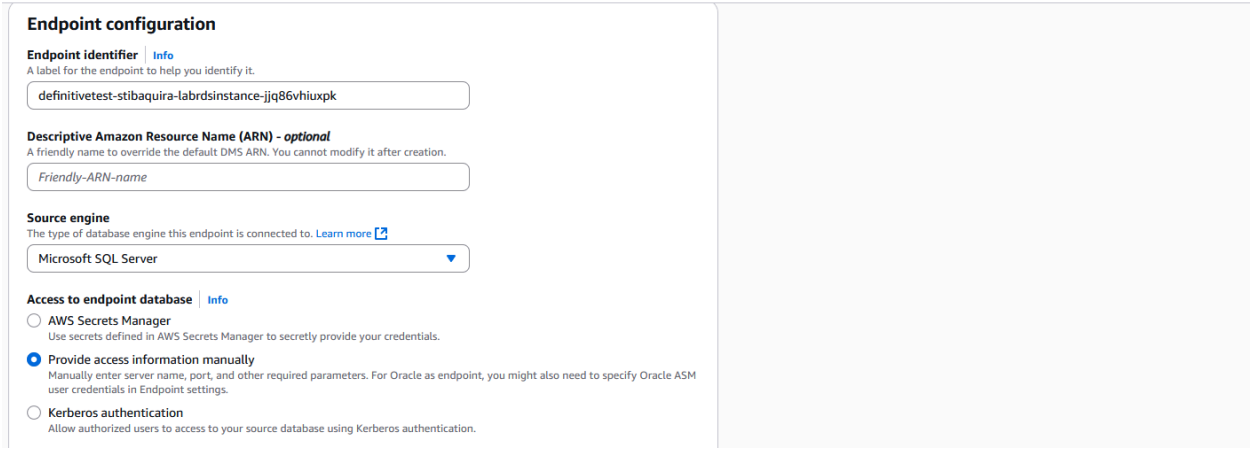
8.1. Endpoint 1: Origen (SQL Server en RDS)

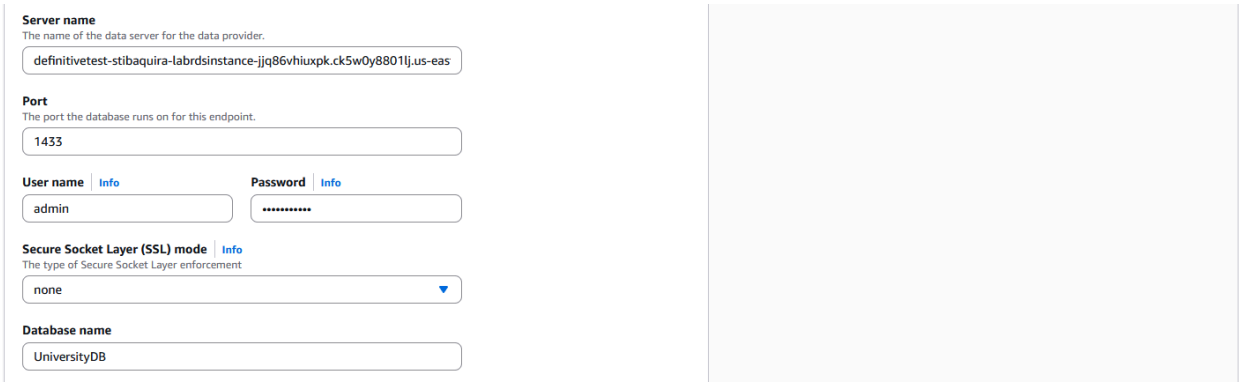
1. En DMS, ve a **Endpoints** y haz clic en **Create endpoint**.
2. **Endpoint type:** Selecciona Source endpoint.
3. **Endpoint identifier:** lab-ep1 (o el nombre que prefieras).
4. **Source engine:** Microsoft SQL Server.

5. Marca la casilla **Select RDS DB Instance** y elige tu instancia RDS.



6. **Access to endpoint database:** Selecciona Provide access information manually.
   - **Server name:** Debería llenarse automáticamente con el endpoint de la RDS. Si no, ingrésalo manualmente.
   - **Port:** 1433
   - **User name:** admin
   - **Password:** #LabDBase3!
   - **Database name:** UniversityDB

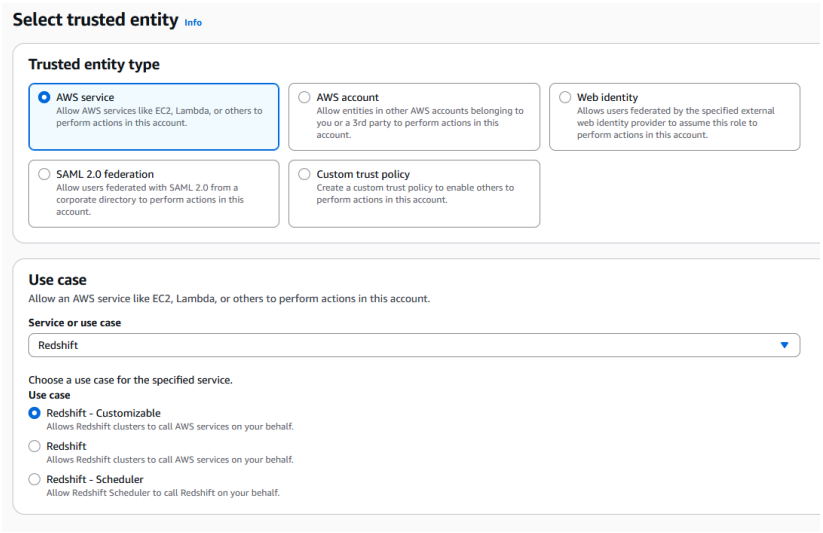




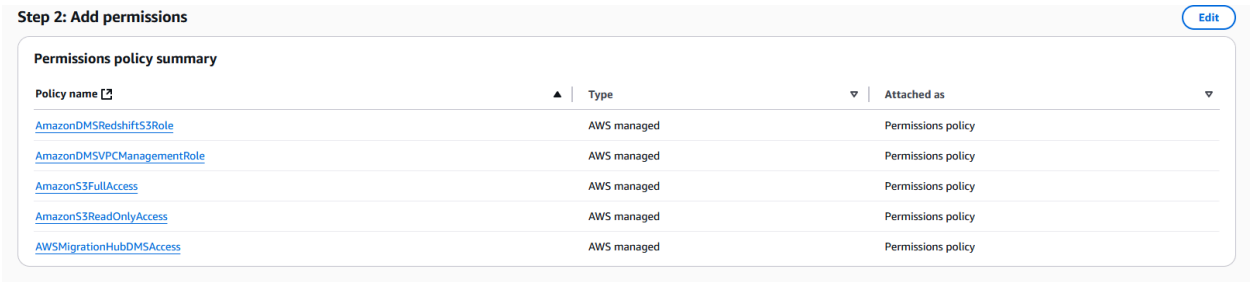
7. **Test endpoint connection:**
   - Selecciona tu instancia de replicación DMS.
   - Haz clic en **Run test**. Verifica que la conexión sea exitosa.
8. Crea el endpoint.

9. Creación de Roles de IAM para AWS Redshift

1. Ve al servicio **IAM** en la AWS Management Console.
2. En el panel de navegación, selecciona **Roles** y haz clic en **Create role**.
3. **Select trusted entity:** Elige AWS service.
4. **Use case:** Selecciona Redshift.
5. **Use case for Redshift:** Selecciona Redshift - Customizable y haz clic en **Next**.



6. **Add permissions:** Busca y selecciona las políticas AmazonS3ReadOnlyAccess, AmazonDMSRedshiftS3Role, AmazonDMSVPCManagementRole, AmazonS3FullAccess, AWSMigrationHubDMSAccess. Haz clic en **Next**.



7. **Role details:**
   - **Role name:** myRedshiftRole
   - **Description:** Deja la descripción automática o modifícala.

8. Añade etiquetas (altamente recomendado) y haz clic en **Create role**.
9. Repite el mismo proceso pero con el servicio de DMS, te quedarás con el siguiente Rol

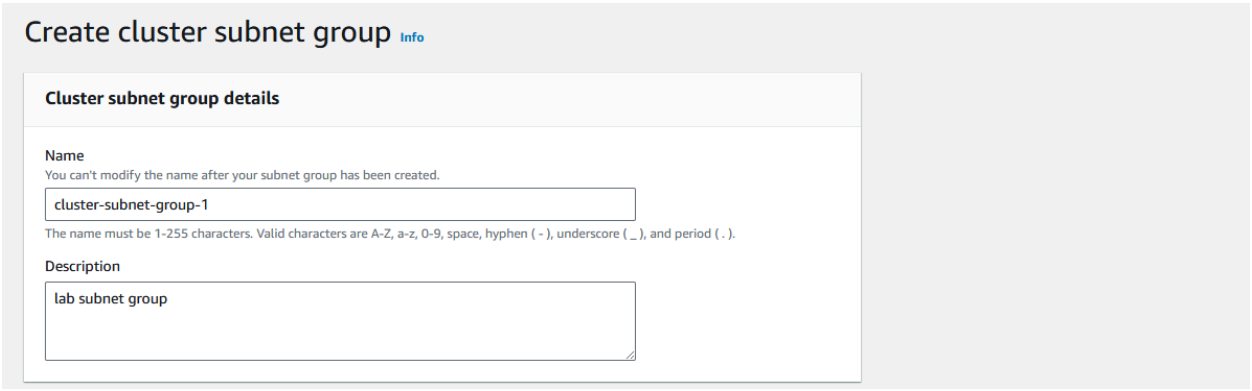
IAM: *AmazonDMSRedshiftS3Role*

10. Configuración de AWS Redshift

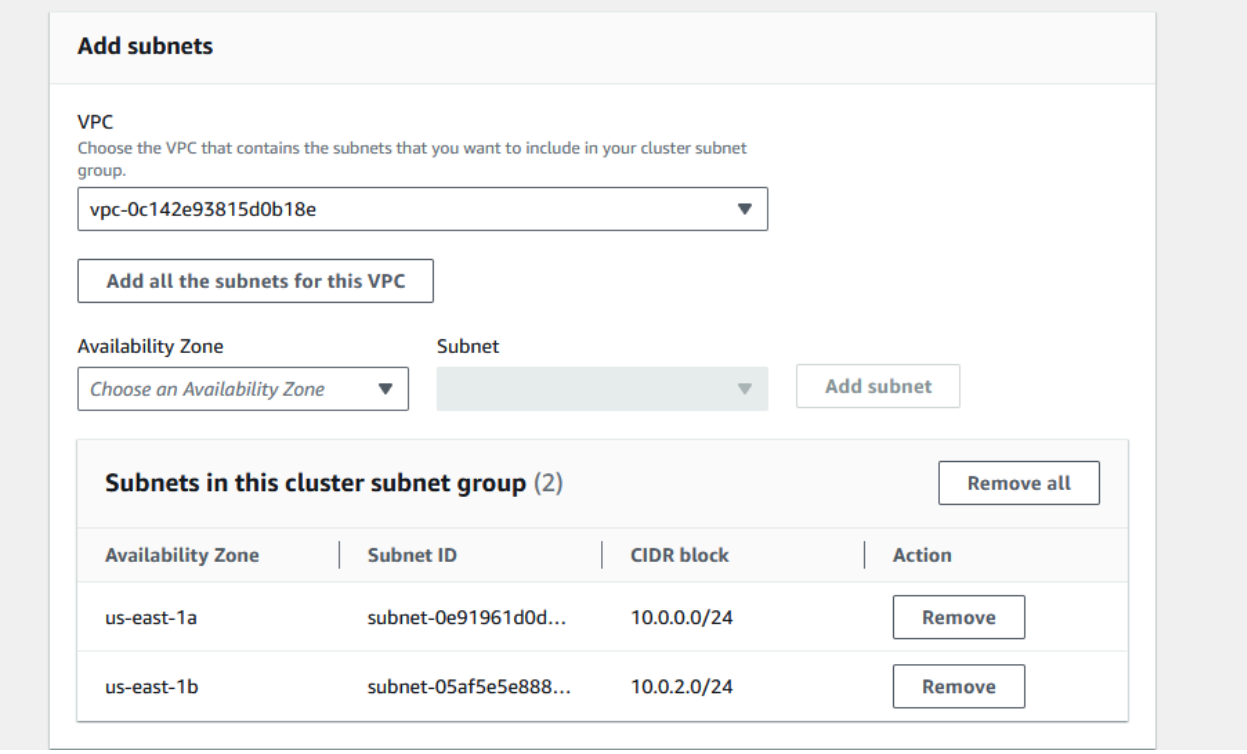
Busca **Redshift** en la AWS Management Console.

10.1. Creación del Cluster Subnet Group en Redshift

1. En el menú de la izquierda de Redshift, bajo **Configurations**, selecciona **Subnet groups**.
2. Haz clic en **Create cluster subnet group**.
3. **Name:** Deja el nombre por defecto o especifica uno (ej: redshift-lab-subnetgroup).
4. **Description:** lab subnet group.

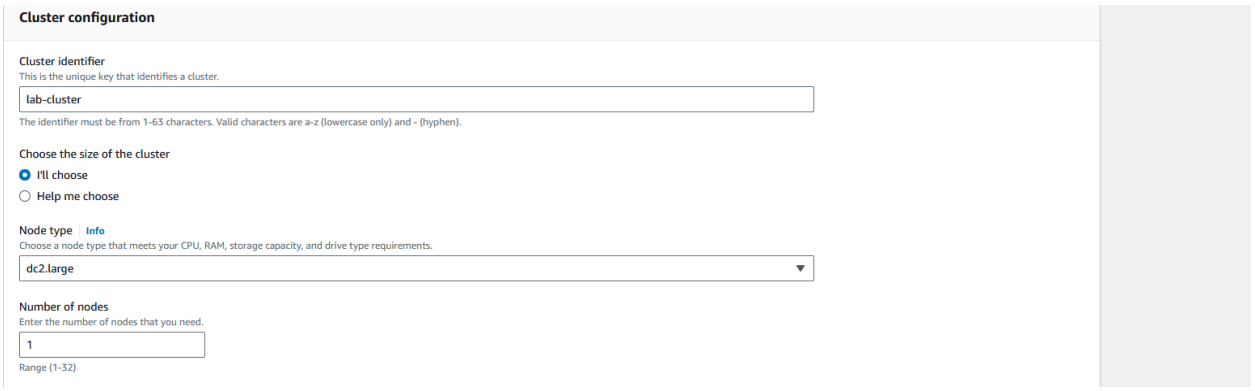


5. **VPC:** Elige la Lab-VPC.
6. **Subnets:** Haz clic en **Add all the subnets for this VPC**.
7. Selecciona las subredes deseadas.
8. Crea el grupo de subredes.

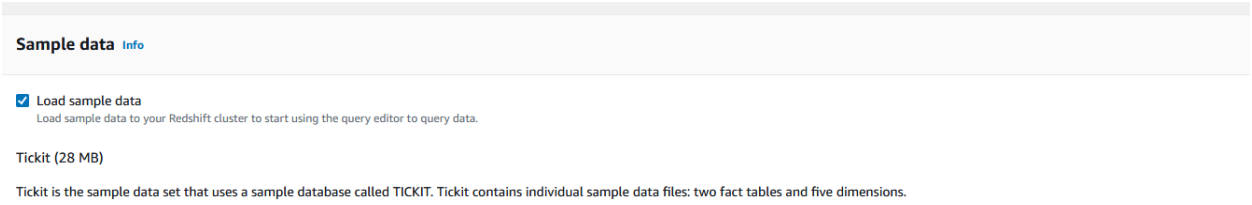


10.2. Creación del Cluster de Redshift

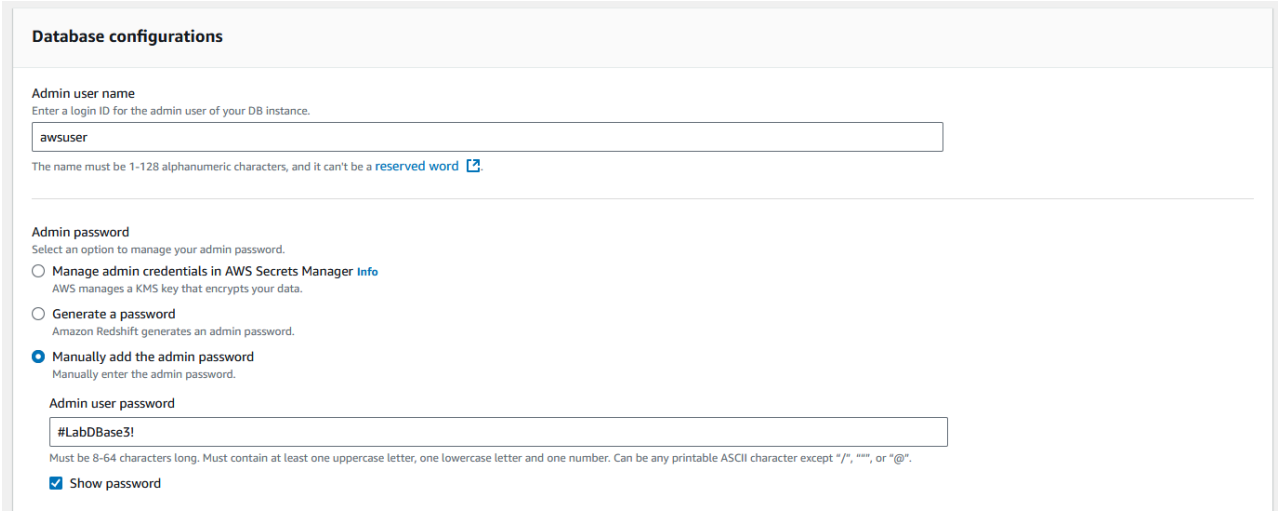
1. En la página de Redshift, haz clic en **Create cluster**.
2. **Cluster configuration:**
   - **Cluster identifier:** lab-cluster.
3. **Nodes:**
   - Elige la opción **I'll choose**
   - **Node type:** dc2.large
   - **Number of nodes:** 1



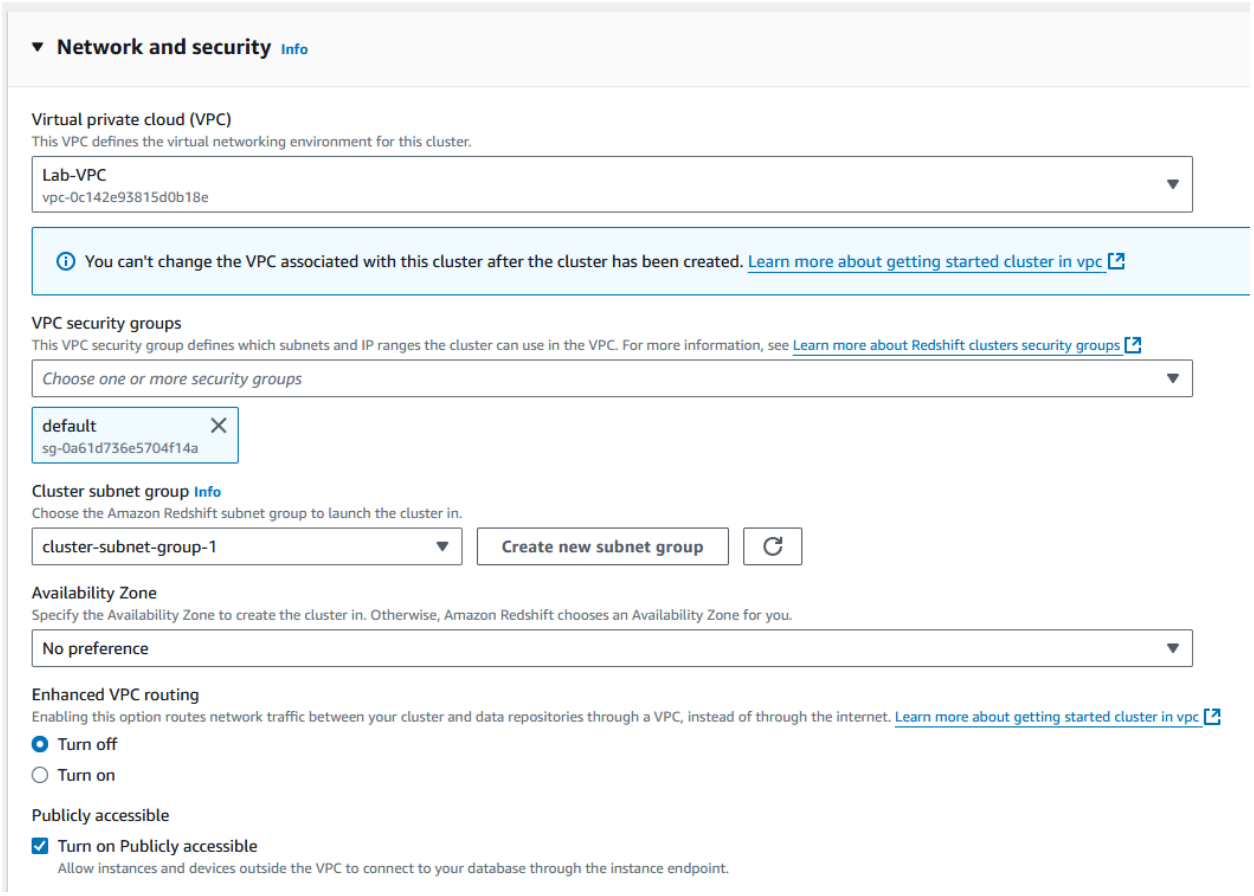
4. **Load sample data:** Marca esta opción.



5. **Database configurations:**
   - **Admin user name:** awsuser
   - **Admin user password:** Selecciona Manually add the admin password (o "Manual password") y pon la contraseña: #LabDBase3!

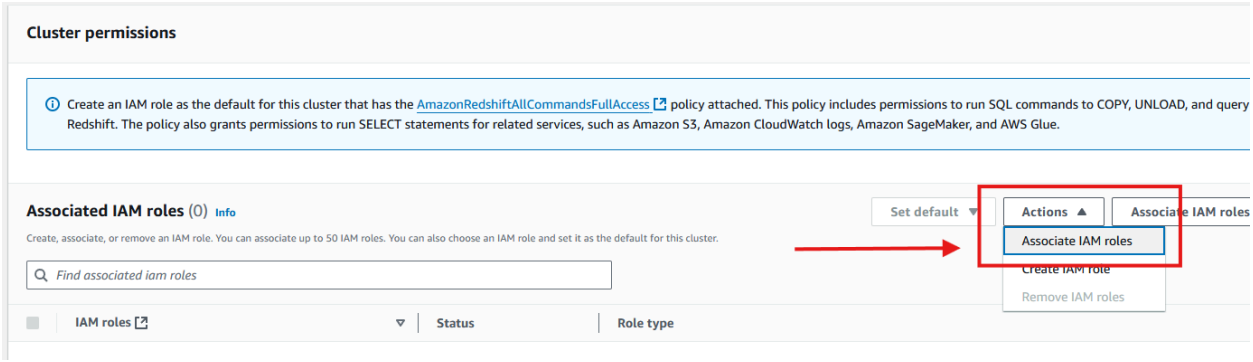


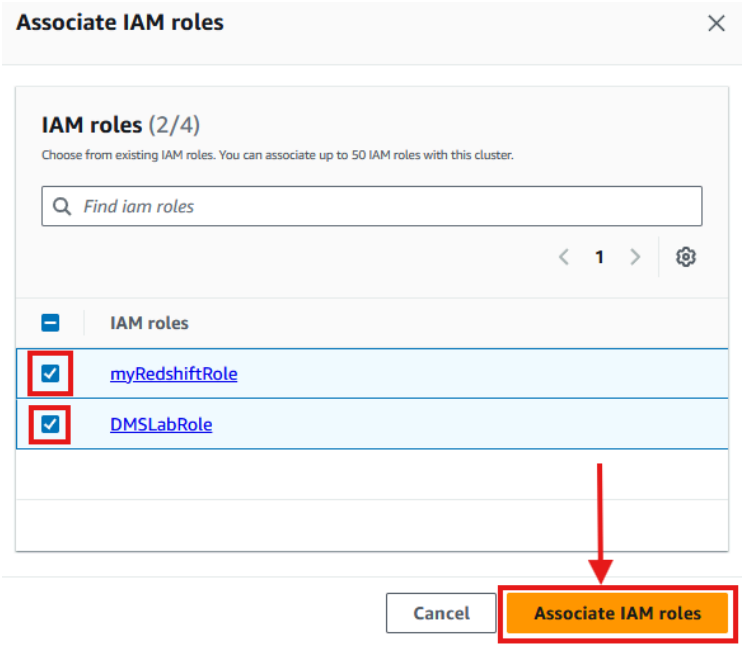
6. **Network and security:**
   - **Virtual private cloud (VPC):** Elige Lab-VPC.
   - **VPC security groups:** Elige el security group para Redshift (si no aparece crealo, ó deja el default).
   - **Cluster subnet group:** Elige el grupo de subredes de cluster que creaste antes.
   - **Publicly accessible:** Activa esta opción (Turn on).

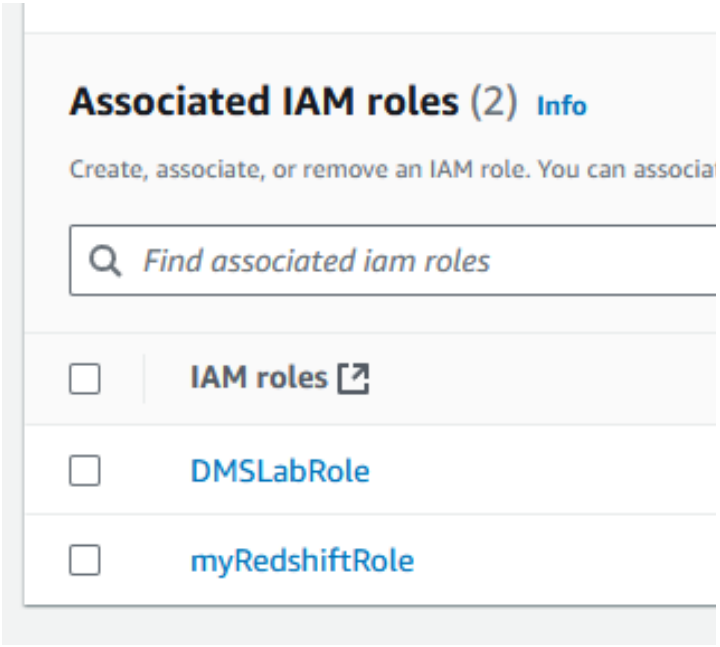


7. **Cluster Permissions:**

   En esta parte bajas a donde sale **Associated IAM roles**, y seleccionas en Actions la opción para asociar los roles de IAM que hemos creado para nuestro Cluster.



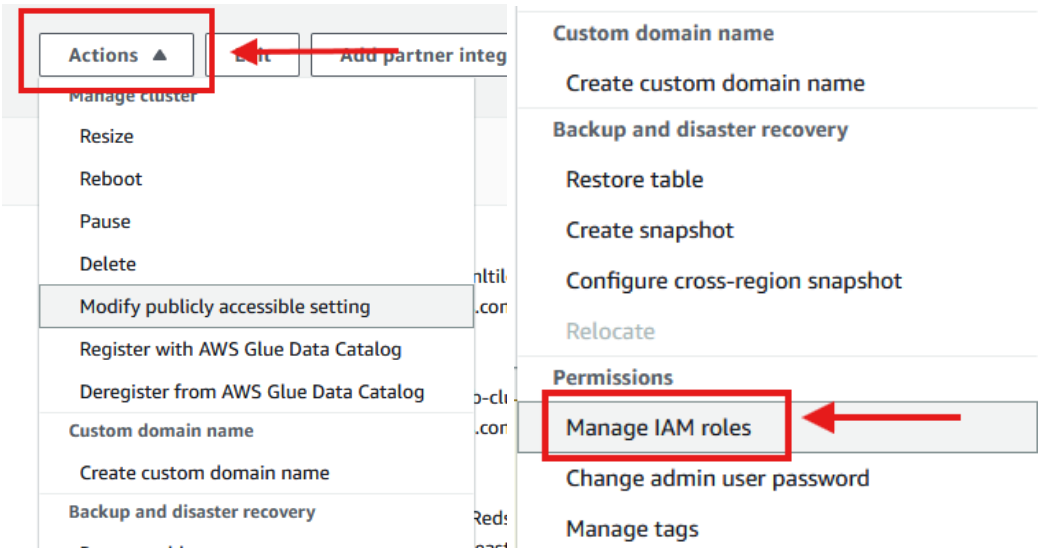




8. Crea el cluster.

10.3. Asociación del Rol IAM al Cluster de Redshift (Si no lo asociaste antes de crearlo)

1. Una vez creado el cluster, selecciónalo en la lista.
2. Ve al menú desplegable **Actions**.
3. Selecciona **Manage IAM roles**.

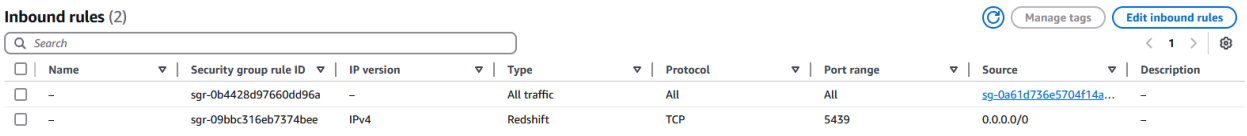


4. Debería aparecer los roles myRedshiftRole y DMSLabRole que creaste. Seleccionalos.
5. Haz clic en **Associate IAM role**.

IMPORTANTE: Si en alguna fase del laboratorio hay algun error de network (específicamente en el endpoint target de DMS), deberás revisar el security group y verificar que este tenga la regla de entrada para Redshift, sí no lo está, sigue estos pasos:

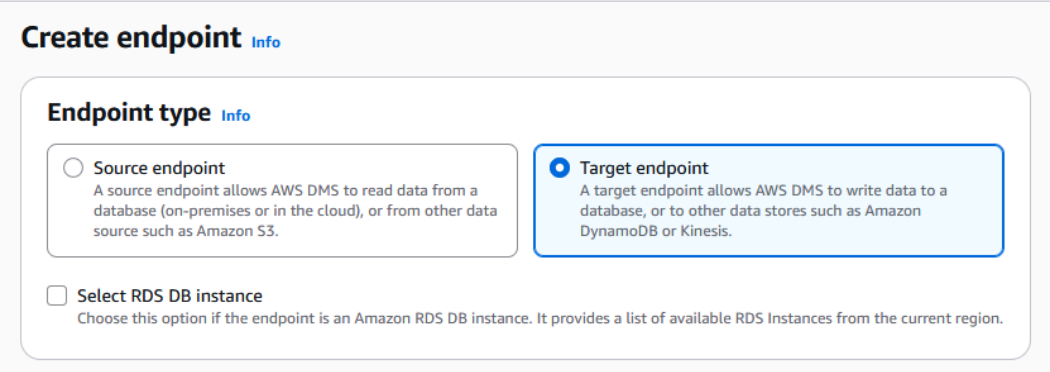
Regla de Security Group para Redshift:
Si el cluster de Redshift quedó conectado al security group default de la Lab-VPC, tendrás que añadirle una regla de entrada:

1. Ve al Security Group asignado al cluster.
2. Selecciona la pestaña **Inbound rules**.
3. Haz clic en **Edit inbound rules**.
4. Haz clic en **Add rule**.
5. Configura la regla:
   - **Type:** Redshift.
   - **Protocol:** TCP (automático).
   - **Port range:** 5439 (automático).
   - **Source:** 0.0.0.0/0.
6. Guarda las reglas.

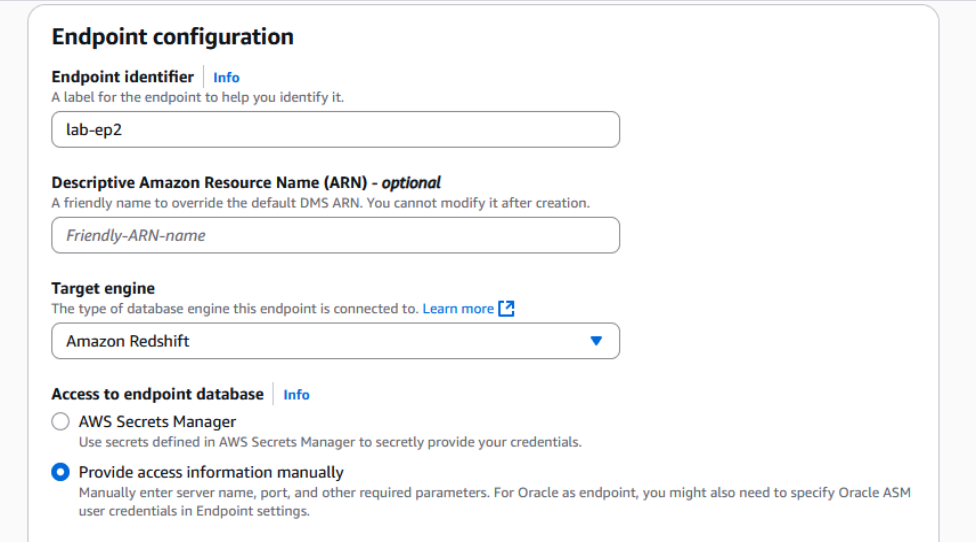


11. Creación del Endpoint 2 en DMS: Destino (Redshift)

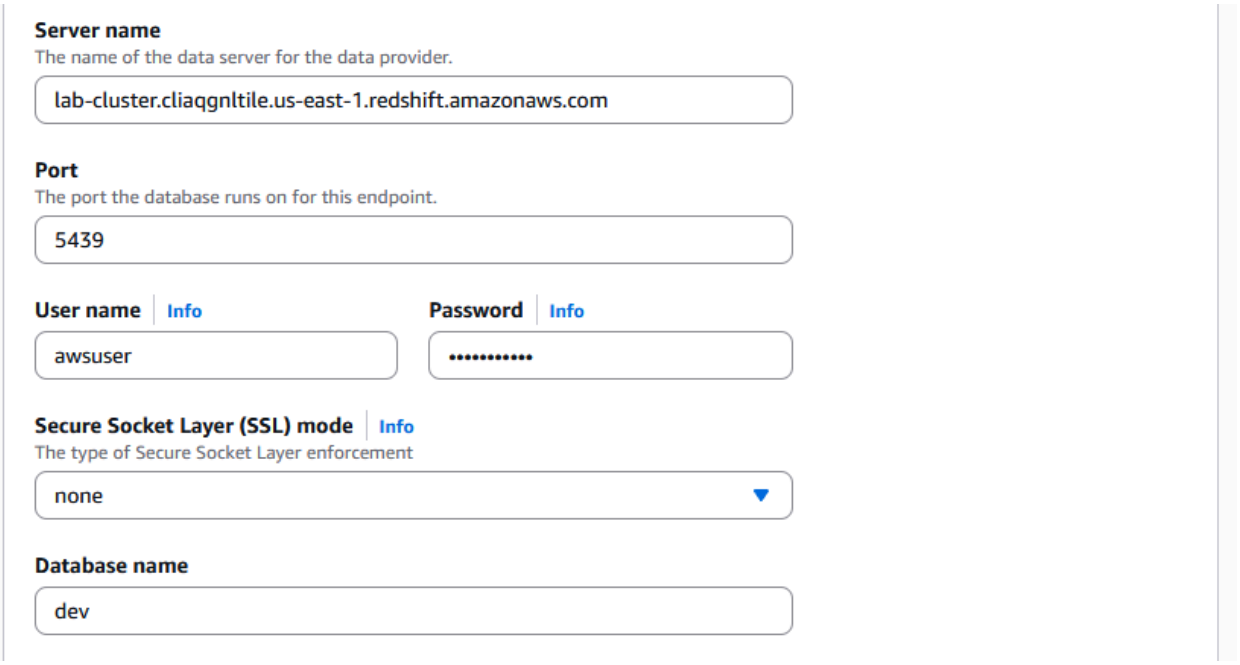
1. Vuelve a **DMS** en la AWS Management Console.
2. Ve a **Endpoints** y haz clic en **Create endpoint**.
3. **Endpoint type:** Selecciona Target endpoint.



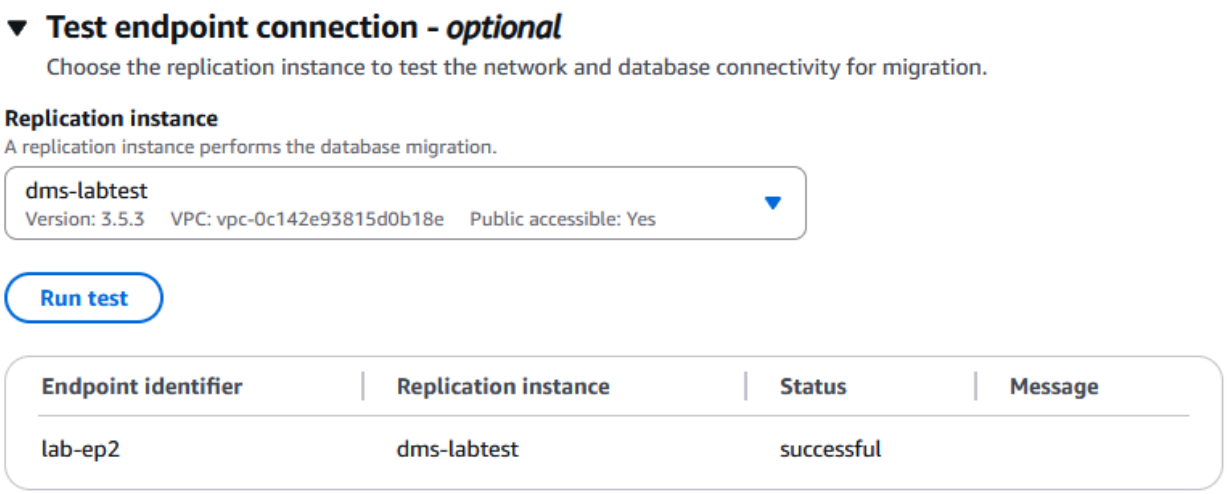
4. **Endpoint identifier:** lab-ep2 (o el nombre que prefieras).
5. **Target engine:** Amazon Redshift.
6. Marca la casilla **Select Redshift cluster** y elige tu cluster lab-cluster.



7. **Access to endpoint database:** Selecciona Provide access information manually.
   - **Server name:** Debería llenarse automáticamente con el endpoint del cluster de Redshift. Si no, ingrésalo manualmente.
   - **Port:** 5439
   - **User name:** awsuser
   - **Password:** #LabDBase3!
   - **Database name:** dev (esta es la base de datos por defecto que se crea en Redshift).



8. **Test endpoint connection:**
   - Selecciona tu instancia de replicación DMS.
   - Haz clic en **Run test**. Verifica que la conexión sea exitosa. (Si falla por error de red, verifica la regla del Security Group del cluster de Redshift).



9. Crea el endpoint.

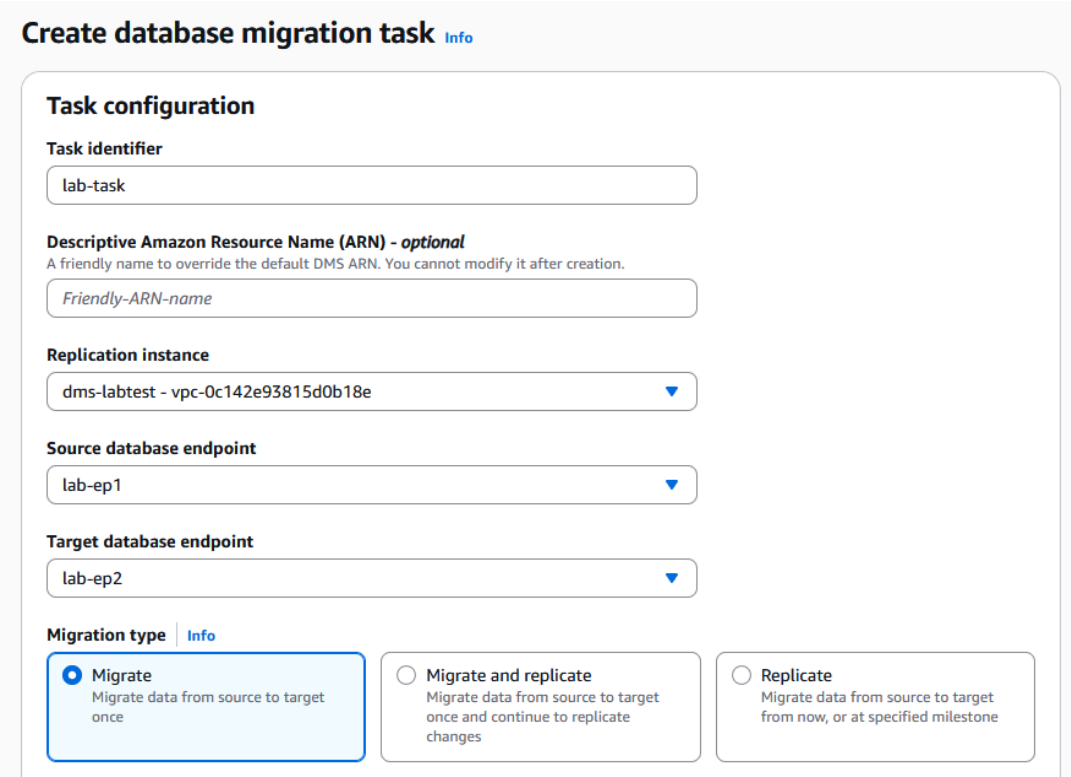
(si sale failed por network error, es por que el security group del cluster no tiene el inbound rule de Redshift:5439)
(en caso de no resolverse el error de conexión entre el endpoint y redshift, ve al cluster, luego a modify, y vuelve a poner la contraseña de awsuser. Sí no, mira de nuevo en modify y ajusta el servername de forma que termine en ...amazonaws.com, y no en ...amazonaws.com:5439/dev)

12. Creación de la Tarea de Migración de Base de Datos en DMS

1. En la página de DMS, ve a la sección **Database migration tasks**.
2. Haz clic en el botón **Create task**.

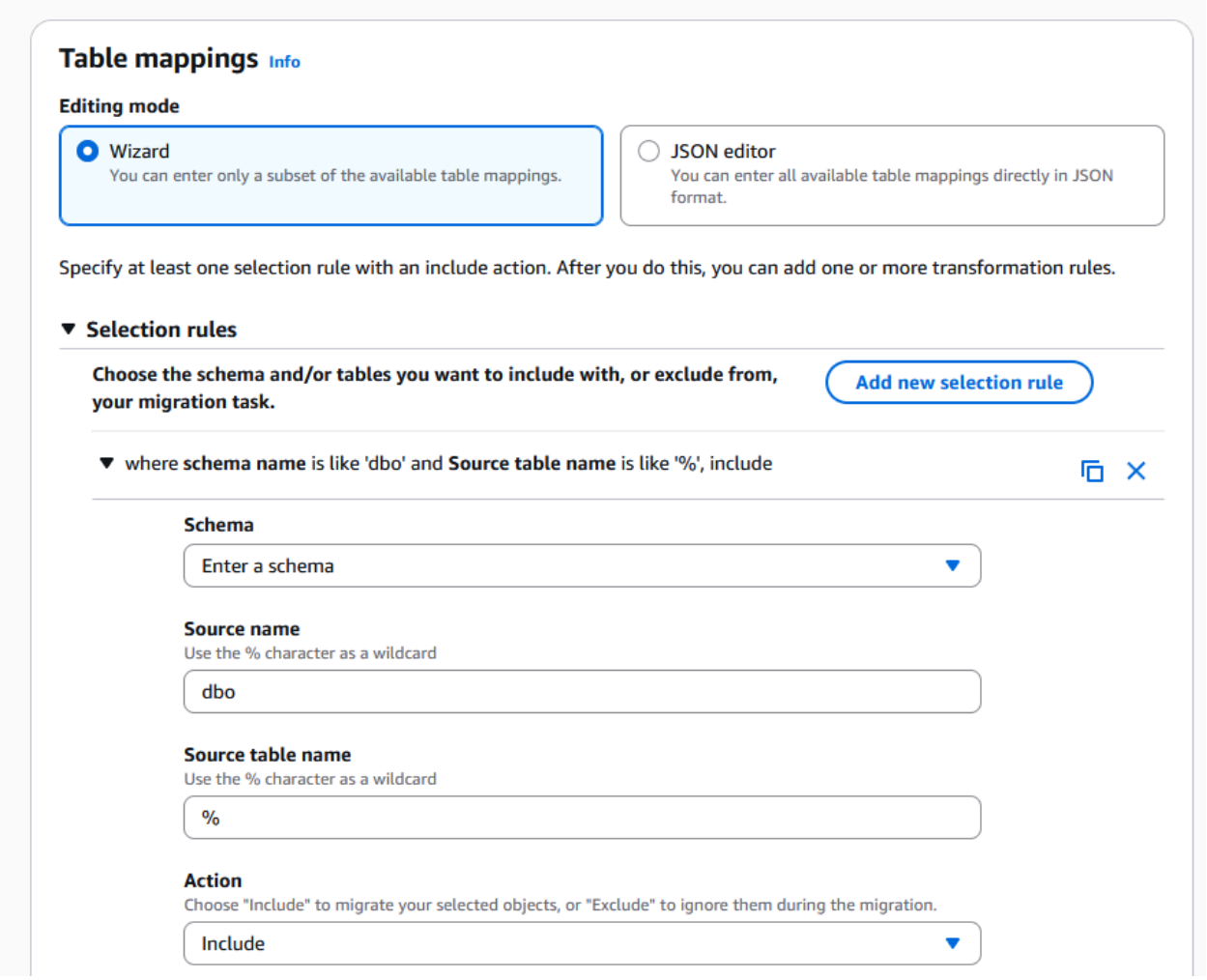
12.1. Configuración de la Tarea

- **Task identifier:** lab-task (o el nombre que prefieras).
- **Replication instance:** Selecciona la instancia DMS ya creada.
- **Source database endpoint:** Elige tu endpoint de origen (lab-ep1 - RDS SQL Server).
- **Target database endpoint:** Elige tu endpoint de destino (lab-ep2 - Redshift).
- **Migration type:** Deja seleccionada la opción Migrate existing data (Migrar datos existentes).



12.2. Mapeo de Tablas (Table Mappings)

1. Baja hasta la sección **Table mappings**.
2. Expande **Selection rules** y haz clic en **Add new selection rule**.
3. Configura la regla de selección:
   - **Schema:** Elige Enter a schema.
   - **Source name (Schema name):** dbo
   - **Table name:** Deja el predeterminado: % (para incluir todas las tablas del esquema dbo).
   - **Action:** Include.



12.3. Evaluación Previa a la Migración (Premigration Assessment)

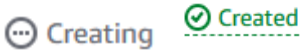
- Asegúrate de que la opción **Turn on premigration assessment** (o similar) **NO** esté activada, a menos que se indique lo contrario para el laboratorio.

12.4. Creación de la Tarea

- Finalmente, haz clic en **Create task**.

13. Monitoreo del Estado de la Tarea

Espera y monitorea los estados de la tarea de migración. Debería pasar por los siguientes estados:

1. Creating 

2. Starting 

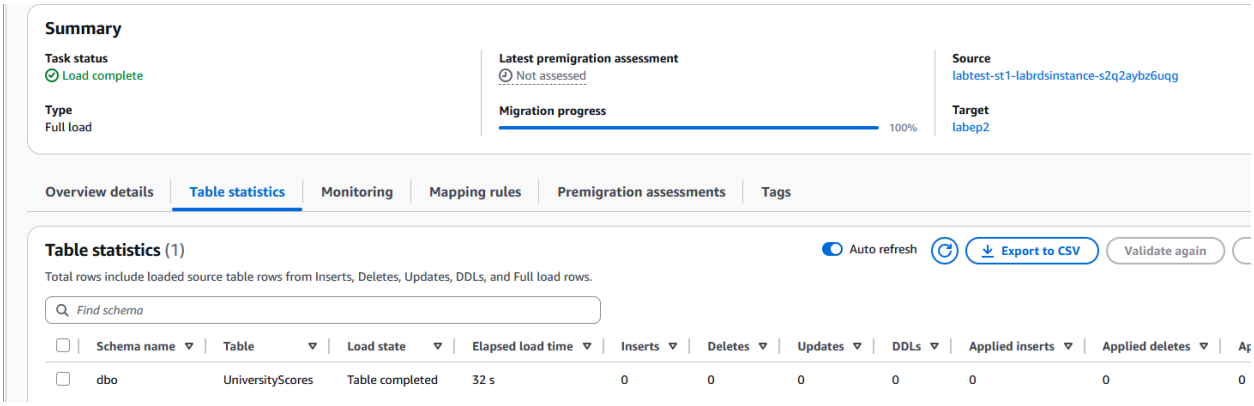
3. Running (o Full load in progress) 

4. Finalmente, Load complete. Load complete

14. Verificación de Tablas Migradas en la Tarea

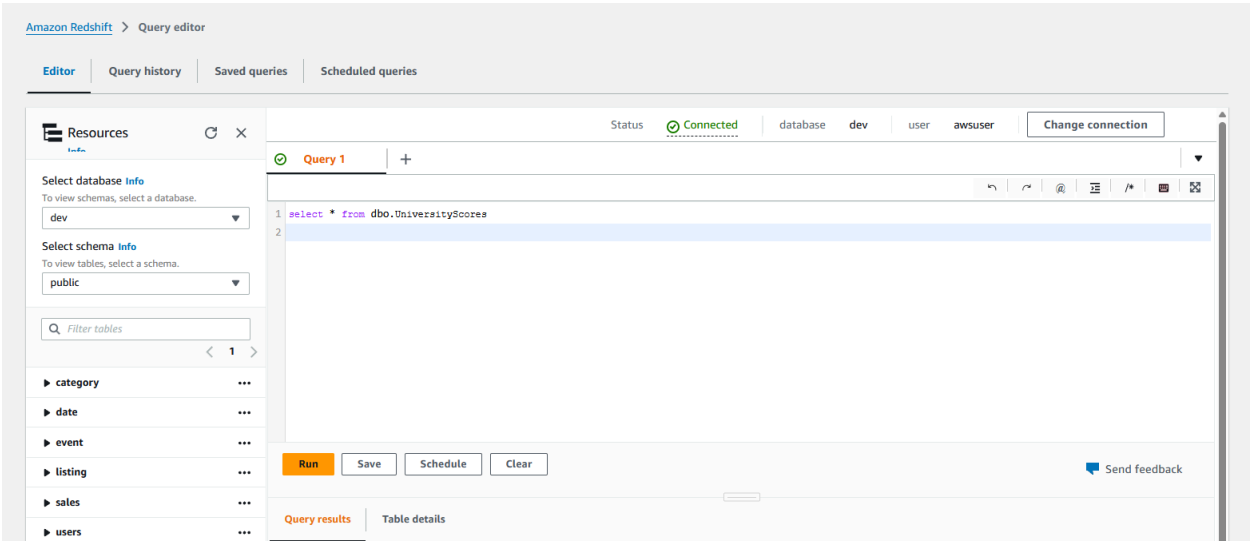
Dentro de los detalles de la tarea de migración completada (en la pestaña **Table statistics** o similar), verifica que la tabla se haya migrado:

- Debería aparecer una tabla con:
  - **Schema:** dbo
  - **Table name:** UniversityScores (o el nombre de tu tabla)
  - **Load state:** Table completed (o similar)
  - Se deberían mostrar estadísticas de las filas cargadas.



15. Validación de la Migración en AWS Redshift

1. Ve al servicio **Redshift** en la AWS Management Console.
2. Selecciona tu cluster (lab-cluster).
3. Haz clic en **Query data** y luego elige **Query editor** (o la versión del editor de consultas disponible).



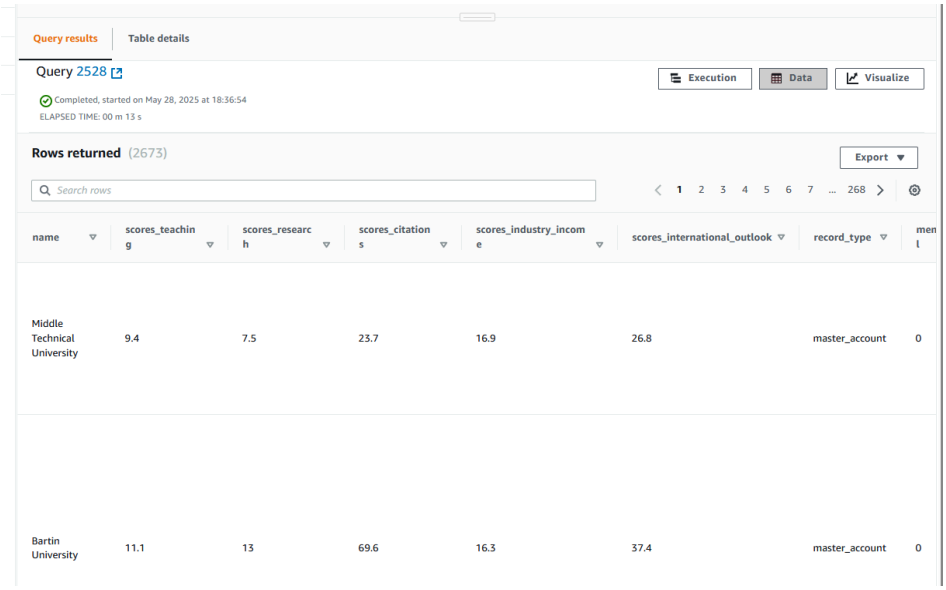
4. Si es la primera vez o necesitas configurar la conexión:

- Haz clic en **Create connection** (o configura la conexión existente).
- **Authentication:** Temporary credentials o Database user name and password.
- **Cluster:** Selecciona tu cluster lab-cluster.
- **Database name:** dev.
- **Database user:** awsuser.
- (Ingresa la contraseña que asignamos en el lab).
- Conéctate.

5. Ejecuta una consulta de prueba para visualizar que todas las columnas de la tabla se migraron:

   SELECT * FROM dbo.UniversityScores LIMIT 10;

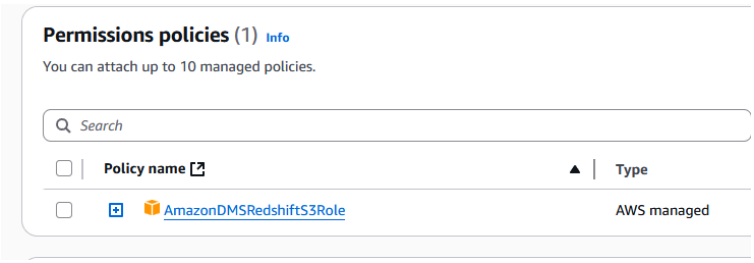
6. Ejecuta la consulta. Deberías ver los datos de la tabla UniversityScores.



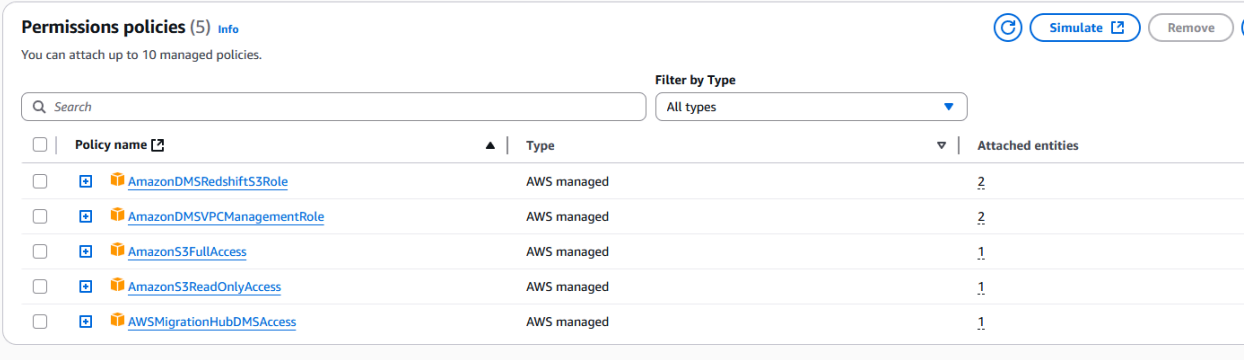
16. Notas Finales (por si hay dudas)

Dentro del cluster en Redshift se utilizó los siguientes roles:

El rol `dms-access-for-endpoint`:



En `myRedshiftRole`:



Recuerda que añadir etiquetas a cada recurso que levantes, hace parte de las buenas prácticas que recomienda AWS.

¡Felicidades!

Ha completado el laboratorio de migración de datos con AWS DMS.