



Class: TY-IT, Semester: VI Subject: Big Data Lab

## Experiment – 6: TO STUDY AND IMPLEMENT PIG COMMANDS

Name: Anish Sharma

Roll no: I011

1. **Aim:** To study and implement Pig commands.

2. **Requirements:** PC, Internet and VMWare software, Cloudera.

3. **Theory:**

Students need to write the theory on following points.

- Overview of Pig
- Features of Pig

4. **Procedure**

Open VMWare Workstation > select Cloudera > open virtual machine > in terminals

Perform the Pig commands.

(Minimum 20 commands should be executed)

1.

```
ne.HExecutionEngine - Connecting to hadoop file system at: hdfs://
grunt> [training@localhost ~]$ pig -x mapreduce
2025-03-04 00:11:52,769 [main] INFO org.apache.pig.Main - Logging error message to: /home/training/pig_1741075912768.log
2025-03-04 00:11:52,927 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:802
2025-03-04 00:11:53,145 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
grunt>
```

2. `students = LOAD 'student_data.txt' USING PigStorage(',') AS (id:int, name:chararray, marks:int);`

```
ne.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2025-03-04 00:11:53,145 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
grunt> students = LOAD 'student_data.txt' USING PigStorage(',') AS (id:int, name:chararray, marks:int);
grunt>
```

3. `dump`



```
ne.util.MapRedUtil - Total in|
(1,Neha,20)
(2,Falak,80)
(3,Isha,90)
(4,Dhruv,60)
grunt> █
```

4.high\_salary = FILTER data BY salary > 50000;  
DUMP high\_salary;

```
2025-03-04 01:32:26,401 [main] INFO org.apache.pig.backend.nadoop.executi
ne.util.MapRedUtil - Total input paths to process : 1
(2,Falak,80)
(3,Isha,90)
(4,Dhruv,60)
grunt>
```

5.name\_salary = FOREACH data GENERATE name, salary; DUMP name\_salary;

```
ne.util.MapRedUtil - Tot
(Neha,20)
(Falak,80)
(Isha,90)
(Dhruv,60)
grunt> █
```

6.max\_salary = FOREACH (GROUP data ALL) GENERATE MAX(data.salary); DUMP  
max\_salary;

```
(90)
grunt> [training@localhost ~]$
```

7.sorted\_salaries = ORDER data BY salary DESC;  
DUMP sorted\_salaries;

```
2025-03-04 01:39:34,710 [main] INFO org
ne.mapReduceLayer.MapReduceLauncher - Su
2025-03-04 01:39:34,714 [main] INFO org
nputFormat - Total input paths to proces
2025-03-04 01:39:34,715 [main] INFO org
ne.util.MapRedUtil - Total input paths t
(3,Isha,90)
(2,Falak,80)
(4,Dhruv,60)
(1,Neha,20)
grunt> █
```

8.avg\_salary = FOREACH (GROUP data ALL) GENERATE AVG(data.salary); DUMP  
avg\_salary;



```
2025-03-04 01:43:54
ne.util.MapRedUtil
(62.5)
grunt>
```

9.min\_salary = FOREACH (GROUP data ALL) GENERATE MIN(data.salary); DUMP min\_salary;

```
2025-03-04 01:45:55,200 [main]
ne.util.MapRedUtil - Total input
(20)
grunt> only_names = FOREACH data
grunt> DUMP only_names
2025-03-04 01:47:10,235 [main]
```

min\_salary = FOREACH (GROUP data ALL) GENERATE MIN(data.salary); DUMP min\_salary;

10.only\_names = FOREACH data GENERATE name;  
DUMP only\_names;

```
ne.util.MapRedUtil - Total
(Neha)
(Falak)
(Isha)
(Dhruv)
grunt>
```

11.low\_salary = FILTER data BY salary < 30000;  
DUMP low\_salary;

```
InputFormat - Total input paths to process : 1
2025-03-04 01:49:14,756 [main] INFO org.apache
ne.util.MapRedUtil - Total input paths to process
(1,Neha,20)
(2,Falak,80)
(3,Isha,90)
(4,Dhruv,60)
grunt>
```

12.i\_names = FILTER data BY name MATCHES 'I.\*';  
DUMP i\_names;

```
(2,Falak,80)
grunt>
```

13.



```
2025-03-04 01:54:20,489
ne.util.MapRedUtil - Tot
(1,Neha,20)
(4,Dhruv,60)
grunt>
```

```
(250)
grunt>
```

14.

```
inputFormat - Total input paths to process : 1
2025-03-04 01:57:06,765 [main] INFO  org.apache.pig.backend.hadoop.execut
ne.util.MapRedUtil - Total input paths to process : 1
(3,Isha,90)
(2,Falak,80)
(4,Dhruv,60)
(1,Neha,20)
grunt>
```

## 5. Result/Observation/Program code

Take the snapshots of each successfully executed commands and with the explanation along with syntax.

## 6. Conclusion:

- Write what was performed in the experiment.
- Mention few applications of what was studied.
- Write the significance of the topic studied in the experiment.

## Resources :

<https://pig.apache.org/docs/latest/basic.html>

<https://pig.apache.org/docs/latest/start.html> <https://www.educba.com/pig-commands/>