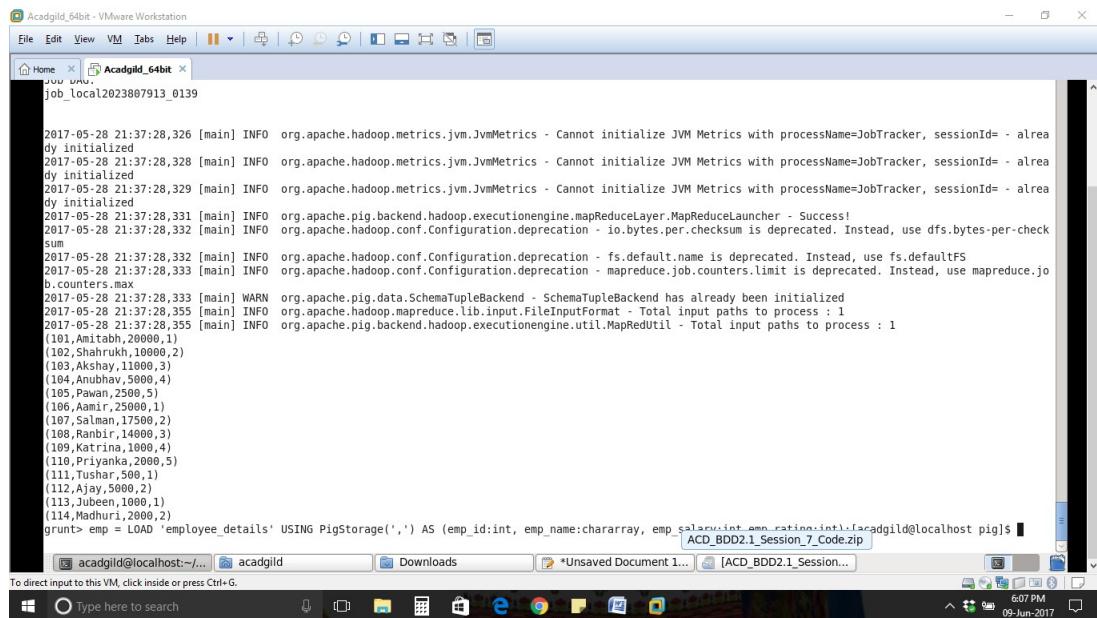


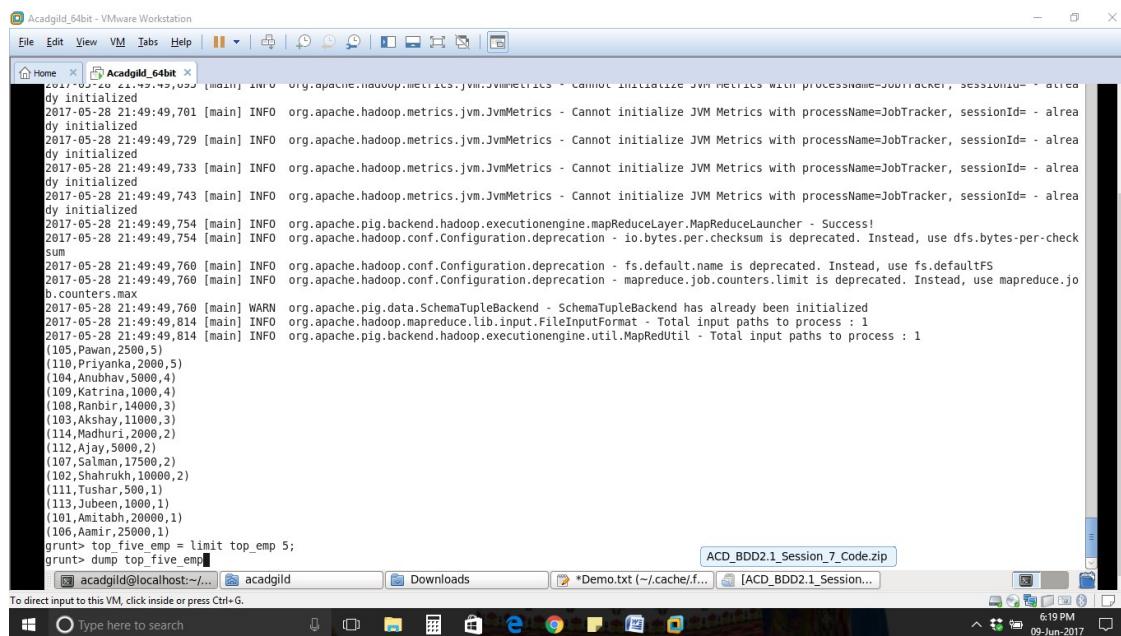
## ASSIGNMENT 7.2

- a) Top 5 employees (employee id and employee name) with highest rating. (In case two employees have same rating, employee with name coming first in dictionary should get preference)



```
2017-05-28 21:37:28,326 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:37:28,328 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:37:28,329 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:37:28,331 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 21:37:28,332 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 21:37:28,333 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.b.counters.max
2017-05-28 21:37:28,333 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 21:37:28,355 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 21:37:28,355 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh,20000,1)
(102,Shahrukh,10000,2)
(103,Akshay,11000,3)
(104,Anubhav,5000,4)
(105,Pawan,2500,5)
(106,Aamir,25000,1)
(107,Salman,17500,2)
(108,Ranbir,14000,3)
(109,Madhuri,15000,4)
(110,Priyanka,20000,5)
(111,Tushar,500,1)
(112,Ajay,5000,2)
(113,Jubeen,1000,1)
(114,Madhuri,20000,2)
grunt> emp = LOAD 'employee_details' USING PigStorage(',') AS (emp_id:int, emp_name:chararray, emp_salary:int, emp_rating:int);[acadgild@localhost pig]$ █
To direct input to this VM, click inside or press Ctrl+G.
acadgild@localhost:~... acadgild Downloads *Unsaved Document 1... [ACD_BDD2.1_Session...
Windows Taskbar: Type here to search, File Explorer, Task View, Start, Taskbar icons, Date/Time: 6:07 PM 09-Jun-2017
```

Employee Record arranged by Rating



```
2017-05-28 21:49:49,701 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:49:49,729 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:49:49,733 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:49:49,743 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 21:49:49,754 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 21:49:49,754 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 21:49:49,760 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 21:49:49,760 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.b.counters.max
2017-05-28 21:49:49,760 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 21:49:49,814 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 21:49:49,814 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(105,Pawan,2500,5)
(110,Priyanka,20000,5)
(104,Anubhav,5000,4)
(109,Katrina,10000,4)
(108,Ranbir,14000,3)
(103,Akshay,11000,3)
(114,Madhuri,20000,2)
(112,Ajay,5000,2)
(107,Salman,17500,2)
(102,Shahrukh,10000,2)
(111,Tushar,500,1)
(113,Jubeen,1000,1)
(101,Amitabh,20000,1)
(106,Aamir,25000,1)
grunt> top_five_emp = limit top_emp 5;
grunt> dump top_five_emp;[acadgild@localhost:~... acadgild Downloads *Demo.txt (~/.cache/f... [ACD_BDD2.1_Session...
Windows Taskbar: Type here to search, File Explorer, Task View, Start, Taskbar icons, Date/Time: 6:19 PM 09-Jun-2017
```

## Top Five Employees according to Rating

Acadgild\_64bit - VMware Workstation

File Edit View VM Tabs Help || |

Home Acadgild\_64bit Acadgild\_64bit\_1

2017-05-28 22:01:39,734 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,734 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,734 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,735 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,738 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,743 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,744 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,744 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,751 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,752 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:01:39,759 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2017-05-28 22:01:39,767 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2017-05-28 22:01:39,768 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-05-28 22:01:39,768 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max is deprecated. Instead, use mapreduce.job.b.counters.max  
2017-05-28 22:01:39,768 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2017-05-28 22:01:39,789 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2017-05-28 22:01:39,789 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
grunt> [Acadgild@localhost:~...][acadgild][Downloads][Unsaved Document 1...][ACD\_BDD2.1\_Session...]

Displaying Emp\_Id and Emp\_name of Top Five employees

**7.2.B** Top 3 employees (employee id and employee name) with highest salary, whose employee id is an odd number. (In case two employees have same salary, employee with name coming first in dictionary should get preference)

Employees sorted by salary

```

2017-05-28 22:05:20,469 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:05:20,470 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:05:20,480 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:05:20,484 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:05:20,490 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:05:20,495 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 22:05:20,500 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:05:20,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:05:20,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max is deprecated. Instead, use mapreduce.job.counters.max
2017-05-28 22:05:20,501 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:05:20,529 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:05:20,529 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh,20000,1)
(107,Salman,17500,2)
(108,Ranbir,14000,3)
(103,Akshay,11000,3)
(102,Shahrukh,10000,2)
(112,Ajay,5000,2)
(104,Anubhav,5000,4)
(105,Pawan,2500,5)
(110,Priyanka,2000,5)
(114,Madhuri,2000,2)
(109,Katrina,1000,4)
(113,Jubeen,1000,1)
(111,Tushar,500,1)
grun> 

```

Top 3 employees by salary whose ID is Odd number;

```

2017-05-28 22:12:33,387 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,392 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,393 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,394 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,400 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,400 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,401 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,400 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,410 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,417 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:12:33,428 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 22:12:33,428 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:12:33,429 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:12:33,432 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.max is deprecated. Instead, use mapreduce.job.counters.max
2017-05-28 22:12:33,432 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:12:33,461 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:12:33,462 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh,20000,1)
(107,Salman,17500,2)
(108,Ranbir,14000,3)
(103,Akshay,11000,3)
grun> 

```

## Displaying Top 3 employees employee id and Name by salary

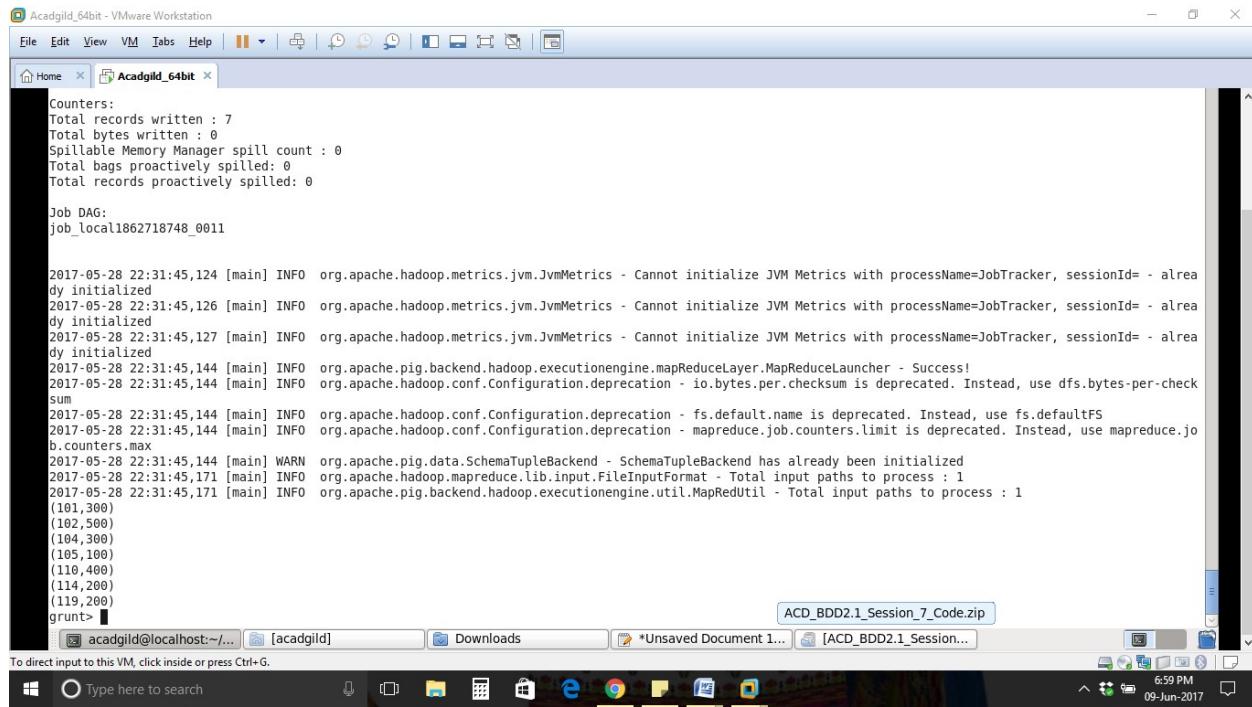
The screenshot shows a Windows desktop environment with a terminal window open in a VMware Workstation window. The terminal window displays the following log output:

```
2017-05-28 22:16:48,750 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,763 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,763 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,763 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,764 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,791 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,792 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,792 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,808 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,819 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,820 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:16:48,842 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 22:16:48,843 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:16:48,844 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:16:48,844 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-28 22:16:48,844 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:16:48,859 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:16:48,859 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh)
(107,Salman)
(103,Akshay)
grunt> *
```

The terminal prompt shows the user has run a command and is awaiting further input. The desktop taskbar at the bottom shows several open windows and the system clock.

**7.2.C Employee (employee id and employee name) with maximum expense (In case two employees have same expense, employee with name coming first in dictionary should get preference)**

Total expense of each employee

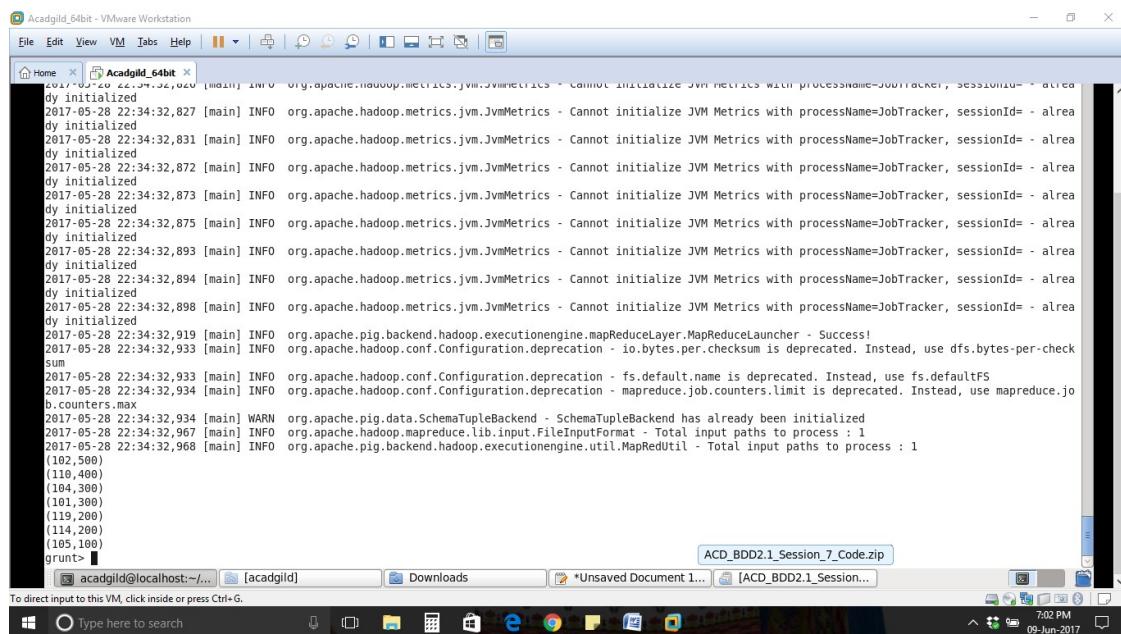


```
Counters:
Total records written : 7
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1862718748_0011

2017-05-28 22:31:45,124 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:31:45,126 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:31:45,127 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:31:45,144 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLauncher - Success!
2017-05-28 22:31:45,144 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:31:45,144 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:31:45,144 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.b.counters.max
2017-05-28 22:31:45,144 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:31:45,171 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:31:45,171 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,300)
(102,500)
(104,300)
(105,100)
(110,400)
(114,200)
(119,200)
grunt> 
```

Total expenses sorted in descending order



```
2017-05-28 22:34:32,827 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,831 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,872 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,873 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,875 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,893 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,894 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,898 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:34:32,919 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLauncher - Success!
2017-05-28 22:34:32,933 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:34:32,934 [main] INFO org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:34:32,967 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:34:32,968 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(102,500)
(110,400)
(104,300)
(101,300)
(119,200)
(114,200)
(105,100)
grunt> 
```

Employee ID and name of employee with top expense

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help || + - ×

Home Acadgild_64bit
Input(s):
Successfully read 14 records from: "file:///home/acadgild/pig/employee_details"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp373925203/tmp497028409"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local958291750_0022

2017-05-28 22:41:31,687 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:41:31,689 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:41:31,690 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:41:31,699 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 22:41:31,700 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:41:31,701 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:41:31,701 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-28 22:41:31,701 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:41:31,720 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:41:31,720 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(102, Shahrukh)
grunt> *Unsaved Document 1 - edit
acadgild@localhost:~/... acadgild Downloads *Unsaved Document 1... ACD_BDD2.1_Session...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 7:05 PM 09-Jun-2017
```

## 7.2 D&E.

(d) List of employees (employee id and employee name) having entries in employee\_expenses file.

(e) List of employees (employee id and employee name) having no entry in employee\_expenses

### Employee Details with expenses

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help ||| 
Home Acadgild_64bit
2017-05-28 22:47:49,570 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:47:49,578 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:47:49,578 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:47:49,583 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 22:47:49,584 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:47:49,584 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:47:49,584 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-28 22:47:49,584 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:47:49,617 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:47:49,617 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh,,101,100)
(101,Amitabh,,101,200)
(102,Shahrukh,,102,400)
(102,Shahrukh,,102,100)
(103,Akshay,,)
(104,Anubhav,,104,300)
(105,Pawan,,105,100)
(106,Aamir,,)
(107,Salman,,)
(108,Tanvir,,)
(109,Katrina,,)
(110,Priyanka,,110,400)
(111,Tushar,,)
(112,Ajay,,)
(113,Jubeen,,)
(114,Madhuri,,114,200)
grunt> 
[ACD_BDD2.1_Session_7_Code.zip]
acadgild@localhost:~/... acadgild Downloads *Unsaved Document 1... ACD_BDD2.1_Session_...
To direct input to this VM, click inside or press Ctrl+G.
Windows Type here to search 7:15 PM 09-Jun-2017
```

### Employee whose entry are in expenses file

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help ||| 
Home Acadgild_64bit
COUNTERS:
Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1699450150_0003

2017-05-28 22:53:30,098 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:53:30,102 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:53:30,103 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-05-28 22:53:30,119 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-05-28 22:53:30,120 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-05-28 22:53:30,120 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-05-28 22:53:30,120 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-05-28 22:53:30,120 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-05-28 22:53:30,150 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-05-28 22:53:30,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh,,101,100)
(101,Amitabh,,101,200)
(102,Shahrukh,,102,400)
(102,Shahrukh,,102,100)
(104,Anubhav,,104,300)
(105,Pawan,,105,100)
(110,Priyanka,,110,400)
(114,Madhuri,,114,200)
grunt> 
[ACD_BDD2.1_Session_7_Code.zip]
acadgild@localhost:~/... acadgild Downloads *Unsaved Document 1... ACD_BDD2.1_Session_...
To direct input to this VM, click inside or press Ctrl+G.
Windows Type here to search 7:21 PM 09-Jun-2017
```

## Employee ID and Name of employees whose entry in expenses file.

```
Job DAG:  
job_local1527808699_0008 -> job_local1787543095_0009,  
  
2017-05-29 01:17:47,127 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-29 01:17:47,128 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-29 01:17:47,129 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-29 01:17:47,154 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-29 01:17:47,161 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-29 01:17:47,162 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-29 01:17:47,177 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2017-05-29 01:17:47,177 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2017-05-29 01:17:47,177 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-05-29 01:17:47,177 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max  
2017-05-29 01:17:47,177 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2017-05-29 01:17:47,210 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2017-05-29 01:17:47,210 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(101,Amitabh)  
(102,Shahrukh)  
(104,Anubhav)  
(105,Pawan)  
(110,Priyanka)  
(114,Madhuri)  
grunt>   
Click to switch to "Workspace 2"  
acadgild@localhost:~/... acadgild Downloads *Unsaved Document 1... ACD_BDD2.1_Session_7_Code.zip  
To direct input to this VM, click inside or press Ctrl+G.  
Type here to search 9:49 PM 09-Jun-2017
```

## Employee ID and Name of employees whose name are not in expenses file.

```
Counters:  
Total records written : 8  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local992786292_0005  
  
2017-05-28 22:57:52,443 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:57:52,444 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:57:52,445 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2017-05-28 22:57:52,465 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2017-05-28 22:57:52,465 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2017-05-28 22:57:52,469 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-05-28 22:57:52,469 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max  
2017-05-28 22:57:52,469 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2017-05-28 22:57:52,486 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2017-05-28 22:57:52,486 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(103,Akshay)  
(106,Aamir)  
(107,Salman)  
(108,Ranbir)  
(109,Katrina)  
(111,Tushar)  
(112,Ajay)  
(113,Jubeen)  
grunt>   
Click to switch to "Workspace 2"  
acadgild@localhost:~/... acadgild Downloads *Unsaved Document 1... ACD_BDD2.1_Session_7_Code.zip  
To direct input to this VM, click inside or press Ctrl+G.  
Type here to search 7:25 PM 09-Jun-2017
```