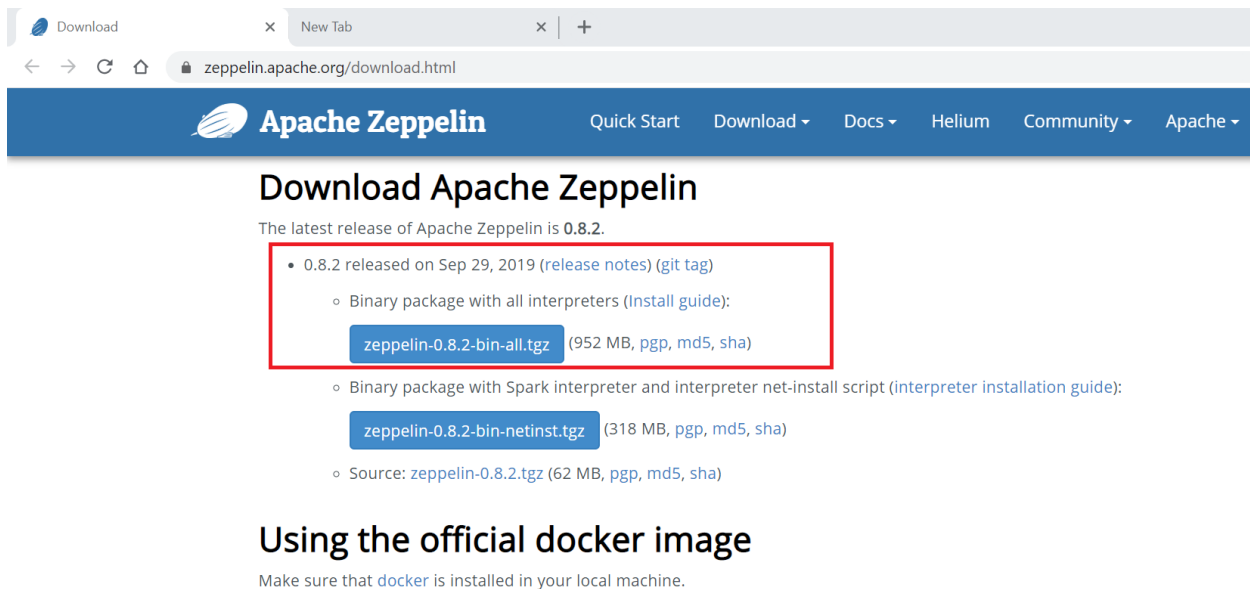


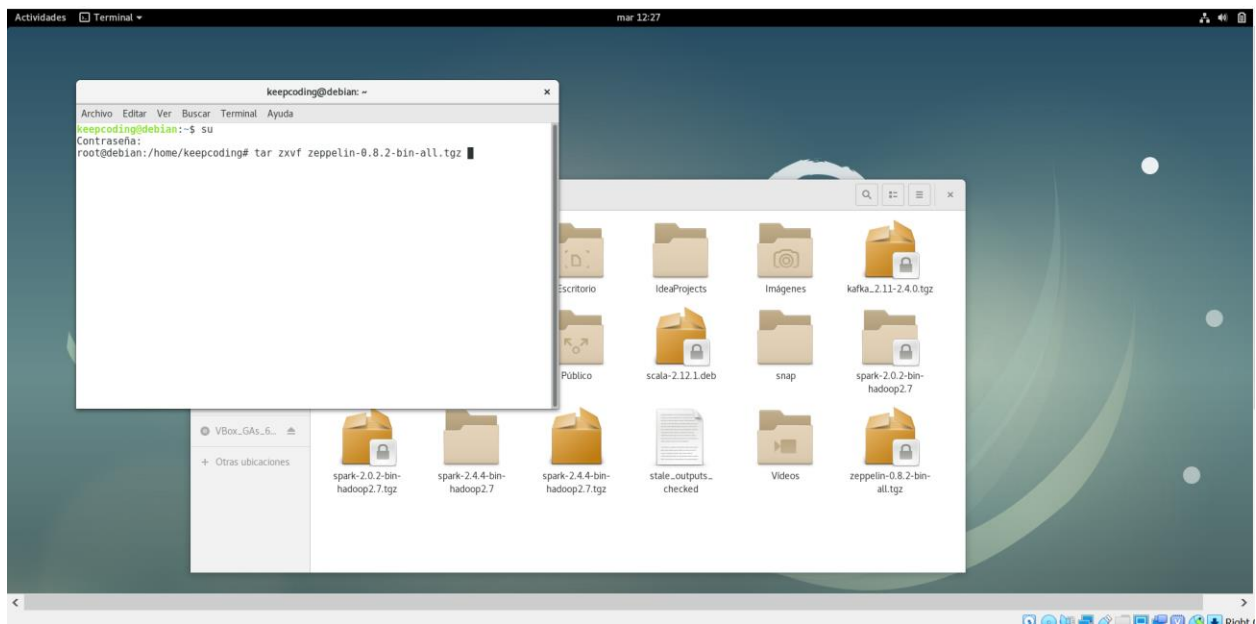
Scala y Spark en Zeppelin

1. Instalacion de Zeppelin

Entrar al navegador: <https://zeppelin.apache.org/download.html>



Una vez descargado, descomprimos en la terminal:

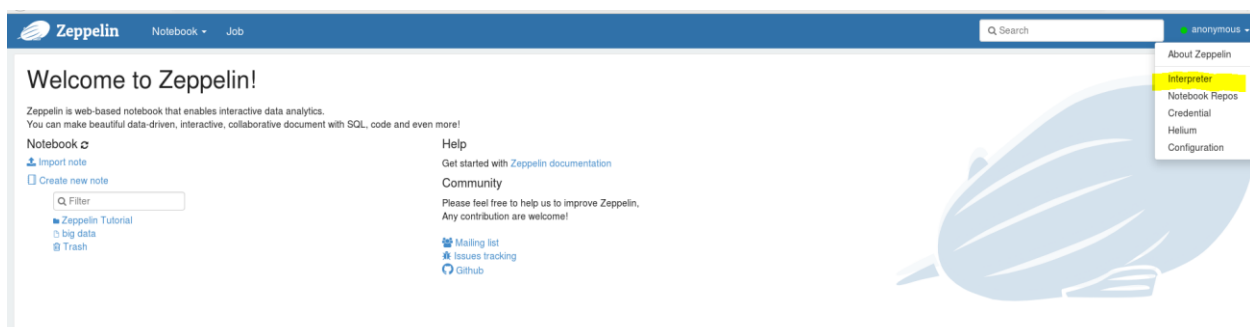


Y al finalizar creamos una carpeta en una ruta escogida por nosotros, copiamos el contenido y ejecutamos el comando para iniciar el servicio de zeppelin:

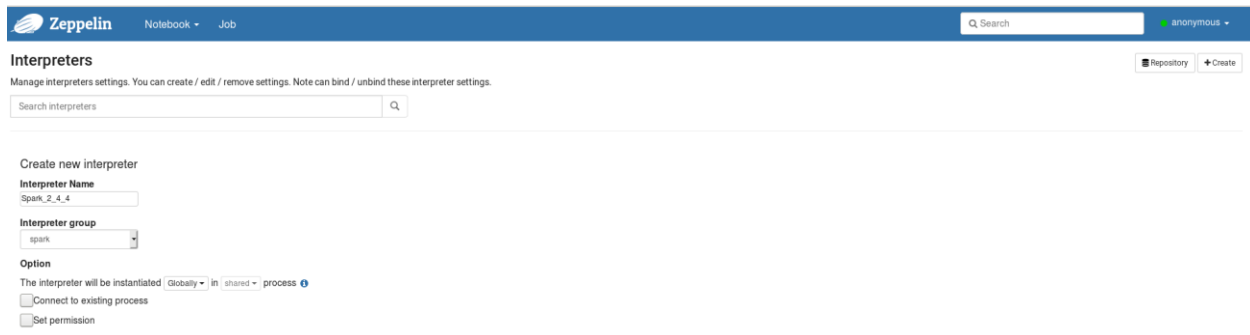
```
keepcoding@debian: ~
Archivo Editar Ver Buscar Terminal Ayuda
zeppelin-0.8.2-bin-all/interpreter/neo4j/log4j-1.2.17.jar
zeppelin-0.8.2-bin-all/interpreter/neo4j/jackson-databind-2.8.9.jar
zeppelin-0.8.2-bin-all/interpreter/neo4j/jackson-core-2.8.9.jar
zeppelin-0.8.2-bin-all/interpreter/neo4j/zeppelin-neo4j-0.8.2.jar
zeppelin-0.8.2-bin-all/interpreter/neo4j/slf4j-log4j12-1.7.10.jar
zeppelin-0.8.2-bin-all/interpreter/neo4j/neo4j-java-driver-1.4.3.jar
zeppelin-0.8.2-bin-all/interpreter/neo4j/slf4j-api-1.7.10.jar
root@debian:/home/keepcoding# mkdir /opt/zeppelin
root@debian:/home/keepcoding# cd /opt/zeppelin
root@debian:/opt/zeppelin# cp -r /home/keepcoding/zeppelin-0.8.2-bin-all/* .
root@debian:/opt/zeppelin# ls
bin interpreter LICENSE notebook README.md
conf lib licenses NOTICE zeppelin-web-0.8.2.war
root@debian:/opt/zeppelin# cd bin/
root@debian:/opt/zeppelin/bin# ls
common.cmd functions.sh interpreter.sh zeppelin-daemon.sh
common.sh install-interpreter.sh stop-interpreter.sh zeppelin.sh
functions.cmd interpreter.cmd zeppelin.cmd
root@debian:/opt/zeppelin/bin# ./zeppelin-daemon.sh start &
[1] 2787
root@debian:/opt/zeppelin/bin# /opt/zeppelin-0.8.2-bin-all/conf/zeppelin-env.sh: línea
113: export: `spark-shell': no es un identificador válido
/opt/zeppelin-0.8.2-bin-all/conf/zeppelin-env.sh: línea 113: export: `--jars': no es u
n identificador válido
/opt/zeppelin-0.8.2-bin-all/conf/zeppelin-env.sh: línea 113: export: `twitter4j-core-4
.0.6.jar,sparkstreaming-': no es un identificador válido
/opt/zeppelin-0.8.2-bin-all/conf/zeppelin-env.sh: línea 114: twitter_2.11-1.6.3.jar,co
mmons-dbutils-1.6.jar,spark-core 2.11-: no se encontró la orden
/opt/zeppelin-0.8.2-bin-all/conf/zeppelin-env.sh: línea 115: 1.5.2.jar,twitter4j-strea
m-4.0.6.jar: no se encontró la orden
Zeppelin start [ OK ]
root@debian:/opt/zeppelin/bin#
```

Una vez iniciado el servicio abrimos en el navegador la consola web de zeppelin usando la siguiente direccion: <http://localhost:8081/#/>

Dentro vamos a la opcion de interpreter:



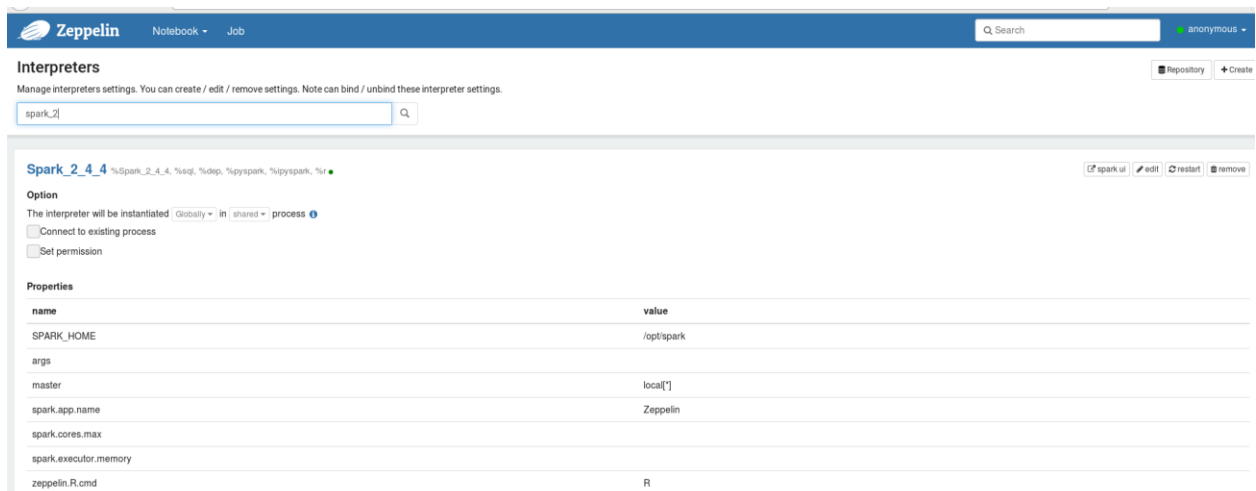
Agregamos uno nuevo, con nombre spark 2_4_4 dentro del grupo spark



The screenshot shows the Zeppelin 'Interpreters' management page. At the top, there's a search bar and a 'Repository' button. Below, a 'Search interpreters' input field is present. The main section is titled 'Create new interpreter'. It includes a text input for 'Interpreter Name' (containing 'Spark_2_4_4'), a dropdown for 'Interpreter group' (set to 'spark'), and an 'Option' section. The 'Option' section has a note: 'The interpreter will be instantiated Globally in shared process'. There are two checkboxes: 'Connect to existing process' (unchecked) and 'Set permission' (unchecked).

Al final del apartado properties agregar la propiedad SPARK_HOME y le asignamos la ruta donde esta instalado la version de spark que utilizaremos:

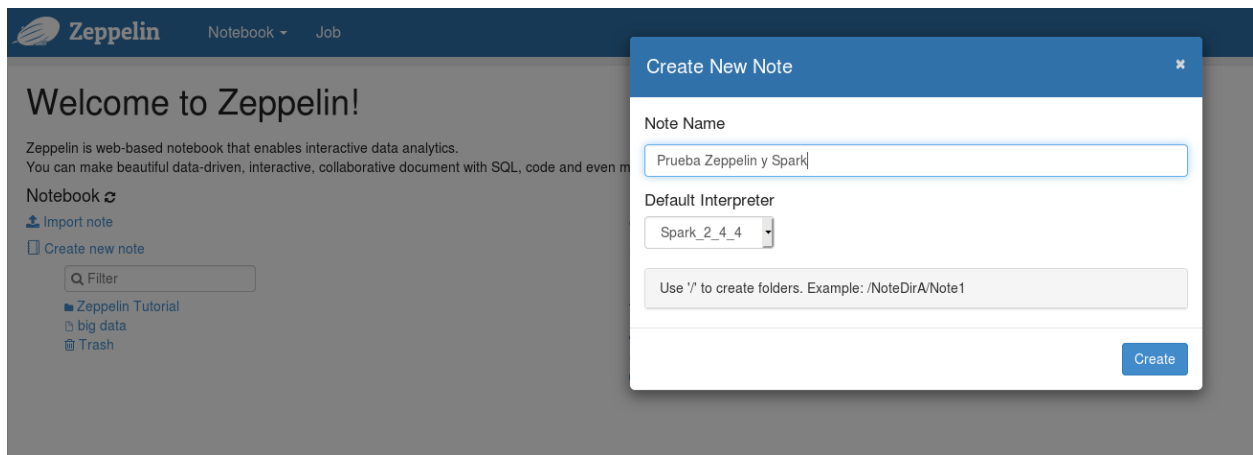
zeppelin.spark.concurrentSQL	<input type="checkbox"/>	x	Execute multiple SQL concurrently if set true.
zeppelin.spark.enableSupportedVersionCheck	<input checked="" type="checkbox"/>	x	Do not change - developer only setting, not for production use
zeppelin.spark.importImplicit	<input checked="" type="checkbox"/>	x	Import implicits, UDF collection, and sql if set true, true by default.
zeppelin.spark.maxResult	<input type="text" value="1000"/>	x	Max number of Spark SQL result to display.
zeppelin.spark.printREPLOutput	<input checked="" type="checkbox"/>	x	Print REPL output
zeppelin.spark.sql.interpolation	<input type="checkbox"/>	x	Enable ZeppelinContext variable interpolation into paragraph text
zeppelin.spark.sql.stacktrace	<input type="checkbox"/>	x	Show full exception stacktrace for SQL queries if set to true.
zeppelin.spark.ui.hidden	<input type="checkbox"/>	x	Whether to hide spark ui in zeppelin ui
zeppelin.spark.ui.WebUrl	<input type="text"/>	x	Override Spark UI default URL
zeppelin.spark.useHiveContext	<input checked="" type="checkbox"/>	x	Use HiveContext instead of SQLContext if it is true.
zeppelin.spark.useNew	<input checked="" type="checkbox"/>	x	Whether use new spark interpreter implementation
SPARK_HOME	<input type="text" value="/opt/spark"/>	textarea +	



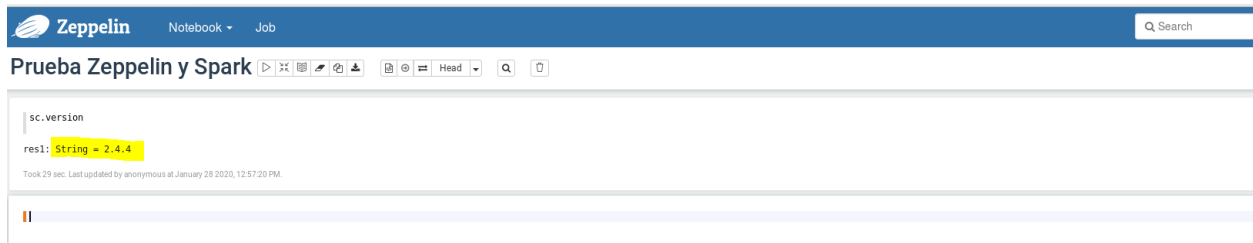
The screenshot shows the Zeppelin 'Interpreters' page with the 'spark_2' interpreter selected. The 'Spark_2_4_4' configuration is displayed. It includes the 'Option' section with the same 'Globally in shared process' note and unchecked checkboxes. Below is the 'Properties' section, which is a table with the following data:

name	value
SPARK_HOME	/opt/spark
args	
master	local[*]
spark_app.name	Zeppelin
spark_cores.max	
spark_executor.memory	
zeppelin.R.cmd	R

Y para verificar el resultado creamos una nueva nota usando el nuevo interprete:



Y ejecutamos la sentencia version



2. Ejercicio: Deberemos trabajar con el csv "amigos.csv" y de él Calcular el número de registros que tenemos:

