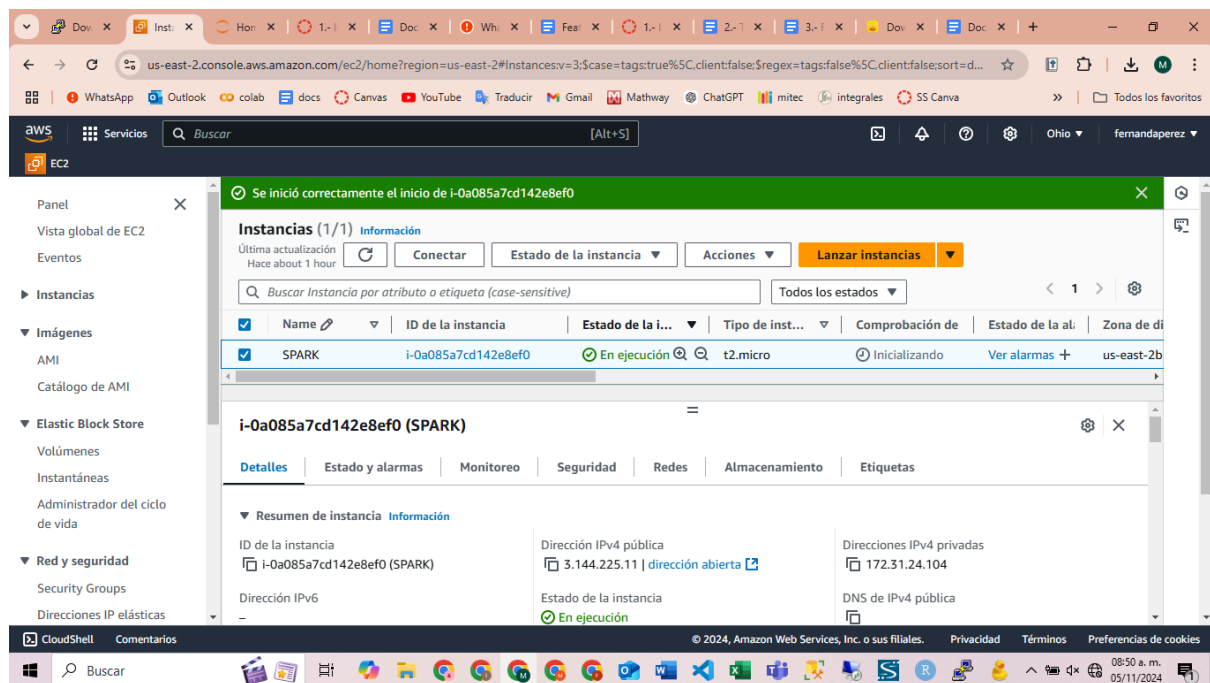


## 1.- Instalación de Spark en AWS

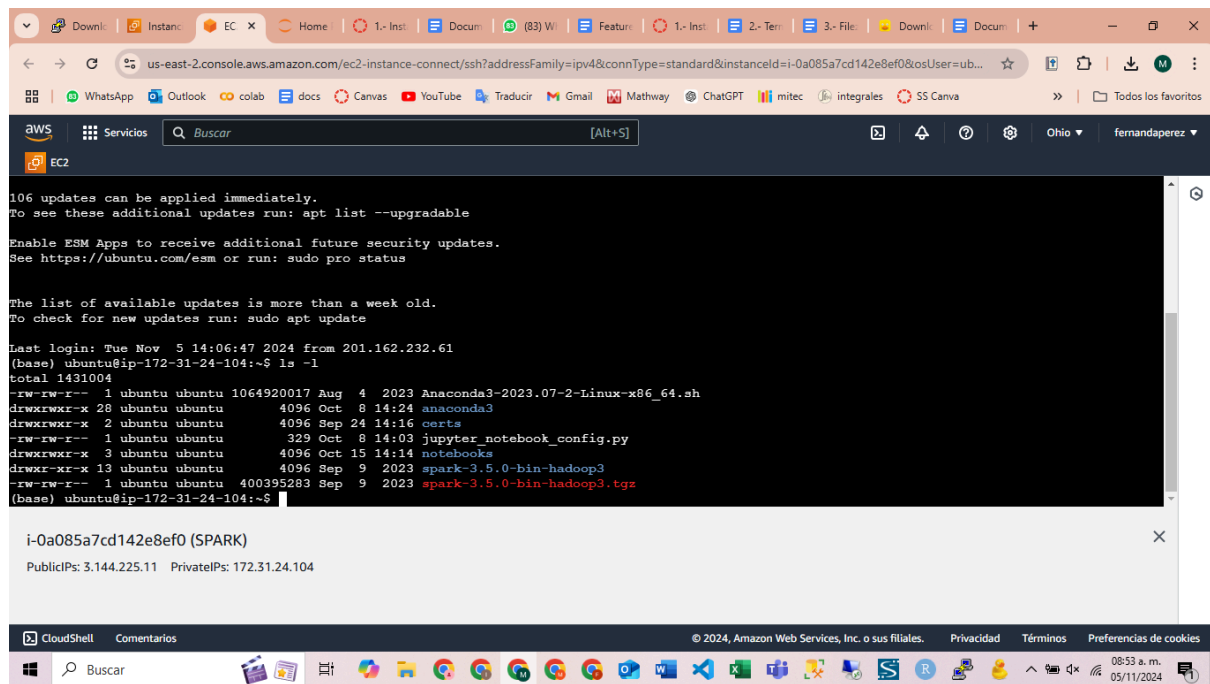
María Fernanda Pérez Ruiz A01742102

Se instalará Spark en una instancia EC2 de AWS acorde a lo visto en clase y se entregará lo siguiente para comprobar su instalación correcta.

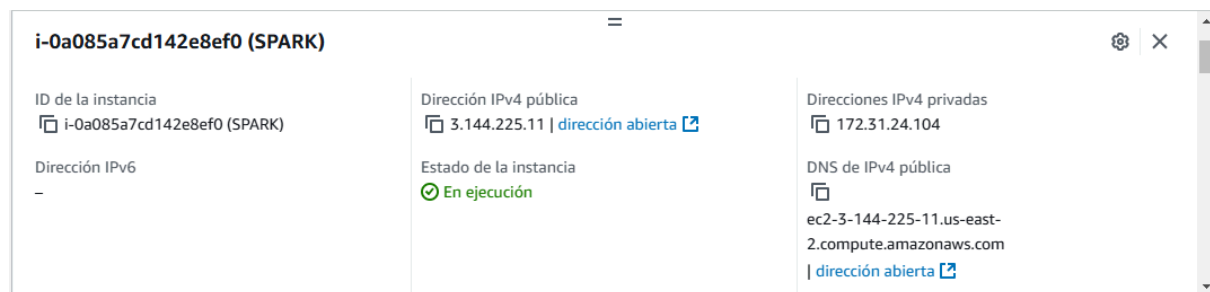
1.- Impresión de pantalla del listado de instancias de EC2 de AWS en donde se muestre la instancia creada.



2.- Impresión de pantalla conectado al servidor ya sea por Terminal o Putty, ya una vez dentro, ejecutar el comando `ls -l` para la toma de la impresión de pantalla.



3.- Impresión de pantalla de la pestaña Detalles para que se vea la ip pública, la ip privada y el DNS público de la instancia (es necesario que la instancia esté Running).



4.- Impresión de pantalla de jupyter notebook visualizando el listado de los notebooks que se proporcionaron como ejemplos.

Documento sin título - Docu... x 1.- Instalación de Spark en / x Instancias | EC2 | us-east-2 x EC2 Instance Connect x Home Page - Select or creat... x

No seguro ec2-18-117-123-110.us-east-2.compute.amazonaws.com:8888/tree?

UPDATE Read the [migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Quit Logout

Files Running Clusters

Select items to perform actions on them. Upload New

|                          | Name                              | Last Modified | File size |
|--------------------------|-----------------------------------|---------------|-----------|
| <input type="checkbox"/> | /                                 |               |           |
| <input type="checkbox"/> | 1.- spark.ipynb                   | hace un mes   | 1.82 kB   |
| <input type="checkbox"/> | 10.- persistencia.ipynb           | hace 23 días  | 5.92 kB   |
| <input type="checkbox"/> | 11.- spark_sql.ipynb              | hace 23 días  | 24.8 kB   |
| <input type="checkbox"/> | 12.- spark_sql_fecha.ipynb        | hace 23 días  | 13.1 kB   |
| <input type="checkbox"/> | 13.- spark_sql_agrupaciones.ipynb | hace 23 días  | 9.91 kB   |
| <input type="checkbox"/> | 14.- mllib.ipynb                  | hace 23 días  | 18.2 kB   |
| <input type="checkbox"/> | 15.- regresion_lineal.ipynb       | hace 23 días  | 17.1 kB   |
| <input type="checkbox"/> | 2.- intro_spark.ipynb             | hace 23 días  | 4.33 kB   |
| <input type="checkbox"/> | 3.- expresiones_lambda.ipynb      | hace 23 días  | 4.24 kB   |
| <input type="checkbox"/> | 4.- acciones.ipynb                | hace 23 días  | 16.9 kB   |
| <input type="checkbox"/> | 5.- transformaciones.ipynb        | hace 23 días  | 12.5 kB   |

Buscar

07:40 a. m. 07/11/2024

Documento sin título - Docu... x 1.- Instalación de Spark en / x Instancias | EC2 | us-east-2 x EC2 Instance Connect x Home Page - Select or creat... x

No seguro ec2-18-117-123-110.us-east-2.compute.amazonaws.com:8888/tree?

UPDATE Read the [migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Quit Logout

|                          |                                   |              |         |
|--------------------------|-----------------------------------|--------------|---------|
| <input type="checkbox"/> | 3.- expresiones_lambda.ipynb      | hace 23 días | 4.24 kB |
| <input type="checkbox"/> | 4.- acciones.ipynb                | hace 23 días | 16.9 kB |
| <input type="checkbox"/> | 5.- transformaciones.ipynb        | hace 23 días | 12.5 kB |
| <input type="checkbox"/> | 6.- ejemplo.ipynb                 | hace 23 días | 44.6 kB |
| <input type="checkbox"/> | 7.- valores_nulos.ipynb           | hace 23 días | 10.1 kB |
| <input type="checkbox"/> | 8.- pair_rdd.ipynb                | hace 23 días | 26.7 kB |
| <input type="checkbox"/> | 9.- particionado.ipynb            | hace 23 días | 8.94 kB |
| <input type="checkbox"/> | AAPL.csv                          | hace 23 días | 192 kB  |
| <input type="checkbox"/> | customers.csv                     | hace 23 días | 86.9 kB |
| <input type="checkbox"/> | ejemplo.txt                       | hace 23 días | 55 B    |
| <input type="checkbox"/> | LaCelestina.txt                   | hace 23 días | 688 kB  |
| <input type="checkbox"/> | Null.csv                          | hace 23 días | 78 B    |
| <input type="checkbox"/> | personas.json                     | hace 23 días | 80 B    |
| <input type="checkbox"/> | sample_linear_regression_data.txt | hace 23 días | 119 kB  |
| <input type="checkbox"/> | ventas.csv                        | hace 23 días | 212 B   |

Buscar

07:40 a. m. 07/11/2024

5.- Crear un notebook con su nombre y colocar el llamado a Pyspark para visualizar la versión instalada.

Documento x 1.- Instalaci x Instancias | E x EC2 Instance x Home Page x 1.- spark - Ju x Fernanda - J x ChatGPT x +

No seguro ec2-18-117-123-110.us-east-2.compute.amazonaws.com:8888/notebooks/Fernanda.ipynb

UPDATE Read the [migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Fernanda Last Checkpoint: hace 3 minutos (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [1]:

```
from pyspark.sql import SparkSession

# Create SparkSession
spark = SparkSession.builder \
    .master("local[*]") \
    .appName("BigData-ETL.com") \
    .getOrCreate()

print(f'The PySpark {spark.version} version is running...')
```

Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
24/11/07 13:44:01 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

The PySpark 3.5.0 version is running...

In [ ]:

07:45 a.m. 07/11/2024