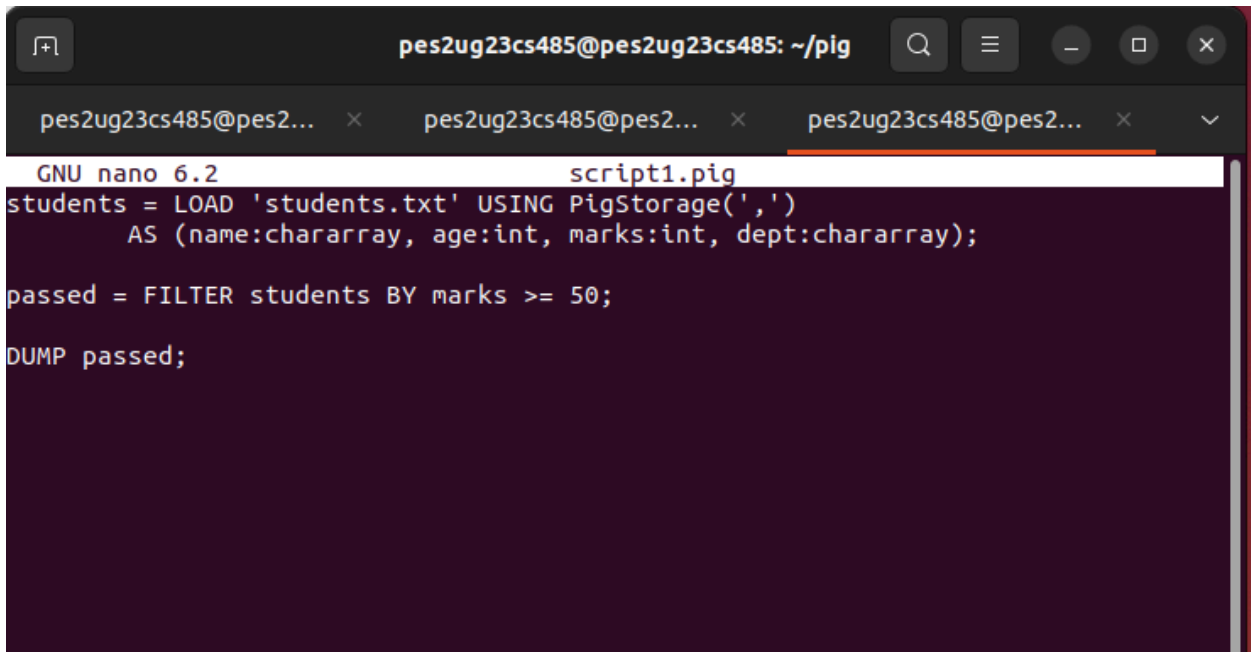# BD hands on 2

Name: Rithvik Rajesh Matta

SRN: PES2UG23CS485

Class: 5H

1)


```
pes2ug23cs485@pes2ug23cs485:~$ pig -version
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58
pes2ug23cs485@pes2ug23cs485:~$
```

2)


```
  GNU nano 6.2                        script1.pig
students = LOAD 'students.txt' USING PigStorage(',')
        AS (name:chararray, age:int, marks:int, dept:chararray);

passed = FILTER students BY marks >= 50;

DUMP passed;
```

3)

```
2025-09-11 10:08:24,726 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2025-09-11 10:08:24,731 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2025-09-11 10:08:24,734 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2025-09-11 10:08:24,748 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2025-09-11 10:08:24,774 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2025-09-11 10:08:24,877 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2025-09-11 10:08:24,893 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(John,25,85,CSE)
(Alice,23,92,ISE)
(Charlie,22,60,ECE)
(Evelyn,24,78,CSE)
(Frank,22,55,ECE)
(Grace,21,88,ISE)
(Ian,23,95,CSE)
(Jack,24,67,ME)
(Kathy,22,82,ECE)
(Leo,21,73,ISE)
(Mona,23,59,CSE)
(Nina,24,90,EEE)
(Paul,25,50,ECE)
(Quinn,23,77,CSE)
(Rachel,22,61,ISE)
(Tina,23,81,CSE)
(Uma,22,93,ECE)
(Wendy,23,70,CSE)
(Zack,24,88,ISE)
2025-09-11 10:08:25,464 [main] INFO  org.apache.pig.Main - Pig script completed in 8 seconds and 149 milliseconds (8149 ms)
[2]+  Done                    gedit script1.pig
pes2ug23cs485@pes2ug23cs485:~/pig$
```

4)

```
script2.pig
~/pig

1 students = LOAD '/pigdata/students.txt' USING PigStorage(',')
2 AS (name:chararray, age:int, marks:int, dept:chararray);
3
4 passed = FILTER students BY marks >= 50;
5 DUMP passed;
```

Loading file "/home/pes2ug23cs485/pig/script2.pig"...   Pig ∨   Tab Width: 8 ∨   Ln 5, Col 13   ∨   INS

5)

```
pes2ug23cs485@pes2ug23cs485:~/pig$ jps
12517 Jps
8072 NameNode
8809 NodeManager
8219 DataNode
12460 JobHistoryServer
8414 SecondaryNameNode
8687 ResourceManager
pes2ug23cs485@pes2ug23cs485:~/pig$
```

6)

```
Input(s):
Successfully read 25 records (756 bytes) from: "/pigdata/students.txt"

Output(s):
Successfully stored 19 records (408 bytes) in: "hdfs://127.0.0.1:9000/tmp/temp1508391691/tmp-1660396209"

Counters:
Total records written : 19
Total bytes written : 408
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1757563974127_0001


2025-09-11 10:13:38,335 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /127.0.0.1:8032
2025-09-11 10:13:38,349 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-09-11 10:13:38,432 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /127.0.0.1:8032
2025-09-11 10:13:38,440 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-09-11 10:13:38,482 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /127.0.0.1:8032
2025-09-11 10:13:38,497 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-09-11 10:13:38,786 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2025-09-11 10:13:38,798 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.e
led
2025-09-11 10:13:38,801 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2025-09-11 10:13:38,863 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2025-09-11 10:13:38,864 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(John,25,85,CSE)
(Alice,23,92,ISE)
(Charlie,22,60,ECE)
(Evelyn,24,78,CSE)
(Frank,22,55,ECE)
(Grace,21,88,ISE)
(Ian,23,95,CSE)
(Jack,24,67,ME)
(Kathy,22,82,ECE)
(Leo,21,73,ISE)
(Mona,23,59,CSE)
(Nina,24,90,EEE)
(Paul,25,50,ECE)
(Quinn,23,77,CSE)
(Rachel,22,61,ISE)
(Tina,23,81,CSE)
(Uma,22,93,ECE)
(Wendy,23,70,CSE)
(Zack,24,88,ISE)
2025-09-11 10:13:39,154 [main] INFO  org.apache.pig.Main - Pig script completed in 58 seconds and 590 milliseconds (58590 ms)
pes2ug23cs485@pes2ug23cs485:~/pig$
```

7)

```
script3.pig
~/pig

1 students = LOAD '/pigdata/students.txt' USING PigStorage(',')
2     AS (name:chararray, age:int, marks:int, dept:chararray);
3
4 valid_records = FILTER students BY (marks IS NOT NULL) AND (dept IS NOT NULL);
5
6 grouped = GROUP valid_records BY dept;
7
8 avg_marks = FOREACH grouped GENERATE group AS dept, AVG(valid_records.marks) AS avg_marks;
9
10 DUMP avg_marks;
```

8)

```
    pes2ug23cs485@pes2...    ✕       pes2ug23cs485@pes2...    ✕       pes2ug23cs485@pes2...    ✕       ⌄

Total records proactively spilled: 0

Job DAG:
job_1757563974127_0002


2025-09-11 10:20:09,751 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARM
FailoverProxyProvider - Connecting to ResourceManager at /127.0.0.1:8032
2025-09-11 10:20:09,793 [main] INFO  org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2025-09-11 10:20:09,859 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARM
FailoverProxyProvider - Connecting to ResourceManager at /127.0.0.1:8032
2025-09-11 10:20:09,869 [main] INFO  org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2025-09-11 10:20:09,957 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARM
FailoverProxyProvider - Connecting to ResourceManager at /127.0.0.1:8032
2025-09-11 10:20:09,986 [main] INFO  org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2025-09-11 10:20:10,089 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2025-09-11 10:20:10,092 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Ins
tead, use yarn.system-metrics-publisher.enabled
2025-09-11 10:20:10,095 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2025-09-11 10:20:10,157 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input files to process : 1
2025-09-11 10:20:10,159 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(ME,50.5)
(CSE,72.5)
(ECE,68.0)
(EEE,56.0)
(ISE,80.4)
2025-09-11 10:20:10,888 [main] INFO  org.apache.pig.Main - Pig script completed
in 1 minute, 8 seconds and 73 milliseconds (68073 ms)
pes2ug23cs485@pes2ug23cs485:~/pig$ s
```