

Creación y Configuración de la Máquina Virtual en Azure

1. Crear la máquina virtual en Azure:

- Entra al portal de Azure y crea una máquina virtual con las especificaciones:
 - Nombre: **P2-tu_boleta** (por ejemplo, **P2-2022630280**).
 - RAM: 4 GB
 - CPU: 2 núcleos virtuales
 - Disco: SSD estándar
- Una vez creada, toma una captura de la última pantalla de confirmación de la creación para incluirla en tu PDF.

The screenshot displays the Azure portal interface. At the top, the browser address bar shows the URL: `portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F73...`. The main header indicates the deployment name: **CreateVm-canonical.ubuntu-24_04-lts-server-20241106184304 | Información general**. Below this, a green checkmark icon and the text **Se completó la implementación** (Implementation completed) are visible. The deployment details include:

- Nombre de implementación: `CreateVm-canonical.ubuntu-24_04-lts...`
- Suscripción: `Azure for Students`
- Id. de correlación: `8a9ad17a-8cf3-439c-a0c9-a469b8adb2f6`

Below the deployment status, a table titled **Detalles de implementación** (Implementation details) lists the resources created:

Recurso	Tipo	Estado	Detalles de la operación
P2-2022630280	Microsoft.Compute/virtualMachines	OK	Detalles de la operación
p2-2022630280225_z1	Microsoft.Network/networkInterfaces	Created	Detalles de la operación
P2-2022630280-vnet	Microsoft.Network/virtualNetworks	OK	Detalles de la operación
P2-2022630280-ip	Microsoft.Network/publicIPAddresses	OK	Detalles de la operación
P2-2022630280-nsg	Microsoft.Network/networkSecurityGroups	OK	Detalles de la operación

Below the table, the **Pasos siguientes** (Next steps) section provides recommendations:

- [Configurar el apagado automático](#) (Recommended)
- [Supervisar el estado, el rendimiento y las dependencias de red de la máquina virtual](#) (Recommended)
- [Ejecutar un script dentro de la máquina virtual](#) (Recommended)

At the bottom of the next steps, there are buttons for [Ir al recurso](#) and [Crear otra VM](#).

On the right side of the screen, there are several informational panels:

- Cost Management**: Obtenga una notificación para permanecer dentro del presupuesto y evitar cargos inesperados en su factura. [Configurar alertas de costo >](#)
- Microsoft Defender for Cloud**: Proteja sus aplicaciones e infraestructura. [Ir a Microsoft Defender for Cloud >](#)
- Tutoriales gratuitos de Microsoft**: Comience a aprender hoy >
- Trabajar con un experto**: Los expertos de Azure son asociados proveedores de servicios que pueden ayudar a administrar sus recursos en Azure y ser la primera línea de soporte técnico.

Instalación de MySQL y Creación de la Base de Datos

2. Instalar MySQL en la máquina virtual ejecutando:

```
sudo apt update
```

```
sudo apt install mysql-server
```

```
18:51:47
Wednesday, 6 November 2024
(alfredo) 52.247.229.251 — Konsole

New Tab Split View
Copy Paste Find...

reading /usr/share/mecab/dic/ipadic/Noun.org.csv ... 16668
reading /usr/share/mecab/dic/ipadic/Adnominal.csv ... 135
reading /usr/share/mecab/dic/ipadic/Noun.nal.csv ... 42
reading /usr/share/mecab/dic/ipadic/Postp.col.csv ... 91
reading /usr/share/mecab/dic/ipadic/Prefix.csv ... 221
reading /usr/share/mecab/dic/ipadic/Symbol.csv ... 208
reading /usr/share/mecab/dic/ipadic/Filler.csv ... 19
reading /usr/share/mecab/dic/ipadic/Conjunction.csv ... 171
reading /usr/share/mecab/dic/ipadic/Noun.number.csv ... 42
reading /usr/share/mecab/dic/ipadic/Noun.adverbial.csv ... 795
reading /usr/share/mecab/dic/ipadic/Noun.proper.csv ... 27328
reading /usr/share/mecab/dic/ipadic/Noun.place.csv ... 72999
reading /usr/share/mecab/dic/ipadic/Noun.demonst.csv ... 120
reading /usr/share/mecab/dic/ipadic/Noun.name.csv ... 34202
reading /usr/share/mecab/dic/ipadic/Noun.csv ... 60477
reading /usr/share/mecab/dic/ipadic/Noun.verb.csv ... 12146
reading /usr/share/mecab/dic/ipadic/Noun.others.csv ... 151
reading /usr/share/mecab/dic/ipadic/Postp.csv ... 146
reading /usr/share/mecab/dic/ipadic/Adverb.csv ... 3032
reading /usr/share/mecab/dic/ipadic/Suffix.csv ... 1393
reading /usr/share/mecab/dic/ipadic/Interjection.csv ... 252
reading /usr/share/mecab/dic/ipadic/Auxil.csv ... 199
reading /usr/share/mecab/dic/ipadic/Verb.csv ... 130750
emitting double-array: 100% [#####]
reading /usr/share/mecab/dic/ipadic/matrix.def ... 1316x1316
emitting matrix : 100% [#####]

done!
update-alternatives: using /var/lib/mecab/dic/ipadic-utf8 to provide /var/lib/mecab/dic/debian (mecab-dictionary) in auto mode
Setting up libhtml-parser-perl:amd64 (3.81-1build3) ...
Setting up libhttp-message-perl (6.45-1ubuntu1) ...
Setting up mysql-server (8.0.39-0ubuntu0.24.04.2) ...
Setting up libcgl-pm-perl (4.63-1) ...
Setting up libhtml-template-perl (2.39-2) ...
Setting up libcgl-fast-perl (1:2.17-1) ...
Processing triggers for man-db (2.12.0-4build2) ...
Processing triggers for libc-bin (2.39-0ubuntu8.3) ...
Scanning processes...
Scanning linux images...

Running kernel seems to be up-to-date.

No services need to be restarted.

No containers need to be restarted.

No user sessions are running outdated binaries.

No VM guests are running outdated hypervisor (qemu) binaries on this host.
alfredo@P2-2022630280:~$
```

3. Crear la base de datos **practica_olap**:

- Abre el monitor de MySQL:

mysql -u root -p

- Ejecuta el siguiente comando para crear la base de datos:

create database practica_olap;

```
18:56:32
Wednesday, 6 November 2024
(alfredo) 52.247.229.251 — Konsole

New Tab Split View
Copy Paste Find...

ERROR 1698 (28000): Access denied for user 'root'@'localhost'
alfredo@P2-2022630280:~$ sudo mysql
Maquinas1234: command not found
alfredo@P2-2022630280:~$ mysql -u root -p
Enter password:
ERROR 1698 (28000): Access denied for user 'root'@'localhost'
alfredo@P2-2022630280:~$ sudo mysql
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 15
Server version: 8.0.39-0ubuntu0.24.04.2 (Ubuntu)

Copyright (c) 2000, 2024, 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> USE mysql;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY 'Maquinas1234';
Query OK, 0 rows affected (0.01 sec)

mysql> FLUSH PRIVILEGES;
Query OK, 0 rows affected (0.00 sec)

mysql> exit
Bye
alfredo@P2-2022630280:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 16
Server version: 8.0.39-0ubuntu0.24.04.2 (Ubuntu)

Copyright (c) 2000, 2024, 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.

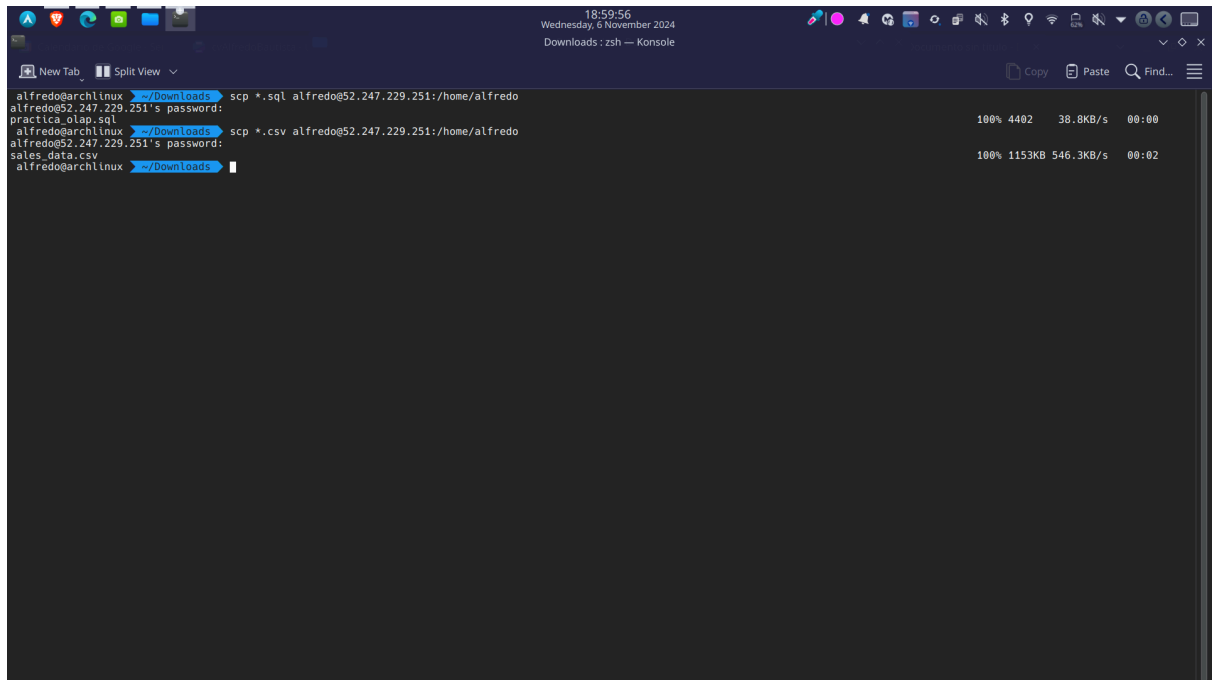
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create database practica_olap;
Query OK, 1 row affected (0.02 sec)

mysql>
```

Carga de Datos en la Base de Datos

4. **Descargar los archivos `practica_olap.sql` y `sales_data.csv` desde Moodle a la máquina virtual.**



5. **Crear tablas con el script SQL:**
 - Ejecuta el script en MySQL:

```
mysql -u root -p practica_olap < practica_olap.sql
```

6. **Configurar MySQL para cargar datos desde archivo local:**

```
mysql --local-infile=1 -u root -p practica_olap
```

7. **Cargar datos en la tabla `sales_data`:**
 - Ejecuta los siguientes comandos en MySQL:

```
SET GLOBAL local_infile = 'ON';
```

```
load data local infile '/home/alfredo/sales_data.csv' into table sales_data
```

```
fields terminated by ',' enclosed by '"' lines terminated by '\n'
```

```
ignore 1 lines
```

```
(sales,@order_date,product,customer,country,region,employee,category,weekday,month,quarter,semester)
```

```
SET order_date=STR_TO_DATE(@order_date,'%Y-%m-%d');
```

```
19:05:55
Wednesday, 6 November 2024
(alfredo) 52.247.229.251 — Konsole

New Tab Split View
Copy Paste Find...

alfredo@P2-2022630280:~$ mysql -u root -p practica_olap < practica_olap.sql
Enter password:
alfredo@P2-2022630280:~$ mysql --local-infile=1 -u root -p practica_olap
Enter password:
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 18
Server version: 8.0.39-0ubuntu0.24.04.2 (Ubuntu)

Copyright (c) 2000, 2024, 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> SET GLOBAL local_infile = 'ON';
Query OK, 0 rows affected (0.00 sec)

mysql> load data local infile 'ruta-archivo-sales_data.csv' into table sales_data
->
->
-> exit
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'exit' at line 4
mysql> load data local infile 'ruta-archivo-sales_data.csv' into table sales_data
-> fields terminated by ',' enclosed by '"' lines terminated by '\n'
-> ignore 1 lines (sales,@order_date,product,customer,country,region,employee,category,weekday,month,quarter,semester)
-> SET order_date=STR_TO_DATE(@order_date,'%d/%m/%Y');
ERROR 2 (HY000): File 'ruta-archivo-sales_data.csv' not found (OS errno 2 - No such file or directory)
mysql> load data local infile '/home/alfredo/sales_data.csv' into table sales_data
-> fields terminated by ',' enclosed by '"' lines terminated by '\n'
-> ignore 1 lines (sales,@order_date,product,customer,country,region,employee,category,weekday,month,quarter,semester)
-> SET order_date=STR_TO_DATE(@order_date,'%d/%m/%Y');
Query OK, 10000 rows affected, 10000 warnings (0.52 sec)
Records: 10000 Deleted: 0 Skipped: 0 Warnings: 10000

mysql>
```

Preparación de Tablas para MapReduce

8. Cargar tablas de dimensiones y la tabla **fact_table** usando los datos de **sales_data**.
9. Verificar el directorio de escritura en MySQL:
 - Ejecuta el siguiente comando para conocer el directorio:

SELECT @@secure_file_priv;

```
19:52:14
Wednesday, 6 November 2024
(alfredo) 52.247.229.251 — Konsole

New Tab Split View
Copy Paste Find...

189.96 | 2001-06-04 | USB Keyboard | Innovate Unlimited | India | Asia | Sarah Moore | Computer Accessories | Monday | June | Q2 | 1st sem |
494.71 | 2001-06-04 | LED Desk Lamp | Pinnacle Tech | Italy | América | Sandra Anderson | Lighting | Monday | June | Q2 | 1st sem |
240.95 | 2001-06-05 | Men's Sneakers | Future Gadgets | Italy | Europa | Joseph Taylor | Footwear | Tuesday | June | Q2 | 1st sem |
913.73 | 2001-06-05 | External Hard Drive | Global Corp | Kenya | Africa | John Smith | Storage Devices | Tuesday | June | Q2 | 1st sem |
630.58 | 2001-06-05 | Hiking Boots | Pinnacle Tech | Egypt | Africa | Sandra Anderson | Outdoor Gear | Tuesday | June | Q2 | 1st sem |
393.08 | 2001-06-05 | Hiking Boots | Elite Tech | Germany | Europa | John Smith | Outdoor Gear | Tuesday | June | Q2 | 1st sem |
295.77 | 2001-06-05 | Smartphone Beta | Elite Tech | France | Europa | Robert Johnson | Mobile Devices | Tuesday | June | Q2 | 1st sem |
667.55 | 2001-06-05 | Acoustic Piano | Tech Pioneers | South Africa | Africa | John Smith | Musical Instruments | Tuesday | June | Q2 | 1st sem |
636.71 | 2001-06-06 | Electric Guitar | Reboot IT | India | Asia | Laura Adams | Musical Instruments | Wednesday | June | Q2 | 1st sem |
724.77 | 2001-06-06 | Desktop Beta | Nebula Innovations | Italy | Europa | Linda Miller | Computers | Wednesday | June | Q2 | 1st sem |
488.82 | 2001-06-06 | Camping Tent | Future Gadgets | South Africa | Africa | Joseph Taylor | Outdoor Gear | Wednesday | June | Q2 | 1st sem |
145.83 | 2001-06-06 | Dumbbell Set | Nebula Innovations | Germany | Europa | Michael Davis | Fitness Equipment | Wednesday | June | Q2 | 1st sem |
483.68 | 2001-06-06 | Pet Bed | Pinnacle Tech | Brazil | América | Sarah Moore | Pet Supplies | Wednesday | June | Q2 | 1st sem |
687.66 | 2001-06-06 | Monitor 4K | Elite Tech | Nigeria | Africa | Robert Johnson | Monitors | Wednesday | June | Q2 | 1st sem |
244.73 | 2001-06-07 | Air Purifier | Tech Pioneers | South Africa | Africa | Sarah Moore | Home Appliances | Thursday | June | Q2 | 1st sem |
388.58 | 2001-06-07 | Fitness Tracker | Pioneer Tech | India | Asia | Patricia Brown | Wearables | Thursday | June | Q2 | 1st sem |
324.93 | 2001-06-07 | Women's Sandals | Global Corp | Egypt | Africa | Sandra Anderson | Footwear | Thursday | June | Q2 | 1st sem |
56.94 | 2001-06-07 | Smart Watch | Tech Trends | Brazil | América | Patricia Brown | Wearables | Thursday | June | Q2 | 1st sem |
342.97 | 2001-06-07 | 4K Television | Ultra Tech ^C | 45.64 | 2001-06-15 | LED Desk Lamp | Pioneer Tech | Nigeria | Africa | Sarah Moore ||
Lighting | Friday | June | Q2 | 1st sem.
^C -- query aborted

20000 rows in set (0.03 sec)

mysql> ;
ERROR:
No query specified

mysql> ;
ERROR:
No query specified

mysql> ;
ERROR:
No query specified

mysql> SELECT @@secure_file_priv;
+-----+
| @@secure_file_priv |
+-----+
| /var/lib/mysql-files/ |
+-----+
1 row in set (0.00 sec)

mysql>
```

10. Generar el archivo CSV necesario para el procesamiento en Hadoop:

```

SELECT a.id_country, b.id_category, a.id_product, a.sales

INTO OUTFILE '/ruta-directorio/country_category_product.csv'

FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'

LINES TERMINATED BY '\n'

FROM fact_table a, product b

WHERE a.id_product = b.id_product;

```

```

19:56:29
Wednesday, 6 November 2024
(alfredo) 52.247.229.251 — Konsole

New Tab Split View Copy Paste Find...

-> WHERE a.id_product = b.id_product;
ERROR 1290 (HY000): The MySQL server is running with the --secure-file-priv option so it cannot execute this statement
mysql> SELECT a.id_country, b.id_category, a.id_product, a.sales
->
-> INTO OUTFILE '/ruta-directorio/country_category_product.csv'
->
-> FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'
->
-> LINES TERMINATED BY '\n'
->
-> FROM fact_table a, product b
->
-> WHERE a.id_product = b.id_product;
ERROR 1290 (HY000): The MySQL server is running with the --secure-file-priv option so it cannot execute this statement
mysql> SHOW VARIABLES LIKE 'secure_file_priv';
+-----+-----+
| Variable_name | Value |
+-----+-----+
| secure_file_priv | /var/lib/mysql-files/ |
+-----+-----+
1 row in set (0.01 sec)

mysql> SELECT a.id_country, b.id_category, a.id_product, a.sales INTO OUTFILE '/ruta-directorio/country_category_product.csv' FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"' LINES
TERMINATED BY '\n' FROM fact_table a, product b WHERE a.id_product = b.id_product;
ERROR 1290 (HY000): The MySQL server is running with the --secure-file-priv option so it cannot execute this statement
mysql> SELECT a.id_country, b.id_category, a.id_product, a.sales
->
->
-> INTO OUTFILE '/var/lib/mysql-files/country_category_product.csv'
->
-> FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'
->
-> LINES TERMINATED BY '\n'
->
-> FROM fact_table a, product b
->
-> WHERE a.id_product = b.id_product;
Query OK, 0 rows affected (0.00 sec)

mysql>
mysql>

```

Instalación y Configuración de Hadoop

11. Descargar e instalar Hadoop:

wget

<https://dlcdn.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0-aarch64.tar.gz>

gunzip hadoop-3.4.0-aarch64.tar.gz

tar xvf hadoop-3.4.0-aarch64.tar

12. Instalar JDK requerido para Hadoop:

sudo apt install openjdk-16-jdk

```
01:09:38
Thursday, 7 November 2024
~: zsh — Konsole

New Tab Split View
Copy Paste Find...

~: zsh x ~: zsh x
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/pro

System information as of Thu Nov 7 05:05:25 UTC 2024

System load: 0.00 Processes: 124
Usage of /: 8.1% of 28.02GB Users logged in: 1
Memory usage: 19% IPv4 address for eth0: 10.0.0.4
Swap usage: 0%

Expanded Security Maintenance for Applications is not enabled.
0 updates can be applied immediately.
Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

Last login: Thu Nov 7 05:03:48 2024 from 187.131.251.142
alfredo@P2-2022630280:~$ sudo apt install openjdk-16-jdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
E: Unable to locate package openjdk-16-jdk
alfredo@P2-2022630280:~$ sudo apt install openjdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
No apt package "openjdk", but there is a snap with that name.
Try "snap install openjdk"
E: Unable to locate package openjdk
alfredo@P2-2022630280:~$ sudo apt get-install openjdk
E: Invalid operation get-install
alfredo@P2-2022630280:~$ sudo apt get install openjdk
E: Invalid operation get
alfredo@P2-2022630280:~$ sudo apt install jdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
E: Unable to locate package jdk
alfredo@P2-2022630280:~$ sudo apt install openjdk-16-jdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
E: Unable to locate package openjdk-16-jdk
```

13. Configurar **JAVA_HOME** en Hadoop:

- Edita el archivo **hadoop-env.sh**:

```
nano hadoop-3.4.0/etc/hadoop/hadoop-env.sh
```

- Descomenta y cambia **JAVA_HOME**:

```
export JAVA_HOME=/usr
```

Configuración de la Aplicación MapReduce en Hadoop

14. Crear directorio y clases Java para MapReduce:

- Define las variables de entorno:

```
export JAVA_HOME=/usr
```

```
export HADOOP_HOME=/home/usuario/hadoop-3.4.0
```

```
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin
```

- Crea el directorio y los archivos Java:

```
mkdir prueba
```

```
nano prueba/AggregationMapper.java
```

```
nano prueba/AggregationReducer.java
```

```
nano prueba/AggregationDriver.java
```

15. Compilar y crear el archivo **.jar**:

```
javac -classpath ` $HADOOP_HOME/bin/hadoop classpath ` -d prueba prueba/*.java
```

```
jar -cvf Aggregation.jar -C prueba .
```

16. Configurar el directorio de entrada para Hadoop:

- Crea el directorio **input**:

```
mkdir prueba/input
```

- Copia el archivo **country_category_product.csv** en el directorio **input**.

17. Ejecutar el job de Hadoop:

```
hadoop jar Aggregation.jar AggregationDriver prueba/input prueba/output
```

- Visualiza el resultado en el directorio **prueba/output**.

Carga del Resultado de MapReduce en MySQL

18. Convertir el archivo de salida a CSV:

```
sed -i 's/\t/,/g' prueba/output/part-r-00000
```

19. Cargar el archivo en MySQL:

- Habilita la carga de datos locales y ejecuta:

```
load data local infile 'ruta-absoluta-del-archivo-part-r-00000' into table  
country_category_product
```

```
fields terminated by ',' enclosed by '"' lines terminated by '\n'
```

```
(id_country,id_category,id_product,sales);
```

Documentación y Conclusión

- **Captura los comandos ejecutados y resultados.**
- **Redacta tus conclusiones** sobre el proceso y resultados obtenidos.

¡Espero que estos pasos te sean útiles para completar tu práctica!