ADVANCE BIG DATA ANALYTICS LAB
Experiment - 2
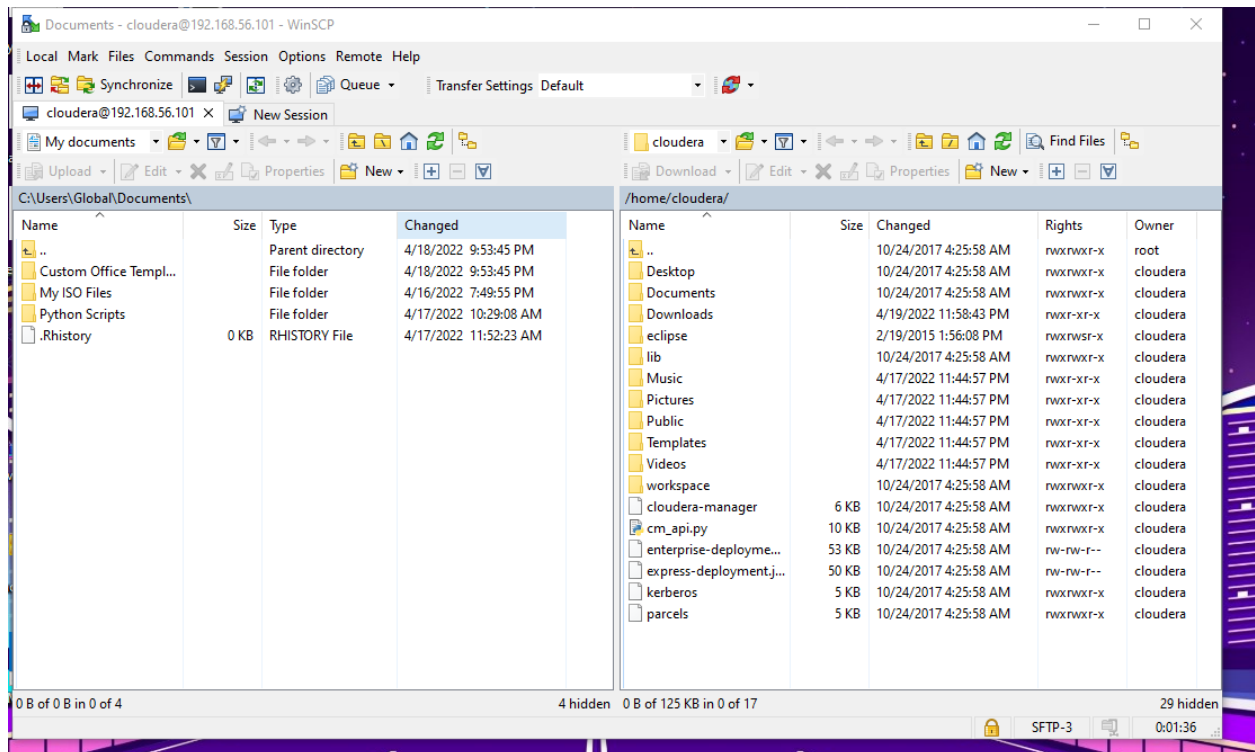
Sai Sharanya Y
19btrcr043

In this task you have 2 files named as Student and Results. You need to use PIG commands for this task.

Step1: Upload this file to Lab through winSCP.
Student: Contains names and roll number of students.
Results: Contains roll number and results of students whether they passed or failed.
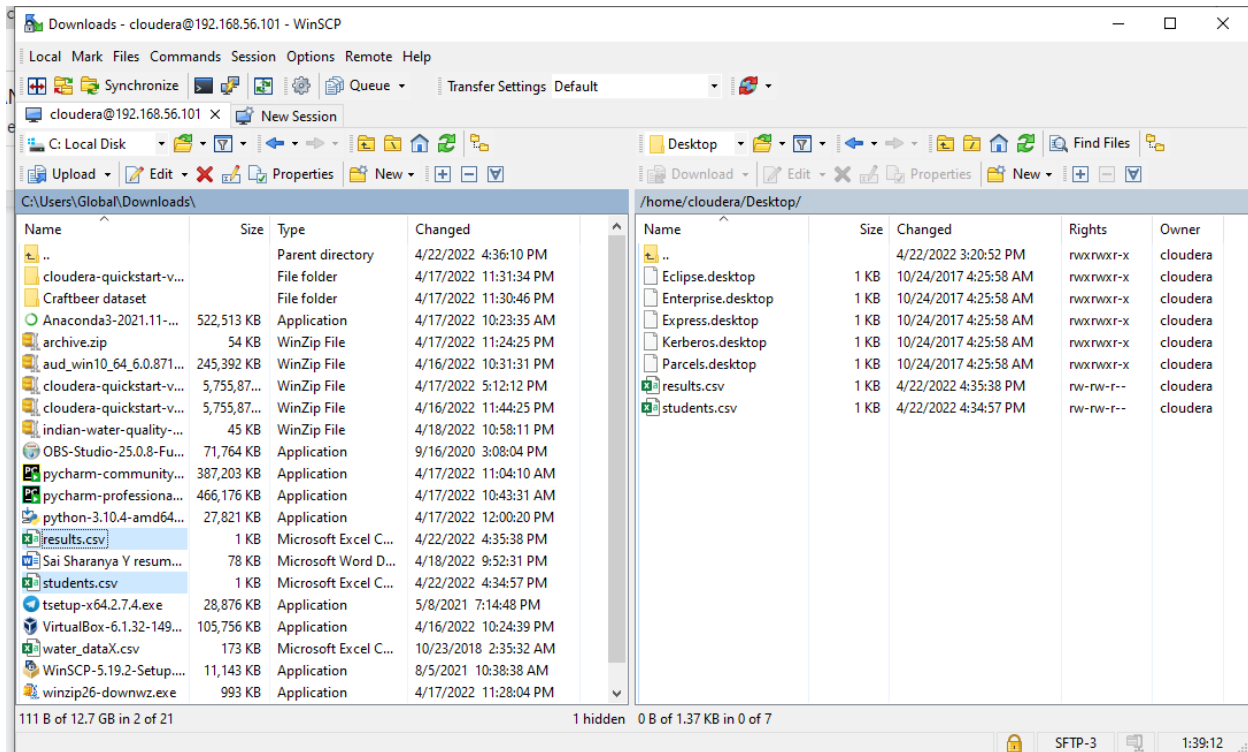
Problem Statement: You need to print the name of all the students who failed or passed in the exam based on the given data.

Results: Contains roll number and results of students whether they passed or failed.

root@quickstart:~   _ □ ✕

File   Edit   View   Search   Terminal   Help

```
2022-04-22 04:07:13,211 [main] INFO   org.apache.hadoop.conf.Configuration.dep
2022-04-22 04:07:13,215 [main] INFO   org.apache.hadoop.conf.Configuration.dep
2022-04-22 04:07:13,229 [main] INFO   org.apache.hadoop.conf.Configuration.dep
2022-04-22 04:07:13,453 [main] INFO   org.apache.hadoop.conf.Configuration.dep
2022-04-22 04:07:13,461 [main] INFO   org.apache.hadoop.conf.Configuration.dep
2022-04-22 04:07:13,467 [main] INFO   org.apache.hadoop.conf.Configuration.dep
grunt> student = load '/home/cloudera/Desktop/students.csv' as (Roll_no : int
, name : chararray);
grunt> result = load '/home/cloudera/Desktop/results.csv' as (Roll_no : int,
status : chararray);
grunt> PassStudents = join student by Roll_no, result by Roll_no;
grunt> PassStudentNames = foreach PassStudents generate name,status;
grunt> Passed = filter PassStudentNames by status matches 'pass';
grunt> dump Passed;
2022-04-22 04:11:01,231 [main] INFO   org.apache.pig.tools.pigstats.ScriptStat
e - Pig features used in the script: HASH_JOIN,FILTER
2022-04-22 04:11:01,356 [main] INFO   org.apache.pig.newplan.logical.optimizer
.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, Duplic
ateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, L
imitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionF
                                        atten, PushUpFilter, SplitFilter, StreamType
                                        terLogicExpressionSimplifier, PartitionFilt
```

Cloudera Live : Welcome! - Cloudera
Live Beginner Tutorial - Mozilla Firefox

🦊 [Cloudera Li...   🖥 root@quick...   🏠 cloudera   📁 home   🖥

🔄 ⊙ 🖥 🖵 🖵 🖥 🖥 🖫 🧰 ⬇ Right Ctrl

Applications   Places   System   Fri Apr 22, 4:13 AM   **cloudera**

root@quickstart:~   _   □   ✕

File   Edit   View   Search   Terminal   Help

```
2022-04-22 04:11:17,404 [main] INFO   org.apache.pig.tools.pigstats.SimplePigS
tats - Script Statistics:

HadoopVersion    PigVersion        UserId   StartedAt        FinishedAt        Featu
res
2.6.0-cdh5.13.0 0.12.0-cdh5.13.0         root     2022-04-22 04:11:01      2022-
04-22 04:11:17   HASH_JOIN,FILTER

Success!

Job Stats (time in seconds):
JobId    Alias    Feature Outputs
job_local1781787715_0001          PassStudentNames,PassStudents,Passed,result,s
tudent   HASH_JOIN       file:/tmp/temp895151890/tmp-199274933,

Input(s):
Successfully read records from: "/home/cloudera/Desktop/results.csv"
Successfully read records from: "/home/cloudera/Desktop/students.csv"

Output(s):
Successfully stored records in: "file:/tmp/temp895151890/tmp-199274933"
```
                                                                          home

[Cloudera Li...]   [root@quick...]   cloudera   home

Right Ctrl