

```
Big Data — ubuntu@ip-172-31-16-124: ~ — ssh -i Spark.pem ubuntu@18.218.221.213 — 106x36

System load: 0.07          Processes:           106
Usage of /:  40.7% of 28.02GB Users logged in:        0
Memory usage: 20%          IPv4 address for enX0: 172.31.16.124
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.

101 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

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

Last login: Tue Oct 15 13:51:45 2024 from 3.16.146.4
(base) ubuntu@ip-172-31-16-124:~$ ls -l
ls-l: command not found
(base) ubuntu@ip-172-31-16-124:~$ ls -l
total 1432092
[-rw-rw-r-- 1 ubuntu ubuntu 1064920017 Aug  4 2023 Anaconda3-2023.07-2-Linux-x86_64.sh
drwxrwxr-x 28 ubuntu ubuntu    4096 Oct  8 13:38 anaconda3
drwxrwxr-x 2 ubuntu ubuntu    4096 Sep 24 13:42 certs
drwxrwxr-x 3 ubuntu ubuntu    4096 Oct 15 14:12 notebooks
drwxr-xr-x 13 ubuntu ubuntu    4096 Sep  9 2023 spark-3.5.0-bin-hadoop3
-rw-rw-r-- 1 ubuntu ubuntu  400395283 Sep  9 2023 spark-3.5.0-bin-hadoop3.tgz
-rw-rw-r-- 1 ubuntu ubuntu    331463 Oct  8 14:11 wget-log
-rw-rw-r-- 1 ubuntu ubuntu    785336 Oct  8 14:30 wget-log.1
(base) ubuntu@ip-172-31-16-124:~$
```

i-0b3d5fc32b3acc946 (Spark)

Details

Status and alarms

Monitoring

Security


Networking

Storage

Tags

▼ Instance summary [Info](#)

Instance ID

 i-0b3d5fc32b3acc946

IPv6 address

–


Hostname type

IP name: ip-172-31-16-124.us-east-2.compute.internal

Answer private resource DNS name

IPv4 (A)


Auto-assigned IP address

 18.218.221.213 [Public IP]


IAM Role

–


Public IPv4 address

 18.218.221.213 [open address](#)

Instance state

 Running


Private IP DNS name (IPv4 only)

 ip-172-31-16-124.us-east-2.compute.internal


Instance type

t2.micro


VPC ID

 vpc-0682b01c215928179 [open address](#)


Subnet ID

 subnet-03e919060b9e0340b [open address](#)

Private IPv4 addresses

 172.31.16.124


Public IPv4 DNS

 ec2-18-218-221-213.us-east-2.compute.amazonaws.com | [open address](#)

Elastic IP addresses

–

AWS Compute Optimizer finding

 [Opt-in to AWS Compute Optimizer for recommendations.](#) | [Learn more](#)

Auto Scaling Group name

–

Chrome File Edit View History Bookmarks Profiles Tabs Window Help

18.218.221.213:8888/tree/notebooks

Not Secure

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

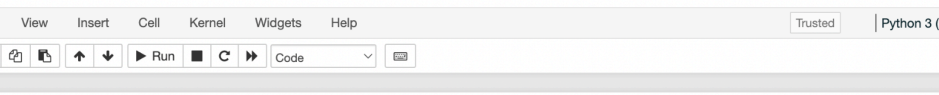
Quit Logout

Files Running Clusters

Select items to perform actions on them.

Upload New

Name	Last Modified	File size
..	seconds ago	
1.- spark.ipynb	a month ago	1.82 kB
10.- persistencia.ipynb	a month ago	5.92 kB
11.- spark_sql.ipynb	a month ago	24.8 kB
12.- spark_sql_fecha.ipynb	a month ago	13.1 kB
13.- spark_sql_agrupaciones.ipynb	a month ago	9.91 kB
14.- mlb.ipynb	a month ago	18.2 kB
15.- regresion_lineal.ipynb	a month ago	17.1 kB
2.- intro_spark.ipynb	a month ago	4.33 kB
3.- expresiones_lambda.ipynb	a month ago	4.24 kB
4.- acciones.ipynb	a month ago	16.9 kB
5.- transformaciones.ipynb	a month ago	12.5 kB
6.- ejemplo.ipynb	a month ago	44.6 kB
7.- valores_nulos.ipynb	a month ago	10.1 kB
8.- pair_rdd.ipynb	a month ago	26.7 kB
9.- particionado.ipynb	a month ago	8.94 kB
test.ipynb	a month ago	911 B
AAPL.csv	a month ago	192 kB
customers.csv	a month ago	86.9 kB
ejemplo.txt	a month ago	55 B
1.-Coleccion.txt	a month ago	658 kB



The screenshot shows a JupyterLab interface. At the top, the header bar includes the Jupyter logo, the username 'Oscar_Gtz', and the text 'Last Checkpoint: a minute ago (unsaved changes)'. Below the header is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. To the right of the menu bar are buttons for 'Trusted' and 'Python 3 (ipykernel)'. Below the menu bar is a toolbar with icons for saving, undo, redo, running, and other actions. The main area displays a code cell with the following code:

```
In [3]: 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...
```