

## PRACTICAL 9

### To implement Word Count problem using Pig

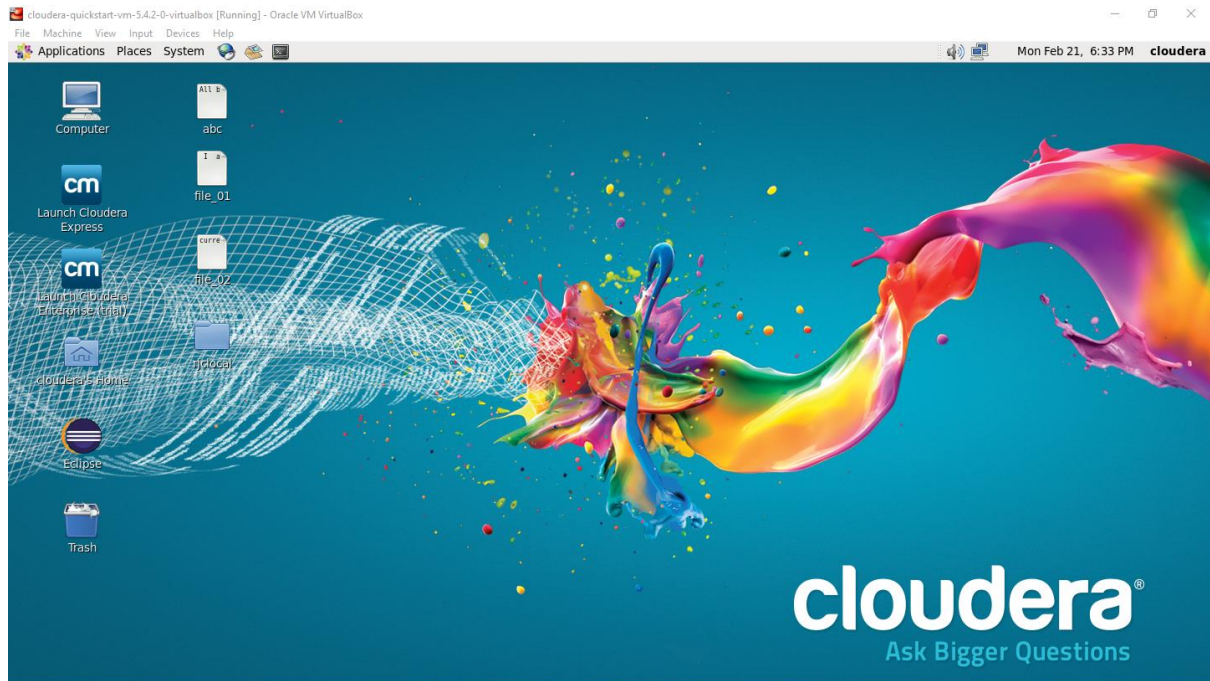
#### Apache Pig

- **Apache Pig** is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.
  - The language used for Pig is Pig Latin. The Pig scripts get internally converted to Map Reduce jobs and get executed on data stored in HDFS.
  - Apart from that, Pig can also execute its job in Apache Tez or Apache Spark.
  - Pig can handle any type of data, i.e., structured, semi-structured or unstructured and stores the corresponding results into Hadoop Data File System.
  - Every task which can be achieved using PIG can also be achieved using java used in MapReduce.
- Pig's language layer currently consists of a textual language called Pig Latin, which has the following key properties:
- **Ease of programming.** It is trivial to achieve parallel execution of simple, "embarrassingly parallel" data analysis tasks. Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.
  - **Optimization opportunities.** The way in which tasks are encoded permits the system to optimize their execution automatically, allowing the user to focus on semantics rather than efficiency.
  - **Extensibility.** Users can create their own functions to do special-purpose processing.

## To implement Word Count problem using Pig

### Steps:

1) Start the cloudera.



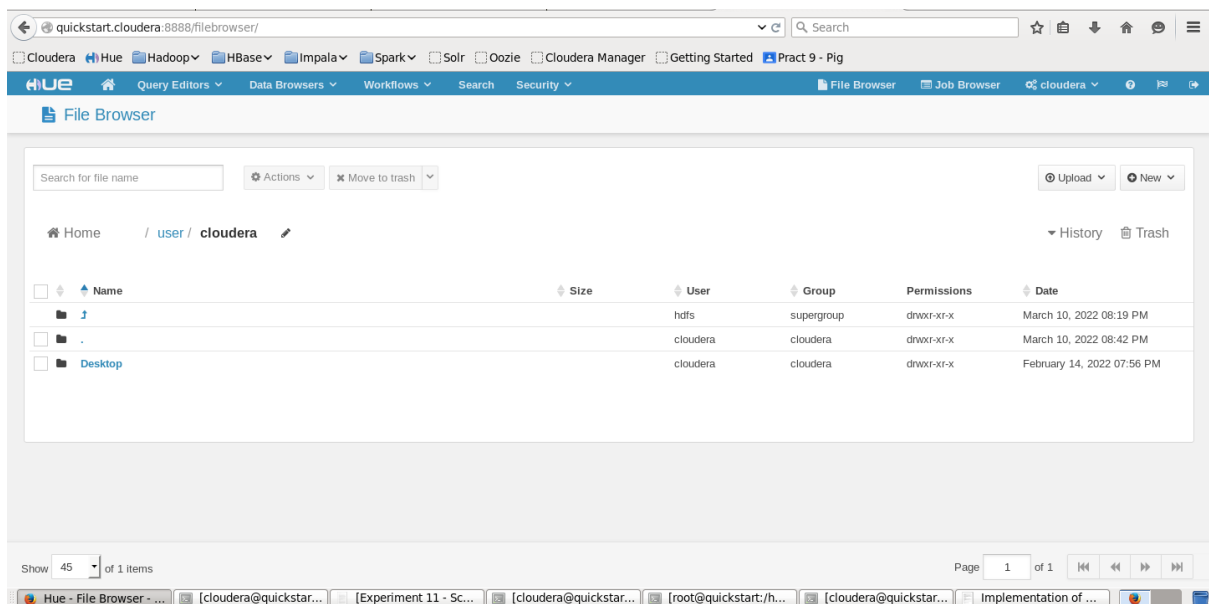
2) Open the browser. And then open Hue and login.



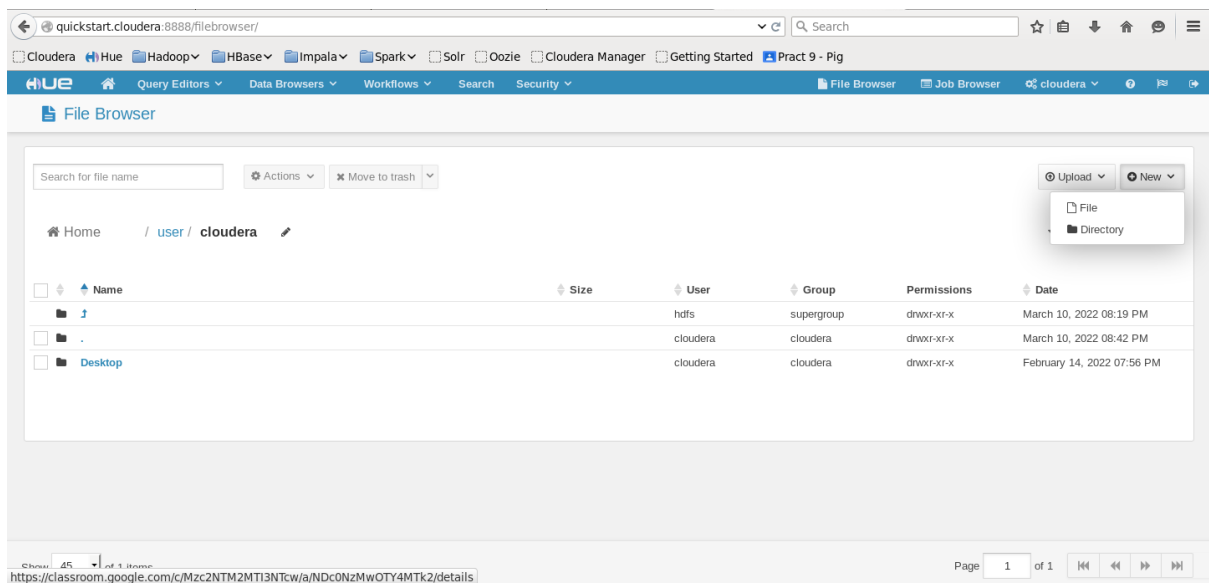
Sign in to continue to Hue

	<input type="text" value="cloudera"/>
	<input type="password" value="....."/>

### 3) In Hue Go to file browser and Now open the directory /user/cloudera



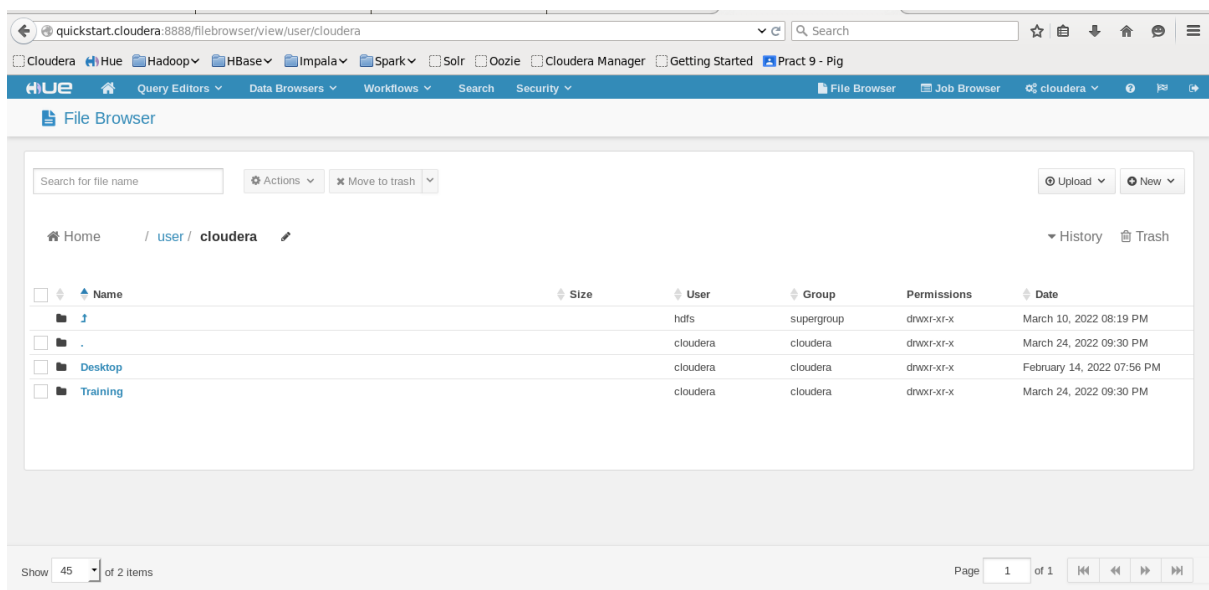
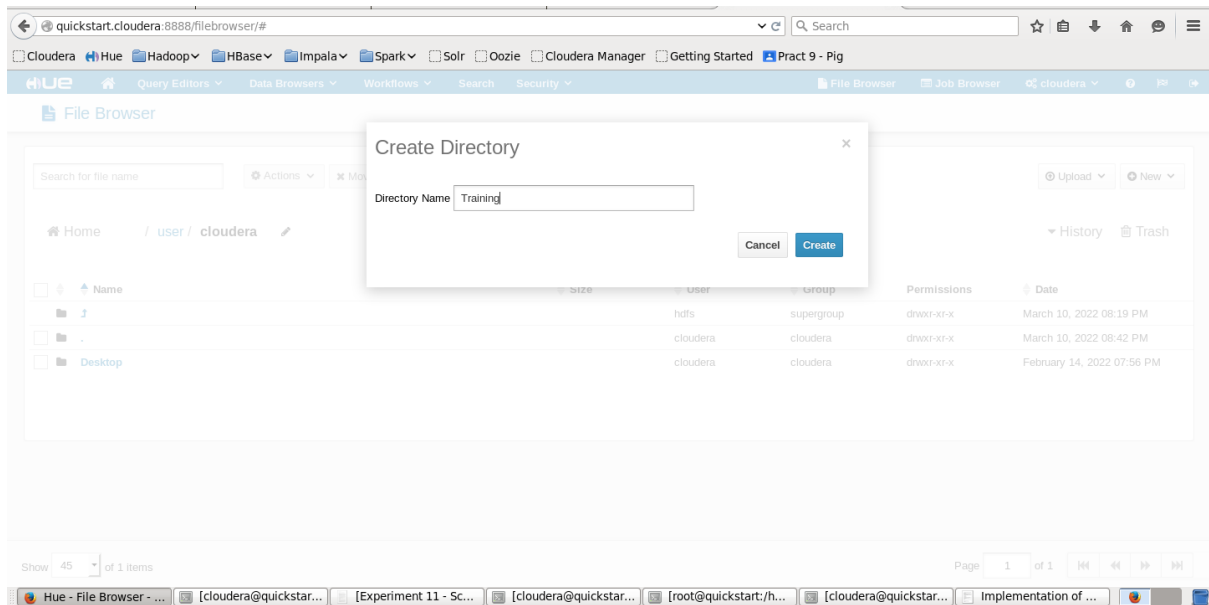
### 4) Now we are creating the directory as Training inside /user/cloudera



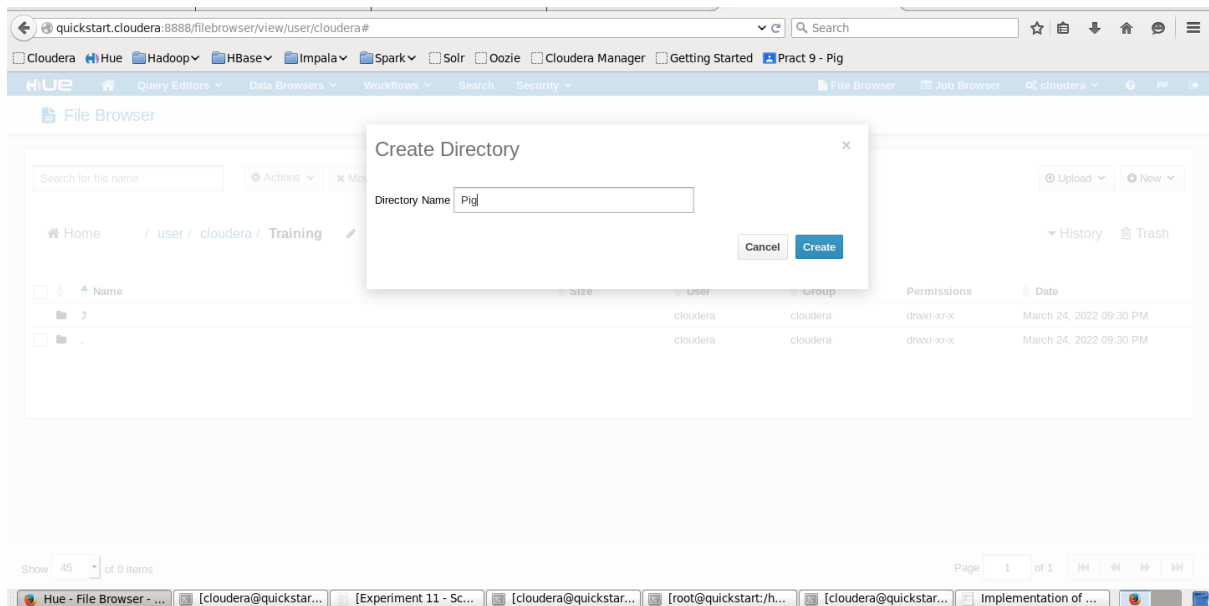
In File Browser we have New option in right corner

Click on New → Directory

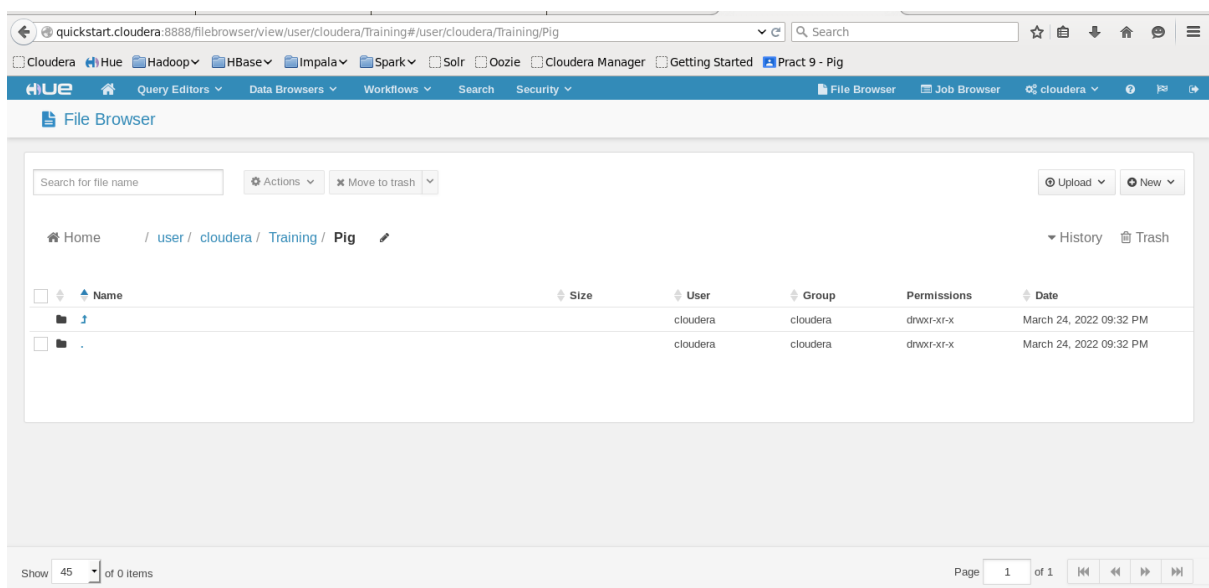
Give the directory name And click on Create



5) After creating Training directory now creating the Pig directory inside Training.

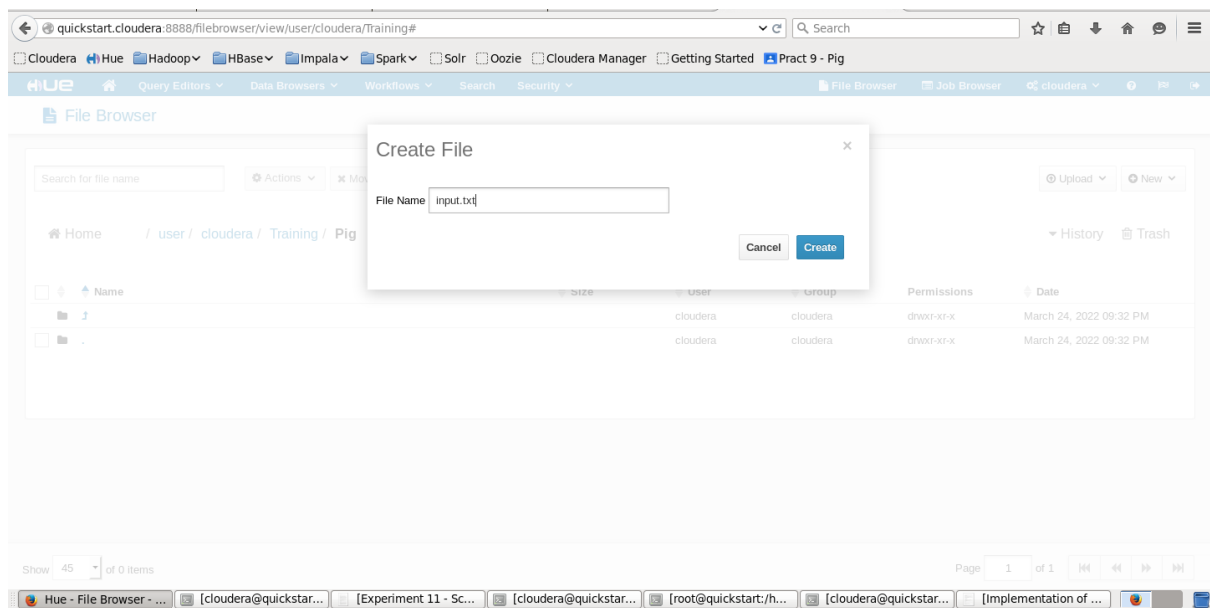


6) Pig directory has been created inside /user/cloudera/Training

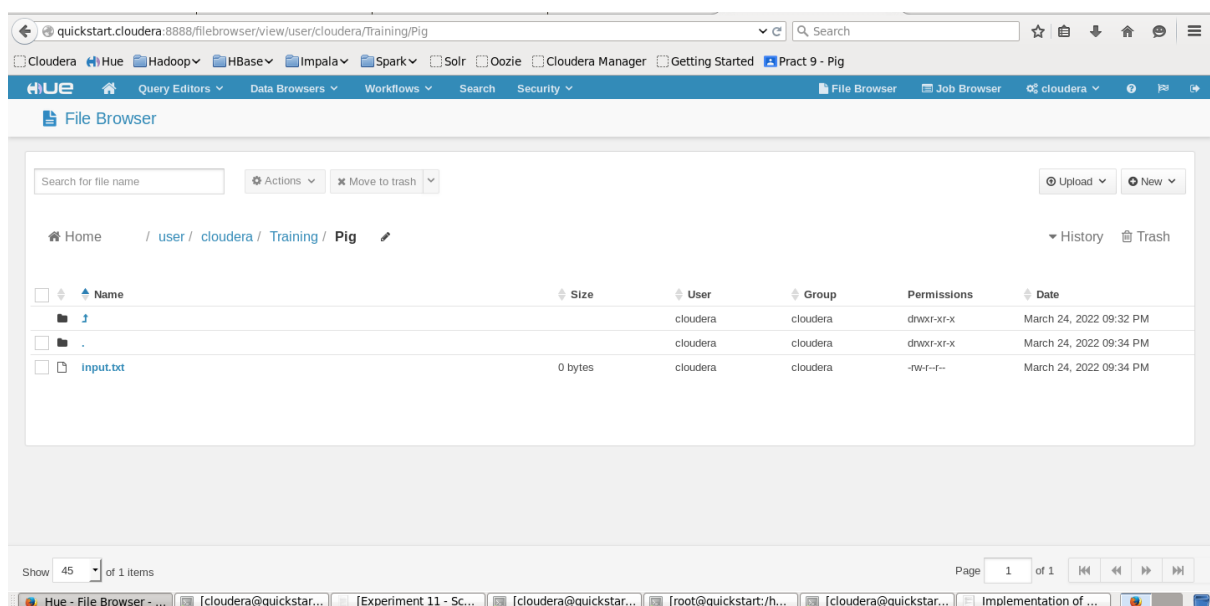


7) Creating input.txt file inside /usr/cloudera/Training/Pig directory

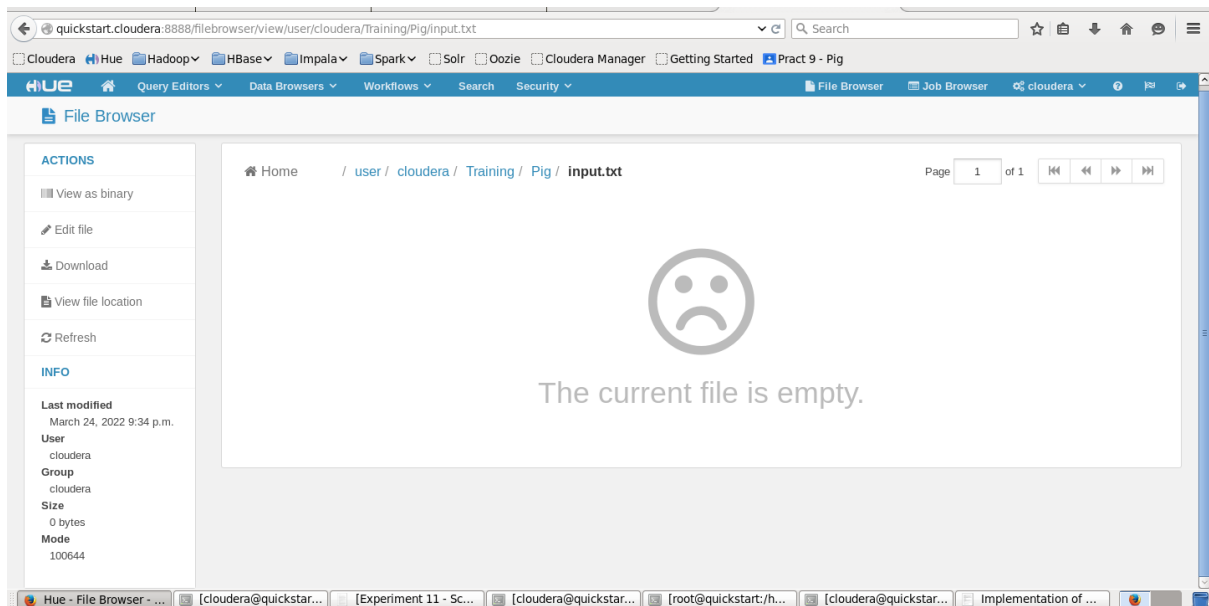
Again inside the Pig directory click on New and create file as 'input.txt'



Once the file has been created click on 'input.txt' to add the content in it

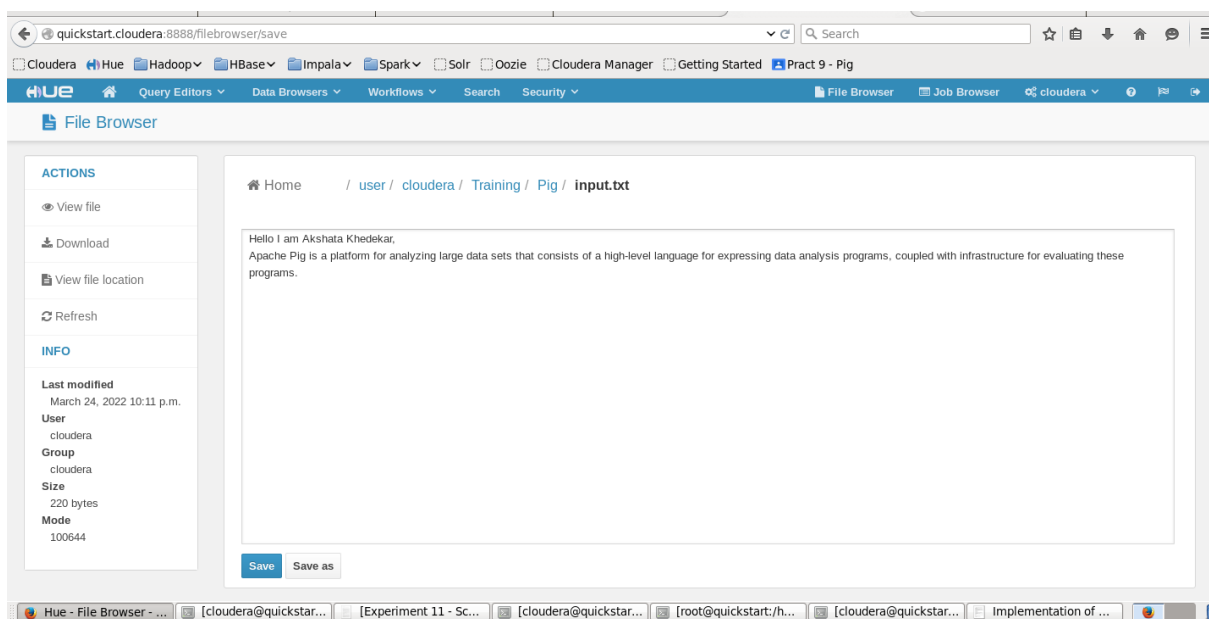


8) Adding some contents to this input.txt file.



For adding content in the input file, Click on ‘Edit file’ option then add the content.

Save the input.txt file



9) Now Open the terminal. And start Pig by typing pig on terminal.

```
[cloudera@quickstart ~]$ pig
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
2022-03-24 22:22:07,847 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.4.2 (reexported) compiled May 19 2015, 17:03:41
2022-03-24 22:22:07,847 [main] INFO org.apache.pig.Main - Logging error messages to: /home/cloudera/pig/1644894610889.log
2022-03-24 22:22:07,869 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/cloudera/pig/BOOTUP not found
2022-03-24 22:22:08,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:08,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:08,544 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020
2022-03-24 22:22:09,886 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:09,887 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2022-03-24 22:22:09,887 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:09,982 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:09,986 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:10,077 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:10,082 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:10,182 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:10,182 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:10,290 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:10,290 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:10,388 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:10,388 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:10,403 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-24 22:22:10,404 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - MR plan size before optimization: 1
2022-03-24 22:22:10,583 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - MR plan size after optimization: 1
2022-03-24 22:22:10,584 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 22:22:10,673 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-24 22:22:10,673 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-24 22:22:10,762 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job2342891949082825398.jar
2022-03-24 22:22:10,725 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job2342891949082825398.jar created
2022-03-24 22:22:10,736 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-24 22:22:10,737 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-24 22:22:10,737 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-24 22:22:10,737 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2022-03-24 22:22:10,755 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-24 22:22:10,755 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:22:10,759 [JobControl] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 22:22:10,789 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:22:10,801 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-24 22:22:10,801 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-03-24 22:22:10,813 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-03-24 22:22:10,813 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-24 22:22:10,813 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644894610889_0027
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644894610889_0027
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644894610889_0027
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases input1
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: input1[3,9],input1[-1,-1] C: R:
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?
jobid=job_1644894610889_0027
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-24 22:22:10,845 [JobControl] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:
```

- 10) Now we have to load that input file where ever it is stored. By typing the command  
**Input1 = LOAD '/usr/cloudera/Training/pig/input.txt' AS (fl:chararray);**

- 11) Now we are dumping the data. It will do the MapReduce task. The Dump operator is used to run the Pig Latin statements and display the results on the screen. It is generally used for debugging Purpose.

**DUMP input1;**

```
File Edit View Search Terminal Help
grunt> input1 = LOAD '/usr/cloudera/Training/Pig/input.txt' AS (fl:chararray);
grunt> DUMP input1;
2022-03-24 22:16:52,286 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-24 22:16:52,286 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2022-03-24 22:16:52,288 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-24 22:16:52,289 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - MR plan size before optimization: 1
2022-03-24 22:16:52,289 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - MR plan size after optimization: 1
2022-03-24 22:16:52,322 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 22:16:52,324 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-24 22:16:52,327 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-24 22:16:52,762 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job2342891949082825398.jar
2022-03-24 22:16:55,725 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job2342891949082825398.jar created
2022-03-24 22:16:55,736 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-24 22:16:55,737 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-24 22:16:55,737 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-24 22:16:55,737 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2022-03-24 22:16:55,755 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-24 22:16:55,755 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:16:55,759 [JobControl] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 22:16:55,789 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:16:56,081 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-24 22:16:56,081 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-03-24 22:16:56,113 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-03-24 22:16:56,163 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-24 22:16:56,164 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644894610889_0027
2022-03-24 22:16:57,245 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644894610889_0027
2022-03-24 22:16:57,265 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644894610889_0027
2022-03-24 22:16:57,266 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases input1
2022-03-24 22:16:57,266 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: input1[3,9],input1[-1,-1] C: R:
2022-03-24 22:16:57,266 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?
jobid=job_1644894610889_0027
2022-03-24 22:16:57,333 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-24 22:17:15,753 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-24 22:17:18,086 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2022-03-24 22:17:18,071 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-24 22:17:18,073 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:
```



```

2022-03-24 22:17:18,073 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.4.2 0.12.0-cdh5.4.2 cloudera 2022-03-24 22:16:52 2022-03-24 22:17:18 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature 0
utputs
job_1644894610889_0027 1 0 5 5 5 5 n/a n/a n/a n/a input1 MAP_ONLY hdfs://quickstart.cloudera:8020/tmp/temp-2010201162/tmp181056881,

Input(s):
Successfully read 3 records (607 bytes) from: "/user/cloudera/Training/Pig/input.txt"

Output(s):
Successfully stored 3 records (239 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-2010201162/tmp181056881"

Counters:
Total records written : 3
Total bytes written : 239
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644894610889_0027

2022-03-24 22:17:18,186 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-24 22:17:18,187 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:17:18,187 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 22:17:18,188 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-24 22:17:18,211 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-24 22:17:18,211 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Hello I am Akshata Khedekar,)
(Apache Pig is a platform for analyzing large data sets that consists of a high-level language)
(For expressing data analysis programs, coupled with infrastructure for evaluating these programs.)
grunt> █

```

12) Here we are counting the words in each line for that we are using the following command

wordsInEachLine = FOREACH input1 GENERATE  
flatten(TOKENIZE(f1)) as word;

```

grunt> wordsInEachLine = FOREACH input1 GENERATE flatten(TOKENIZE(f1)) as word;
2022-03-24 22:48:33,226 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 22:48:33,226 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt> █

```

13) Again, we are dumping the data. It will do the MapReduce task. dump wordsInEachLine;

```

File Edit View Search Terminal Help

grunt> dump wordsInEachLine;
2022-03-24 23:11:33,367 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-24 23:11:33,369 [main] INFO org.apache.pig.newplan.LogicalOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2022-03-24 23:11:33,380 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCCompiler - File concatenation threshold: 100 optimistic? false
2022-03-24 23:11:33,380 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-24 23:11:33,380 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-24 23:11:33,401 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 23:11:33,404 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-24 23:11:33,415 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-24 23:11:33,817 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job5225789320897719770.jar
2022-03-24 23:11:36,815 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job5225789320897719770.jar created
2022-03-24 23:11:36,835 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-24 23:11:36,836 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-24 23:11:36,836 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-24 23:11:36,836 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2022-03-24 23:11:36,857 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-24 23:11:36,857 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 23:11:36,859 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 23:11:36,863 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 23:11:36,989 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-24 23:11:36,989 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-03-24 23:11:36,991 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-03-24 23:11:37,020 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-24 23:11:37,097 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644894610889_0030
2022-03-24 23:11:37,137 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644894610889_0030
2022-03-24 23:11:37,140 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644894610889_0030/
2022-03-24 23:11:37,357 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644894610889_0030
2022-03-24 23:11:37,357 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases input1,wordsInEachLine
2022-03-24 23:11:37,357 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: input1[1,9],wordsInEachLine[-1,-1] C
: R;
2022-03-24 23:11:37,357 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?
jobid=job_1644894610889_0030
2022-03-24 23:11:37,404 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-24 23:11:54,322 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-24 23:11:57,611 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-24 23:11:57,612 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

```

```
File Edit View Search Terminal Help
: R:
2022-03-24 23:11:37,357 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?
jobid=job_1644894610889_0030
2022-03-24 23:11:37,404 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-24 23:11:54,322 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-24 23:11:57,611 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-24 23:11:57,612 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.4.2 0.12.0-cdh5.4.2 cloudera 2022-03-24 23:11:33 2022-03-24 23:11:57 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature 0
utputs
job_1644894610889_0030 1 0 7 7 7 7 n/a n/a n/a n/a input1,wordsInEachLine MAP_ONLY hdfs://quickstart.cloudera:8020/tmp/
tmp858098770/tmp1962244065,

Input(s):
Successfully read 3 records (607 bytes) from: "/user/cloudera/Training/Pig/input.txt"

Output(s):
Successfully stored 33 records (417 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/858098770/tmp1962244065"

Counters:
Total records written : 33
Total bytes written : 417
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644894610889_0030

2022-03-24 23:11:57,703 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-24 23:11:57,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 23:11:57,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 23:11:57,704 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
```

```
File Edit View Search Terminal Help
2022-03-24 23:11:57,703 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-24 23:11:57,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 23:11:57,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 23:11:57,704 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-24 23:11:57,711 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-24 23:11:57,711 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

(Hello)
(I)
(am)
(Akshata)
(Khedeekar)
(Apache)
(Pig)
(is)
(a)
(platform)
(for)
(analyzing)
(large)
(data)
(sets)
(that)
(consists)
(of)
(a)
(high-level)
(language)
(For)
(expressing)
(data)
(analysis)
(programs)
(coupled)
(with)
(infrastructure)
(for)
(evaluating)
(these)
(programs.)
grunt> █

[Hue - File Brow... cloudera@quick... [Experiment 11 -... cloudera@quick... [root@quickstart... cloudera@quick... [Implementation... cloudera@quicks...]
```

- 14) Now grouping the words present in each line.  
groupedWords = group wordsInEachLine by word;

```
grunt> groupedWords = group wordsInEachLine by word;
grunt> █
```

And then dumping the data by the following command.

dump groupedWords;

```

File Edit View Search Terminal Help
grunt> dump groupedWords;
2022-03-24 23:18:29,191 [main] INFO org.apache.pig.tools.pigstats.ScriptState -
Pig features used in the script: GROUP_BY
2022-03-24 23:18:29,192 [main] INFO org.apache.pig.newplan.logical.optimizer.Lo
gicalPlanOptimizer - [RULES_ENABLED=[AddForEach, ColumnKeyPrune, DuplicateFor
EachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptim
izer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimiz
er, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter],
RULES_DISABLED=[FilterLogicExpressionsSimplifier, PartitionFilterOptimizer]]
2022-03-24 23:18:29,198 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.MRCompiler - File concatenation threshold: 100 optimistic? fal
se
2022-03-24 23:18:29,205 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-24 23:18:29,205 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-24 23:18:29,245 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Con
necting to ResourceManager at /0.0.0.0:8032
2022-03-24 23:18:29,249 [main] INFO org.apache.pig.tools.pigstats.ScriptState -
Pig script settings are added to the job
2022-03-24 23:18:29,257 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percen
t is not set, set to default 0.3
2022-03-24 23:18:29,257 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.JobControlCompiler - Reduce phase detected, estimating # of re
quired reducers.
2022-03-24 23:18:29,258 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.JobControlCompiler - Using reducer estimator: org.apache.pig.b
ackend.hadoop.executionengine.mapreduce.layer.InputSizeReducerEstimator
2022-03-24 23:18:29,264 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxRedu
cers=999 totalInputFileSize=220
2022-03-24 23:18:29,264 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.JobControlCompiler - Setting Parallelism to 1
2022-03-24 23:18:29,707 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapreduce.layer.JobControlCompiler - creating jar file Job6681608896550704445.
jar
2022-03-24 23:18:32,714 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.JobControlCompiler - jar file Job6681608896550704445.jar created
2022-03-24 23:18:32,722 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.JobControlCompiler - Setting up single store job
2022-03-24 23:18:32,722 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

File Edit View Search Terminal Help
2022-03-24 23:18:33,265 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?
jobid=job_1644894610889_0031
2022-03-24 23:18:33,302 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.MapReduceLauncher - 0% complete
2022-03-24 23:18:50,950 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.MapReduceLauncher - 50% complete
2022-03-24 23:19:03,524 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.MapReduceLauncher - 100% complete
2022-03-24 23:19:03,525 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.4.2 0.12.0-cdh5.4.2 cloudera 2022-03-24 23:18:29 2022-03-24 23:19:03 GROUP_BY

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature 0
outputs
job_1644894610889_0031 1 1 7 7 7 6 6 6 6 groupedWords,input1,wordsInEachLine GROUP_BY hdf://quickstart
.cloudera:8020/tmp/temp850898770/tmp-2146257697,

Input(s):
Successfully read 3 records (607 bytes) from: "/user/cloudera/Training/Pig/input.txt"

Output(s):
Successfully stored 30 records (766 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp850898770/tmp-2146257697"

Counters:
Total records written : 30
Total bytes written : 766
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644894610889_0031

2022-03-24 23:19:03,595 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.layer.MapReduceLauncher - Success!
2022-03-24 23:19:03,595 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 23:19:03,596 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 23:19:03,596 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-24 23:19:03,602 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

```

```

File Edit View Search Terminal Help
2022-03-24 23:19:03,595 [main] INFO org.apache.pig.backend.hadoop.executionengi
2022-03-24 23:19:03,595 [main] INFO org.apache.hadoop.conf.Configuration.deprec
2022-03-24 23:19:03,596 [main] INFO org.apache.hadoop.conf.Configuration.deprec
2022-03-24 23:19:03,596 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
2022-03-24 23:19:03,602 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
2022-03-24 23:19:03,603 [main] INFO org.apache.pig.backend.hadoop.executionengi
(I,{{I}})
(a,{{(a),(a)}})
(am,{{(am)}})
(is,{{(is)}})
(of,{{(of)}})
(For,{{(For)}})
(Pig,{{(Pig)}})
(for,{{(for),(for)}})
(data,{{(data),(data)}})
(sets,{{(sets)}})
(that,{{(that)}})
(with,{{(with)}})
(Hello,{{(Hello)}})
(large,{{(large)}})
(these,{{(these)}})
(Apache,{{(Apache)}})
(Akshata,{{(Akshata)}})
(coupled,{{(coupled)}})
(Khedekar,{{(Khedekar)}})
(analysis,{{(analysis)}})
(consists,{{(consists)}})
(language,{{(language)}})
(platform,{{(platform)}})
(programs,{{(programs)}})
(analyzing,{{(analyzing)}})
(programs.,{{(programs.)}})
(evaluating,{{(evaluating)}})
(expressing,{{(expressing)}})
(high-level,{{(high-level)}})
(infrastructure,{{(infrastructure)}})
grunt> █

```

15) Now we count those words. For each group we count words in each line.

countedWords = foreach groupedWords generate group,  
COUNT(wordsInEachLine);

```

grunt> countedWords = foreach groupedWords generate group, COUNT(wordsInEachLine);
grunt> █

```

- 16) After every counting of words commands, we are dumping the data dump countedWords;  
Now the Final Output we are getting as word count for every word.

```

grunt> dump countedWords;
2022-03-24 23:27:59,297 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig
features used in the script: GROUP_BY
2022-03-24 23:27:59,298 [main] INFO org.apache.pig.newplan.logical.optimizer.Logica
lPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColu
mnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTy
peCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownFore
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[Filt
erLogicExpressionSimplifier, PartitionFilterOptimizer]}
2022-03-24 23:27:59,305 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-24 23:27:59,306 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.CombinerOptimizer - Choosing to move algebraic foreach to combiner
2022-03-24 23:27:59,314 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-24 23:27:59,314 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-24 23:27:59,338 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connect
ing to ResourceManager at /0.0.0.0:8032
2022-03-24 23:27:59,349 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig
script settings are added to the job
2022-03-24 23:27:59,367 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not
set, set to default 0.3
2022-03-24 23:27:59,367 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required r
educers.
2022-03-24 23:27:59,367 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.h
adoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2022-03-24 23:27:59,369 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999
totalInputFileSize=220
2022-03-24 23:27:59,369 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - Setting Parallelism to 1
2022-03-24 23:27:59,786 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - creating jar file Job128247859528418909.jar
2022-03-24 23:28:02,769 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - jar file Job128247859528418909.jar created
2022-03-24 23:28:02,782 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
apReduceLayer.JobControlCompiler - Setting up single store job
2022-03-24 23:28:02,782 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-24 23:28:02,782 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-24 23:28:02,782 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

[Hue - File Brow... [cloudera@quick... [Experiment 11 ... [cloudera@quick... [root@quickstart... [cloudera@quick... Implementation ... cloudera@quicks...

EachLine
2022-03-24 23:28:03,316 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: input1[1,9],wordsInEachLine[-1,-1],c
ountedWords[4,15],groupedWords[3,15] C: countedWords[4,15],groupedWords[3,15] R: countedWords[4,15]
2022-03-24 23:28:03,316 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?
jobid=job_1644894610889_0032
2022-03-24 23:28:03,349 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-24 23:28:19,129 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-24 23:28:33,274 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-24 23:28:33,274 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.4.2 0.12.0-cdh5.4.2 cloudera 2022-03-24 23:27:59 2022-03-24 23:28:33 GROUP_BY

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature 0
utputs
job_1644894610889_0032 1 1 5 5 5 5 6 6 6 6 countedWords,groupedWords,input1,wordsInEachLine GROUP_BY,COMBINER
hdfs://quickstart.cloudera:8020/tmp/temp858098770/tmp-1216859044,

Input(s):
Successfully read 3 records (607 bytes) from: "/user/cloudera/Training/Pig/input.txt"

Output(s):
Successfully stored 30 records (421 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp858098770/tmp-1216859044"

Counters:
Total records written : 30
Total bytes written : 421
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644894610889_0032

2022-03-24 23:28:33,382 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-24 23:28:33,382 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 23:28:33,382 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 23:28:33,383 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

```

```

cloudera@quickstart:~
job_1644894610889_0032

2022-03-24 23:28:33,382 [main] INFO org.apache.pig.backend.hadoop.executionengine.m
2022-03-24 23:28:33,382 [main] INFO org.apache.hadoop.conf.Configuration.deprecatio
2022-03-24 23:28:33,382 [main] INFO org.apache.hadoop.conf.Configuration.deprecatio
2022-03-24 23:28:33,383 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [p
2022-03-24 23:28:33,390 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInput
2022-03-24 23:28:33,390 [main] INFO org.apache.pig.backend.hadoop.executionengine.u
(I,1)
(a,2)
(am,1)
(is,1)
(of,1)
(for,1)
(Pig,1)
(for,2)
(data,2)
(sets,1)
(that,1)
(with,1)
(Hello,1)
(large,1)
(these,1)
(Apache,1)
(Akshata,1)
(coupled,1)
(Khedekar,1)
(analysis,1)
(consists,1)
(language,1)
(platform,1)
(programs,1)
(analyzing,1)
(programs.,1)
(evaluating,1)
(expressing,1)
(high-level,1)
(infrastructure,1)
grunt>

```

As we can see from above image the Word “a” occurred twice, word “for, data” start with small w occurred twice, word “I” occurred once, and so on.

17) Now Exit from the grunt shell using quit command.

```

grunt> quit
[cloudera@quickstart ~]$

```