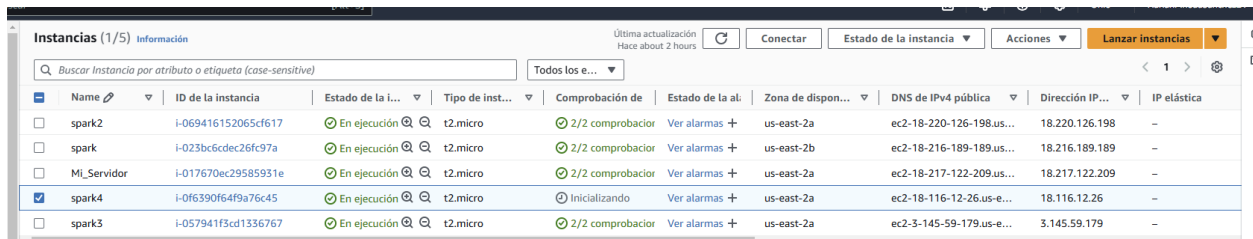


1.- Instalación de Spark en AWS

Adrian Pineda Sanchez

A00834710

1.- Impresión de pantalla del listado de instancias de EC2 de AWS en donde se muestre la instancia creada. (Spark2)



Instancias (1/5) Información										
Última actualización: Hace about 2 hours										
Conectar Estado de la instancia Acciones Lanzar instancias										
Buscar Instancia por atributo o etiqueta (case-sensitive) Todos los e...										
	Name	ID de la instancia	Estado de la i...	Tipo de inst...	Comprobación de	Estado de la al...	Zona de dispon...	DNS de IPv4 pública	Dirección IP...	IP elástica
<input type="checkbox"/>	spark2	i-069416152065c617	En ejecución	t2.micro	2/2 comprobador	Ver alarmas +	us-east-2a	ec2-18-220-126-198.us...	18.220.126.198	-
<input type="checkbox"/>	spark	i-023bc6cdec26fc97a	En ejecución	t2.micro	2/2 comprobador	Ver alarmas +	us-east-2b	ec2-18-216-189-189.us...	18.216.189.189	-
<input type="checkbox"/>	Mi_Servidor	i-017670ec29585931e	En ejecución	t2.micro	2/2 comprobador	Ver alarmas +	us-east-2a	ec2-18-217-122-209.us...	18.217.122.209	-
<input checked="" type="checkbox"/>	spark4	i-0f6390f64f9a76c45	En ejecución	t2.micro	Iniciando	Ver alarmas +	us-east-2a	ec2-18-116-12-26.us-e...	18.116.12.26	-
<input type="checkbox"/>	spark3	i-057941f3cd1336767	En ejecución	t2.micro	2/2 comprobador	Ver alarmas +	us-east-2a	ec2-3-145-59-179.us-e...	3.145.59.179	-


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.

```
[W 04:10:39.588 NotebookApp] Notebook 2.- intro_spark.ipynb is not trusted
[W 04:12:20.162 NotebookApp] SSL Error on 7 ('201.172.206.7', 60527): [SSL: SSLV3_ALERT_CERTIFICATE_EXPIRED]
[W 04:12:20.189 NotebookApp] SSL Error on 8 ('201.172.206.7', 60528): [SSL: SSLV3_ALERT_CERTIFICATE_EXPIRED]
^C[I 04:12:24.258 NotebookApp] interrupted
Serving notebooks from local directory: /home/ubuntu
0 active kernels
Jupyter Notebook 6.5.4 is running at:
https://ip-172-31-13-36:8888/?token=3161f536bacb264a77ed2bdf7e41003a428e29871886954
or https://127.0.0.1:8888/?token=3161f536bacb264a77ed2bdf7e41003a428e29871886954
Shutdown this notebook server (y/[n])? y
[C 04:12:26.428 NotebookApp] Shutdown confirmed
[I 04:12:26.429 NotebookApp] Shutting down 0 kernels
[I 04:12:26.429 NotebookApp] Shutting down 0 terminals
(base) ubuntu@ip-172-31-13-36:~$ ls -l
total 2472832
-rw-rw-r-- 1 ubuntu ubuntu      1823 Nov 11 04:10 '1.- spark.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     5925 Nov 11 04:10 '10.- persistencia.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    24755 Nov 11 04:10 '11.- spark_sql.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    13138 Nov 11 04:10 '12.- spark_sql fecha.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     9911 Nov 11 04:10 '13.- spark_sql agrupaciones.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    18182 Nov 11 04:10 '14.- mllib.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    17150 Nov 11 04:10 '15.- regresion lineal.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     4331 Nov 11 04:10 '2.- intro spark.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     4242 Nov 11 04:10 '3.- expresiones lambda.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    16928 Nov 11 04:10 '4.- acciones.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    12509 Nov 11 04:10 '5.- transformaciones.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     44630 Nov 11 04:10 '6.- ejemplo.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    10051 Nov 11 04:10 '7.- valores nulos.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     26687 Nov 11 04:10 '8.- pair rdd.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu     8935 Nov 11 04:10 '9.- particionado.ipynb'
-rw-rw-r-- 1 ubuntu ubuntu    192227 Nov 11 04:10 AAPL.csv
-rw-rw-r-- 1 ubuntu ubuntu     2008 Nov 11 03:34 'AdrianPineda .ipynb'
-rw-rw-r-- 1 ubuntu ubuntu 1064920017 Aug  4 2023 Anaconda3-2023.07-2-Linux-x86_64.sh
-rw-rw-r-- 1 ubuntu ubuntu 1064920017 Aug  4 2023 Anaconda3-2023.07-2-Linux-x86_64.sh.1
-rw-rw-r-- 1 ubuntu ubuntu    223179 Nov 11 04:10 'Configuración Python con Spark.docx'
-rw-rw-r-- 1 ubuntu ubuntu    308114 Nov 11 04:10 'Configuración Python con Spark.pdf'
-rw-rw-r-- 1 ubuntu ubuntu    687503 Nov 11 04:10 LaCelestina.txt
-rw-rw-r-- 1 ubuntu ubuntu       78 Nov 11 04:10 Null.csv
drwxrwxr-x 28 ubuntu ubuntu     4096 Nov 11 03:09 anaconda3
drwxrwxr-x 2 ubuntu ubuntu     4096 Nov 11 02:18 certs
-rw-rw-r-- 1 ubuntu ubuntu    86871 Nov 11 04:10 customers.csv
-rw-rw-r-- 1 ubuntu ubuntu       55 Nov 11 04:10 ejemplo.txt
-rw-rw-r-- 1 ubuntu ubuntu       80 Nov 11 04:10 personas.json
-rw-rw-r-- 1 ubuntu ubuntu   119069 Nov 11 04:10 sample_linear_regression_data.txt
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       212 Nov 11 04:10 ventas.csv
(base) ubuntu@ip-172-31-13-36:~$
```

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).

<div> <div>i-0f6390f64f9a76c45 (spark4)</div> <div> <div>▼ Resumen de instancia</div> <div>Información</div> </div> </div>		
ID de la instancia	Dirección IPv4 pública	Direcciones IPv4 privadas
<div> <div>i-0f6390f64f9a76c45</div> </div>	<div> <div>18.116.12.26</div> <div>dirección abierta</div> </div>	<div> <div>172.31.13.36</div> </div>
Dirección IPv6	Estado de la instancia	DNS de IPv4 pública
-	<div> <div>En ejecución</div> </div>	<div> <div>ec2-18-116-12-26.us-east-2.compute.amazonaws.com</div> <div>dirección abierta</div> </div>
Tipo de nombre de anfitrión	Nombre DNS de IP privada (solo IPv4)	Direcciones IP elásticas
Nombre de IP: ip-172-31-13-36.us-east-2.compute.internal	<div> <div>ip-172-31-13-36.us-east-2.compute.internal</div> </div>	
Responder al nombre DNS de recurso privado IPv4 (A)	Tipo de instancia	
	t2.micro	-
Dirección IP asignada automáticamente	ID de VPC	Hallazgo de AWS Compute Optimizer
<div> <div>18.116.12.26</div> <div>[IP pública]</div> </div>	<div> <div>vpc-07e45a092c8f0cc6e</div> </div>	<div> <div>Suscribirse a AWS Compute Optimizer para recibir recomendaciones.</div> <div>Más información</div> </div>

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



Quit

Logout

Select items to perform actions on them.

0

▼

/

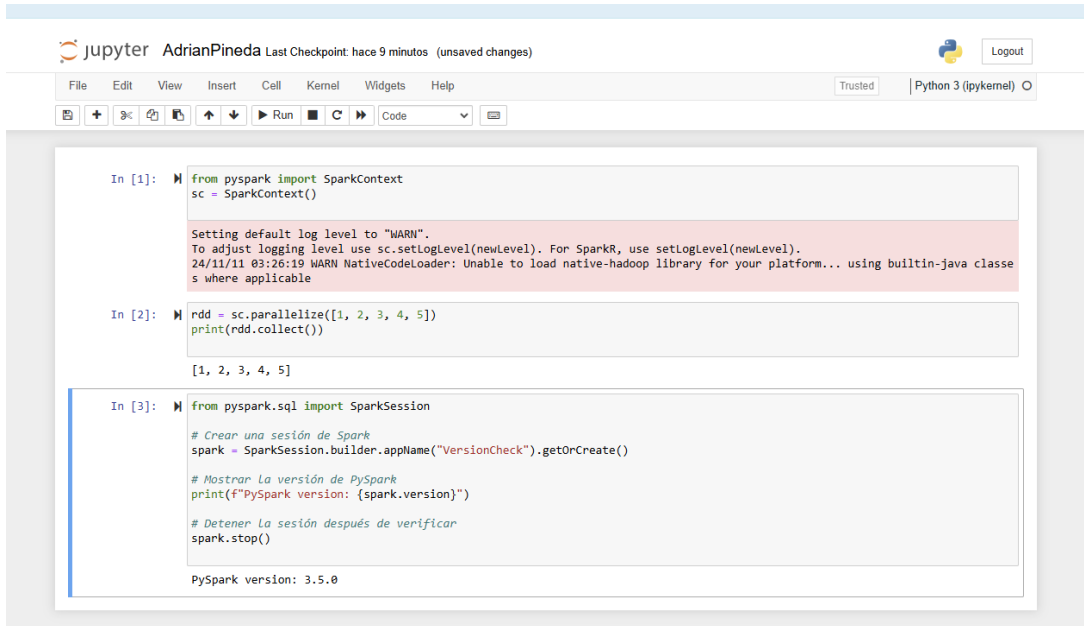
Name ▼

Last Modified

File size

<input type="checkbox"/>	anaconda3	hace una hora	
<input type="checkbox"/>	certs	hace 2 horas	
<input type="checkbox"/>	spark-3.5.0-bin-hadoop3	hace un año	
<input type="checkbox"/>	1.- spark.ipynb	hace un minuto	1.82 kB
<input type="checkbox"/>	10.- persistencia.ipynb	hace un minuto	5.92 kB
<input type="checkbox"/>	11.- spark_sql.ipynb	hace un minuto	24.8 kB
<input type="checkbox"/>	12.- spark_sql_fecha.ipynb	hace unos segundos	13.1 kB
<input type="checkbox"/>	13.- spark_sql_agrupaciones.ipynb	hace un minuto	9.01 kB
<input type="checkbox"/>	14.- mllib.ipynb	hace un minuto	18.2 kB
<input type="checkbox"/>	15.- regresion_lineal.ipynb	hace un minuto	17.1 kB
<input type="checkbox"/>	2.- intro_spark.ipynb	hace unos segundos	4.33 kB
<input type="checkbox"/>	3.- expresiones_lambda.ipynb	hace unos segundos	4.24 kB
<input type="checkbox"/>	4.- acciones.ipynb	hace unos segundos	16.9 kB
<input type="checkbox"/>	5.- transformaciones.ipynb	hace unos segundos	12.5 kB
<input type="checkbox"/>	6.- ejemplo.ipynb	hace unos segundos	44.6 kB
<input type="checkbox"/>	7.- valores_nulos.ipynb	hace unos segundos	10.1 kB
<input type="checkbox"/>	8.- pair_rdd.ipynb	hace unos segundos	26.7 kB
<input type="checkbox"/>	9.- particionado.ipynb	hace unos segundos	8.04 kB
<input type="checkbox"/>	AdrianPineda .ipynb	hace 36 minutos	2.01 kB
<input type="checkbox"/>	AAPL.csv	hace un minuto	102 kB
<input type="checkbox"/>	Anaconda3-2023.07-2-Linux-x86_64.sh	hace un año	1.06 GB
<input type="checkbox"/>	Anaconda3-2023.07-2-Linux-x86_64.sh.1	hace un año	1.06 GB
<input type="checkbox"/>	Configuración Python con Spark.docx	hace un minuto	223 kB
<input type="checkbox"/>	Configuración Python con Spark.pdf	hace un minuto	308 kB
<input type="checkbox"/>	customers.csv	hace un minuto	86.9 kB
<input type="checkbox"/>	ejemplo.txt	hace un minuto	55 B
<input type="checkbox"/>	LaCelestina.txt	hace un minuto	688 kB
<input type="checkbox"/>	Null.csv	hace un minuto	78 B
<input type="checkbox"/>	personas.json	hace un minuto	80 B
<input type="checkbox"/>	sample_linear_regression_data.txt	hace un minuto	119 kB
<input type="checkbox"/>	spark-3.5.0-bin-hadoop3.tgz	hace un año	400 MB
<input type="checkbox"/>	ventas.csv	hace un minuto	212 B

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



```
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).
24/11/11 03:26:19 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

In [2]: rdd = sc.parallelize([1, 2, 3, 4, 5])
print(rdd.collect())

[1, 2, 3, 4, 5]

In [3]: from pyspark.sql import SparkSession

# Crear una sesión de Spark
spark = SparkSession.builder.appName("VersionCheck").getOrCreate()

# Mostrar la versión de PySpark
print(f"PySpark version: {spark.version}")

# Detener la sesión después de verificar
spark.stop()

PySpark version: 3.5.0
```

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