

BDA Lab 7

Counting Profile Views

You are given a dataset in .txt format. The file contains 90 lines where a sample line looks like this:

“8 Razi Haider Musa Dildar”

Where 8 is the line number and it shows that Razi Haider Visited Musa Dildar's Profile. This is called a Profile View of Musa Dildar and can be shown in Linked in notification that how many Profile Views Musa Dildar has.



I want to calculate all the Profile Views for the students whose profile was visited so that we can show a notification such as above.

REQUIREMENTS

Requirement-1:

Write a MapReduce Program in Python that calculates the Profile Views for all the Visited Profiles.

Requirement-2:

Create a Directory “ProfileViews/Input” on your Hadoop File System and put the file there.

Requirement-3:

Run your MapReduce Program in Python (as done in previous lab) your output folder name must be your roll number (So I can mark your task just by looking at the images).

Requirement-4:

Once all done create a MS Document file and paste the following in the file:

Hadoop	Overview	Datanodes	Datanode Volume Failures	Snapshot	Startup Progress	Utilities
Overview 'localhost:8020' (active)						
Started:	Sat Apr 10 13:07:54 +0500 2021					
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0b13842					
Compiled:	Tue Sep 10 20:56:00 +0500 2019 by rohitsharmaks from branch-3.2.1					
Cluster ID:	CID-b3272eb6-82f1-47c7-a331-9cc7a3974d10					
Block Pool ID:	BP-42924340-127.0.0.1-1611566886266					

Image 1: Overview (Cluster ID, Block Pool ID)

```
2021-04-10 13:31:04,167 INFO mapred.LocalJobRunner: Finishing task: attempt_local1409834900_0001_r_000000_0
2021-04-10 13:31:04,167 INFO mapred.LocalJobRunner: reduce task executor complete.
2021-04-10 13:31:04,746 INFO mapreduce.Job: Job job_local1409834900_0001 running in uber mode : false
2021-04-10 13:31:04,748 INFO mapreduce.Job: map 100% reduce 100%
2021-04-10 13:31:04,750 INFO mapreduce.Job: Job job_local1409834900_0001 completed successfully
2021-04-10 13:31:04,759 INFO mapreduce.Job: Counters: 36
  File System Counters
    FILE: Number of bytes read=14468
    FILE: Number of bytes written=1871925
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=5168
    HDFS: Number of bytes written=279
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
    HDFS: Number of bytes read erasure-coded=0
  Map-Reduce Framework
    Map input records=90
    Map output records=90
    Map output bytes=1575
    Map output materialized bytes=1761
    Input split bytes=123
    Combine input records=0
    Combine output records=0
    Reduce input groups=10
    Reduce shuffle bytes=1761
    Reduce input records=90
    Reduce output records=18
    Spilled Records=180
    Shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=0
    Total committed heap usage (bytes)=553648128
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=2880
  File Output Format Counters
    Bytes Written=279
saadnaeem@Macbook-Pro: /usr/local/cellar/hadoop/3.2.1/libexec/sbin
```

Image 2:Terminal Success starting from Red underline all the way down to your terminal username

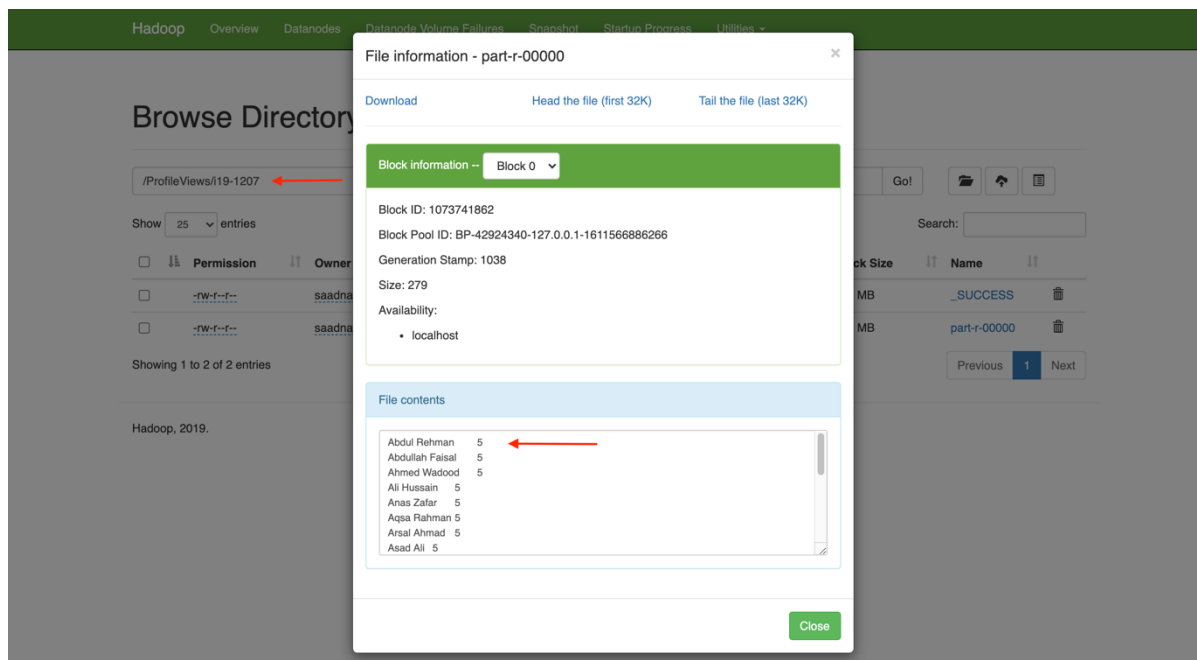


Image 3:Final GUI Output that shows the directory structure in the background

- Paste Your Mapper logic

- Paste Your Reducer Logic
- Finally:

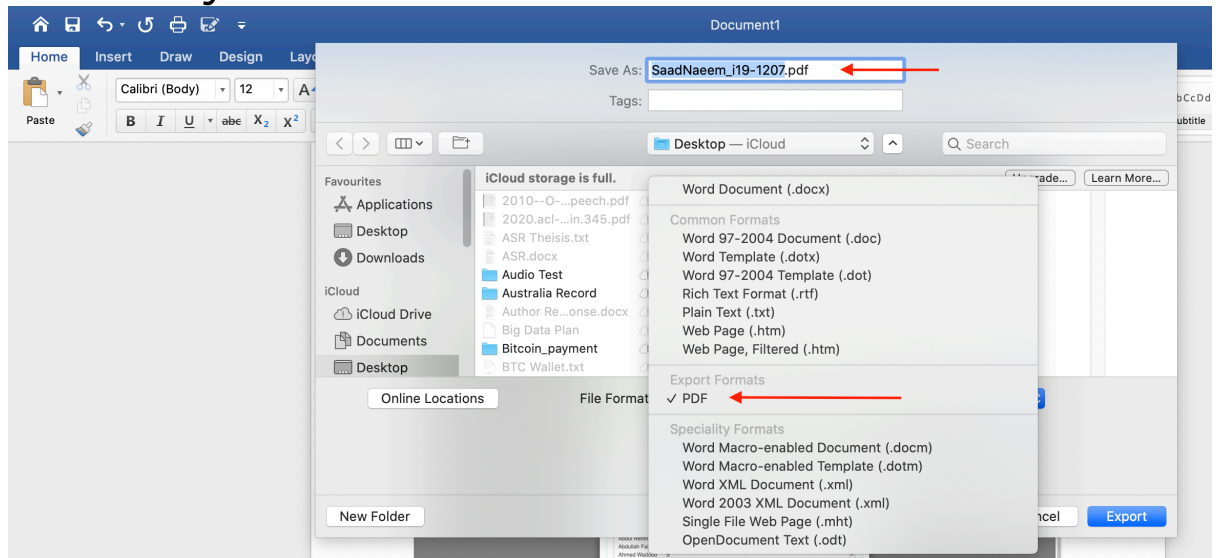


Image 4: Must Save as PDF with your Name_Rollnumber.pdf