MMD Quiz 1 - 15th February 2024

Advertisments.txt

File is structered such that first comes user ID then advertisement ID and then user action. User action can be of 3 types : skip, click, watch for n seconds

e.g

Sample Input

1033,203,watch for 20 secs
1011,203,skip
1022,204,click
1033,205,watch for 5 secs
1044,205,skip
1055,206,click
1066,206,click
1088,206,click
1099,203,skip
1113,205,watch for 5 secs
1116,205,watch for 2 secs
1119,205,watch for 9 secs

Expected Output

"AD ID 203 -> " "Skipped 2 times, Clicked 0 times, Watched on Average 20.0 secs"
"AD ID 204 -> " "Skipped 0 times, Clicked 1 times, Watched on Average 0 secs"
"AD ID 205 -> " "Skipped 1 times, Clicked 0 times, Watched on Average 5.25 secs"
"AD ID 206 -> " "Skipped 0 times, Clicked 3 times, Watched on Average 0 secs"

Find Ad Id , times skipped, clicked and average time watched by users

In [8]:
```
%%file quiz1.py
from mrjob.job import MRJob
from mrjob.step import MRStep

class findAdsInformation(MRJob):

    def mapper_init(self):
        self.adsInfo={}
    def mapper(self,_,line):
        userID,adsID,action=line.split(',')
        self.adsInfo.setdefault(adsID,[0,0,(0,0)]) #List of Skipped, Clicked & Watched(Duration watched,Numbe
r of times watched)
        if(action=='skip'):
            self.adsInfo[adsID][0]=self.adsInfo[adsID][0]+1
        elif(action=='click'):
            self.adsInfo[adsID][1]=self.adsInfo[adsID][1]+1
        elif('watch' in action ):
            time,count=self.adsInfo[adsID][2]
            self.adsInfo[adsID][2]=(time+int(action.split('for')[1].split('secs')[0].strip()),count+1)

    def mapper_final(self):
        for adsID,count in self.adsInfo.items():
            yield adsID,count

    def combiner_reducer(self,key,value):
        skippedCount=0
        clickedCount=0
        watchCount=0
        watchSum=0

        for scount,ccount,wCount in value:
            skippedCount=skippedCount+scount
            clickedCount=clickedCount+ccount
            sumW,countW=wCount
            watchCount=watchCount+countW
            watchSum=watchSum+sumW

        yield key,(skippedCount,clickedCount,(watchCount,watchSum))
    def printReducer(self,key,value):
        skippedCount=0
        clickedCount=0
        watchCount=0
```

```
        watchSum=0

        for scount,ccount,wCount in value:
            skippedCount=skippedCount+scount
            clickedCount=clickedCount+ccount
            sumW,countW=wCount
            watchCount=watchCount+countW
            watchSum=watchSum+sumW

        if(watchCount>0):
            avgWatched=watchSum/watchCount
        else:
            avgWatched=0
        information=('Skipped '+str(skippedCount)+' times, Clicked '+ str(clickedCount) + ' times, Watched on
Average '+str(avgWatched)+' secs')
        key='AD ID ' + key + ' -> '

        yield key,information

    def steps(self):
        return [
            MRStep(mapper_init=self.mapper_init,
            mapper=self.mapper,
            mapper_final=self.mapper_final,
            combiner=self.combiner_reducer,
            reducer=self.combiner_reducer),
            MRStep(reducer=self.printReducer)
        ]

if __name__=='__main__':
    findAdsInformation.run()
```

```
Overwriting quiz1.py
```

In [9]:
```
!python quiz1.py Advertisments.txt
```

```
"AD ID 203 -> " "Skipped 2 times, Clicked 0 times, Watched on Average 20.0 secs"
"AD ID 204 -> " "Skipped 0 times, Clicked 1 times, Watched on Average 0 secs"
"AD ID 205 -> " "Skipped 1 times, Clicked 0 times, Watched on Average 5.25 secs"
"AD ID 206 -> " "Skipped 0 times, Clicked 3 times, Watched on Average 0 secs"

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.104745.983858
Running step 1 of 2...
Running step 2 of 2...
job output is in C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.104745.983858\output
Streaming final output from C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.104745.983858\output...
Removing temp directory C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.104745.983858...
```