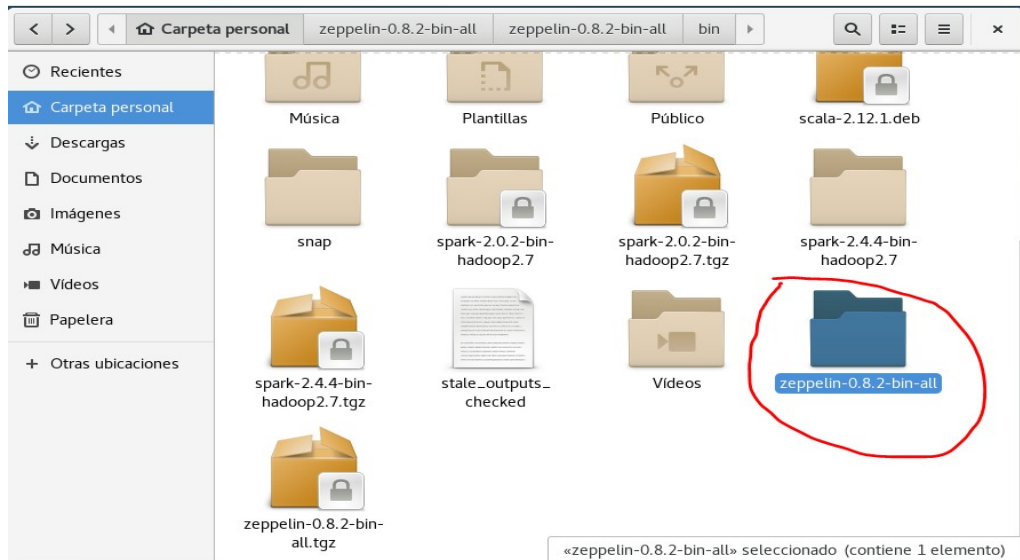


Pasos para configurar nuestro spark con el notebook de zeppelin

Descomprimos el zeppelin que tenemos en la maquina virtual.

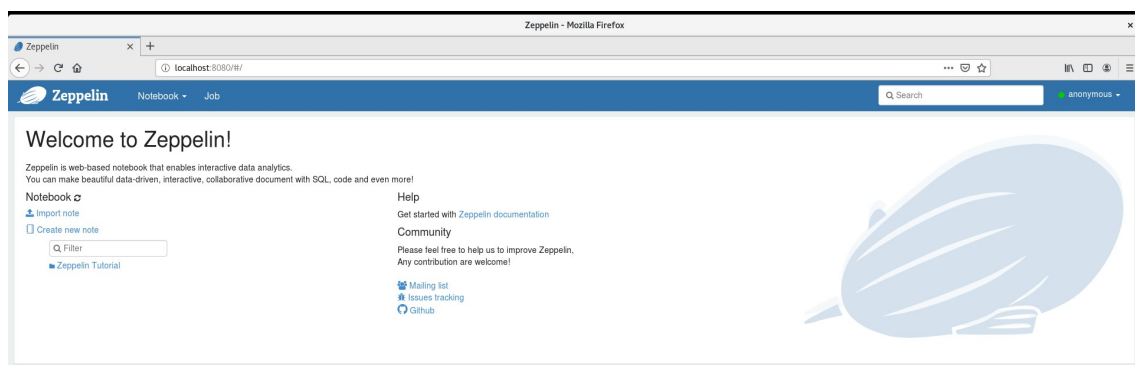


Una vez descomprimido vamos al terminal y desde el directorio donde zeppelin ejecutamos el siguiente comando: `bin/zeppelin-daemon.sh start`

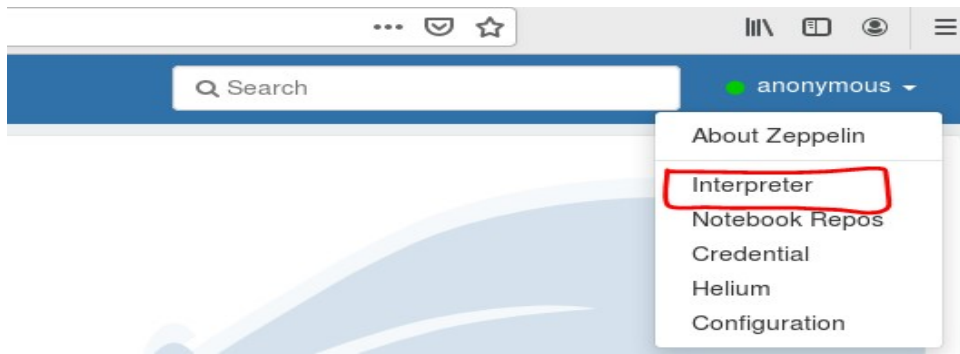
```
keepcoding@debian:~/zeppelin-0.8.2-bin-all/zeppelin-0.8.2-bin-all$ bin/zeppelin-daemon.sh start
Log dir doesn't exist, create /home/keepcoding/zeppelin-0.8.2-bin-all/zeppelin-0.8.2-bin-all/logs
Pid dir doesn't exist, create /home/keepcoding/zeppelin-0.8.2-bin-all/zeppelin-0.8.2-bin-all/run
Zeppelin start
keepcoding@debian:~/zeppelin-0.8.2-bin-all/zeppelin-0.8.2-bin-all$ [ OK ]
```

Una vez hecho esto vamos al navegador y en la barra de búsqueda ponemos localhost:8080.

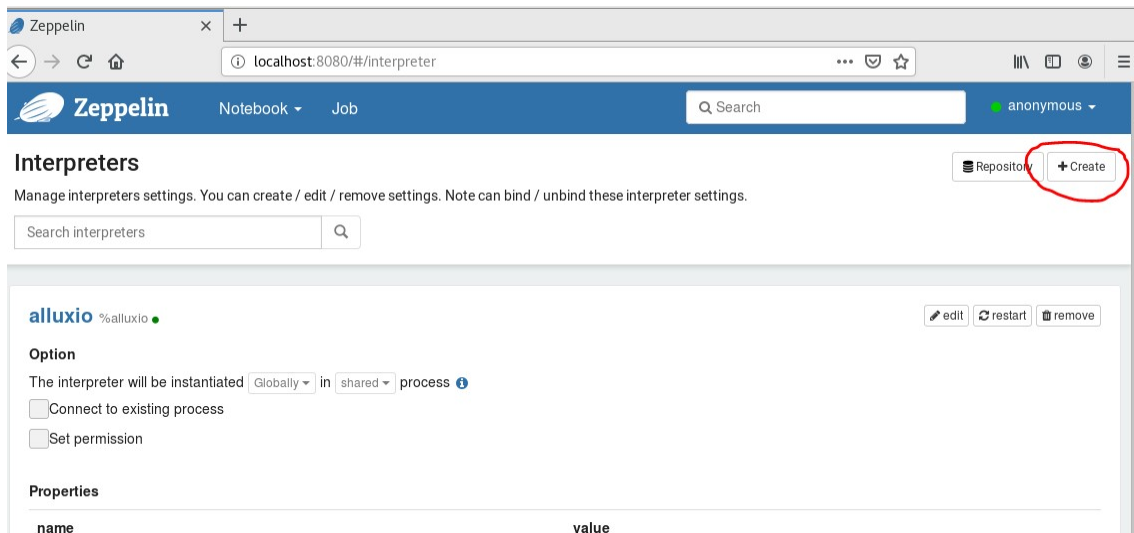
Ya podemos usar zeppelin, ahora vamos a configurarlo.



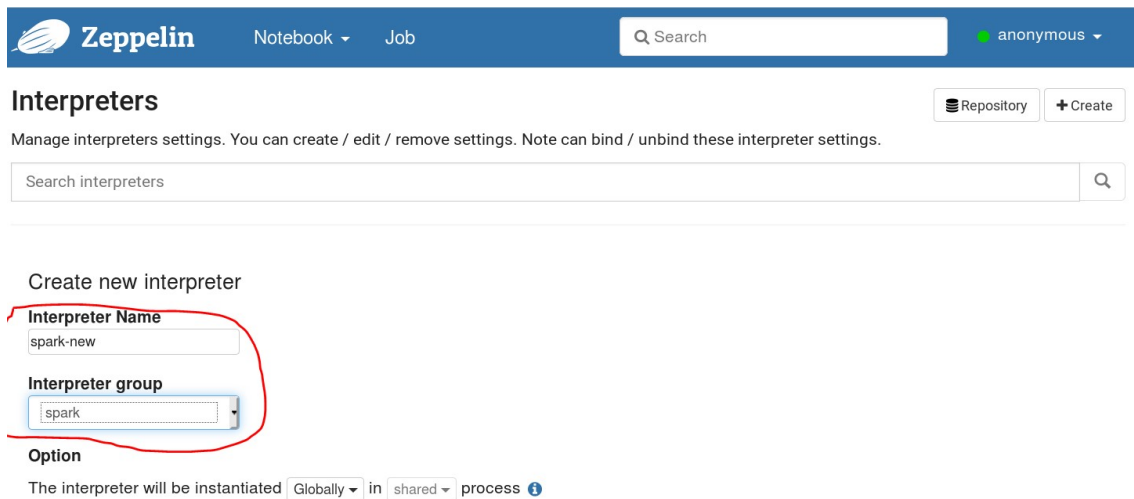
Vamos al desplegable de anonymous y pinchamos en interpreter.



Vamos a crear uno nuevo, para ello pulsamos créate



Le ponemos un nombre al intérprete que vamos a crear y lo metemos en el grupo spark.



En la parte de abajo antes de guardar tenemos que poner SPARK_HOME y la ruta a nuestra carpeta spark y una vez hecho pulsamos en Save y ya hemos creado el interprete.

The screenshot shows the Zeppelin interpreter configuration interface. At the bottom, the **SPARK_HOME** field is highlighted with a red circle and contains the value `/opt/spark/spark-2.4.4-bin-hadoop2.7`. Above this field is a list of configuration options with checkboxes and a 'maxResult' input field. Below the highlighted field is a 'Dependencies' section with input fields for 'artifact' and 'exclude', and an 'action' button. At the very bottom are 'Save' and 'Cancel' buttons.

Configuration Option	Value / State
zeppelin.pyspark.usePython	<input checked="" type="checkbox"/>
zeppelin.spark.concurrentSQL	<input type="checkbox"/>
zeppelin.spark.enableSupportedVersionCheck	<input checked="" type="checkbox"/>
zeppelin.spark.importImplicit	<input checked="" type="checkbox"/>
zeppelin.spark.maxResult	1000
zeppelin.spark.printREPLOutput	<input checked="" type="checkbox"/>
zeppelin.spark.sql.interpolation	<input type="checkbox"/>
zeppelin.spark.sql.stacktrace	<input type="checkbox"/>
zeppelin.spark.ui.hidden	<input type="checkbox"/>
zeppelin.spark.ui.WebUrl	
zeppelin.spark.useHiveContext	<input checked="" type="checkbox"/>
zeppelin.spark.useNew	<input checked="" type="checkbox"/>
SPARK_HOME	/opt/spark/spark-2.4.4-bin-hadoop2.7

Dependencies

artifact	exclude	action
<input type="text" value="groupid:artifactid:version or local file path"/>	<input type="text" value="(Optional) comma separated groupid:artifactid list"/>	<input type="button" value="+"/>

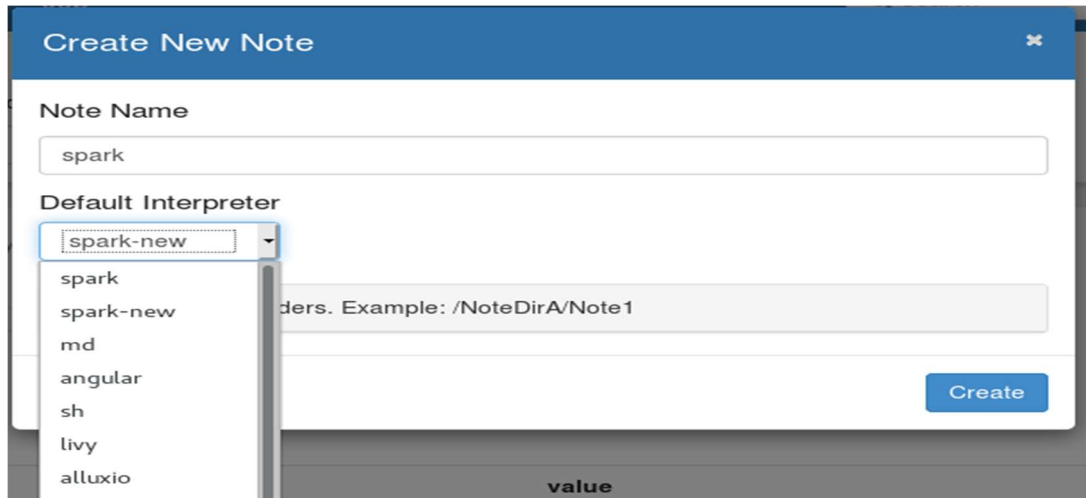
Si queremos ver que está bien creado en Iterpreters buscamos el que acabamos de crear y vemos que está bien.

The screenshot shows the Zeppelin 'Interpreters' page in a Mozilla Firefox browser. The page title is 'Zeppelin - Mozilla Firefox'. The URL is 'localhost:8080/#/interpreter'. The page has a search bar and a 'Repository' button. Below the search bar, there is a list of interpreters. The 'spark-new' interpreter is selected, and its configuration is shown. The 'Option' section indicates that the interpreter will be instantiated 'Globally' in a 'shared' process. The 'Properties' section shows a table with the following data:

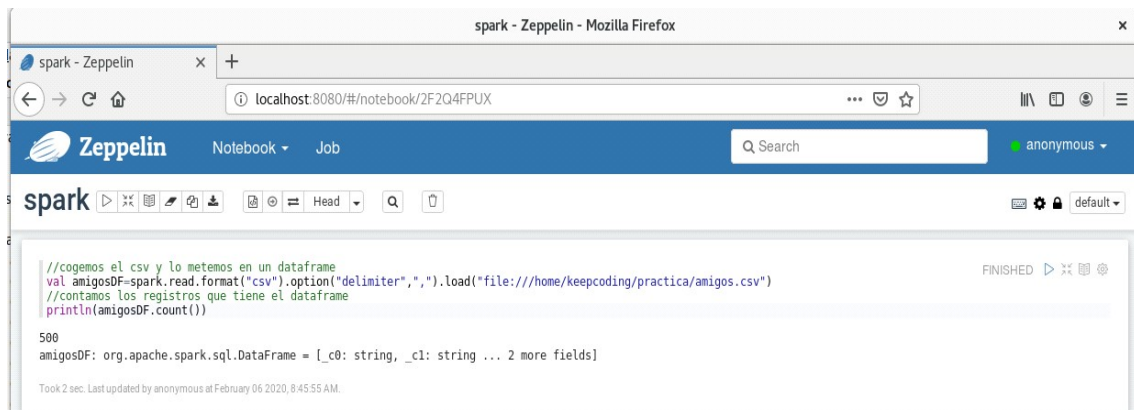
name	value
SPARK_HOME	/opt/spark/spark-2.4.4-bin-hadoop2.7
args	
master	local[*]
spark.app.name	Zeppelin
spark.cores.max	

At the top right of the 'spark-new' configuration, there are buttons for 'spark ui', 'edit', 'restart', and 'remove'.

Ahora vamos a crear un nuevo notebook para realizar la práctica, para ello pinchamos en el desplegable que pone notebook y pinchamos en créate new note, le damos un nombre y buscamos el interprete que acabamos de crear.



Una vez creado vamos a realizar la practica contando los registros de amigos.csv



como podemos ver ha salido todo correcto.