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
1011,203,skip
1022,204,skip
1033,205,watch for 5 secs
1044,205,watch for 6 secs
1055,200,click

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

```python
In [106]: %%file quiz1.py
          from mrjob.job import MRJob
          from mrjob.step import MRStep

          class findAdsInformation(MRJob):

              def mapper_init(self):
                  self.adsSkipped={}
                  self.adsClicked={}
                  self.adsWatched={}
              def mapper(self,_,line):
                  userID,adsID,action=line.split(',')
                  if(action=='skip'):
                      self.adsSkipped.setdefault(adsID,0)
                      self.adsSkipped[adsID]=self.adsSkipped[adsID]+1
                  elif(action=='click'):
                      self.adsClicked.setdefault(adsID,0)
                      self.adsClicked[adsID]=self.adsClicked[adsID]+1
                  elif('watch' in action ):
                      self.adsWatched.setdefault(adsID,(0,0))
                      time,count=self.adsWatched[adsID]
                      self.adsWatched[adsID]=(time+int(action.split('for')[1].split('secs')[0].strip()),count+1)

              def mapper_final(self):
                  adsCatred=[]
                  for adsID,count in self.adsSkipped.items():
                      self.adsClicked.setdefault(adsID,0)
                      self.adsWatched.setdefault(adsID,(0,0))
                      adsCatred.append(adsID)
                      yield adsID,(count,self.adsClicked[adsID],self.adsWatched[adsID])
                  for adsID,count in self.adsClicked.items():
                      if(adsID not in adsCatred):
                          adsCatred.append(adsID)
                          self.adsWatched.setdefault(adsID,(0,0))
                          self.adsSkipped.setdefault(adsID,0)
                          yield adsID,(self.adsSkipped[adsID],count,self.adsWatched[adsID])

                  for adsID,count in self.adsWatched.items():
                      if(adsID not in adsCatred):
                          self.adsClicked.setdefault(adsID,0)
                          self.adsSkipped.setdefault(adsID,0)
                          yield adsID,(self.adsSkipped[adsID],self.adsClicked[adsID],count)
```

```python
    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 [107]:
```
!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.20240215.094715.382417
Running step 1 of 2...
Running step 2 of 2...
job output is in C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.094715.382417\output
Streaming final output from C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.094715.382417\output...
Removing temp directory C:\Users\DELL\AppData\Local\Temp\quiz1.DELL.20240215.094715.382417...
```