Name: Arun Kumar N
Student ID : 2017CBDE030
Saavn Assignment

Note: Please open Top 100 zip folder contents in Notepad++. Opening in notepad not showing the entries in next line.

Steps Involved in Getting the Trending Songs using the Given Dataset:

Step 1: Exported the JAR file from the code and placed on S3 Bucket

Step 2: From S3 bucket, took the JAR File to EC2 instance using the command

Aws s3 cp s3://BucketName/JarFileName.jar /root

Step 3: Ran the JAR file using the command

Hadoop jar JarFileName MainClassPath InputFolderName S3 bucket OutputFolderName

Step 4: Once the execution is completed, took the results from the s3 bucket

Step 5: Sort the results using the command

sort -r -t: -nk3 OutputFolderName >NewFileName.txt

Step 6: Get the top 100 using the command

Vi NewFileName.txt | head -100

Step 7: Get the results and store it in a separate file.

Step 8: Compare the Saavn Results data with the data which we generated in the previous step.

Step 9: Calculate the overlap between both the results in Excel file using the command

=IF(ISERROR(MATCH(A3,$E$3:$E$102,0)),"",A3)

Step 10: Overlap between Saavn results and the outputted results is:

Date	Overlap %
25th December	66
26th December	65
27th December	68
28th December	71
29th December	72
30th December	72
31st December	70