

README FILE

Steps for making the jar file.

- 1) Import the whole project in intellij.
- 2) Then go to view-> tools windows and then click sbt
- 3) On the “sbt projects” pane on the right, expand the sbt tasks and then go down to the “package” task and double click on it.
- 4) If everything goes ok you will a success message.
- 5) You can then see the jar file under target/scala-2.11/classes folder

Steps for creating the bucket in S3 in AWS and you can run the jar file as below:

- 1) Login to your s3 account and create the bucket in the same region as your cluster and key.
- 2) Give the unique bucket name and accept all the default values and create the bucket
- 3) Upload the jar file and all the input files required to run all the 3 parts in the same bucket.
- 4) Now, go back to the EMR cluster and click on the cluster name and then click on the “steps” tab.
- 5) Click on “add step” and then fill the values as follows:
 - a. Step type: Spark application
 - b. Spark submit option: --class “Part1” //the name of the object file for the first part
 - c. Application location: go to the bucket which you have created and select the jar file
 - d. Arguments: give the input as input file and output will be the location where we need to save the output file

Finally run the step and once completed, you will see the output folder in your bucket which you have created previously.