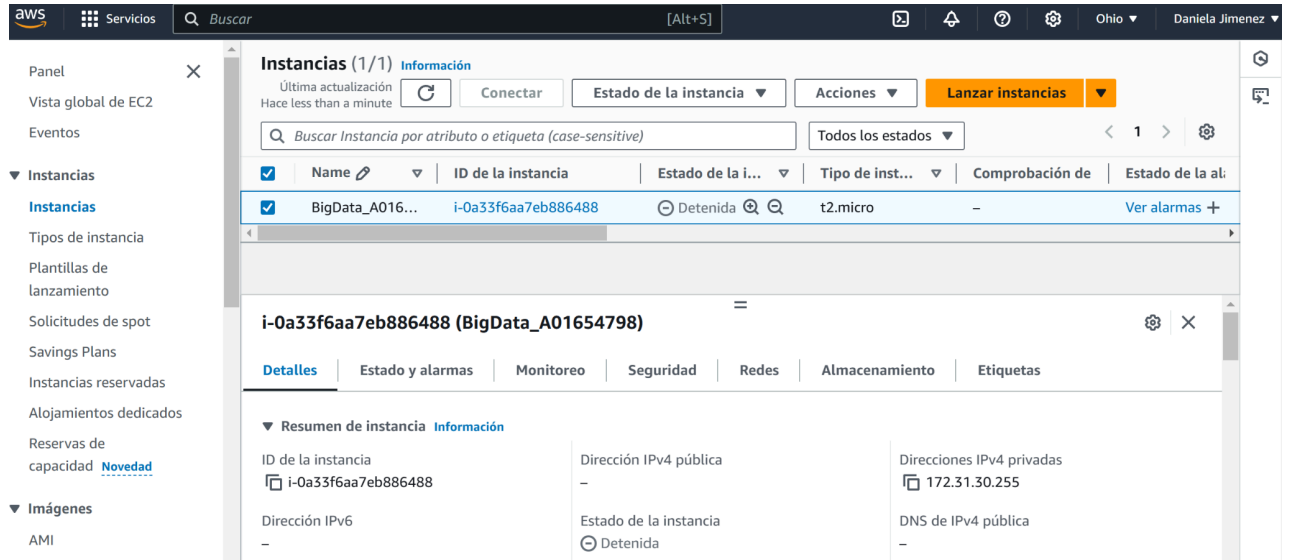
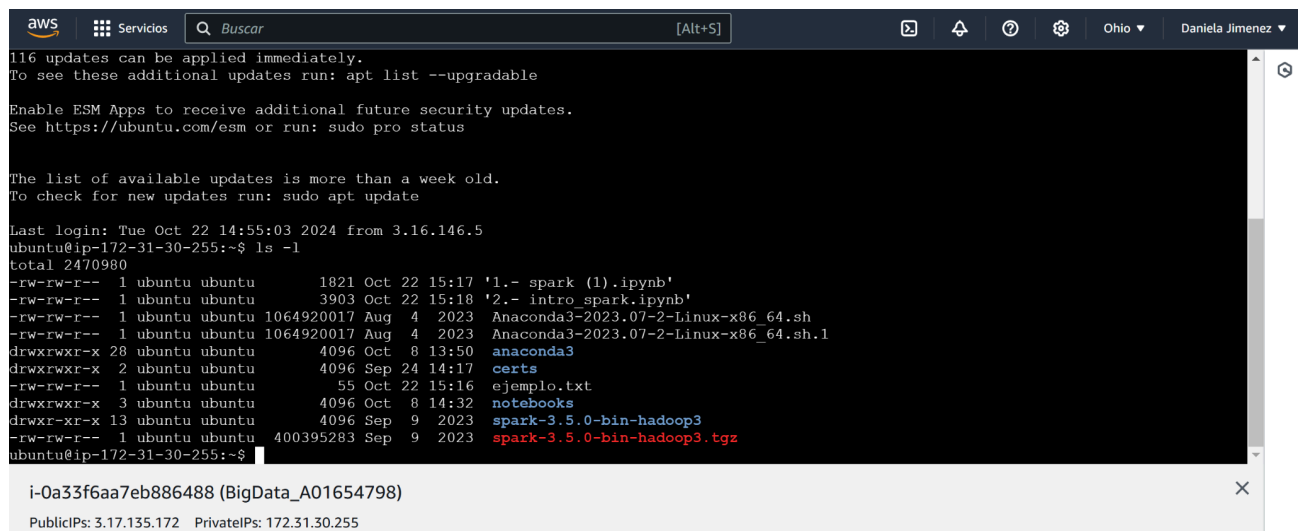


I. Instalación de Spark en AWS

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).

EC2 > Instancias > i-0a33f6aa7eb886488


Resumen de instancia de i-0a33f6aa7eb886488 (BigData_A01654798) [Información](#)

 [Conectar](#) [Estado de la instancia ▼](#) [Acciones ▼](#)

Se ha actualizado hace less than a minute


ID de la instancia  i-0a33f6aa7eb886488	Dirección IPv4 pública  3.17.135.172 dirección abierta 	Direcciones IPv4 privadas  172.31.30.255
Dirección IPv6 –	Estado de la instancia  En ejecución	DNS de IPv4 pública  ec2-3-17-135-172.us-east-2.compute.amazonaws.com dirección abierta 
Tipo de nombre de anfitrión Nombre de IP: ip-172-31-30-255.us-east-2.compute.internal	Nombre DNS de IP privada (solo IPv4)  ip-172-31-30-255.us-east-2.compute.internal	Direcciones IP elásticas –
Responder al nombre DNS de recurso privado IPv4 (A)	Tipo de instancia t2.micro	Hallazgo de AWS Compute Optimizer  Suscribirse a AWS Compute Optimizer par
Dirección IP asignada automáticamente  3.17.135.172 [IP pública]	ID de VPC  vpc-07a17fc7fc98bdf9 	

4. Impresión de pantalla de Jupyter Notebook visualizando el listado de los notebooks que se proporcionaron como ejemplos.

 jupyter

File View Settings Help

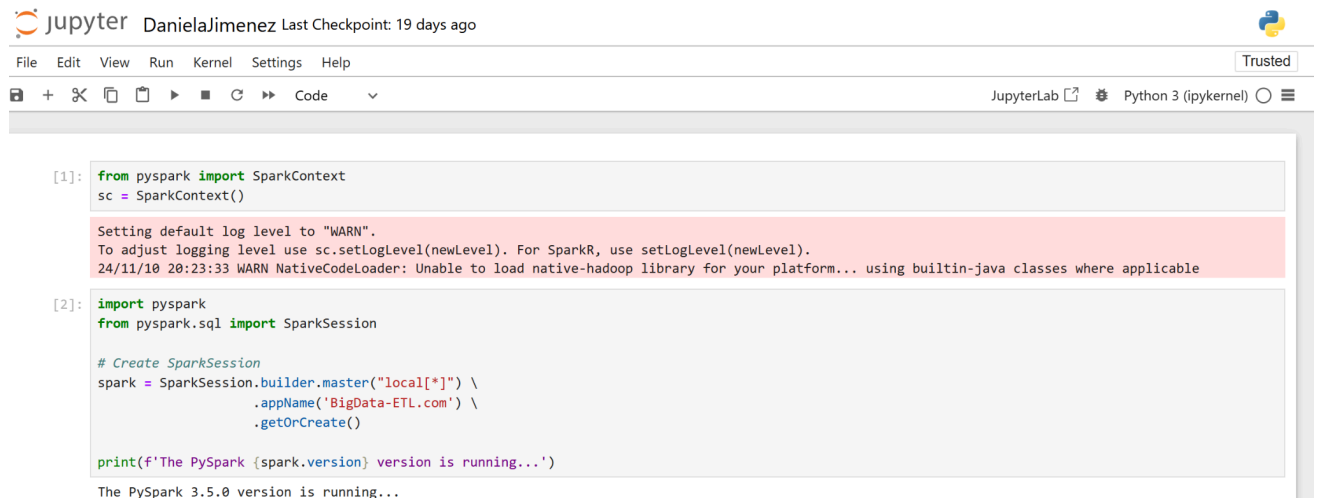
Files **Running**

Select items to perform actions on them. [New](#) [Upload](#) 

/

<input type="checkbox"/> Name	Last Modified	File Size
<input type="checkbox"/> anaconda3	last month	
<input type="checkbox"/> certs	last month	
<input type="checkbox"/> myenv	3 minutes ago	
<input type="checkbox"/> notebooks	last month	
<input type="checkbox"/> spark-3.5.0-bin-hadoop3	last year	
<input type="checkbox"/> 1.- spark (1).ipynb	19 days ago	1.8 KB
<input type="checkbox"/> 2.- intro_spark.ipynb	19 days ago	3.8 KB
<input type="checkbox"/> Anaconda3-2023.07-2-Linux-x86_64.sh	last year	1015.6 MB
<input type="checkbox"/> Anaconda3-2023.07-2-Linux-x86_64.sh.1	last year	1015.6 MB
<input type="checkbox"/> ejemplo.txt	19 days ago	55 B
<input type="checkbox"/> spark-3.5.0-bin-hadoop3.tgz	last year	381.8 MB

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



The image shows a JupyterLab interface. At the top, the Jupyter logo is followed by the username 'DanielaJimenez' and the text 'Last Checkpoint: 19 days ago'. On the right, there is a Python logo and a 'Trusted' badge. Below the header is a menu bar with 'File', 'Edit', 'View', 'Run', 'Kernel', 'Settings', and 'Help'. A toolbar contains icons for saving, opening, and running code. The main area displays a notebook with two code cells. The first cell contains code to create a SparkContext, and the second cell contains code to create a SparkSession and print its version. The output of the first cell shows a warning about the native-hadoop library. The output of the second cell shows the PySpark version 3.5.0.

```
[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).
24/11/10 20:23:33 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

[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 3.5.0 version is running...
```