Task 1

```
[acadgild@localhost ~]$ cat wc.pig
loaded = LOAD 'wc.txt' AS (word : chararray);
grouped = GROUP loaded by word;
wc_group = FOREACH grouped GENERATE group, COUNT(loaded.word);
dump wc_group;
[acadgild@localhost ~]$
```

Executing

```
[acadgild@localhost ~]$ pig -x local -f wc.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/sha
re/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.
class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/s
lf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/06/05 19:24:25 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/06/05 19:24:25 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2018-06-05 19:24:26,070 [main] INFO org.apache.pig.Main - Apache Pig version 0.
16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-06-05 19:24:26,076 [main] INFO org.apache.pig.Main - Logging error message
s to: /home/acadgild/pig_1528206866057.log
2018-06-05 19:24:26,280 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - user.name is deprecated. Instead, use mapreduce.job.user.name
2018-06-05 19:24:27,819 [main] INFO org.apache.pig.impl.util.Utils - Default bo
otup file /home/acadgild/.pigbootup not found
2018-06-05 19:24:28,893 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
2018-06-05 19:24:28.893 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-06-05 19:24:28,913 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to hadoop file system at: file:///
2018-06-05 19:24:29,154 [main] INFO org.apache.pig.PigServer - Pig Script ID fo
r the session: PIG-wc.pig-74e41d56-ecda-4b67-9aa7-1ab16670c9fe
2018-06-05 19:24:29,154 [main] WARN org.apache.pig.PigServer - ATS is disabled
```

Output

```
2018-06-05 19:24:50,532 [main] INFO org.apache.pig.backend.hadoop.executionengi ne.util.MapRedUtil - Total input paths to process : 1 (abc,3) (hfg,1) (jkh,1) (lmn,1) (xyz,2) 2018-06-05 19:24:50,818 [main] INFO org.apache.pig.Main - Pig script completed in 29 seconds and 457 milliseconds (29457 ms)
```

```
a)
```

```
Script
| lacadgild@localhost ~J$ cat task2.pig
| empdetails = LOAD 'employee_details.txt' USING PigStorage (',') AS (empid : int, name : chararray, salary : int, deptid : int
orderemp = ORDER empdetails by deptid desc, name asc;
Limit5 = LIMIT orderemp 5;
dump Limit5;
[acadgild@localhost ~]$
```

```
[acadglld@localhost ~]$
[acadgild@localhost ~]$ pig -x local -f task2.pig
```

Output

```
2018-06-05 22:50:40,762 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized 2018-06-05 22:50:40,836 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process: 1 2018-06-05 22:50:40,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
(105,Pawan,2500,5)
(110,Priyanka,2000,5)
 (104, Anubhav, 5000, 4)
 (109.Katrina.1000.4)
(103, Akshay, 11000, 3)
2018-06-05 22:50:41,192 [main] INFO org.apache.pig.Main - Pig script completed in 37 seconds and 261 milliseconds (37261 ms)
You have new mail in /var/spool/mail/acadgild
```

B)

Script

```
[acadgild@localhost ~]$ cat task2b.pig
empdetails = LOAD 'employee_details.txt' USING PigStorage (',') AS (empid : int, name : chararray, salary : int, deptid : int
);
ordersal = ORDER empdetails by salary desc, name asc;
oddempid = FILTER ordersal BY empid % 2 == 1;
Limit3 = LIMIT oddempid 3;
dump Limit3;
[acaddild@localhost ~l$
```

```
[acadgild@localhost ~]$ pig -x local -f task2b.pig
```

```
(101, Amitabh, 20000, 1)
(107, Salman, 17500, 2)
(103, Akshay, 11000, 3)
2018-06-05 23:09:04,842 [main] INFO org.apache.pig.Main - Pig script completed in 35 seconds and 801 milliseconds (35801 ms) [acadgild@localhost ~]$
```

C)

Script

```
[acadgild@localhost ~]$ cat taskc.txt
joined = JOIN empdetails BY empid, empexpense BY eid;
orderexp = ORDER joined by expense desc, name asc;
limit1 = LIMIT orderexp 1;
dump limit1;
[acadgild@localhost ~]$ pig -x local -f taskc.txt
```

Output

```
2018-06-05 23:35:43,472 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to 2018-06-05 23:35:43,472 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input
cess : 1
(110, Priyanka, 2000, 5, 110, 400)
```

D)

Script

```
[acadgild@localhost ~]$ cat taskd.txt
joined = JOIN empdetails BY empid, empexpense BY eid;
final = FOREACH joined GENERATE empid, name;
dump final;
```

```
Output
2018-06-05 23:49:34,880 [main] INFO org.apacne.pig.backend.nadoop.executionengine.util.mapkedutil - lotal input paths to process: 1
[101,Amitabh)
(101,Amitabh)
(102,Shahrukh)
(102, Shahrukh)
(102, Shahrukh)
(104, Anubhav)
(105, Pawan)
(110, Priyanka)
(114, Madhuri)
```