

Zeppelin

Para instalar zeppelin deberemos, en primera instancia, descomprimir el archivo:

```
tar -zxvf zeppelin-0.8.2-bin-all.tgz
```

Una vez descomprimido entraremos en la carpeta con dicho nombre que se ha generado, dentro de esta habrá una carpeta llamada *bin* que contiene todos los scripts, para iniciar *zeppelin* usaremos el daemon:

```
bin/zeppelin-daemon.sh start
```

Esto abre el puerto 8080 por defecto, éste podemos cambiarlo en la carpeta *conf/* en una copia del archivo *zeppelin-env.sh.template* en que indicaremos en la propiedad *ZEPPELIN_PORT*.

Por defecto spark ya viene configurado, pero podremos poner la versión que tenemos en nuestro ordenador, modificando el mismo archivo poniendo la ruta de nuestra versión

```
export SPARK_HOME=/opt/spark/spark-2.4.4-bin-hadoop2.7
```

Ahora para hacer el desarrollo de la práctica usaremos las siguientes líneas:

1. Leer el archivo csv

```
val text = sc.textFile("/home/keepcoding/Descargas/amigos.csv")
```

2. Definir nuestra estructura *Friend*

```
case class Friend(id:Integer, name:String, age : Integer, friends : Integer)
```

3. Mapear lo leído a nuestro dominio

```
val friends = text.map(s=>s.split(",")).map(  
  s=>Friend(s(0).toInt, s(1), s(2).toInt, s(3).toInt )  
)
```

4. Crear nuestra vista para que sea accesible

```
friends.toDF().registerTempTable("friends")
```


5. Hacer la consulta SQL

```
%sql select count(*) from friends -> 500
```

En la imagen inferior podemos ver los resultados

test1 - Zeppelin

localhost:8080/#/notebook/2F336UHRM

 Zeppelin Notebook Job

test1

val text = sc.textFile("/home/keepcoding/Descargas/amigos.csv")

text: org.apache.spark.rdd.RDD[String] = /home/keepcoding/Descargas/amigos.csv MapPartitionsRDD[5] at textFile at <console>:26

Took 1 sec. Last updated by anonymous at February 08 2020, 10:01:07 PM.

case class Fiend(id:Integer, name:String, age : Integer, firends : Integer)

defined class Fiend

Took 1 sec. Last updated by anonymous at February 08 2020, 9:59:42 PM.

val fiends = text.map(s=>s.split(",")).map(s=>Fiend(s(0).toInt, s(1), s(2).toInt, s(3).toInt))

fiends: org.apache.spark.rdd.RDD[Fiend] = MapPartitionsRDD[7] at map at <console>:28

Took 2 sec. Last updated by anonymous at February 08 2020, 10:02:41 PM.

fiends.toDF().registerTempTable("friends")

warning: there was one deprecation warning; re-run with -deprecation for details

Took 6 sec. Last updated by anonymous at February 08 2020, 10:03:49 PM.

%sql select count(*) from friends

settings

count(1)

500