**Process a log file & list which users are logged for maximum times using hadoop**

**Prerequisites:** We have consider that Hadoop 2.9.0 single node cluster is installed with dedicated hadoop user with name hduser.

1. Login with hadoop user (hduser)
2. Create directory in home folder with name as **analyzelogs**
3. Change the owner of this folder using following command:

**$ sudo chown -R hduser analyzelogs/**

1. Copy all six files (Driver, Reducer, Mapper accesslog\_short.csv & example file) in this analyzelogs folder.
2. Give permission to all file in this analyzelogs folder using following command:

**$ cd analyzelogs/**

**~/analyzelogs $ sudo chmod +r \*.\***

1. Now export classpath to set all the paths (execute following command in terminal.

**~/analyzelogs $ export CLASSPATH="$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.9.0.jar:$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.9.0.jar:$HADOOP\_HOME/share/hadoop/common/hadoop-common-2.9.0.jar:/home/hduser/analyzelogs/SalesCountry/\*:$HADOOP\_HOME/lib/\*"**

Change the path according to your hadoop installation directory

1. Compile all the java files (Mapper, Reducer, Driver):

**~/analyzelogs $ javac -d . SalesMapper.java SalesCountryReducer.java SalesCountryDriver.java**

This will create a folder in analyzelogs with name **SalesCountry** which contain all the compiled i.e. class files.

1. Create Manifest.txt file in analyzelogs folder and paste follwing content in that file and at the end of line press enter (this is important).

**Main-Class: SalesCountry.SalesCountryDriver**

1. Now create jar file

**~/analyzelogs $ jar -cfm analyzelogs.jar Mainfest.txt SalesCountry/\*.class**

1. Go to home folder

**~/analyzelogs $ CD ..**

1. Start HADOOP

**$ start -dfs.sh**

**$ start-yarn.sh**

Make sure that NameNode, DataNode, ResourceManager, NodeManager & SecondaryNameNode is running.

1. Create input directory and copy log.csv file in that folder

**$ cd analyzelogs**

**~/analyzelogs $ sudo mkdir ~/input**

**~/analyzelogs $ sudo cp access\_log\_short.csv ~/input/**

**~/analyzelogs $ $HADOOP\_HOME/bin/hdfs dfs -put ~/input/**

**~/analyzelogs $ $HADOOP\_HOME/bin/hadoop jar analyzelogs.jar /input /output**

1. This will create output file in the output folder located in hadoop home directory with file name as part-00000

**~/analyzelogs $ $HADOOP\_HOME/bin/hdfs dfs -cat /output/part-00000**