

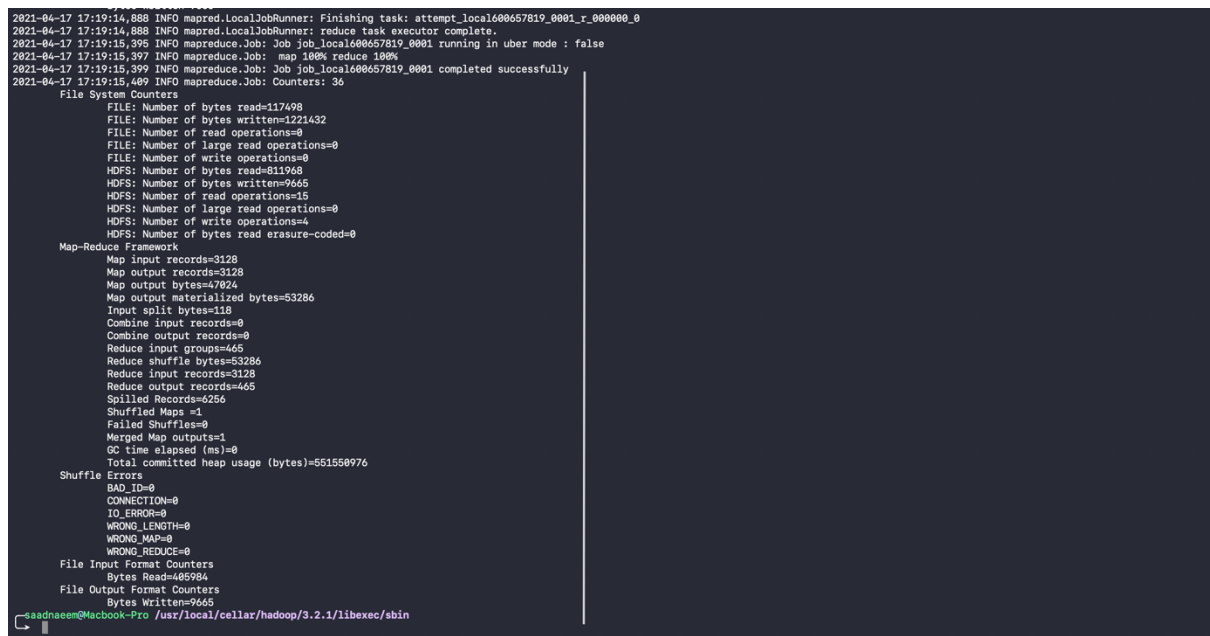
BDA Lab 8

There is a text file (“papers.txt” is uploaded in Class Room) about computer science bibliography. Each line of this file describes the details of one paper in the following format: Authors|Title|Conference|Year. The different fields are separated by the “|” character, and the list of authors are separated by commas (“,”). An example line is given below:
D Zhang, J Wang, D Cai, J Lu|Self-Taught Hashing for Fast Similarity Search|SIGIR|2010
You can assume that there are no duplicate records, and each distinct author or conference as a different name.

TASK: Write a MapReduce program to calculate for each conference the number of papers. You are required to name the output directory as ConferenceCount/RollNumber

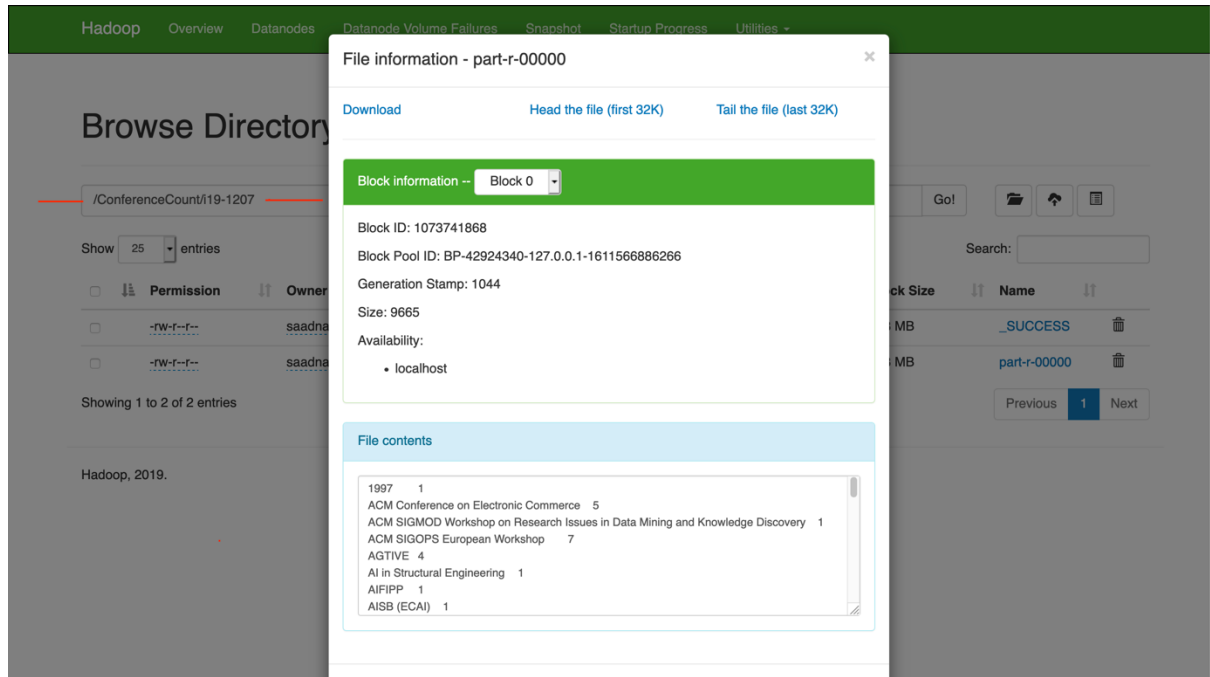
Requirements:

I am reducing the screenshots to two in this lab but I need the exact area as shown in the images you can split that into multiple images if your pixel density is less but **failure to do so will result in zero marks for that task.**



```
2021-04-17 17:19:14,888 INFO mapred.LocalJobRunner: Finishing task: attempt_local600657819_0001_r_000000_0
2021-04-17 17:19:14,888 INFO mapred.LocalJobRunner: reduce task executor complete.
2021-04-17 17:19:15,395 INFO mapreduce.Job: Job job_local600657819_0001 running in uber mode : false
2021-04-17 17:19:15,397 INFO mapreduce.Job: map 100% reduce 100%
2021-04-17 17:19:15,399 INFO mapreduce.Job: Job job_local600657819_0001 completed successfully
2021-04-17 17:19:15,409 INFO mapreduce.Job: Counters: 36
  File System Counters
    FILE: Number of bytes read=117498
    FILE: Number of bytes written=1221432
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=811968
    HDFS: Number of bytes written=9665
    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=3128
    Map output records=3128
    Map output bytes=47024
    Map output materialized bytes=53286
    Input split bytes=119
    Combine input records=0
    Combine output records=0
    Reduce input groups=465
    Reduce shuffle bytes=53286
    Reduce input records=3128
    Reduce output records=465
    Spilled Records=6256
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=0
    Total committed heap usage (bytes)=551550976
  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=405984
  File Output Format Counters
    Bytes Written=9665
saadnaem@Macbook-Pro /usr/local/cellar/hadoop/3.2.1/libexec/sbin
```

Follow the white line and your terminal username must be visible Name: “Rollno-1.png”



The red line's part must be visible along with the rest of the window Name: "Rollno-2.png"

Upload the above two screenshots into google classroom without zipping them Name them according to the naming convention in the images.