



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Actividad. Instalación de Spark en AWS

**TC3007C.501 Inteligencia artificial avanzada para la ciencia de
datos II**

Profesores:

Iván Mauricio Amaya Contreras

Blanca Rosa Ruiz Hernández

Félix Ricardo Botello Urrutia

Edgar Covantes Osuna

Felipe Castillo Rendón

Hugo Terashima Marín

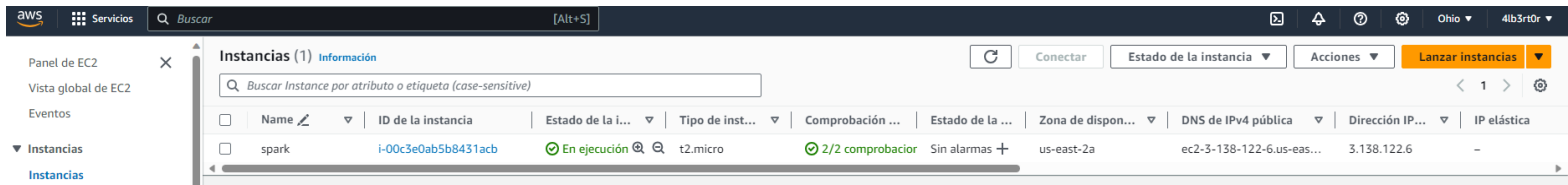
Alumno:

Alberto H Orozco Ramos – A00831719

30 de Octubre de 2023

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.

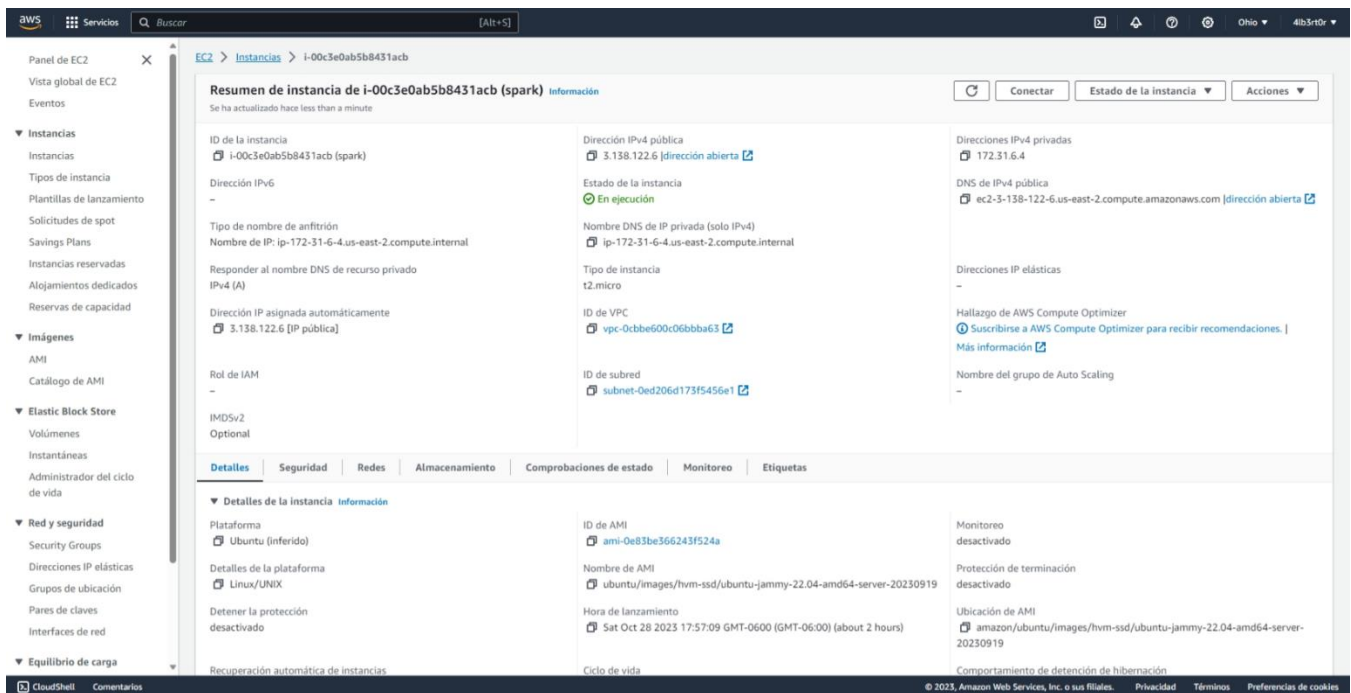
1. **Impresión de pantalla del listado de instancias de EC2 de AWS en donde se muestre la instancia creada.**



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.**

```
ubuntu@ip-172-31-6-4: ~  
(base) ubuntu@ip-172-31-6-4:~$ ls -l  
total 1517404  
-rw-rw-r-- 1 ubuntu ubuntu 1153404010 Sep 29 15:47 Anaconda3-2023.09-0-Linux-x86_64.sh  
drwxrwxr-x 32 ubuntu ubuntu 4096 Oct 29 00:35 anaconda3  
drwxrwxr-x 2 ubuntu ubuntu 4096 Oct 29 00:57 certs  
drwxr-xr-x 13 ubuntu ubuntu 4096 Sep 9 02:08 spark-3.5.0-bin-hadoop3  
-rw-rw-r-- 1 ubuntu ubuntu 400395283 Sep 9 02:10 spark-3.5.0-bin-hadoop3.tgz  
(base) ubuntu@ip-172-31-6-4:~$
```

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 la terminal o putty una vez que se ejecuta el comando *jupyter notebook*.

```

ubuntu@ip-172-31-6-4: ~
(base) ubuntu@ip-172-31-6-4:~$ jupyter notebook


JupyterLab
Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions.
https://jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html
Please note that updating to Notebook 7 might break some of your extensions.

[W 01:03:14.075 NotebookApp] WARNING: The notebook server is listening on all IP addresses and not using encryption. This is not recommended.
[W 01:03:14.346 NotebookApp] Loading JupyterLab as a classic notebook (v6) extension.
[W 2023-10-29 01:03:14.349 LabApp] 'ip' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2023-10-29 01:03:14.349 LabApp] 'ip' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2023-10-29 01:03:14.349 LabApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2023-10-29 01:03:14.349 LabApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[I 2023-10-29 01:03:14.352 LabApp] JupyterLab extension loaded from /home/ubuntu/anaconda3/lib/python3.11/site-packages/jupyterlab
[I 2023-10-29 01:03:14.353 LabApp] JupyterLab application directory is /home/ubuntu/anaconda3/share/jupyter/lab
[I 01:03:15.699 NotebookApp] Serving notebooks from local directory: /home/ubuntu
[I 01:03:15.699 NotebookApp] Jupyter Notebook 6.5.4 is running at:
[I 01:03:15.700 NotebookApp] http://ip-172-31-6-4:8888/?token=4c7870a222ec73b006falbc0231b71aa5b2f52a356a3abc0
[I 01:03:15.700 NotebookApp] or http://127.0.0.1:8888/?token=4c7870a222ec73b006falbc0231b71aa5b2f52a356a3abc0
[I 01:03:15.700 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 01:03:15.702 NotebookApp]



To access the notebook, open this file in a browser:
file:///home/ubuntu/.local/share/jupyter/runtime/nbserver-5956-open.html
Or copy and paste one of these URLs:
http://ip-172-31-6-4:8888/?token=4c7870a222ec73b006falbc0231b71aa5b2f52a356a3abc0
or http://127.0.0.1:8888/?token=4c7870a222ec73b006falbc0231b71aa5b2f52a356a3abc0



```
























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

 Quit Logout

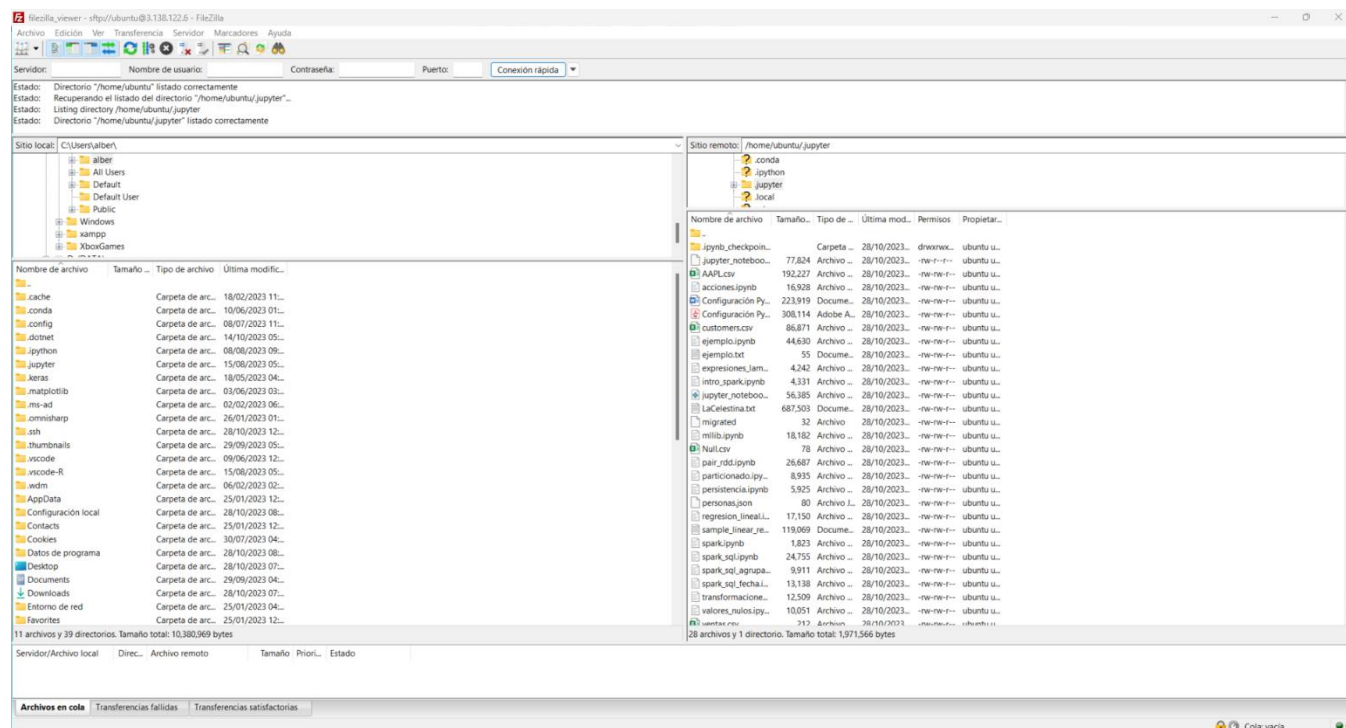
Files Running Clusters

Duplicate Rename Move Download View Edit  Upload New 

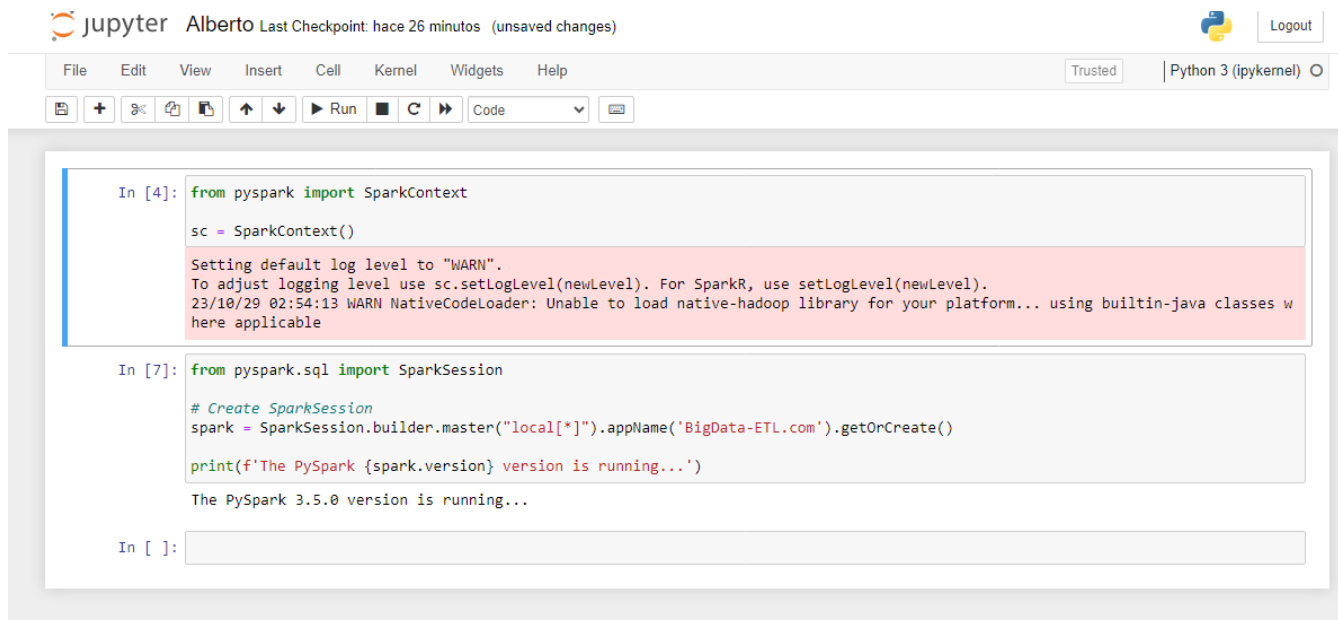
 1  /

	Name	Last Modified	File size
<input checked="" type="checkbox"/>	 acciones.ipynb	hace unos segundos	16.9 kB
<input type="checkbox"/>	 ejemplo.ipynb	hace unos segundos	44.6 kB
<input type="checkbox"/>	 expresiones_lambda.ipynb	hace unos segundos	4.24 kB
<input type="checkbox"/>	 intro_spark.ipynb	hace unos segundos	4.33 kB
<input type="checkbox"/>	 mlilb.ipynb	hace unos segundos	18.2 kB
<input type="checkbox"/>	 pair_rdd.ipynb	hace unos segundos	26.7 kB
<input type="checkbox"/>	 particionado.ipynb	hace unos segundos	8.94 kB
<input type="checkbox"/>	 persistencia.ipynb	hace unos segundos	5.92 kB
<input type="checkbox"/>	 regresion_lineal.ipynb	hace unos segundos	17.1 kB
<input type="checkbox"/>	 spark.ipynb	hace unos segundos	1.82 kB
<input type="checkbox"/>	 spark_sql.ipynb	hace unos segundos	24.8 kB
<input type="checkbox"/>	 spark_sql_agrupaciones.ipynb	hace unos segundos	9.91 kB
<input type="checkbox"/>	 spark_sql_fecha.ipynb	hace unos segundos	13.1 kB
<input type="checkbox"/>	 transformaciones.ipynb	hace unos segundos	12.5 kB
<input type="checkbox"/>	 valores_nulos.ipynb	hace unos segundos	10.1 kB
<input type="checkbox"/>	 AAPL.csv	hace unos segundos	192 kB
<input type="checkbox"/>	 Configuración Python con Spark.docx	hace unos segundos	224 kB
<input type="checkbox"/>	 Configuración Python con Spark.pdf	hace unos segundos	308 kB
<input type="checkbox"/>	 customers.csv	hace unos segundos	86.9 kB
<input type="checkbox"/>	 ejemplo.txt	hace unos segundos	55 B
<input type="checkbox"/>	 jupyter_notebook_config.py	hace 20 minutos	56.4 kB
<input type="checkbox"/>	 LaCelestina.txt	hace unos segundos	688 kB
<input type="checkbox"/>	 migrated	hace una hora	17 B

6. Impresión de pantalla de la conexión abierta al servidor utilizando Cyberduck o Filezilla (ver listado de archivos).



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



Todas las impresiones de pantalla se entregarán en un zip o en un documento de Word