

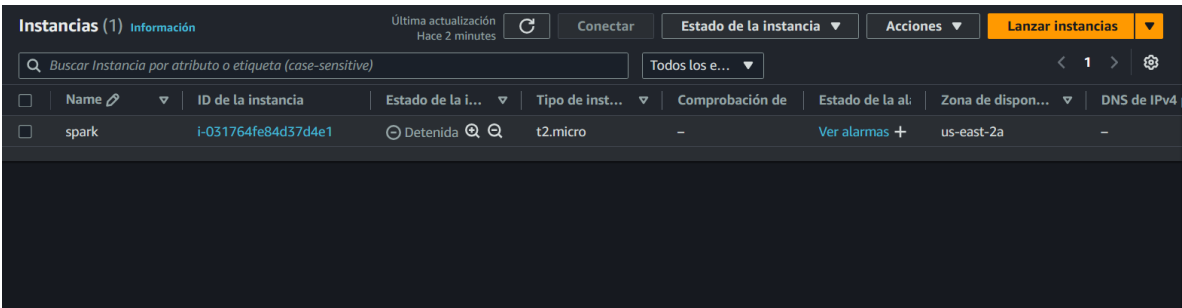
# “Módulo 1 tarea 1: Instalación de Spark en AWS”

Rodolfo Jesús Cruz Rebollar

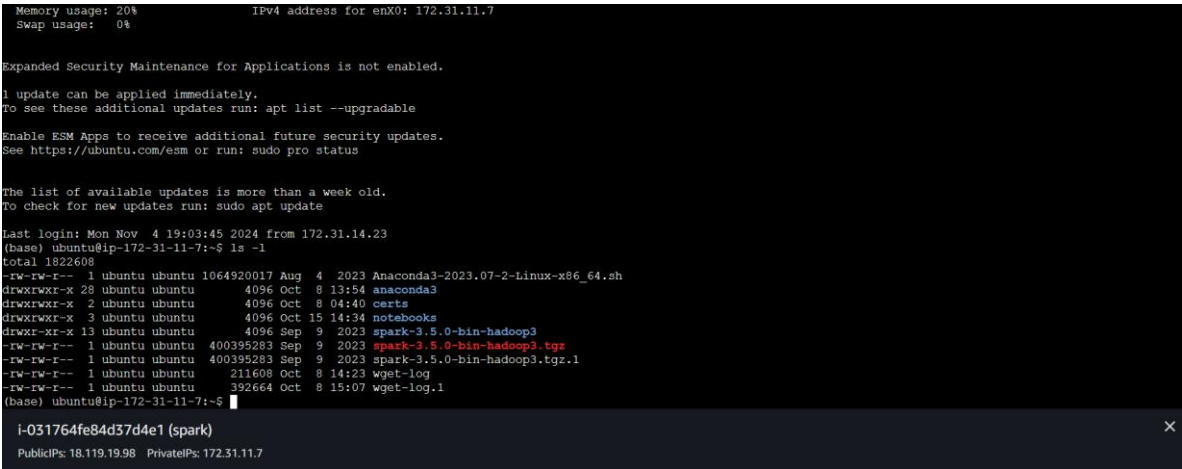
A01368326

Grupo 101

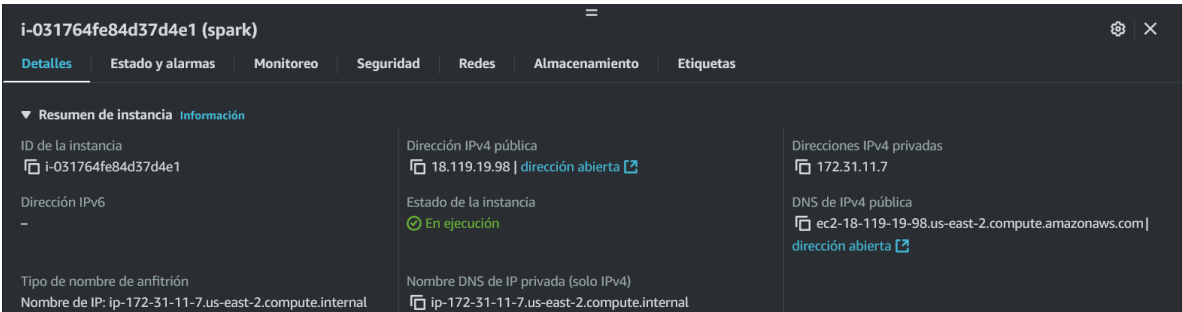
## 1. Listado de instancias de EC2 en AWS:



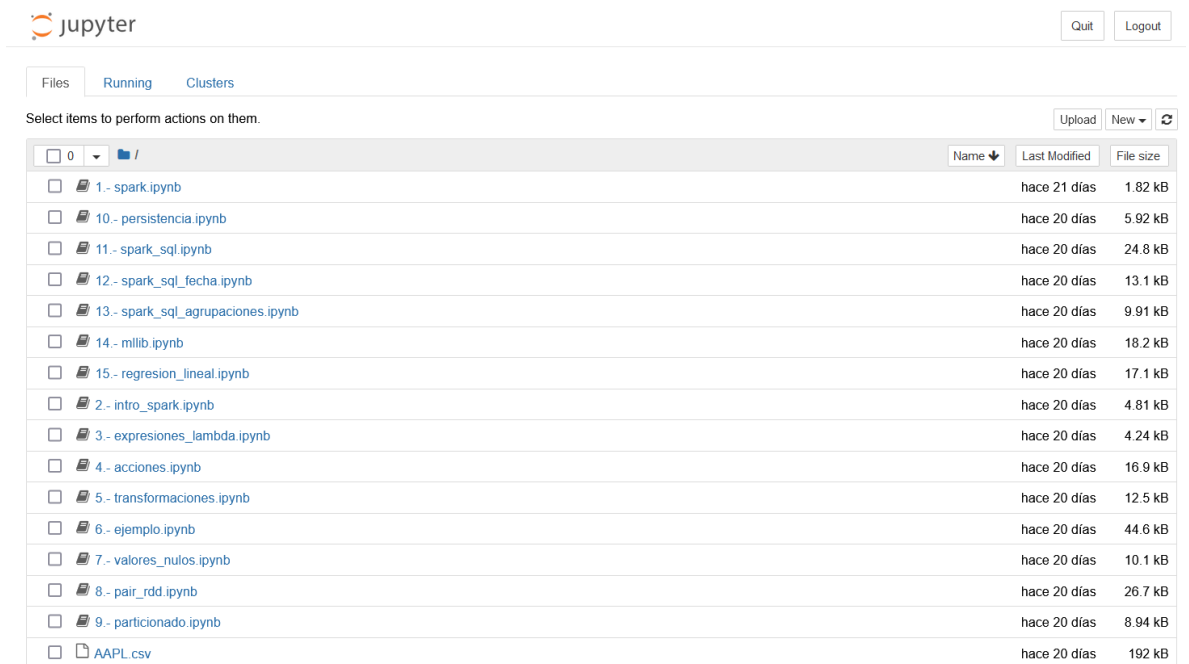
## 2. Conexión al servidor a través de terminal o Putty:



## 3. Pestaña Detalles con IP pública, IP privada y DNS público de la instancia:



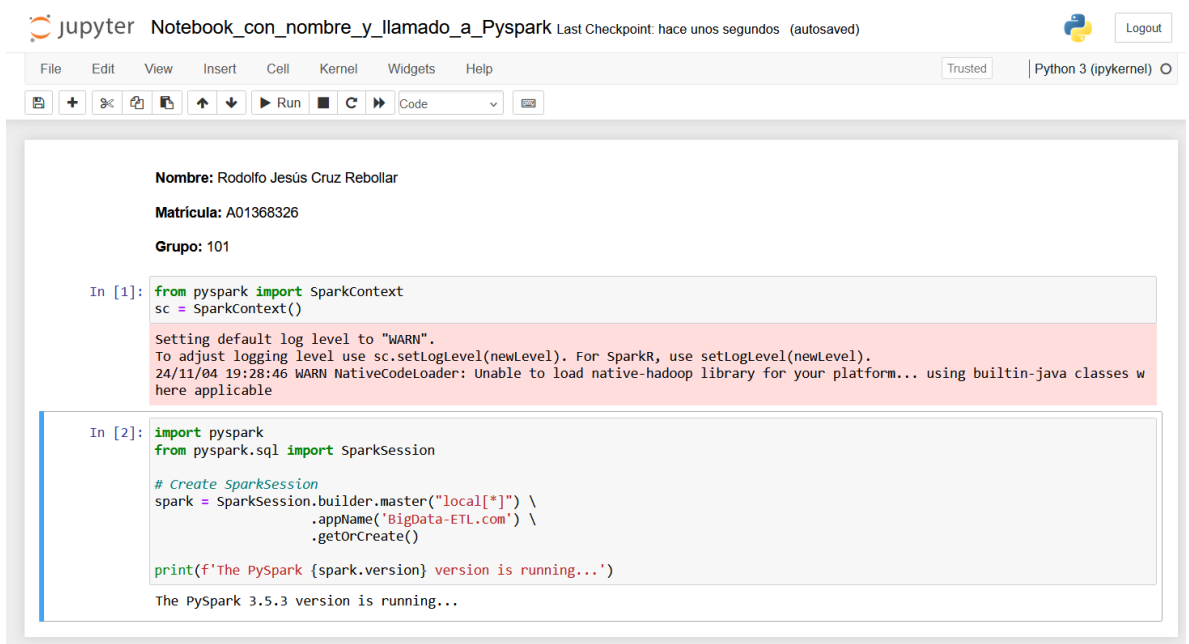
#### 4. Jupyter Notebook con el listado de los notebooks:



The screenshot shows the Jupyter Notebook interface with the 'Files' tab selected. At the top, there are 'Quit' and 'Logout' buttons. Below the tabs, a message says 'Select items to perform actions on them.' with 'Upload', 'New', and a refresh icon. A table lists the files and notebooks in the current directory.

	Name	Last Modified	File size
<input type="checkbox"/>	/		
<input type="checkbox"/>	1.- spark.ipynb	hace 21 días	1.82 kB
<input type="checkbox"/>	10.- persistencia.ipynb	hace 20 días	5.92 kB
<input type="checkbox"/>	11.- spark_sql.ipynb	hace 20 días	24.8 kB
<input type="checkbox"/>	12.- spark_sql_fecha.ipynb	hace 20 días	13.1 kB
<input type="checkbox"/>	13.- spark_sql_agrupaciones.ipynb	hace 20 días	9.91 kB
<input type="checkbox"/>	14.- mlilib.ipynb	hace 20 días	18.2 kB
<input type="checkbox"/>	15.- regresion_lineal.ipynb	hace 20 días	17.1 kB
<input type="checkbox"/>	2.- intro_spark.ipynb	hace 20 días	4.81 kB
<input type="checkbox"/>	3.- expresiones_lambda.ipynb	hace 20 días	4.24 kB
<input type="checkbox"/>	4.- acciones.ipynb	hace 20 días	16.9 kB
<input type="checkbox"/>	5.- transformaciones.ipynb	hace 20 días	12.5 kB
<input type="checkbox"/>	6.- ejemplo.ipynb	hace 20 días	44.6 kB
<input type="checkbox"/>	7.- valores_nulos.ipynb	hace 20 días	10.1 kB
<input type="checkbox"/>	8.- pair_rdd.ipynb	hace 20 días	26.7 kB
<input type="checkbox"/>	9.- particionado.ipynb	hace 20 días	8.94 kB
<input type="checkbox"/>	AAPL.csv	hace 20 días	192 kB

#### 5. Notebook con nombre y con el llamado a Pyspark:



The screenshot shows a Jupyter Notebook titled 'Notebook\_con\_nombre\_y\_llamado\_a\_Pyspark'. The top bar indicates 'Last Checkpoint: hace unos segundos (autosaved)' and a 'Logout' button. The notebook has a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar. The code is as follows:

```
Nombre: Rodolfo Jesús Cruz Rebollar
Matrícula: A01368326
Grupo: 101

In [1]: from pyspark import SparkContext
sc = SparkContext()

Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/11/04 19:28:46 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

In [2]: import pyspark
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...')

The PySpark 3.5.3 version is running...
```