

File Browser

Search for file name

ActionsMove to trash

UploadNew

Home / user / cloudera

HistoryTrash

| Name    | Size | User     | Group      | Permissions | Date                       |
|---------|------|----------|------------|-------------|----------------------------|
| .       |      | hdfs     | supergroup | drwxr-xr-x  | March 10, 2022 08:22 PM    |
| Desktop |      | cloudera | cloudera   | drwxr-xr-x  | March 21, 2022 08:55 PM    |
| vj      |      | cloudera | cloudera   | drwxr-xr-x  | February 14, 2022 06:57 PM |
|         |      | cloudera | cloudera   | drwxr-xr-x  | March 10, 2022 08:07 PM    |

Create Directory

Directory Name

training

Training

CancelCreate

|          | Group      | Permissions |
|----------|------------|-------------|
| hdfs     | supergroup | drwxr-xr-x  |
| cloudera | cloudera   | drwxr-xr-x  |
| cloudera | cloudera   | drwxr-xr-x  |
| cloudera | cloudera   | drwxr-xr-x  |

Create Directory

Directory Name

pig

pig

CancelCreate

|          | Group    | Permissions |
|----------|----------|-------------|
| cloudera | cloudera | drwxr-x     |
| cloudera | cloudera | drwxr-x     |
| cloudera | cloudera | drwxr-x     |

Create File

File Name

Cancel

Create

|         |          | Group    | Permissions |
|---------|----------|----------|-------------|
|         | cloudera | cloudera | drwxr-xr-x  |
|         | cloudera | cloudera | drwxr-xr-x  |
| 0 bytes | cloudera | cloudera | -rwxr-xr-x  |

| <input type="checkbox"/> | Name      | Size    | User     | Group    | Permissions | Date                    |
|--------------------------|-----------|---------|----------|----------|-------------|-------------------------|
| <input type="checkbox"/> | ..        |         | cloudera | cloudera | drwxr-xr-x  | March 24, 2022 09:21 PM |
| <input type="checkbox"/> | .         |         | cloudera | cloudera | drwxr-xr-x  | March 24, 2022 09:22 PM |
| <input type="checkbox"/> | input.txt | 0 bytes | cloudera | cloudera | -rwxr-xr-x  | March 24, 2022 09:22 PM |

```
[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 in
fo.
2022-03-24 21:27:30,534 [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 21:27:30,534 [main] INFO org.apache.pig.Main - Logging error message
s to: /home/cloudera/pig_1648182450505.log
2022-03-24 21:27:30,564 [main] INFO org.apache.pig.impl.util.Utils - Default bo
otup file /home/cloudera/.pigbootup not found
2022-03-24 21:27:31,410 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2022-03-24 21:27:31,411 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:31,411 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to hadoop file system at: hdfs://quickstart.clo
udera:8020
2022-03-24 21:27:33,409 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2022-03-24 21:27:33,409 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2022-03-24 21:27:33,415 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,466 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,468 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2022-03-24 21:27:33,517 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,517 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2022-03-24 21:27:33,571 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,571 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
2022-03-24 21:27:33,670 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2022-03-24 21:27:33,755 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,761 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 21:27:33,832 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,839 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 21:27:33,911 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,911 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 21:27:33,956 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:27:33,957 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunt>
```

```
grunt> Input1 = LOAD '/usr/cloudera/Training/pig/input.txt' AS (f1:chararray);
grunt>
```

```
grunt> Input1 = LOAD '/usr/cloudera/Training/pig/input.txt' AS (f1:chararray);
grunt> DUMP Input1;
2022-03-24 21:34:49,873 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-24 21:34:49,874 [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 21:34:49,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-24 21:34:49,876 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-24 21:34:49,876 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-24 21:34:49,900 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 21:34:49,903 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-24 21:34:49,905 [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 21:34:49,906 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job3515548187496591125.jar
2022-03-24 21:34:50,158 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job3515548187496591125.jar created
2022-03-24 21:34:53,053 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-24 21:34:53,060 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-24 21:34:53,060 [main] INFO org.apache.pig.data.SchemaTupleFrontend - key [pig.schematuple] is false, will not generate code
2022-03-24 21:34:53,060 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-24 21:34:53,060 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2022-03-24 21:34:53,067 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-24 21:34:53,067 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 21:34:53,068 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-24 21:34:53,074 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-24 21:34:53,270 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:34:53,270 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Total input paths to process : 1
2022-03-24 21:34:53,270 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Total input paths to process : 1
2022-03-24 21:34:53,291 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Total input paths (combined) to process : 1
2022-03-24 21:34:53,331 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - number of splits:1
2022-03-24 21:34:53,331 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - number of splits:1
2022-03-24 21:34:53,799 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644548343526_0028
2022-03-24 21:34:53,801 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644548343526_0028
```

```
2022-03-24 21:34:53,331 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-24 21:34:53,799 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644548343526_0020
2022-03-24 21:34:53,981 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644548343526_0020
2022-03-24 21:34:54,049 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8080/proxy/application_1644548343526_0020/
2022-03-24 21:34:54,049 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644548343526_0020
2022-03-24 21:34:54,049 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Input1
2022-03-24 21:34:54,049 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: Input1[13,9],Input1[-1,-1] C: R:
2022-03-24 21:34:54,049 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1644548343526_0020
2022-03-24 21:34:54,098 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-24 21:35:12,995 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-24 21:35:19,500 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2022-03-24 21:35:20,906 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2022-03-24 21:35:20,963 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-24 21:35:20,964 [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 21:34:49 | 2022-03-24 21:35:20 | UNKNOWN  |

Success!

Job Stats (time in seconds):

| JobId                  | Maps | Reducers | MaxMapTime | MinMapTime | AvgMapTime | MedianMapTime | MaxReduceTime | MinReduceTime | AvgReduceTime | MedianReduceTime | Alias   | Feature Outputs |
|------------------------|------|----------|------------|------------|------------|---------------|---------------|---------------|---------------|------------------|---|-----------------|
| job_1644548343526_0020 | 1    | 0        | 6          | 6          | 6          | n/a           | n/a           | n/a           | n/a           | Input1 MAP_ONLY  | hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp-413572416, |                 |

Input(s):

Successfully read 3 records (497 bytes) from: "/user/cloudera/Training/pig/input.txt"

Output(s):

Successfully stored 3 records (129 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp-413572416"

Counters:

Total records written : 3

Total bytes written : 129

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1644548343526\_0020

```
2022-03-24 21:35:21,051 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-24 21:35:21,052 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-24 21:35:21,052 [main] INFO org.apache.pig.backend.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-24 21:35:21,052 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-24 21:35:21,065 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-24 21:35:21,065 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(we are studying BIG DATA TECHNOLOGY!)
(WE HAVE STUDIED HADOOP ECOSYSTEM)
(WE HAVE DONE WORD COUNT USING MAP REDUCE)
grunt>
```

```
grunt> wordsInEachLine = FOREACH Input1 GENERATE flatten(TOKENIZE(f1)) as word;
```

```
2022-03-25 22:09:29,072 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2022-03-25 22:09:29,072 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

grunt>

```
grunt> wordsInEachLine = FOREACH Input1 GENERATE flatten(TOKENIZE(f1)) as word;
```

```
2022-03-25 22:09:29,072 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2022-03-25 22:09:29,072 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

```
grunt> dump wordsInEachLine;
```

```
2022-03-25 22:10:49,971 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-25 22:10:49,986 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES_ENABLED=[AdaptForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOp
timizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachJoin, PushOptimizer, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionsInserter, PartitionFilterOptimizer])
2022-03-25 22:10:41,014 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-25 22:10:41,016 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-25 22:10:41,016 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-25 22:10:41,067 [main] INFO org.apache.hadoop.yarn.client.NMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-25 22:10:41,067 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-25 22:10:41,084 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-25 22:10:42,118 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job66908087589058459951.jar
2022-03-25 22:10:45,225 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job66908087589058459951.jar created
2022-03-25 22:10:45,536 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-25 22:10:45,537 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-25 22:10:45,537 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-25 22:10:45,537 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2022-03-25 22:10:45,546 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-25 22:10:45,546 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:10:45,550 [JobControl] INFO org.apache.hadoop.yarn.client.NMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-25 22:10:45,566 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:10:45,603 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths (combined) to process : 1
2022-03-25 22:10:45,776 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-03-25 22:10:45,887 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-25 22:10:45,928 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted tokens for job: job_1644548343526_0020
2022-03-25 22:10:45,939 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8080/proxy/application_1644548343526_0020/
2022-03-25 22:10:46,046 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644548343526_0020
2022-03-25 22:10:46,046 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Input1,wordsInEachLine
2022-03-25 22:10:46,046 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: Input1[13,9],wordsInEachLine[-1,-1] C: R:
2022-03-25 22:10:46,046 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1644548343526_0020
2022-03-25 22:10:46,100 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-25 22:11:09,026 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-25 22:11:11,459 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-25 22:11:11,459 [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 | 2022-03-25 22:10:41 | 2022-03-25 22:11:11 | UNKNOWN  |

Success!

```
Job Stats (time in seconds):
JobId  Maps    Reduces MaxMapTime    MinMapTime    AvgMapTime    MedianMapTime    MaxReduceTime    MinReduceTime    AvgReduceTime    MedianReduceTime    Alias    Feature Outputs
Job 1644548343526 0029 1    0    5    5    5    5    n/a    n/a    n/a    n/a    Input1,wordsInEachLine    MAP_ONLY    hdf://quickstart.cloudera:8020/tmp/temp-669075149/tmp-341911088,

Input(s):
Successfully read 3 records (497 bytes) from: "/user/cloudera/Training/pig/input.txt"

Output(s):
Successfully stored 19 records (225 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp-341911088"

Counters:
Total records written : 19
Total bytes written : 225
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job 1644548343526 0029

2022-03-25 22:11:11,537 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-25 22:11:11,537 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:11:11,537 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:11:11,538 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig:schematuple] was not set... will not generate code.
2022-03-25 22:11:11,543 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-25 22:11:11,543 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(we)
(are)
(studying)
(BIG)
(DATA)
(TECHNOLOGY))
(WE)
(HAVE)
(STUDIED)
(HADOOP)
(ECOSYSTEM)
(WE)
(HAVE)
(DONE)
(WORD)
(COUNT)
(USING)
(WAP)
(REDUCE)
grunt> █
```

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

```
grunt> dump groupedWords;
2022-03-25 22:12:59,694 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP BY
2022-03-25 22:12:59,695 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, Limit
Optimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter}, RULES_DISABLED=FilterLogicExpressionSimplifier, PartitionFilterOptimizer
}
2022-03-25 22:12:59,702 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCCompiler - File concatenation threshold: 100 optimistic? false
2022-03-25 22:12:59,706 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-25 22:12:59,706 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-25 22:12:59,736 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-25 22:12:59,738 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-25 22:12:59,750 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-25 22:12:59,750 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2022-03-25 22:12:59,752 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2022-03-25 22:12:59,757 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1080000000 maxReducers=599 totalInputFileSize=110
2022-03-25 22:12:59,757 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2022-03-25 22:13:00,128 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Creating jar file Job816439686515970146.jar
2022-03-25 22:13:03,367 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job816439686515970146.jar created
2022-03-25 22:13:03,373 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-25 22:13:03,374 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig:schematuple] is false, will not generate code.
2022-03-25 22:13:03,374 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-25 22:13:03,374 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig:schematuple.classes] with classes to deserialize []
2022-03-25 22:13:03,676 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-25 22:13:03,676 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:13:03,677 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-25 22:13:03,687 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:13:03,825 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-25 22:13:03,825 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-03-25 22:13:03,829 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-03-25 22:13:03,858 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits=1
2022-03-25 22:13:03,905 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644548343526_0030
2022-03-25 22:13:03,951 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644548343526_0030
2022-03-25 22:13:03,953 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer - The url to track the job: http://quickstart.cloudera:8080/proxy/appLocation_1644548343526_0030/
2022-03-25 22:13:04,176 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644548343526_0030
2022-03-25 22:13:04,176 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Input1,groupedWords,wordsInEachLine
2022-03-25 22:13:04,176 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Detailed locations: M: Input[1,9],wordsInEachLine[1,-1],groupedWords[5,15] C: R:
2022-03-25 22:13:04,176 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1644548343526_0030
2022-03-25 22:13:20,489 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-25 22:13:20,489 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-25 22:13:34,144 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-25 22:13:34,144 [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    2022-03-25 22:12:59    2022-03-25 22:13:34    GROUP BY

Success!

Job Stats (time in seconds):
JobId  Maps    Reduces MaxMapTime    MinMapTime    AvgMapTime    MedianMapTime    MaxReduceTime    MinReduceTime    AvgReduceTime    MedianReduceTime    Alias    Feature Outputs
Job 1644548343526 0030 1    1    5    5    5    5    6    6    6    6    Input1,groupedWords,wordsInEachLine    GROUP BY    hdf://quickstart.cloudera:8020/tmp/temp-669075149/tmp325198341,
```



```
Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job 1644548343526_0030 1 1 5 5 5 5 6 6 6 6 Input1,groupedWords,wordsInEachLine GROUP_BY hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp325198341,

Input(s):
Successfully read 3 records (497 bytes) from: "/user/cloudera/Training/pig/input.txt"

Output(s):
Successfully stored 17 records (407 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp325198341"

Counters:
Total records written : 17
Total bytes written : 407
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644548343526_0030

2022-03-25 22:13:34,226 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-25 22:13:34,226 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:13:34,226 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:13:34,227 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-25 22:13:34,235 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-25 22:13:34,235 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(we, ((we)),
(BIG, ((BIG)))
(MAP, ((MAP)))
(re, ((re)))
(DATA, ((DATA)))
(DONE, ((DONE)))
(HAVE, ((HAVE), (HAVE)))
(WORDS, ((WORD)))
(COUNT, ((COUNT)))
(USING, ((USING)))
(HADOOP, ((HADOOP)))
(REDUCE, ((REDUCE)))
(STUDED, ((STUDED)))
(studying, ((studying)))
(ECOSYSTEM, ((ECOSYSTEM)))
(TWECNHOLOGY, ((TWECNHOLOGY)))
grunt>
```

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

2022-03-25 22:17:07,065 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP BY
2022-03-25 22:17:07,066 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES_ENABLED=AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOp
Cmbiner, LoadFingerprinter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachPlattion, PushUpFilter, SplitFilter, StreamTypeCasterInserter), RULES_DISABLED=FilterLogicExpressionsSimplifier, PartitionFilterOptimizer
).
2022-03-25 22:17:07,074 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-25 22:17:07,075 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOptimizer - Choosing to move algebraic foreach to combiner
2022-03-25 22:17:07,080 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-25 22:17:07,080 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-25 22:17:07,110 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /s.s.d.s:8032
2022-03-25 22:17:07,113 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-25 22:17:07,132 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-25 22:17:07,133 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2022-03-25 22:17:07,133 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2022-03-25 22:17:07,136 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1060000000 maxReducers=999 TotalInputFilesSize=110
2022-03-25 22:17:07,136 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2022-03-25 22:17:07,716 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job7925097203331900059.jar
2022-03-25 22:17:11,601 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Jar file Job7925097203331900059.jar created
2022-03-25 22:17:11,636 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-25 22:17:11,637 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-25 22:17:11,637 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-25 22:17:11,637 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.class] with classes to serialize []
2022-03-25 22:17:11,677 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-25 22:17:11,677 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:17:11,678 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /s.s.d.s:8032
2022-03-25 22:17:11,687 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:17:11,215 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-25 22:17:11,215 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
2022-03-25 22:17:11,222 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths (combined) to process : 1
2022-03-25 22:17:11,990 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-25 22:17:12,055 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644548343526_0031
2022-03-25 22:17:12,099 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644548343526_0031
2022-03-25 22:17:12,104 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8080/proxy/application_1644548343526_0031/
2022-03-25 22:17:12,106 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644548343526_0031
2022-03-25 22:17:12,106 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases Input1,countedWords,groupedWords,wordsInEachLine
2022-03-25 22:17:12,106 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: Input1[3,9],wordsInEachLine[1,-,1],countedWords[0,15],groupedWords[0,15],g
roupedWords[5,15] R: countedWords[5,15]
2022-03-25 22:17:12,106 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1644548343526_0031
2022-03-25 22:17:12,141 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-25 22:17:20,842 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
```

```
2022-03-25 22:17:42,771 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-03-25 22:17:42,771 [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 2022-03-25 22:17:07 2022-03-25 22:17:42 GROUP_BY

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job 1644548343526_0031 1 1 5 5 5 5 6 6 6 6 Input1,countedWords,groupedWords,wordsInEachLine GROUP_BY,COMBINER hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp1191620340,

Input(s):
Successfully read 3 records (497 bytes) from: "/user/cloudera/Training/pig/input.txt"

Output(s):
Successfully stored 17 records (224 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-669075149/tmp1191620340"

Counters:
Total records written : 17
Total bytes written : 224
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644548343526_0031

2022-03-25 22:17:42,858 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-25 22:17:42,858 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:17:42,858 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:17:42,858 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-25 22:17:42,866 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-25 22:17:42,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
```

```
(WE,2)
(we,1)
(BIG,1)
(MAP,1)
(are,1)
(DATA,1)
(DONE,1)
(HAVE,2)
(WORD,1)
(COUNT,1)
(USING,1)
(HADOOP,1)
(REDUCE,1)
(STUDED,1)
(studying,1)
(ECOSYSTEM,1)
(TWECHNOLOGY!,1)
grunt> █
```