

Lab Assignment 2

Pig Commands

Submitted by: Aadarsha Chapagain

Execute below commands in Pig and take a screenshot of below and post it.

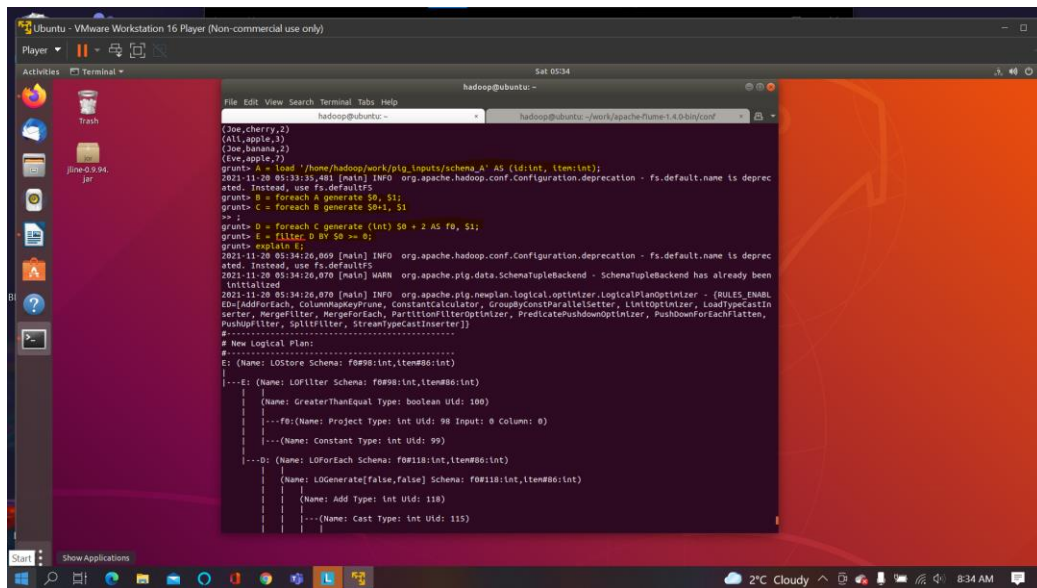
1. Execute Load command with an Example.

```
hadoop@ubuntu:~$ pig -x mapreduce
21/11/20 05:26:11 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/20 05:26:11 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
21/11/20 05:26:11 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2021-11-20 05:26:11,292 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2
015, 11:44:35
2021-11-20 05:26:11,292 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1637414771
289.log
2021-11-20 05:26:11,329 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup
not found
2021-11-20 05:26:11,735 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is dep
recated. Instead, use mapreduce.jobtracker.address
2021-11-20 05:26:11,735 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
2021-11-20 05:26:11,735 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting
to hadoop file system at: hdfs://localhost:8020
2021-11-20 05:26:12,510 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> records = load '/home/hadoop/work/pig_inputs/sample.txt' AS (year, temperature, quality);
2021-11-20 05:26:30,864 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> describe records;
records: {year: bytearray,temperature: bytearray,quality: bytearray}
grunt>
```

2. Execute Dump command with an Example.

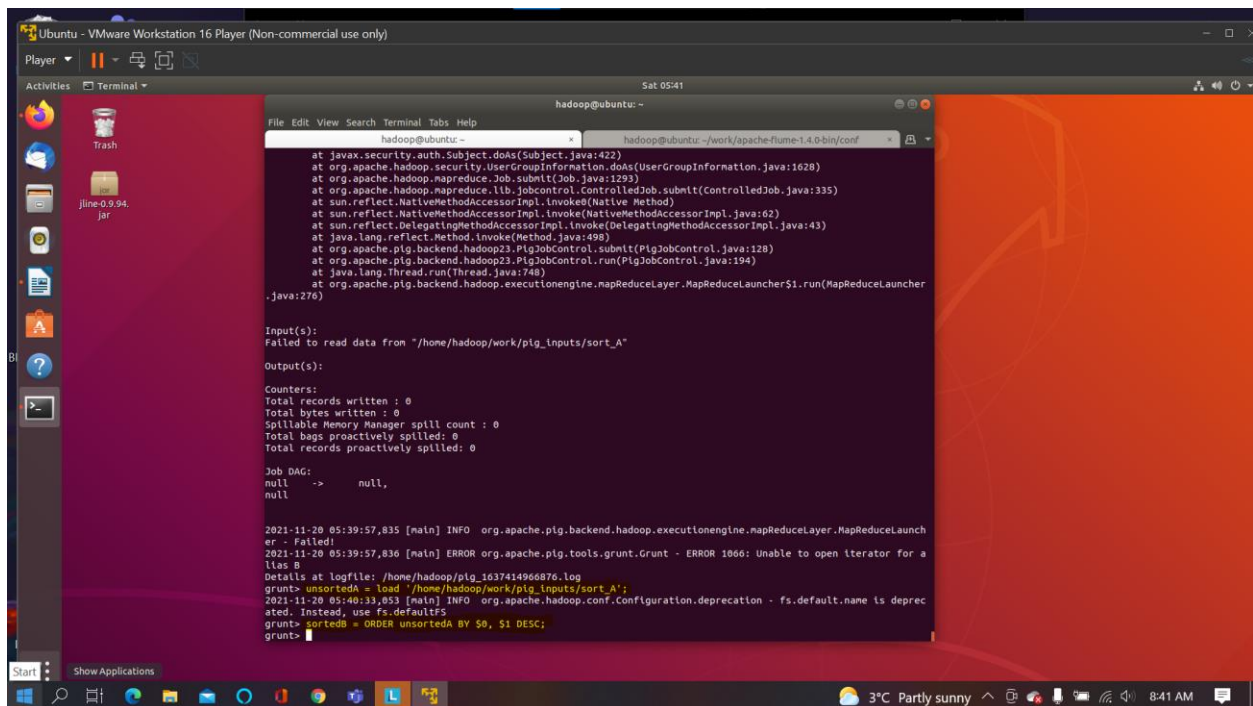
```
hadoop@ubuntu:~$ pig -x mapreduce
21/11/20 05:29:27,308 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
2021-11-20 05:29:27,308 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting
to hadoop file system at: hdfs://localhost:8020
2021-11-20 05:29:28,014 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> A = load '/home/hadoop/work/pig_inputs/foreach.A' AS (f1:chararray, f1:chararray, f2:int);
2021-11-20 05:29:35,271 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> dump A;
2021-11-20 05:29:43,370 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script:
UNKNOWN
2021-11-20 05:29:43,431 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
2021-11-20 05:29:43,438 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set..
Will not generate code.
2021-11-20 05:29:43,476 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABL
ED:[addforeach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelizer, LimitOptimizer, LoadTypeCastIn
serter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten,
PushUpFilter, SplitFilter, StreamTypeCasterInserter]}
2021-11-20 05:29:43,611 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.MRCompiler - Fi
le concatenation threshold: 100 optimize? false
2021-11-20 05:29:43,658 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.MultiQueryOptim
izer - MR plan size before optimization: 1
2021-11-20 05:29:43,658 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.MultiQueryOptim
izer - MR plan size after optimization: 1
2021-11-20 05:29:43,695 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
2021-11-20 05:29:43,754 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at loc
alhost/127.0.0.1:8032
2021-11-20 05:29:44,101 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings
are added to the job
2021-11-20 05:29:44,109 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markres
et.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
2021-11-20 05:29:44,110 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.JobControlCompl
ier - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-20 05:29:44,110 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress
is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2021-11-20 05:29:44,126 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.JobControlCompl
ier - This job cannot be converted run in-process
2021-11-20 05:29:44,437 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.JobControlCompl
ier - Added jar file:/home/hadoop/work/pig-0.15.0/pig-0.15.0-core-h2.jar to DistributedCache through /tmp/temp822
```

3. Execute Filter command with an Example.

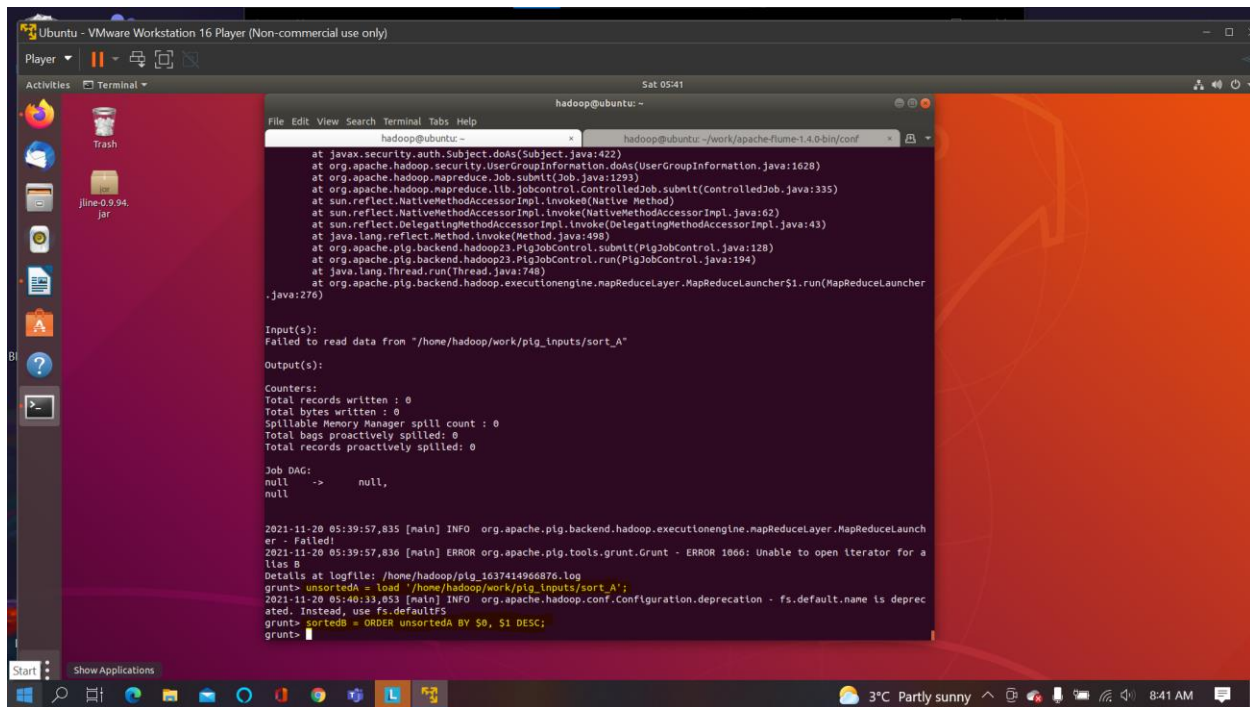


4. Execute Sort command with an Example.

Order is used to sort in pig



5. Execute Order command with an Example.



The screenshot shows a terminal window titled 'hadoop@ubuntu: ~' with a file explorer in the background. The terminal displays the following output:

```
at java.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1628)
at org.apache.hadoop.mapreduce.Job.submit(Job.java:1293)
at org.apache.hadoop.mapreduce.lib.jobcontrol.ControlledJob.submit(ControlledJob.java:335)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethod)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.pig.backend.hadoop23.PigJobControl.submit(PigJobControl.java:128)
at org.apache.pig.backend.hadoop23.PigJobControl.run(PigJobControl.java:194)
at java.lang.Thread.run(Thread.java:748)
at org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher
.java:276)

Input(s):
Failed to read data from "/home/hadoop/work/pig_inputs/sort_A"

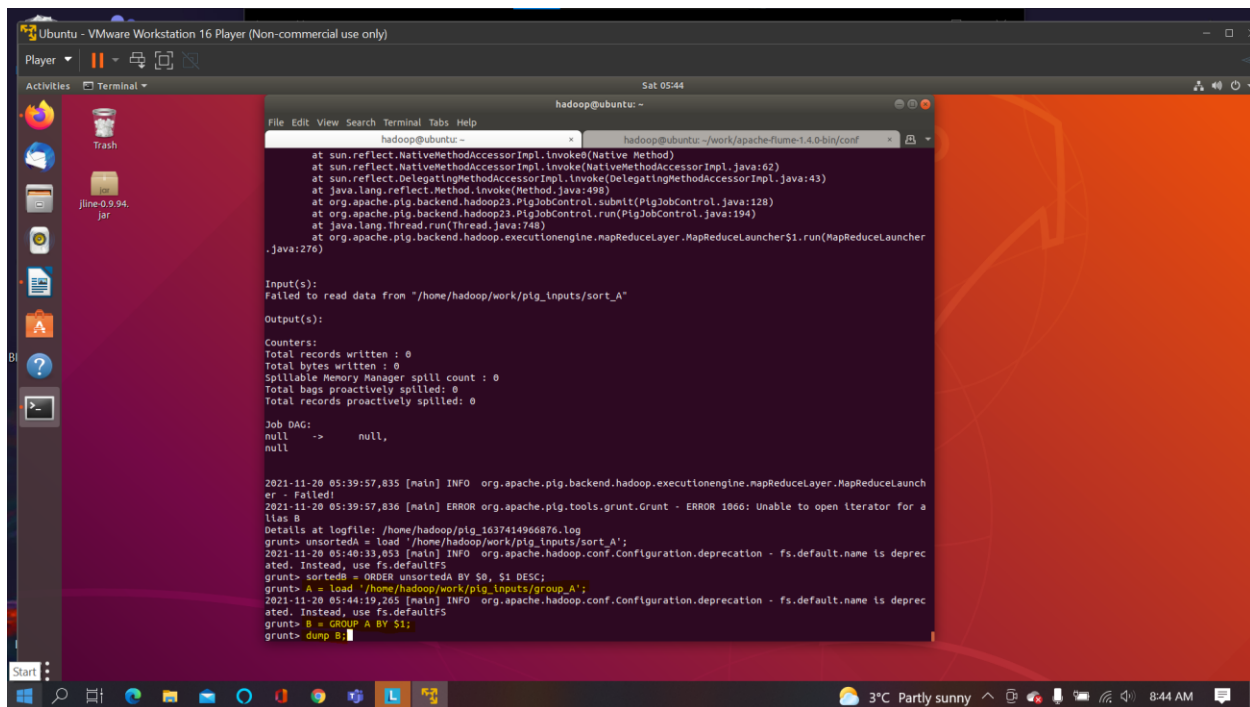
Output(s):

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

Job DAG:
null -> null,
null

2021-11-20 05:39:57,835 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Failed!
2021-11-20 05:39:57,836 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1866: Unable to open iterator for a
lias B
Details at logfile: /home/hadoop/pig_1637414966876.Log
grunt> unsortedA = load '/home/hadoop/work/pig_inputs/sort_A';
2021-11-20 05:40:33,053 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> sortedB = ORDER unsortedA BY $0, $1 DESC;
grunt>
```

6. Execute Group command with an Example.



The screenshot shows a terminal window titled 'hadoop@ubuntu: ~' with a file explorer in the background. The terminal displays the following output:

```
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethod)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.pig.backend.hadoop23.PigJobControl.submit(PigJobControl.java:128)
at org.apache.pig.backend.hadoop23.PigJobControl.run(PigJobControl.java:194)
at java.lang.Thread.run(Thread.java:748)
at org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher
.java:276)

Input(s):
Failed to read data from "/home/hadoop/work/pig_inputs/sort_A"

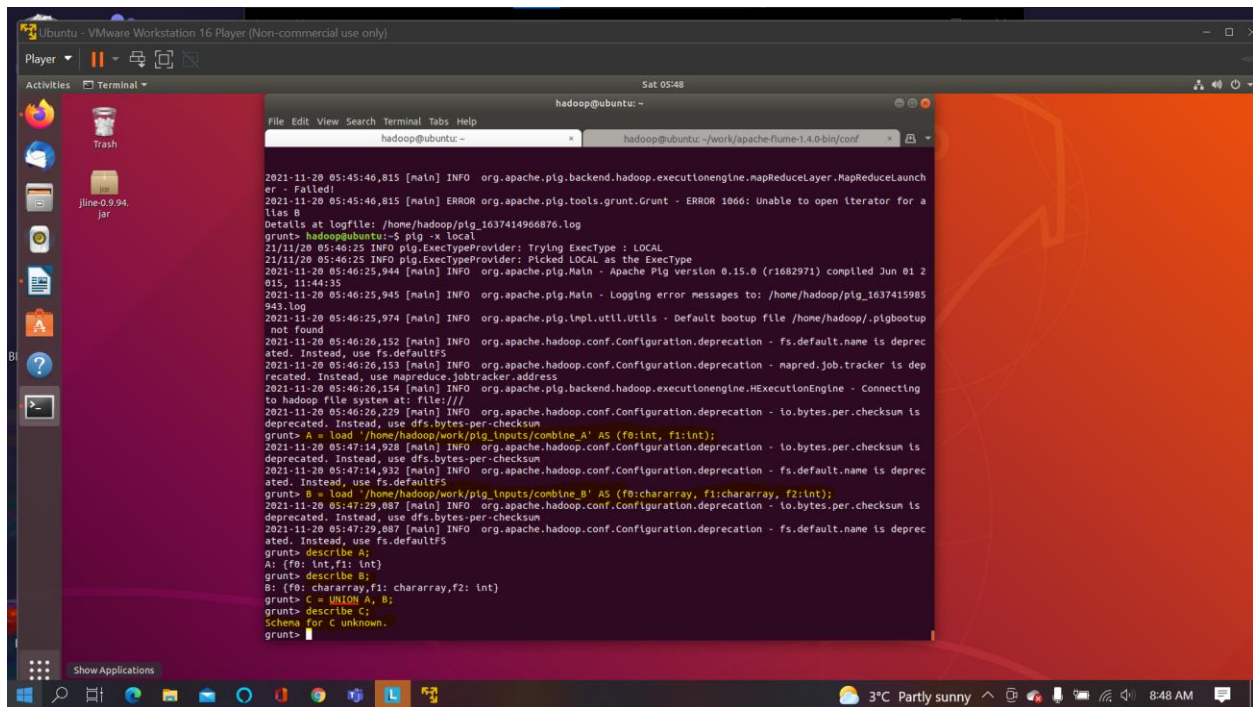
Output(s):

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

Job DAG:
null -> null,
null

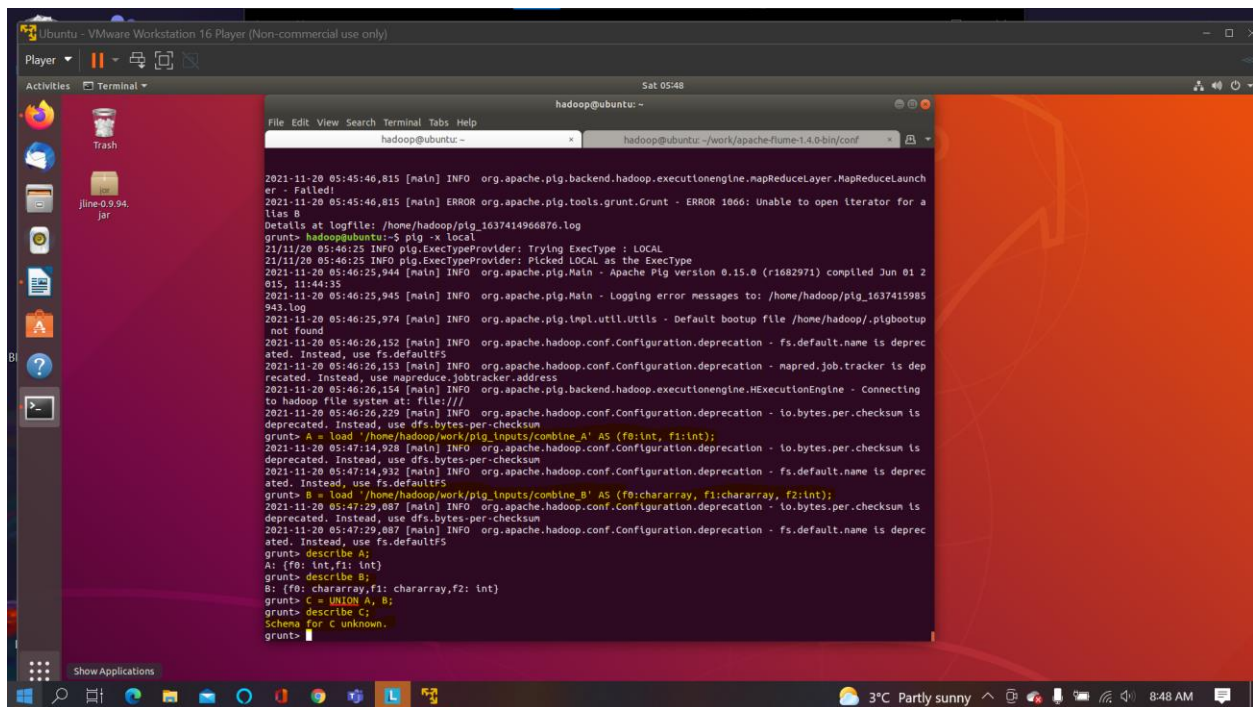
2021-11-20 05:39:57,835 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Failed!
2021-11-20 05:39:57,836 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1866: Unable to open iterator for a
lias B
Details at logfile: /home/hadoop/pig_1637414966876.Log
grunt> unsortedA = load '/home/hadoop/work/pig_inputs/sort_A';
2021-11-20 05:40:33,053 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> sortedB = ORDER unsortedA BY $0, $1 DESC;
2021-11-20 05:44:19,265 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> B = GROUP A BY $1;
grunt> dump B;
```


7. Execute Combine command with an Example.



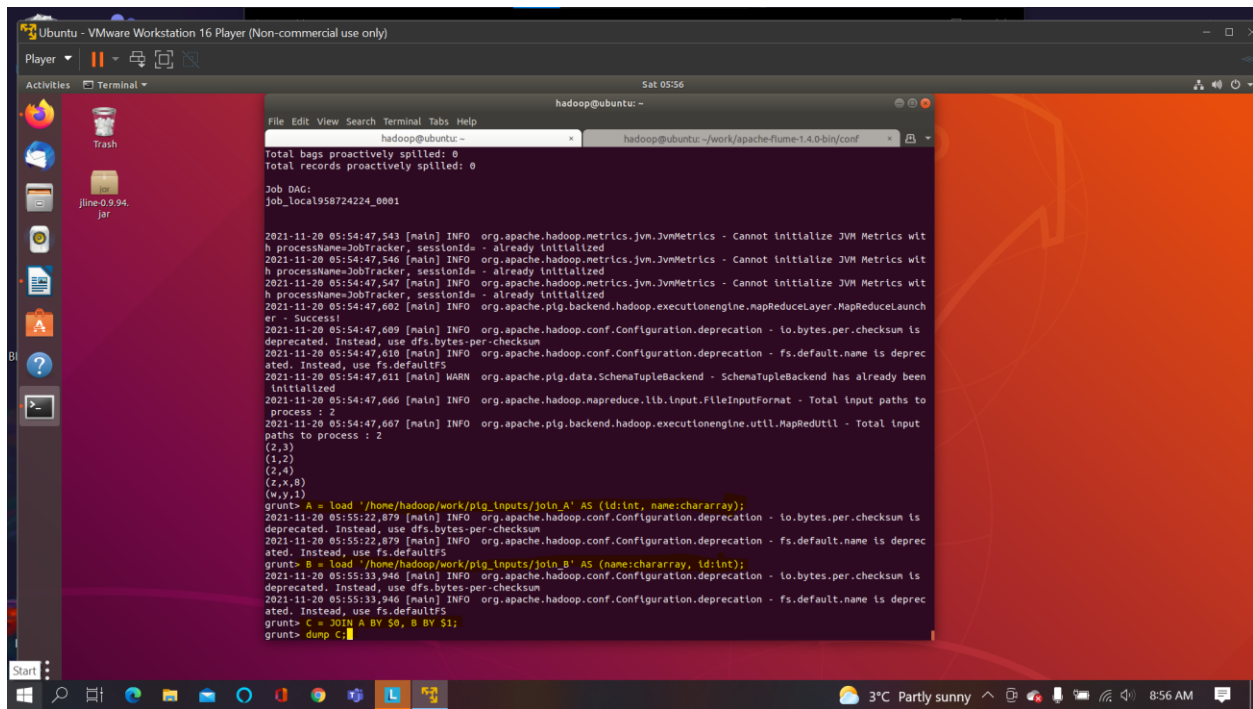
```
2021-11-20 05:45:46,815 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Failed!
2021-11-20 05:45:46,815 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1066: Unable to open iterator for a
lias B
Details at logfile: /home/hadoop/pig_1037414966876.log
grunt> hadoop@ubuntu:~$ pig -x local
21/11/20 05:46:25 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/20 05:46:25 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2021-11-20 05:46:25,944 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2
015, 11:44:35
2021-11-20 05:46:25,945 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1037415985
943.log
2021-11-20 05:46:25,974 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootu
p not found
2021-11-20 05:46:26,152 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
2021-11-20 05:46:26,153 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is dep
recated. Instead, use mapreduce.jobtracker.address
2021-11-20 05:46:26,154 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting
to hadoop file system at: file:///
2021-11-20 05:46:26,229 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is
deprecated. Instead, use dfs.bytes-per-checksum
grunt> A = load '/home/hadoop/work/pig_inputs/combine_A' AS (f0:int, f1:int);
2021-11-20 05:47:14,928 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is
deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:47:14,932 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> B = load '/home/hadoop/work/pig_inputs/combine_B' AS (f0:chararray, f1:chararray, f2:int);
2021-11-20 05:47:29,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is
deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:47:29,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> describe A;
A: (f0: int,f1: int)
grunt> describe B;
B: (f0: chararray,f1: chararray,f2: int)
grunt> C = UNION A, B;
grunt> describe C;
Schema for C unknown.
grunt>
```

8. Execute Union command with an Example.



```
2021-11-20 05:45:46,815 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Failed!
2021-11-20 05:45:46,815 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1066: Unable to open iterator for a
lias B
Details at logfile: /home/hadoop/pig_1037414966876.log
grunt> hadoop@ubuntu:~$ pig -x local
21/11/20 05:46:25 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/20 05:46:25 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2021-11-20 05:46:25,944 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2
015, 11:44:35
2021-11-20 05:46:25,945 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1037415985
943.log
2021-11-20 05:46:25,974 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootu
p not found
2021-11-20 05:46:26,152 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
2021-11-20 05:46:26,153 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is dep
recated. Instead, use mapreduce.jobtracker.address
2021-11-20 05:46:26,154 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting
to hadoop file system at: file:///
2021-11-20 05:46:26,229 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is
deprecated. Instead, use dfs.bytes-per-checksum
grunt> A = load '/home/hadoop/work/pig_inputs/combine_A' AS (f0:int, f1:int);
2021-11-20 05:47:14,928 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is
deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:47:14,932 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> B = load '/home/hadoop/work/pig_inputs/combine_B' AS (f0:chararray, f1:chararray, f2:int);
2021-11-20 05:47:29,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is
deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:47:29,087 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec
ated. Instead, use fs.defaultFS
grunt> describe A;
A: (f0: int,f1: int)
grunt> describe B;
B: (f0: chararray,f1: chararray,f2: int)
grunt> C = UNION A, B;
grunt> describe C;
Schema for C unknown.
grunt>
```

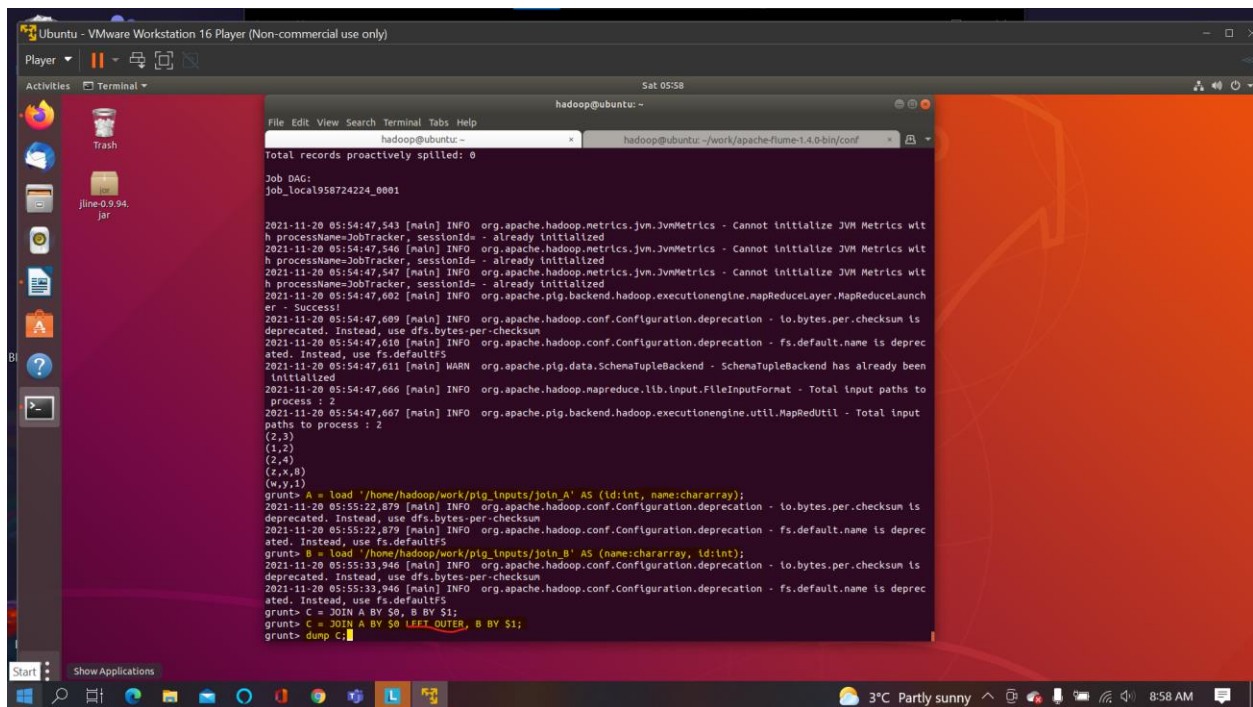
9. Execute Inner Join command with an Example.



The screenshot shows a terminal window titled "hadoop@ubuntu: ~" with a file explorer on the left. The terminal displays the output of a Hadoop job, including logs and a summary of the execution. The job is named "job_local958724224_0001". The logs show various warnings and information messages, including "Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized" and "org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum". The summary shows that the job was successful and that the input paths were processed correctly.

```
hadoop@ubuntu: ~  
File Edit View Search Terminal Tabs Help  
hadoop@ubuntu: ~  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
Job_local958724224_0001  
  
2021-11-20 05:54:47,543 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2021-11-20 05:54:47,546 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2021-11-20 05:54:47,547 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2021-11-20 05:54:47,602 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 05:54:47,610 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2021-11-20 05:54:47,611 [main] WARN org.apache.hadoop.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2021-11-20 05:54:47,666 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2  
2021-11-20 05:54:47,667 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2  
(2,3)  
(1,2)  
(2,4)  
(2,x,8)  
(w,y,1)  
grunt> A = load '/home/hadoop/work/pig_inputs/join_A' AS (id:int, name:chararray);  
2021-11-20 05:55:22,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 05:55:22,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> B = load '/home/hadoop/work/pig_inputs/join_B' AS (name:chararray, id:int);  
2021-11-20 05:55:33,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 05:55:33,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> C = JOIN A BY $0, B BY $1;  
grunt> dump C;
```

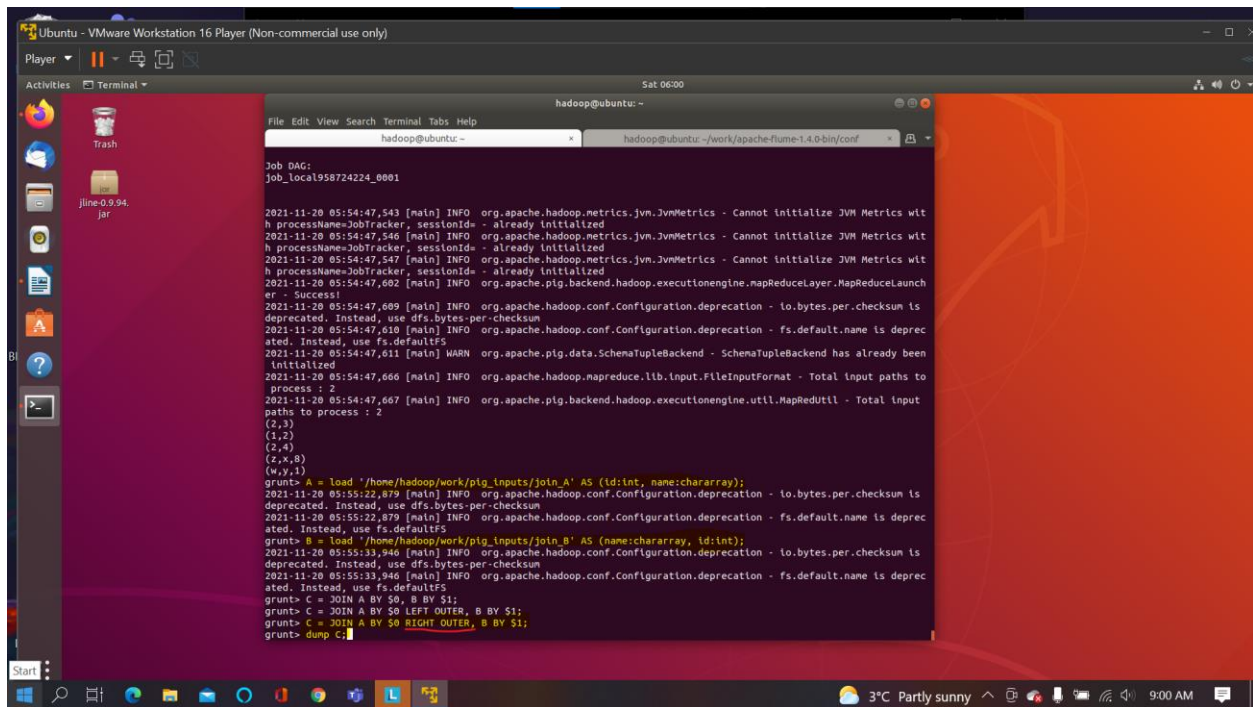
10. Execute Left Outer Join command with an Example.



The screenshot shows a terminal window titled "hadoop@ubuntu: ~" with a file explorer on the left. The terminal displays the output of a Hadoop job, including logs and a summary of the execution. The job is named "job_local958724224_0001". The logs show various warnings and information messages, including "Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized" and "org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum". The summary shows that the job was successful and that the input paths were processed correctly.

```
hadoop@ubuntu: ~  
File Edit View Search Terminal Tabs Help  
hadoop@ubuntu: ~  
Total records proactively spilled: 0  
  
Job DAG:  
Job_local958724224_0001  
  
2021-11-20 05:54:47,543 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2021-11-20 05:54:47,546 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2021-11-20 05:54:47,547 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot Initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2021-11-20 05:54:47,602 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 05:54:47,610 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2021-11-20 05:54:47,611 [main] WARN org.apache.hadoop.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2021-11-20 05:54:47,666 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2  
2021-11-20 05:54:47,667 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2  
(2,3)  
(1,2)  
(2,4)  
(2,x,8)  
(w,y,1)  
grunt> A = load '/home/hadoop/work/pig_inputs/join_A' AS (id:int, name:chararray);  
2021-11-20 05:55:22,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 05:55:22,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> B = load '/home/hadoop/work/pig_inputs/join_B' AS (name:chararray, id:int);  
2021-11-20 05:55:33,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 05:55:33,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> C = JOIN A BY $0, B BY $1;  
grunt> C = JOIN A BY $0 LEFT OUTER, B BY $1;  
grunt> dump C;
```

11. Execute Right outer join command with an Example.

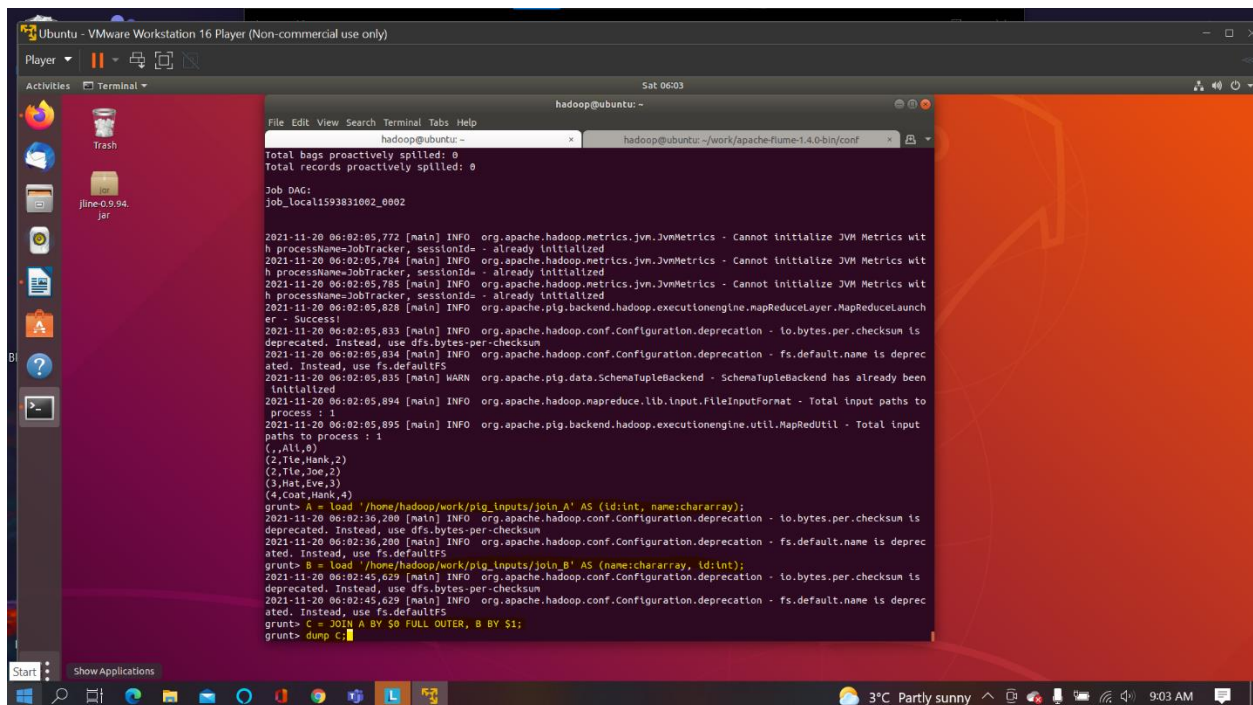


```
hadoop@ubuntu:--
hadoop@ubuntu:~/work/apache-flume-1.4.0-bin/conf

Job DAG:
Job_local958724224_0001

2021-11-20 05:54:47,543 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 05:54:47,546 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 05:54:47,547 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 05:54:47,602 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2021-11-20 05:54:47,609 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:54:47,610 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-20 05:54:47,611 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-20 05:54:47,666 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2
2021-11-20 05:54:47,667 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 2
(2,3)
(1,2)
(2,4)
(2,x,8)
(W,Y,1)
grunt> A = load '/home/hadoop/work/pig_inputs/join_A' AS (id:int, name:chararray);
2021-11-20 05:55:22,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:55:22,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = load '/home/hadoop/work/pig_inputs/join_B' AS (name:chararray, id:int);
2021-11-20 05:55:33,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 05:55:33,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> C = JOIN A BY $0, B BY $1;
grunt> C = JOIN A BY $0 LEFT OUTER, B BY $1;
grunt> C = JOIN A BY $0 RIGHT OUTER, B BY $1;
grunt> dump C;
```

12. Execute Full Join command with an Example.



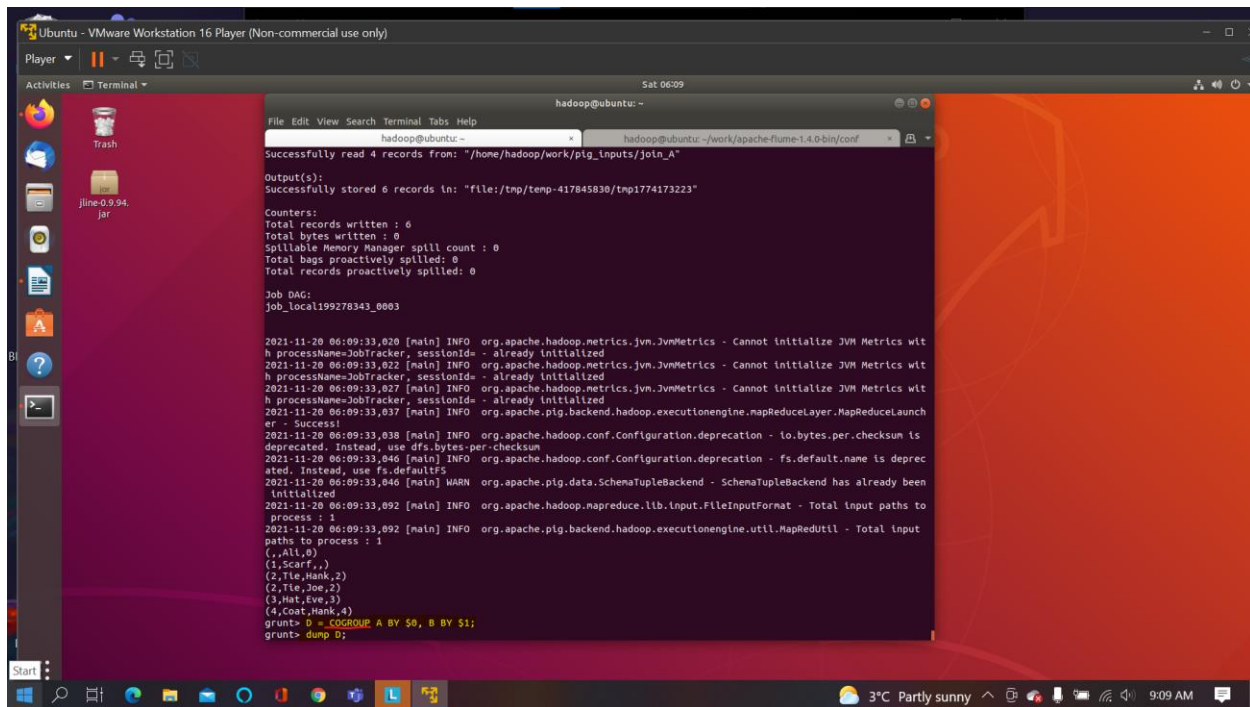
```
hadoop@ubuntu:--
hadoop@ubuntu:~/work/apache-flume-1.4.0-bin/conf

Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job_local1593831002_0002

2021-11-20 06:02:05,772 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:02:05,784 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:02:05,785 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:02:05,828 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2021-11-20 06:02:05,833 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 06:02:05,834 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-20 06:02:05,835 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-20 06:02:05,894 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-20 06:02:05,895 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(, ,All,8)
(2,Tle,Hank,2)
(2,Tle,Joe,2)
(3,Hat,Eve,3)
(4,Coat,Hank,4)
grunt> A = load '/home/hadoop/work/pig_inputs/join_A' AS (id:int, name:chararray);
2021-11-20 06:02:36,280 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 06:02:36,280 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = load '/home/hadoop/work/pig_inputs/join_B' AS (name:chararray, id:int);
2021-11-20 06:02:45,629 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 06:02:45,629 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> C = JOIN A BY $0 FULL OUTER, B BY $1;
grunt> dump C;
```


13. Execute CoGroup command with an Example.



```
hadoop@ubuntu:~$ hadoop@ubuntu:~/work/apache-flume-1.4.0-bin/conf$
Successfully read 4 records from: "/home/hadoop/work/pig_inputs/join_A"

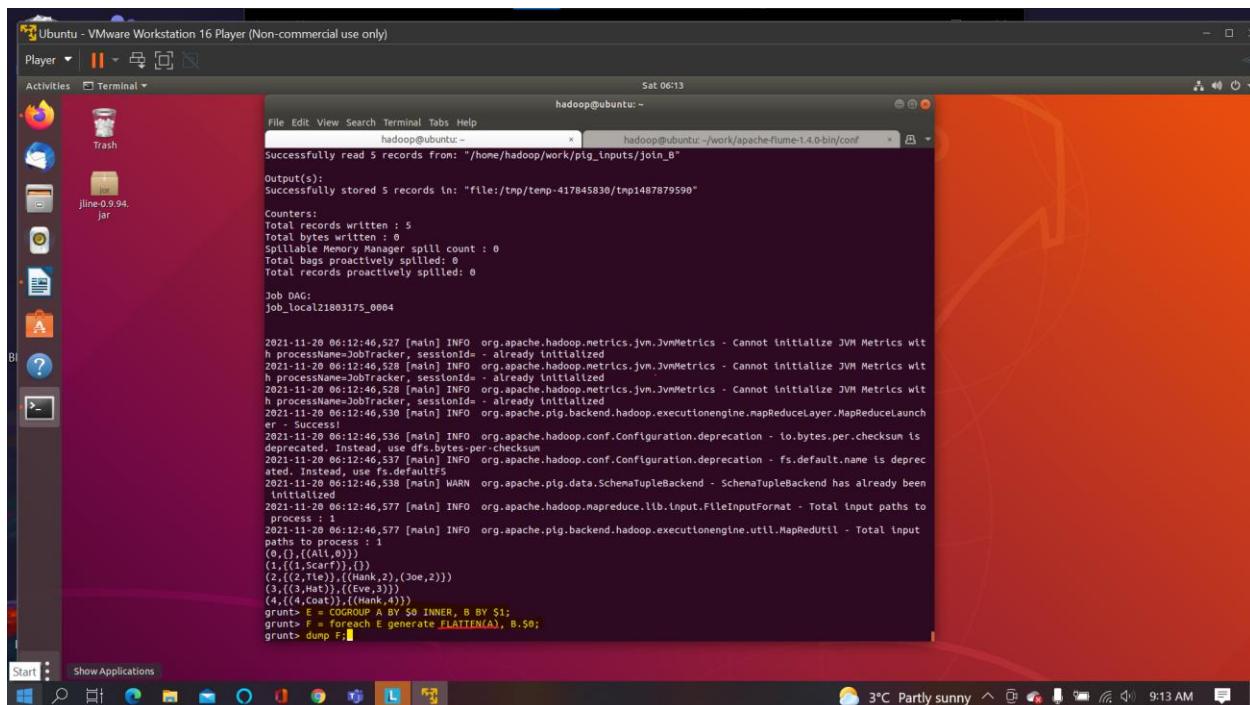
Output(s):
Successfully stored 6 records in: "file:/tmp/temp-417845830/tmp1774173223"

Counters:
Total records written : 6
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job_local199278343_0003

2021-11-20 06:09:33,020 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:09:33,022 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:09:33,027 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:09:33,037 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-20 06:09:33,038 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 06:09:33,046 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-20 06:09:33,046 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-20 06:09:33,092 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-20 06:09:33,092 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(,(All,0))
(1,(Scar,))
(2,(file,Hank,2))
(3,(file,Joe,2))
(3,(Hat,Eve,3))
(4,(Coat,Hank,4))
grunt> D = COGROUP A BY $0, B BY $1;
grunt> dump D;
```

14. Execute Flatten command with an Example.



```
hadoop@ubuntu:~$ hadoop@ubuntu:~/work/apache-flume-1.4.0-bin/conf$
Successfully read 5 records from: "/home/hadoop/work/pig_inputs/join_B"

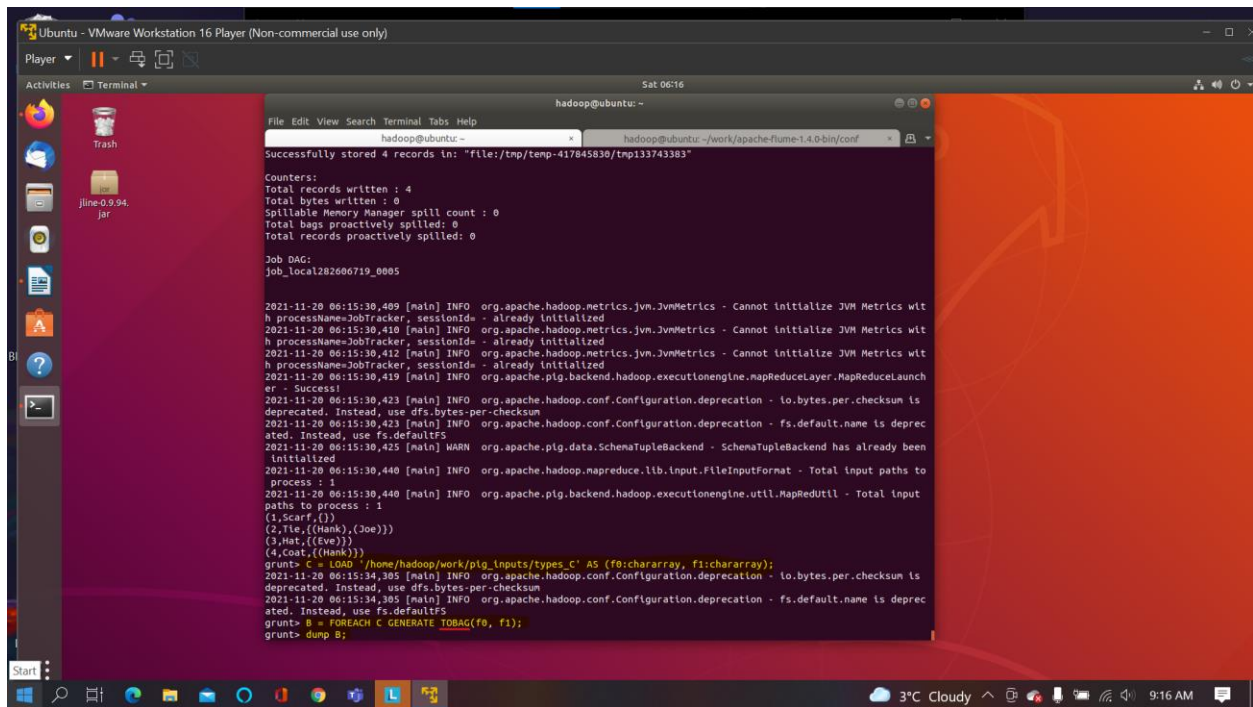
Output(s):
Successfully stored 5 records in: "file:/tmp/temp-417845830/tmp1487879590"

Counters:
Total records written : 5
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job_local21803175_0004

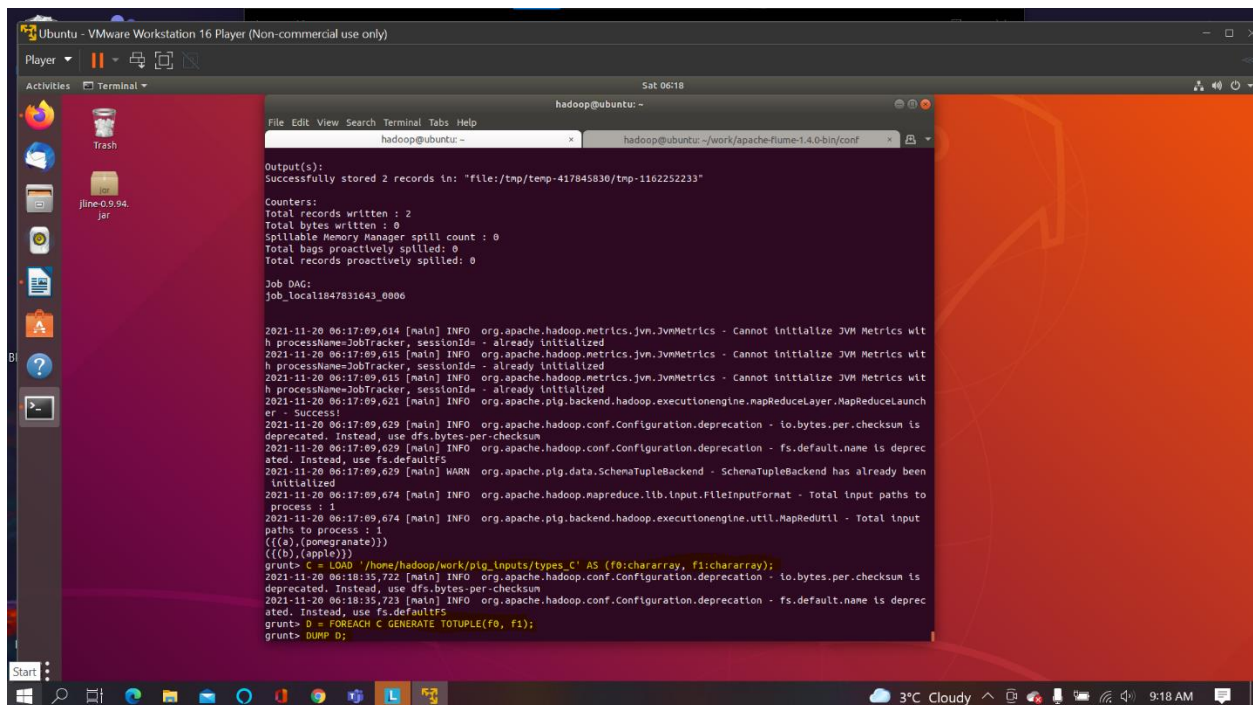
2021-11-20 06:12:46,527 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:12:46,528 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:12:46,528 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-20 06:12:46,530 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-20 06:12:46,536 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-20 06:12:46,537 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-20 06:12:46,538 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-20 06:12:46,577 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-20 06:12:46,577 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(0,(),((All,0)))
(1,((1,(Scar)),))
(2,((2,(file)),((Hank,2)),(Joe,2)))
(3,((3,(Hat)),((Eve,3)))
(4,((4,(Coat)),((Hank,4)))
grunt> E = COGROUP A BY $0 INNER, B BY $1;
grunt> F = foreach E generate FLATTEN(A), B.$0;
grunt> dump F;
```

15. Execute ToBag command with an Example.



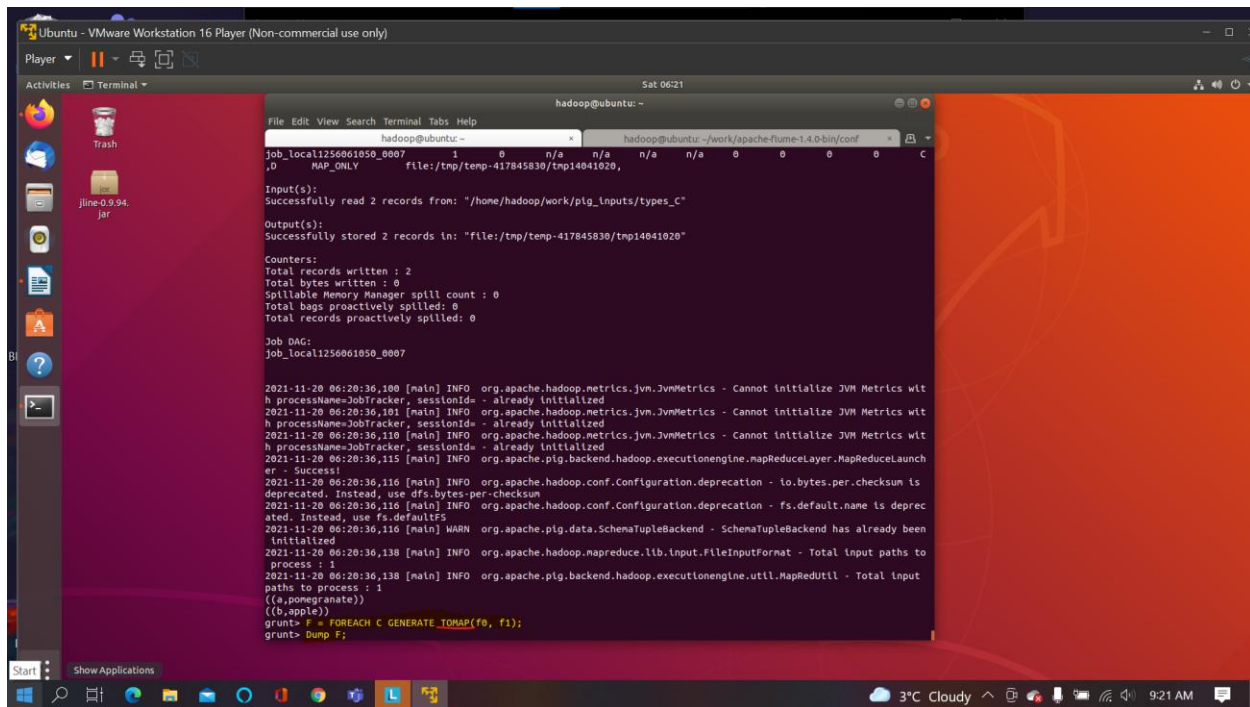
```
hadoop@ubuntu: ~  
File Edit View Search Terminal Tabs Help  
hadoop@ubuntu: ~/work/apache-flume-1.4.0-bin/conf  
Successfully stored 4 records in: "file:/tmp/temp-417845830/tmp133743383"  
  
Counters:  
Total records written : 4  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local282686719_0005  
  
2021-11-20 06:15:30,409 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics wit  
h processName=JobTracker, sessionId= - already initialized  
2021-11-20 06:15:30,410 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics wit  
h processName=JobTracker, sessionId= - already initialized  
2021-11-20 06:15:30,412 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics wit  
h processName=JobTracker, sessionId= - already initialized  
2021-11-20 06:15:30,419 [main] INFO org.apache.hadoop.executionengine.mapreduce_layer.MapReduceLaunch  
er - Success!  
2021-11-20 06:15:30,423 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is  
deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 06:15:30,423 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec  
ated. Instead, use fs.defaultFS  
2021-11-20 06:15:30,425 [main] WARN org.apache.plg.data.SchemaTupleBackend - SchemaTupleBackend has already been  
initialized  
2021-11-20 06:15:30,440 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to  
process : 1  
2021-11-20 06:15:30,440 [main] INFO org.apache.plg.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths to process : 1  
(1,Scarf,())  
(2,Tie,((Hank),(Joe)))  
(3,Mat,((Ive)))  
(4,Coat,((Hank)))  
grunt> C = LOAD "/home/hadoop/work/plg_inputs/types.C" AS (f0:chararray, f1:chararray);  
2021-11-20 06:15:34,305 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is  
deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 06:15:34,305 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec  
ated. Instead, use fs.defaultFS  
grunt> B = FOREACH C GENERATE TOBAG(f0, f1);  
grunt> dump B;
```

16. Execute ToTuple command with an Example.

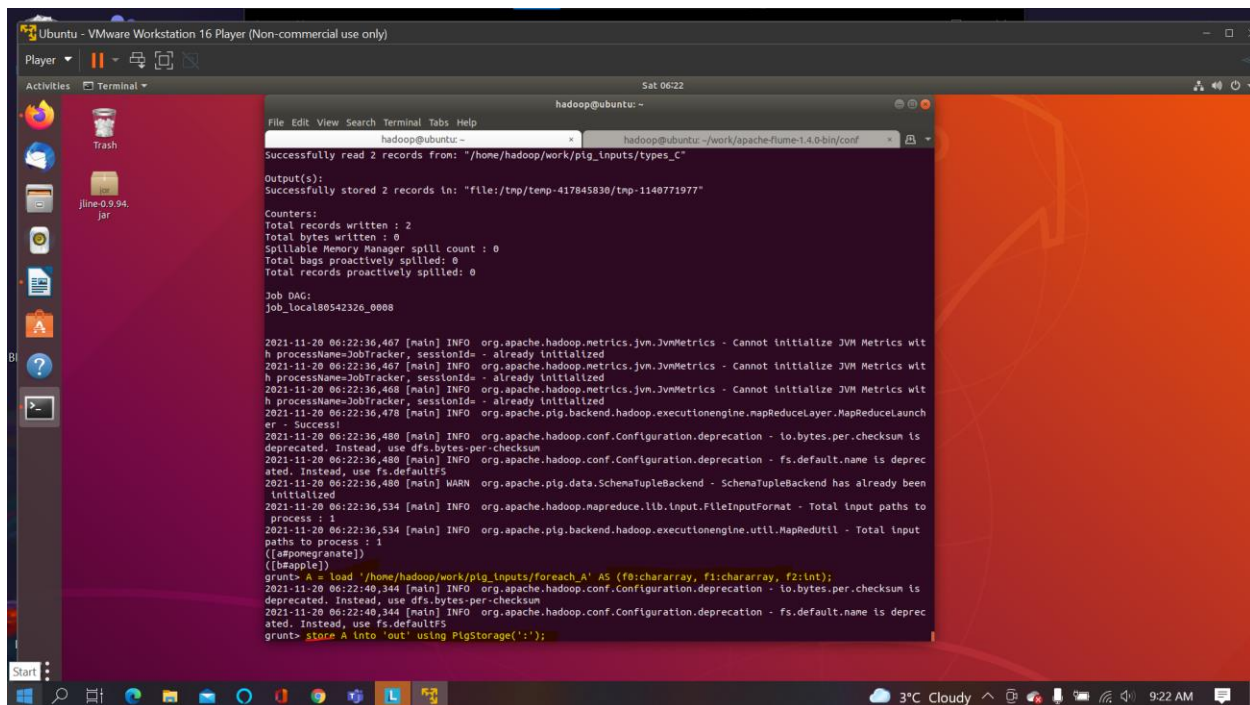


```
hadoop@ubuntu: ~  
File Edit View Search Terminal Tabs Help  
hadoop@ubuntu: ~/work/apache-flume-1.4.0-bin/conf  
Output(s):  
Successfully stored 2 records in: "file:/tmp/temp-417845830/tmp-1102252233"  
  
Counters:  
Total records written : 2  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local1847831643_0006  
  
2021-11-20 06:17:09,614 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics wit  
h processName=JobTracker, sessionId= - already initialized  
2021-11-20 06:17:09,615 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics wit  
h processName=JobTracker, sessionId= - already initialized  
2021-11-20 06:17:09,615 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics wit  
h processName=JobTracker, sessionId= - already initialized  
2021-11-20 06:17:09,621 [main] INFO org.apache.hadoop.executionengine.mapreduce_layer.MapReduceLaunch  
er - Success!  
2021-11-20 06:17:09,629 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is  
deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 06:17:09,629 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec  
ated. Instead, use fs.defaultFS  
2021-11-20 06:17:09,629 [main] WARN org.apache.plg.data.SchemaTupleBackend - SchemaTupleBackend has already been  
initialized  
2021-11-20 06:17:09,674 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to  
process : 1  
2021-11-20 06:17:09,674 [main] INFO org.apache.plg.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths to process : 1  
(((a,(pomegranate))))  
(((b,(apple))))  
grunt> C = LOAD "/home/hadoop/work/plg_inputs/types.C" AS (f0:chararray, f1:chararray);  
2021-11-20 06:18:35,722 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is  
deprecated. Instead, use dfs.bytes-per-checksum  
2021-11-20 06:18:35,723 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprec  
ated. Instead, use fs.defaultFS  
grunt> D = FOREACH C GENERATE TOTUPLE(f0, f1);  
grunt> DUMP D;
```

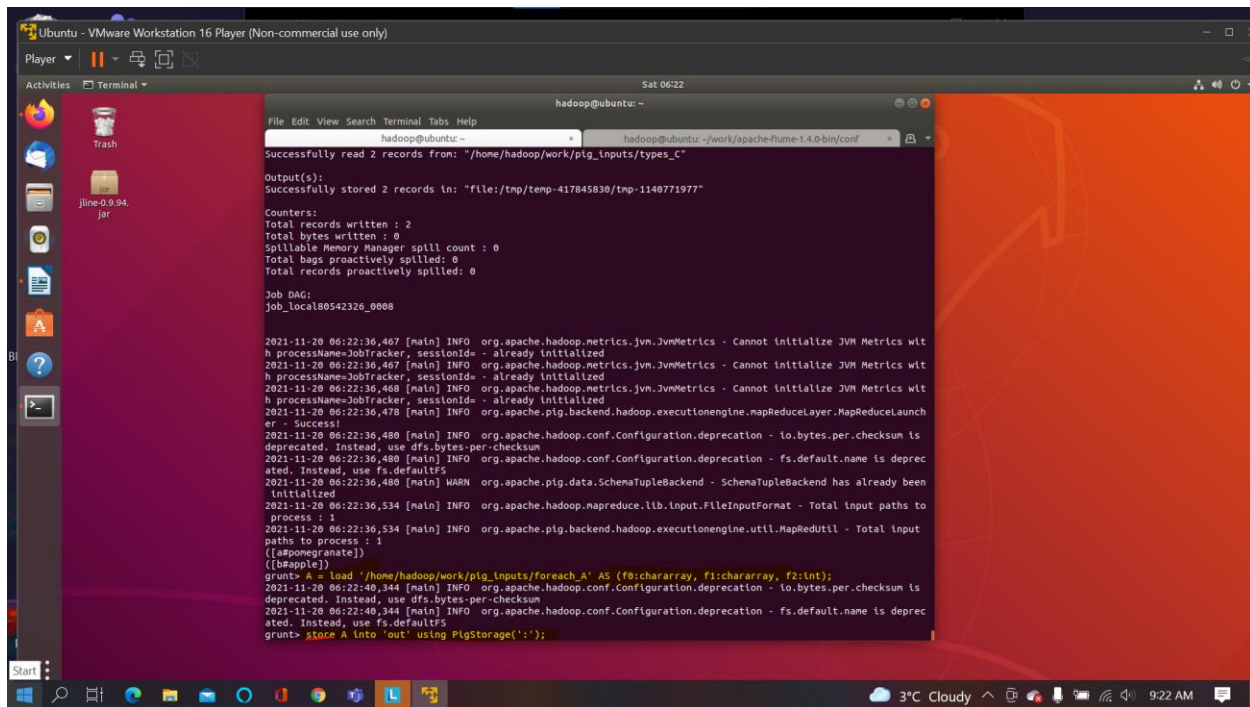

17. Execute ToMap command with an Example.



18. Execute Store command with an Example.



19. Execute PigStorage command with an Example.



20. Execute Stream command with an Example

