<<<<<< Important Book Marks >>>>>>>>

Spark Cluster:-

- **How to start a cluster**

Go to Amazon Web Console and click on EMR -> Create Cluster. Specify the properties that you need and click Create Cluster.

OR

If aws is installed on your system, you can start it through command line by the following command:-

---→ aws emr create-cluster --name "Spark_cluster" --release-label emr-4.7.0 --instance-type c3.xlarge --instance-count 1 --use-default-roles --ec2-attributes KeyName=aws_spark --applications Name=Spark

Where……

emr- Elastic Map Reduce

release label- version which you need to install

instance-type- basically a hardware configuration based on application requirement.

instance-count- Master and slave instances.

***applicationsName- applications which you need to need to install

- Spark-submit command:-

Spark-submit --class class_name --master local /home/hadoop/jar_file

Projects:-

1) **Statistics** (Calculate mean, median and standard Deviation based on the arguments that we pass.)

Spark-submit command:-

spark-submit --class com.nlpcaptcha.histatistics.Statistics --master local /home/hadoop/Statistics-0.0.1-SNAPSHOT.jar arg[0] arg[1] arg[2]

Where

arg[0] → Parameter

arg[1] →limit

arg[2] → condition (e.g. “captchaAccuracy=0”)

- Connecting it to mysql :-

E.g :-> spark-submit --class com.nlpcaptcha.histatistics.Statistics --master local --packages mysql:mysql-connector-java:5.1.6 /home/hadoop/Statistics-0.0.1-SNAPSHOT.jar imp_recall_timeinterval.average all “captchaAccuracy=0”

2) **Report** (Generate the Reports and save it to couchbase/MySQL.)

Spark-submit command:-

Spark-submit --class com.nlpcaptcha.tag.TagData --master local /home/hadoop/report-0.0.1-SNAPSHOT.jar

- For connecting it to couchbase:-

spark-submit --class com.nlpcaptcha.tag.TagData --master local --packages com.couchbase.client:spark-connector_2.11:1.2.1 /home/hadoop/report-0.0.1-SNAPSHOT.jar

- For connecting it to mysql:-

Provide the jar file of my Sql connector:-

1. **For score_reason :-**

spark-submit --class com.nlpcaptcha.tag.TagData --master local --packages mysql:mysql-connector-java:5.1.6 */home/hadoop/report-0.0.1-SNAPSHOT.jar score_reason. “nlpcaptcha4632ff5b195065ab18ad92d47522a690.captchaAccuracy>10”*

2. **For human/bot Probablility**

spark-submit --class com.nlpcaptcha.tag.TagData --master local --packages mysql:mysql-connector-java:5.1.6 */home/hadoop/report-0.0.1-SNAPSHOT.jar humanProbablility*

3. **For sessionId**

spark-submit --class com.nlpcaptcha.tag.TagData --master local --packages mysql:mysql-connector-java:5.1.6 */home/hadoop/report-0.0.1-SNAPSHOT.jar sessionId=’*nlpcaptcha4632ff5b195065ab18ad92d47522a690*’*

4. **For visibliity_percentage :-**

spark-submit --class com.nlpcaptcha.tag.TagData --master local --packages mysql:mysql-connector-java:5.1.6 */home/hadoop/report-0.0.1-SNAPSHOT.jar "tagData.visibility_percentage>=50"*

3) **Score Summary** (Generate the summary report based on score.)

Spark-submit command:-

Spark-submit --class com.nlpcaptcha.scoresummary.ScoreSummary --master local /home/hadoop/ scoresummary-0.0.1-SNAPSHOT.jar

- For connecting it to couchbase:-

spark-submit --class com.nlpcaptcha. scoresummary.ScoreSummary *--master local --packages com.couchbase.client:spark-connector_2.11:1.2.1 /home/hadoop/report-0.0.1-SNAPSHOT.jar*

- For connecting it to mysql:-

Provide the jar file of my Sql connector:-

*spark-submit --class com.nlpcaptcha.*scoresummary.ScoreSummary *--master local* --packages mysql:mysql-connector-java:5.1.6 */home/hadoop/report-0.0.1-SNAPSHOT.jar*

Important :- Connecting to couchbase with spark Application.

Note:- If you need to connect your application with couchbase, then provide the package of spark connector as follows:-

--package *com.couchbase.client:spark-connector_2.11:1.2.1*

Include the following dependency in pom.xml :-

```
<dependency>
    <groupId>com.couchbase.client</groupId>
    <artifactId>spark-connector_2.11</artifactId>
    <version>1.2.1</version>
</dependency>
```

Then include this package in spark-submit command while running as:-

----→ *spark-submit --class org.tag.report.App --master local --packages com.couchbase.client:spark-connector_2.11:1.2.1 /home/hadoop/report-0.0.1-SNAPSHOT.jar*

- **For further information:-**

https://spark-packages.org/package/couchbaselabs/couchbase-spark-connector