

SESSION 4 ASSIGNMENT 2

Contents of the input dataset

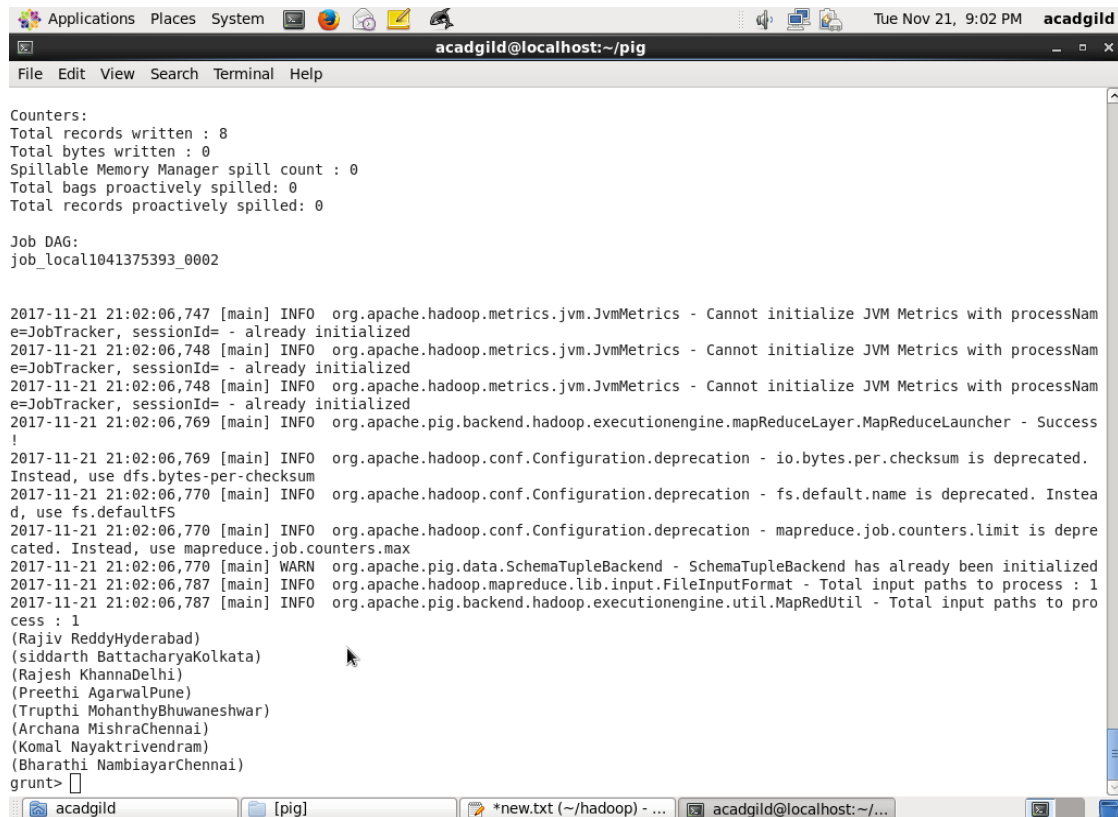
```
001,Rajiv Reddy,21,9848022337,Hyderabad,89
002,siddarth Battacharya,22,9848022338, Bhuwaneshwar,78
003,Rajesh Khanna,22,9848022339,Delhi,90
004,Preethi Agarwal,21,9848022330,Pune,93
005,Trupthi Mohanthy,23,9848022336,Bhuwaneshwar,75
006,Archana Mishra,23,9848022335,Chennai,87
006,Archana Mishra,23,9848022335,Chennai,99
007,Komal Nayak,24,9848022334, Delhi,83
008,Bharathi Nambiayar,24,9848022333,Chennai,72
```

Load Command:

```
grunt> student_details = LOAD 'pig_data/student_details.txt' USING
PigStorage(',') as (id:int, name:chararray, age:int, phone:chararray,
city:chararray, gpa:int);
```

1)Concat:

```
grunt> student_name_concat = foreach student_details Generate CONCAT(name, city);
```



The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The output displays various counters, a job DAG, and a series of log messages from the Hadoop ecosystem. The log messages include warnings about deprecated configurations and a successful completion of the job. The final output of the 'Concat' operation is shown as a list of names and cities concatenated together.

```
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_local1041375393_0002

2017-11-21 21:02:06,747 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:02:06,748 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:02:06,748 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:02:06,769 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-21 21:02:06,769 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-21 21:02:06,770 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-21 21:02:06,770 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-21 21:02:06,770 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 21:02:06,787 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 21:02:06,787 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Rajiv ReddyHyderabad)
(siddarth BattacharyaKolkata)
(Rajesh KhannaDelhi)
(Preethi AgarwalPune)
(Trupthi MohanthyBhuwaneshwar)
(Archana MishraChennai)
(Komal Nayaktrivendram)
(Bharathi NambiayarChennai)
grunt>
```

2)Tokenize:

```
grunt> student_name_tokenize = foreach student_details Generate TOKENIZE(name);
```

```
acadmild@localhost:~/pig
File Edit View Search Terminal Help

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_local369912192_0003

2017-11-21 21:03:09,449 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:03:09,450 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:03:09,450 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:03:09,453 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-21 21:03:09,458 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-21 21:03:09,458 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-21 21:03:09,458 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-21 21:03:09,459 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 21:03:09,475 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 21:03:09,475 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((Rajiv),(Reddy))
((siddarth),(Battacharya))
((Rajesh),(Khanna))
((Preethi),(Agarwal))
((Trupthi),(Mohanthy))
((Archana),(Mishra))
((Komal),(Nayak))
((Bharathi),(Nambiayar))
grunt>
```

3)Sum:

```
grunt> student_gpa_group = Group student_details by city;
```

```
grunt> student_gpa_sum = foreach student_gpa_group Generate (student_details.city, student_details.gpa),SUM(student_details.gpa);
```

```
Applications Places System Tue Nov 21, 9:44 PM acadgild
acadgild@localhost:~/pig
File Edit View Search Terminal Help
Successfully stored 7 records in: "file:/tmp/temp-1841638081/tmp-68818700"

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_local310938594_0004

2017-11-21 21:44:06,550 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:44:06,551 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:44:06,554 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:44:06,560 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-21 21:44:06,560 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-21 21:44:06,565 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-21 21:44:06,565 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-21 21:44:06,565 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 21:44:06,597 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 21:44:06,597 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Pune,{{(4,Preethi Agarwal,21,9848022330,Pune,93)}})
(Delhi,{{(3,Rajesh Khanna,22,9848022339,Delhi,90)}})
(Chennai,{{(8,Bharathi Nambiayar,24,9848022333,Chennai,72)}},{(6,Archana Mishra,23,9848022335,Chennai,87)}})
(Kolkata,{{(2,siddarth Battacharya,22,9848022338,Kolkata,78)}})
(Hyderabad,{{(1,Rajiv Reddy,21,9848022337,Hyderabad,89)}})
(trivendram,{{(7,Komal Nayak,24,9848022334,trivendram,83)}})
(Bhuwaneswar,{{(5,Trupthi Mohanthy,23,9848022336,Bhuwaneswar,75)}})
grunt>
```

```
Applications Places System Tue Nov 21, 9:48 PM acadgild
acadgild@localhost:~/pig
File Edit View Search Terminal Help
Successfully stored 7 records in: "file:/tmp/temp-1841638081/tmp-1258890408"

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

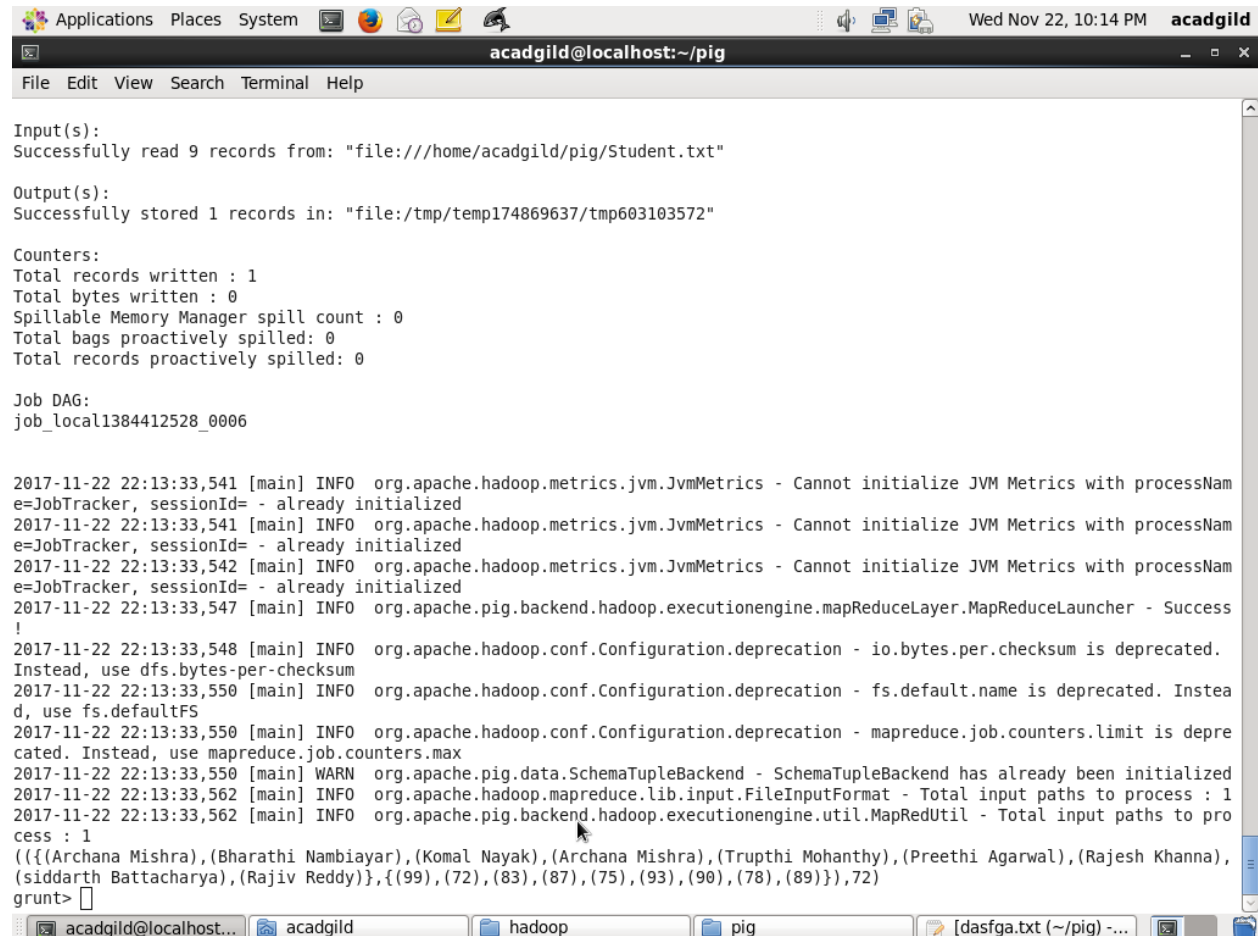
Job DAG:
job_local1743563254_0005

2017-11-21 21:47:46,416 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:47:46,422 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:47:46,423 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-21 21:47:46,429 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-21 21:47:46,430 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-21 21:47:46,430 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-21 21:47:46,430 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-21 21:47:46,430 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-21 21:47:46,448 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-21 21:47:46,448 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(((Pune)),{{(93)}}),93)
(((Delhi)),{{(90)}}),90)
(((Chennai)),{{(Chennai)}},{(72),(87)}),159)
(((Kolkata)),{{(78)}}),78)
(((Hyderabad)),{{(89)}}),89)
(((trivendram)),{{(83)}}),83)
(((Bhuwaneswar)),{{(75)}}),75)
grunt>
```

4)Min:

```
grunt> student_group_all = Group student_details All;

grunt> student_gpa_min = foreach student_group_all Generate
(student_details.firstname, student_details.gpa), MIN(student_details.gpa);
```



```
acadgild@localhost:~/pig
File Edit View Search Terminal Help

Input(s):
Successfully read 9 records from: "file:///home/acadgild/pig/Student.txt"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp174869637/tmp603103572"

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_local1384412528_0006

2017-11-22 22:13:33,541 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:13:33,541 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:13:33,542 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:13:33,547 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-22 22:13:33,548 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-22 22:13:33,550 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:13:33,550 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-22 22:13:33,550 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-22 22:13:33,562 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:13:33,562 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((Archana Mishra),(Bharathi Nambiayar),(Komal Nayak),(Archana Mishra),(Trupthi Mohanthy),(Preethi Agarwal),(Rajesh Khanna),(siddarth Battacharya),(Rajiv Reddy)),{(99),(72),(83),(87),(75),(93),(90),(78),(89)},72)
grunt>
```

5)Max:

```
grunt> student_group_all = Group student_details All;

grunt> student_gpa_max = foreach student_group_all Generate
(student_details.firstname, student_details.gpa), MAX(student_details.gpa);
```

```
Applications Places System [Icons] [System Tray] Wed Nov 22, 10:15 PM acadgild
acadgild@localhost:~/pig
File Edit View Search Terminal Help

Input(s):
Successfully read 9 records from: "file:///home/acadgild/pig/Student.txt"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp174869637/tmp1789084066"

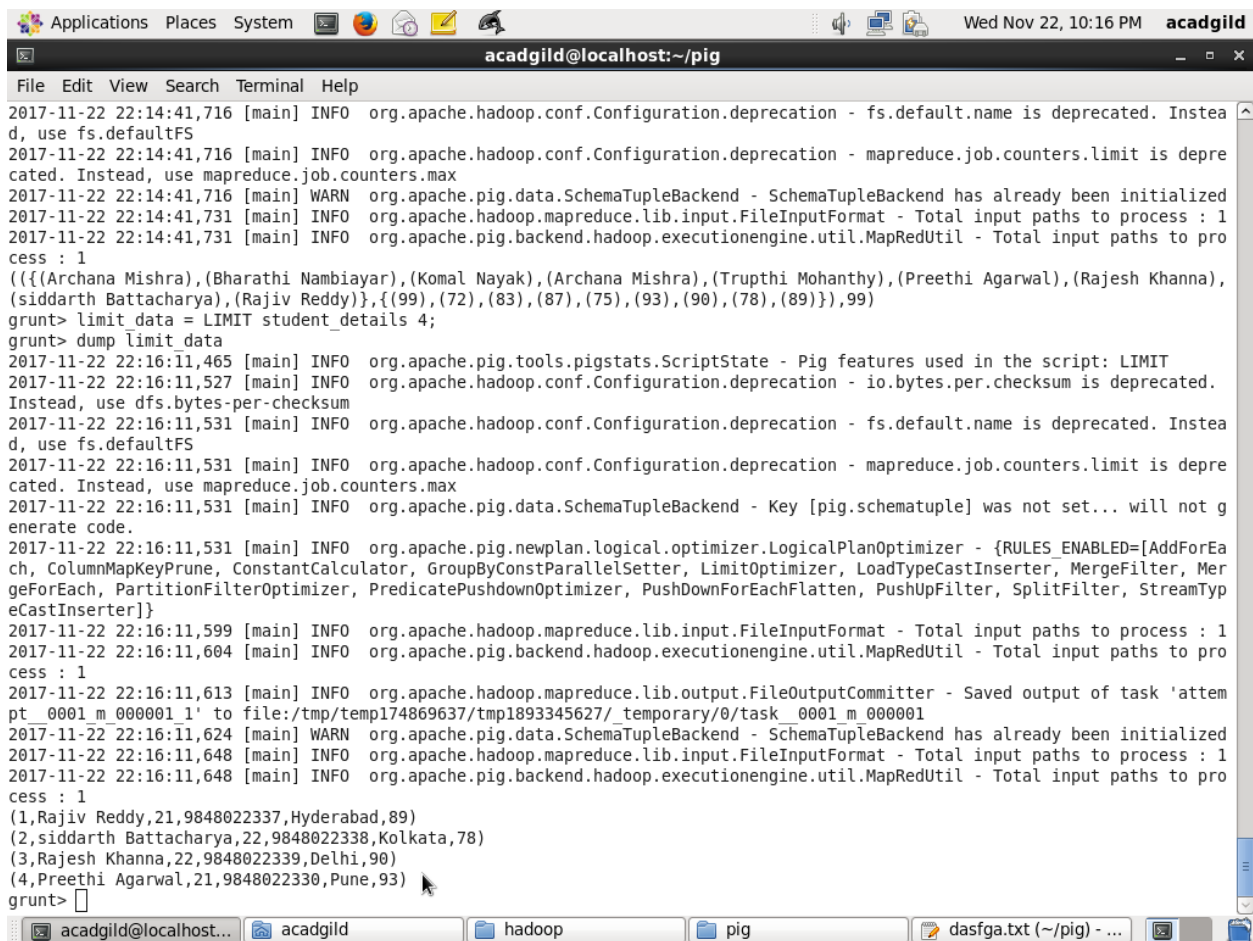
Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local675485429_0007

2017-11-22 22:14:41,687 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:14:41,688 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:14:41,688 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:14:41,710 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-22 22:14:41,711 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-22 22:14:41,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:14:41,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-22 22:14:41,716 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-22 22:14:41,731 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:14:41,731 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
({(Archana Mishra),(Bharathi Nambiayar),(Komal Nayak),(Archana Mishra),(Trupthi Mohanthy),(Preethi Agarwal),(Rajesh Khanna),(siddarth Battacharya),(Rajiv Reddy)},{(99),(72),(83),(87),(75),(93),(90),(78),(89)}},99)
grunt>
```

6)Limit:


```
grunt> limit_data = LIMIT student_details 4;
```



```
2017-11-22 22:14:41,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:14:41,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-22 22:14:41,716 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-22 22:14:41,731 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:14:41,731 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
({(Archana Mishra),(Bharathi Nambiayar),(Komal Nayak),(Archana Mishra),(Trupthi Mohanthy),(Preethi Agarwal),(Rajesh Khanna),(siddarth Battacharya),(Rajiv Reddy)},{(99),(72),(83),(87),(75),(93),(90),(78),(89))},99)
grunt> limit data = LIMIT student_details 4;
grunt> dump limit data
2017-11-22 22:16:11,465 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-11-22 22:16:11,527 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-22 22:16:11,531 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:16:11,531 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-22 22:16:11,531 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-11-22 22:16:11,531 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-11-22 22:16:11,599 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:16:11,604 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-11-22 22:16:11,613 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt_0001_m_000001_1' to file:/tmp/temp174869637/tmp1893345627/_temporary/0/task_0001_m_000001
2017-11-22 22:16:11,624 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-22 22:16:11,648 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:16:11,648 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Rajiv Reddy,21,9848022337,Hyderabad,89)
(2,siddarth Battacharya,22,9848022338,Kolkata,78)
(3,Rajesh Khanna,22,9848022339,Delhi,90)
(4,Preethi Agarwal,21,9848022330,Pune,93)
grunt>
```

7)Store:

```
grunt> STORE student INTO ' hdfs://localhost:9000/pig_Output/ ' USING PigStorage
(',');
```

Applications Places System  Wed Nov 22, 10:23 PM acadgild

acadgild@localhost:~/pig

File Edit View Search Terminal Help

duceTime	MedianReducetime	Alias	Feature	Outputs
job_local1357044580_0008	1	1	n/a	n/a
student_details				
job_local1740695811_0009	1	1	n/a	n/a
ls				

file:///home/acadgild/pig/pig_Output.txt,

Input(s):
Successfully read 4 records from: "file:///home/acadgild/pig/Student.txt"

Output(s):
Successfully stored 4 records in: "file:///home/acadgild/pig/pig_Output.txt"

Counters:
Total records written : 4
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1357044580_0008 -> job_local1740695811_0009,
job_local1740695811_0009

pig_Output.txt

File Edit View Places Help

1, Raj
2, sid
3, Raj
4, Pre

part-r-00000

_SUCCESS

pig_Output.txt "part-r-00000" sele...

```

2017-11-22 22:17:23,739 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:17:23,740 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:17:23,745 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:17:23,751 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:17:23,751 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:17:23,751 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:17:23,766 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>

```

acadgild@local... acadgild hadoop pig [dasfga.txt (~/... pig_Output.txt

8)Distinct:

```
grunt> distinct_data = DISTINCT student_details;
```



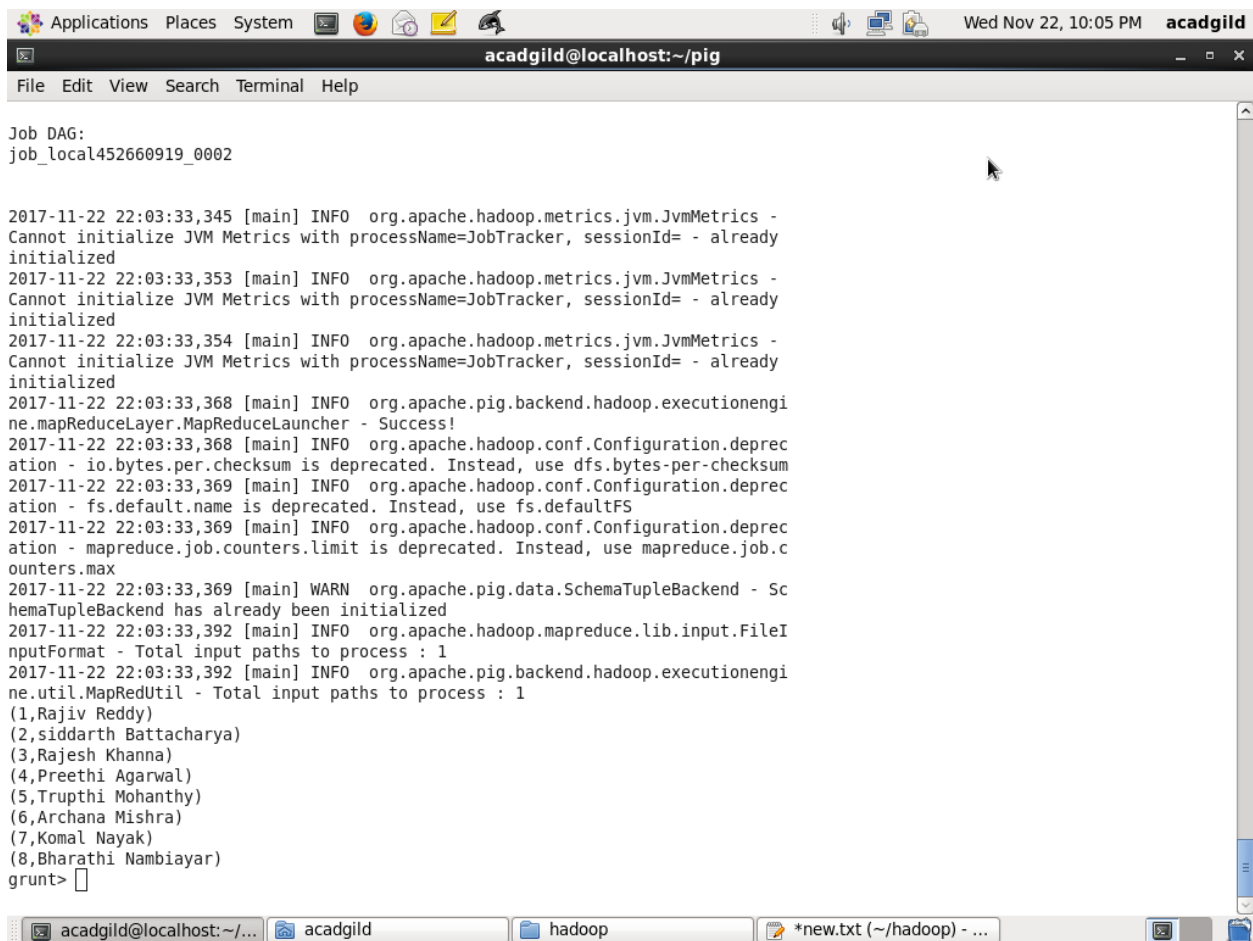
```
acadm@localhost:~/pig
File Edit View Search Terminal Help
Counters:
Total records written : 9
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1799825510_0010

2017-11-22 22:24:25,012 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:24:25,012 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:24:25,013 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:24:25,020 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-22 22:24:25,020 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-22 22:24:25,020 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:24:25,020 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-22 22:24:25,020 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-22 22:24:25,046 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:24:25,046 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Rajiv Reddy,21,9848022337,Hyderabad,89)
(2,siddarth Battacharya,22,9848022338,Kolkata,78)
(3,Rajesh Khanna,22,9848022339,Delhi,90)
(4,Preethi Agarwal,21,9848022330,Pune,93)
(5,Trupthi Mohanthy,23,9848022336,Bhuwaneshwar,75)
(6,Archana Mishra,23,9848022335,Chennai,87)
(6,Archana Mishra,23,9848022335,Chennai,99)
(7,Komal Nayak,24,9848022334,trivendram,83)
(8,Bharathi Nambiayar,24,9848022333,Chennai,72)
grunt>
```

9)Flatten:

```
grunt> student_falppen = FOREACH student_details GENERATE id, flatten(name);
```

The screenshot shows a terminal window titled 'acadgild@localhost:~/pig'. The window contains a series of log messages from Hadoop and Pig. The logs indicate that JVM metrics could not be initialized for the JobTracker. There are also warnings about deprecated configuration options like 'io.bytes.per.checksum' and 'fs.default.name'. The logs show that the Pig backend Hadoop execution engine was successfully initialized. Finally, the logs show the total input paths to the process, which is 1. The terminal prompt is 'grunt>'.

```
Job DAG:
job_local452660919_0002

2017-11-22 22:03:33,345 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics -
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
initialized
2017-11-22 22:03:33,353 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics -
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
initialized
2017-11-22 22:03:33,354 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics -
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
initialized
2017-11-22 22:03:33,368 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-11-22 22:03:33,368 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-22 22:03:33,369 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:03:33,369 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.c
ounters.max
2017-11-22 22:03:33,369 [main] WARN org.apache.pig.data.SchemaTupleBackend - Sc
hemaTupleBackend has already been initialized
2017-11-22 22:03:33,392 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-11-22 22:03:33,392 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(1,Rajiv Reddy)
(2,siddarth Battacharya)
(3,Rajesh Khanna)
(4,Preethi Agarwal)
(5,Trupthi Mohanthy)
(6,Archana Mishra)
(7,Komal Nayak)
(8,Bharathi Nambiayar)
grunt>
```

10) ISEMPY:

Contents of the input dataset

STUDENT_1:

```
001,Rajiv Reddy,21,9848022337,Hyderabad,89
002,siddarth Battacharya,22,9848022338, Bhuwaneshwar,78
003,Rajesh Khanna,22,9848022339,Delhi,90
004,Preethi Agarwal,21,9848022330,Pune,93
005,Trupthi Mohanthy,23,9848022336,Bhuwaneshwar,75
006,Archana Mishra,23,9848022335,Chennai,87
006,Archana Mishra,23,9848022335,Chennai,99
007,Komal Nayak,24,9848022334, Delhi,83
008,Bharathi Nambiayar,24,9848022333,Chennai,72
```

STUDENT_2

```
001,Rajiv Reddy,21,9848022337,Hyderabad,89
002,siddarth Battacharya,22,9848022338, Bhuwaneshwar,78
003,Rajesh Khanna,22,9848022339,Delhi,90
004,Preethi Agarwal,21,9848022330,Pune,93
005,Trupthi Mohanthi,23,9848022336,Bhuwaneshwar,75
006,Archana Mishra,23,9848022335,Chennai,87
006,Archana Mishra,23,9848022335,Chennai,99
007,Komal Nayak,24,9848022334, Delhi,83
008,Bharathi Nambiayar,42,9848022333,Chennai,72
```

```
grunt> student_details = LOAD 'STUDENT_1.txt' USING PigStorage(',') as
(id:int, name:chararray, age:int, phone:chararray, city:chararray, gpa:int);

grunt> student_details_1 = LOAD 'STUDENT_2.txt' USING PigStorage(',') as
(id:int, name:chararray, age:int, phone:chararray, city:chararray, gpa:int);

grunt> cogroup_data = COGROUP student_details by age, student_details_1 by
age;

grunt> isempty_data = filter cogroup_data by IsEmpty(student_details);
```

```
Applications Places System [Icons] [System Tray] Wed Nov 22, 10:47 PM acadgild
acadgild@localhost:~/pig
File Edit View Search Terminal Help

Input(s):
Successfully read 8 records from: "file:///home/acadgild/pig/Student.txt"
Successfully read 8 records from: "file:///home/acadgild/pig/stu.txt"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp174869637/tmp174792832"

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_local1445198437_0023

2017-11-22 22:45:58,251 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:45:58,251 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:45:58,251 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-11-22 22:45:58,255 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-22 22:45:58,255 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-11-22 22:45:58,256 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-11-22 22:45:58,261 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-11-22 22:45:58,261 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-22 22:45:58,278 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-22 22:45:58,278 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(45,{}),{(8,Bharathi Nambiyar,45,9848022333,Chennai,72)}}
grunt>
```