



**Tecnológico
de Monterrey**

Instituto Tecnológico y de Estudios Superiores de Monterrey

AWS E2C y Sparks

Diego Alberto Baños Lopez

A01275100

Monterrey, Nuevo León, México 29 de Octubre del 2023

Inteligencia artificial avanzada para la ciencia de datos II (Gpo 501)

Panel de EC2

Vista global de EC2

Eventos

Instancias

Tipos de Instancia

Plantillas de lanzamiento

Solicitudes de spot

Savings Plans

Instancias reservadas

Alojamientos dedicados

Reservas de capacidad

Imágenes

AMI

Catálogo de AMI

Elastic Block Store

Volúmenes

Instantáneas

Administrador del ciclo de vida

Red y seguridad

Security Groups

Direcciones IP elásticas

Grupos de ubicación

Pares de claves

Interfases de red

Equilibrio de carga

Instancias (1) Información

Buscar instancia por atributo o etiqueta (case-sensitive)

Conectar

Estado de la instancia

Acciones

Lanzar instancias

<input type="checkbox"/>	Name	ID de la instancia	Estado de la instancia	Tipo de inst...	Comprobación ...	Estado de la ...	Zona de dispon...	DNS de IPv4 pública
<input type="checkbox"/>	escuelaInstancia	i-0e4f12370ff71a104	En ejecución	t2.micro	2/2 comprobador	Sin alarmas	us-east-2a	ec2-5-137-210-35-us-east-2.compute.amazonaws.com

Seleccione una instancia

CloudShell Comentarios

© 2023, Amazon Web Services, Inc. o sus filiales. Privacidad Términos Preferencias de cookies

```
(ubuntu) ec2-3-137-210-35-us-east-2.compute.amazonaws.com — Console

Nueva pestaña  Dividir vista  Copiar  Pegar  Buscar  ☰

* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/advantage

System information as of Mon Oct 30 04:42:50 UTC 2023

System load: 0.0      Processes:      95
Usage of /:  6.0% of 25.99GB   Users logged in: 0
Memory usage: 21%      IPv4 address for eth0: 172.31.1.244
Swap usage:  0%

* Ubuntu Pro delivers the most comprehensive open source security and
  compliance features.
  https://ubuntu.com/aws/pro

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

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-172-31-1-244:~$ ls -l
total 0
ubuntu@ip-172-31-1-244:~$
```

```

root@ubuntu[ip-172-31-1-244:~]# jupyter notebook
W 05:15:24.808 NotebookApp] Writing notebook server cookie secret to /home/ubuntu/.local/share/jupyter/runtime/notebook_cookie_secret

[W 05:15:24.808]
Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions.

read //jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html

Please note that updating to Notebook 7 might break some of your extensions.

W 05:15:24.589 NotebookApp] WARNING: The notebook server is listening on all IP addresses and not using encryption. This is not recommended.

W 05:15:25.436 NotebookApp] Loading JupyterLab as a classic notebook (vix) extension.

W 05:15:26.340 NotebookApp] '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 05:15:26.340 NotebookApp] '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 05:15:26.340 NotebookApp] '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 05:15:26.340 NotebookApp] '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 05:15:26.340 NotebookApp] JupyterLab extension loaded from /home/ubuntu/anaconda3/lib/python3.11/site-packages/jupyterlab

W 05:15:26.340 NotebookApp] JupyterLab application directory is /home/ubuntu/anaconda3/share/jupyterlab

W 05:15:27.089 NotebookApp] Serving notebooks from local directory: /home/ubuntu

W 05:15:27.089 NotebookApp] Jupyter Notebook 6.5.4 is running at:

W 05:15:27.089 NotebookApp] http://ip-172-31-1-244:8888/?token=ddc659abf85fd8c4ff08c23f36f9afb3c982a0b186ff78

W 05:15:27.089 NotebookApp] or http://127.0.0.1:8888/?token=ddc659abf85fd8c4ff08c23f36f9afb3c982a0b186ff78

W 05:15:27.089 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).

W 05:15:27.812 NotebookApp]

```

No seguro

ec2-3-137-210-35-us-east-2.compute.amazonaws.com:8888/tree/ejemplos_notebooks

UPDATERead 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

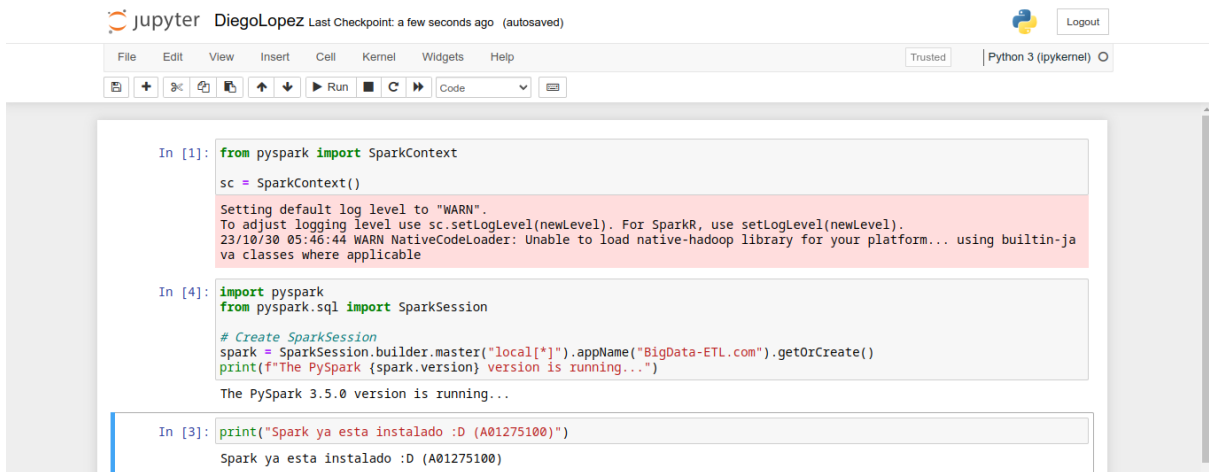
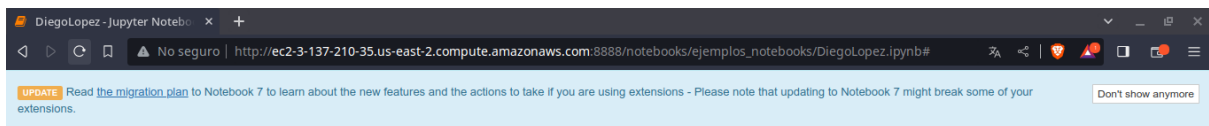
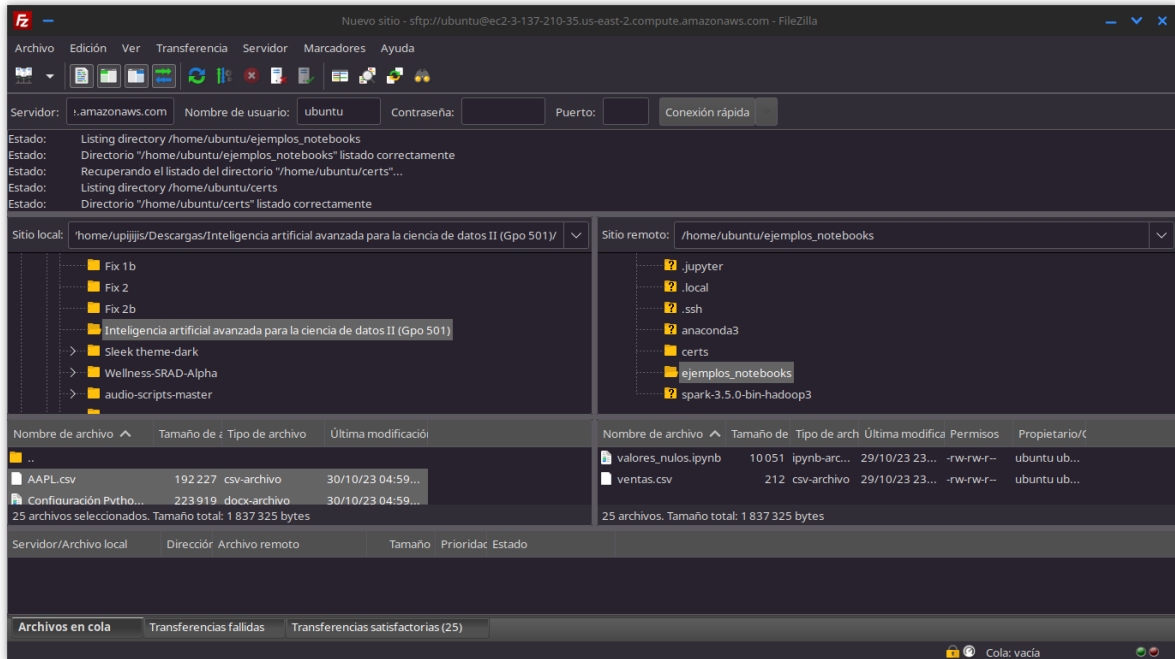
QuitLogout

FilesRunningClusters

Select items to perform actions on them.

UploadNew

Name	Last Modified	File size
/		
hace unos segundos		
acciones.ipynb	hace 10 minutos	16.9 kB
DiegoLopez.ipynb	hace 4 minutos	610 B
ejemplo.ipynb	hace 10 minutos	44.6 kB
expresiones_lambda.ipynb	hace 10 minutos	4.24 kB
Intro_spark.ipynb	hace 10 minutos	4.33 kB
mltib.ipynb	hace 10 minutos	19.2 kB
pair_rdd.ipynb	hace 10 minutos	26.7 kB
particionado.ipynb	hace 10 minutos	8.94 kB
persistencia.ipynb	hace 10 minutos	5.92 kB
regression_lineal.ipynb	hace 10 minutos	17.1 kB
spark.ipynb	hace 10 minutos	1.82 kB
spark_sql.ipynb	hace 10 minutos	24.9 kB
spark_sql_agrupaciones.ipynb	hace 10 minutos	9.91 kB
spark_sql_techa.ipynb	hace 10 minutos	13.1 kB
transformaciones.ipynb	hace 10 minutos	12.5 kB
valores_nulos.ipynb	hace 10 minutos	10.1 kB
AAPL.csv	hace 10 minutos	192 kB
Configuración Python con Spark.docx	hace 10 minutos	224 kB
Configuración Python con Spark.pdf	hace 10 minutos	308 kB
customers.csv	hace 10 minutos	86.9 kB
ejemplo.txt	hace 10 minutos	55 B
LaCelestina.txt	hace 10 minutos	688 kB



```
In [4]: spark = SparkSession.builder.appName('DataFrame').getOrCreate()
```

```
23/11/25 20:19:59 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
```

```
In [5]: df = spark.read.csv('ventas.csv', inferSchema = True, header = True)
```

```
In [6]: df.show()
```

```
+-----+-----+-----+
|Empresa|Persona|Ventas|
+-----+-----+-----+
|GOOG|Carlos|200|
|GOOG|Juan|120|
|GOOG|Felipe|340|
|MSFT|Tina|600|
|MSFT|Andrea|124|
|MSFT|Carla|243|
|FB|Sara|870|
|FB|Ignacio|350|
|APPL|Miguel|250|
|APPL|Oscar|130|
|APPL|Jorge|750|
|APPL|Ivan|350|
+-----+-----+-----+
```

```
In [7]: df.groupBy('Empresa').sum().show()
```

```
[Stage 3:>                                     (0 + 1) / 1]
```

```
+-----+-----+
|Empresa|sum(Ventas)|
+-----+-----+
|APPL|1480|
|GOOG|660|
|FB|1220|
|MSFT|967|
+-----+-----+
```

```
In [8]: df.agg({'Ventas': 'max'}).show()
```

```
+-----+
|max(Ventas)|
+-----+
|870|
+-----+
```

```
In [9]: agrupado = df.groupBy('Empresa')
agrupado.agg({'Ventas': 'max'}).show()
```

```
+-----+-----+
|Empresa|max(Ventas)|
+-----+-----+
|APPL|750|
|GOOG|340|
|FB|870|
|MSFT|600|
+-----+-----+
```

```
In [10]: agrupado.agg({'Ventas': 'min'}).show()
```

```
+-----+-----+
|Empresa|min(Ventas)|
+-----+-----+
|APPL|130|
|GOOG|120|
|FB|350|
|MSFT|124|
+-----+-----+
```

```

In [11]: df_agrupado = agrupado.agg({'Ventas': 'max'})
df_agrupado_pandas = df_agrupado.toPandas()

In [15]: df_agrupado_pandas = df_agrupado_pandas.sort_values('max(Ventas)', ascending=False)
import matplotlib.pyplot as plt

plt.figure(figsize=(10,6))
plt.bar(df_agrupado_pandas['Empresa'], df_agrupado_pandas['max(Ventas)'])
plt.xlabel('Empresa')
plt.ylabel('Máximo de Ventas')
plt.title('Máximo de Ventas por Empresa')
plt.xticks(rotation=45) # Rotar las etiquetas si son muchas o largas
plt.show()

```

