

What is Pig in Hadoop?

Pig is a scripting platform that runs on Hadoop clusters designed to process and analyze large datasets. Pig is extensible, self-optimizing, and easily programmed.

Programmers can use Pig to write data transformations without knowing Java. Pig uses both structured and unstructured data as input to perform analytics and uses HDFS to store the results.

Components of Pig

There are two major components of the Pig:

- Pig Latin script language
- A runtime engine

Pig Latin script language:

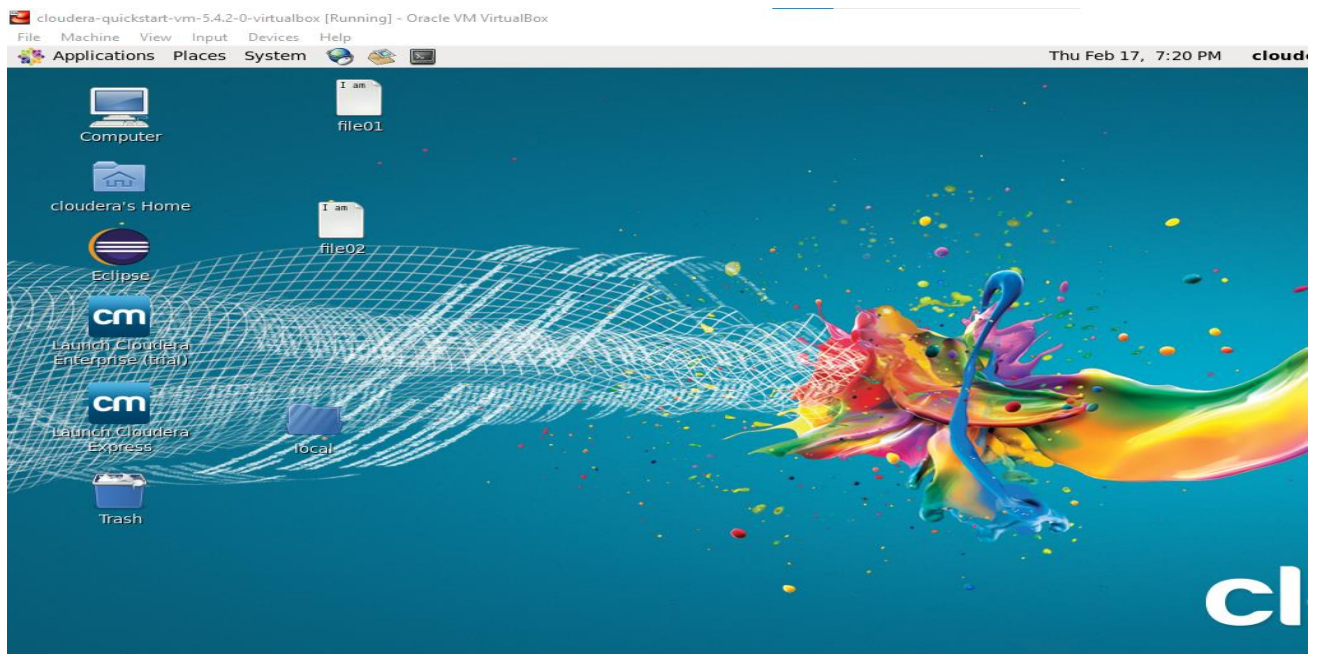
The Pig Latin script is a procedural data flow language. It contains syntax and commands that can be applied to implement business logic. Examples of Pig Latin are LOAD and STORE.

A runtime engine:

The runtime engine is a compiler that produces sequences of MapReduce programs. It uses HDFS to store and retrieve data. It is also used to interact with the Hadoop system (HDFS and MapReduce).

The runtime engine parses, validates, and compiles the script operations into a sequence of MapReduce jobs.

step1) Start the cloudera



Step 2 : Now Open the terminal. And start Pig by typing pig on terminal.



Step 3: now pig get started

Step 4: now load the file mydata.txt file,

A=LOAD '/user/cloudera/Training/pig/mydata.txt' AS (c1:int,c2:int,c3:int);

And then dump,

dump A;

```
grunt> A = LOAD '/user/cloudera/Training/pig/mydata.txt' AS (c1:int,c2:int,c3:int);
grunt> dump A;
2022-03-29 20:16:10,627 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-29 20:16:10,627 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnWrite, GroupByControlParallelSetter, ImplicitSplitInserter, LimitOp
Optimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownMergeFacilitator, PushUpFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer])
2022-03-29 20:16:10,630 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 20:16:10,631 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 20:16:10,631 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 20:16:10,661 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-29 20:16:10,663 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 20:16:10,667 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 20:16:11,048 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - creating jar file Job5189562480314421076.jar
2022-03-29 20:16:14,083 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - jar file Job5189562480314421076.jar created
2022-03-29 20:16:14,082 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - Setting up single store job
2022-03-29 20:16:14,083 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-03-29 20:16:14,083 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-29 20:16:14,083 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2022-03-29 20:16:14,101 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-29 20:16:14,102 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 20:16:14,106 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-29 20:16:14,118 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 20:16:14,232 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-29 20:16:14,232 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
2022-03-29 20:16:14,342 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths (combined) to process : 1
2022-03-29 20:16:14,391 [JobControl] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputFormat - Total output paths to process : 1
2022-03-29 20:16:14,600 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644548343526_0033
2022-03-29 20:16:14,837 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644548343526_0033
2022-03-29 20:16:14,866 [JobControl] INFO org.apache.hadoop.mapreduce.job - The url to track the job: http://quickstart.cloudera:8080/proxy/application_1644548343526_0033/
2022-03-29 20:16:14,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - HadoopJobId: job_1644548343526_0033
2022-03-29 20:16:14,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Processing aliases A
2022-03-29 20:16:14,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - detailed locations: R: A[3..4], A[1..1], C: R
2022-03-29 20:16:14,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - More information at: http://localhost:8080/jobdetails.html?jobid=job_1644548343526_0033
```

```
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.4.2 0.12.0-cdh5.4.2 cloudera 2022-03-29 20:16:10 2022-03-29 20:16:35 UNKNOWN
Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_1644548343526_0033 1 0 5 5 5 5 n/a n/a n/a n/a A MAP_ONLY hdfs://quickstart.cloudera:8020/tmp/temp2008527170/tmp977761476,

Input(s):
Successfully read 3 records (417 bytes) from: "/user/cloudera/Training/pig/mydata.txt"

Output(s):
Successfully stored 3 records (30 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp2008527170/tmp977761476"

Counters:
Total records written : 3
Total bytes written : 30
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644548343526_0033

2022-03-29 20:16:35,540 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Success!
2022-03-29 20:16:35,541 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 20:16:35,541 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 20:16:35,541 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-29 20:16:35,554 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-29 20:16:35,554 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(10,20,30)
(40,50,60)
(70,80,90)
```

Step 5: LOAD B

B=LOAD 'user/cloudera/Training/pig/mydata.txt.

Name: Atharva
suroshe
Roll No: 40

```
grunt> B = LOAD '/user/cloudera/Training/pig/mydata.txt';
2022-03-29 20:19:09,155 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 20:19:09,156 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
grunts dump B;
2022-03-29 20:19:21,128 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-29 20:19:21,129 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptinizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2022-03-29 20:19:21,135 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 20:19:21,136 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 20:19:21,136 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 20:19:21,153 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-29 20:19:21,155 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 20:19:21,157 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 20:19:21,499 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job8711437822572202530.jar
2022-03-29 20:19:24,396 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job8711437822572202530.jar created
2022-03-29 20:19:24,405 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-03-29 20:19:24,414 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-03-29 20:19:24,414 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 20:19:24,417 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-29 20:19:24,436 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 20:19:24,593 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-29 20:19:24,594 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-03-29 20:19:24,596 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2022-03-29 20:19:24,637 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2022-03-29 20:19:24,722 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1644548343526_0034
2022-03-29 20:19:24,758 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1644548343526_0034
2022-03-29 20:19:24,759 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644548343526_0034/
2022-03-29 20:19:24,915 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1644548343526_0034
2022-03-29 20:19:24,915 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases B
2022-03-29 20:19:24,915 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: B[4,4] C: R:
2022-03-29 20:19:24,915 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1644548343526_0034
2022-03-29 20:19:24,959 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2022-03-29 20:19:37,779 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-29 20:19:39,988 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2022-03-29 20:19:39,989 [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  2022-03-29 20:19:21  2022-03-29 20:19:39  UNKNOWN

Success!

Job Stats (time in seconds):
```

```
Counters:
Total records written : 3
Total bytes written : 50
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644548343526_0034

2022-03-29 20:19:40,055 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-29 20:19:40,055 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 20:19:40,055 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 20:19:40,056 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 20:19:40,061 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-29 20:19:40,061 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10, 20, 30,)
(40, 50, 60,)
(70, 80, 90)
```

Step 6:now check Schema of A.

```
grunt> DESCRIBE A;
A: {c1: int,c2: int,c3: int}
grunt> DESCRIBE B;
Schema for B unknown.
```

Step 7: ILLUSTRATE A

```
grunt> ILLUSTRATE A;
2022-03-29 20:25:47,127 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 20:25:47,127 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 20:25:47,127 [main] INFO org.apache.pig.backend.hadoop.executionengine.MRExecutionEngine - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020
2022-03-29 20:25:47,128 [main] INFO org.apache.pig.backend.hadoop.executionengine.MRExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2022-03-29 20:25:47,128 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[DuplicateForEachColumnRewrite, ImplicitSplitInserter, LoadTypeCastInserter, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[AddForEach, ColumnMapKeyPrune, FilterLogicExpressionSimplifier, GroupByConstParallelSetter, LimitOptimizer, MergeFilter, MergeForEach, PartitionFilterOptimizer, PushDownForEachFlatten]}
2022-03-29 20:25:47,136 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 20:25:47,137 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 20:25:47,137 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 20:25:47,138 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 20:25:47,138 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 20:25:47,163 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 20:25:47,200 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Aliases being processed per job phase (AliasName[line,offset]): M: A[3,4] C: R:
2022-03-29 20:25:47,204 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-29 20:25:47,204 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-03-29 20:25:47,223 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - File concatenation threshold: 100 optimistic? false
2022-03-29 20:25:47,230 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 20:25:47,230 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 20:25:47,230 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 20:25:47,230 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 20:25:47,250 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 20:25:47,255 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Aliases being processed per job phase (AliasName[line,offset]): M: A[3,4] C: R:
2022-03-29 20:25:47,258 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 20:25:47,259 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 20:25:47,259 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 20:25:47,259 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 20:25:47,259 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 20:25:47,308 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 20:25:47,311 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Aliases being processed per job phase (AliasName[line,offset]): M: A[3,4] C: R:
(70, 80, 90)
2022-03-29 20:25:47,314 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 20:25:47,315 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
```


A	c1:int	c2:int	c3:int	
	70	80	90	

File name : STUDENT.txt

All the data available in this file

mydata.txt	STUDENT1.txt	student.pig
1,REHAN,22,NEW YORK,20000,2019-05-10		
2,RISHI,21,FLORIDA,15000,2018-09-11		
3,SHIVAM,23,BOSTON,17000,2020-10-10		
4,ANAND,24,CALIFORNIA,25000,2021-10-30		
5,PRINCE,23,CHICAGO,21000,2022-02-15		
6,AMANDEEP,22,SEATTLE,14000,2018-03-14		

Step 8: pig/home/cloudera/Documents/student.pig

Pig script file.

```
[cloudera@quickstart ~]$ pig /home/cloudera/Documents/student.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#config for more info.
2022-03-29 21:28:38.213 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.4.2 (repackaged) compiled May 19 2015, 17:03:41
2022-03-29 21:28:38.213 [main] INFO org.apache.pig.Main - Logging error messages to: /home/cloudera/pig/1640614510198.log
2022-03-29 21:28:31.275 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/cloudera/pig/bootstrap not found
2022-03-29 21:28:31.638 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:31.638 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:31.638 [main] INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2022-03-29 21:28:32.371 [main] INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to map-reduce job tracker at: localhost:8021
2022-03-29 21:28:32.410 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.411 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.463 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.467 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.528 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.529 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.602 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.605 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.680 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.687 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.759 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.765 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.822 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.825 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:32.870 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:28:32.870 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:33.545 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2022-03-29 21:28:33.551 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES ENABLED=[AddJoinEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptim�
nizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]
2022-03-29 21:28:33.609 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for listing: $1
2022-03-29 21:28:33.730 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 21:28:33.754 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 21:28:33.754 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 21:28:33.851 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2022-03-29 21:28:34.360 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 21:28:34.402 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
2022-03-29 21:28:34.463 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 21:28:34.463 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2022-03-29 21:28:35.526 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - creating jar file job27953402452876648.jar
2022-03-29 21:28:35.542 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - jar file job27953402452876648.jar created
2022-03-29 21:28:38.493 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is deprecated. Instead, use mapreduce.jar
2022-03-29 21:28:38.526 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.JobControlCompiler - Setting up single store job
2022-03-29 21:28:38.542 [main] INFO org.apache.pig.data.SchemaTupleFrontend - key [pig.schema tuple] is false. Will not generate code.
2022-03-29 21:28:38.542 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-03-29 21:28:38.542 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schema tuple classes] with classes to decompile []
2022-03-29 21:28:38.684 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 1 map-reduce jobs waiting for submission.
2022-03-29 21:28:38.687 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address
2022-03-29 21:28:38.687 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:28:38.691 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
```

Name: Atharva
suroshe
Roll No: 40

```
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
job_1644548343526_0039 1 0 6 6 6 6 n/a n/a n/a n/a listing_MAP_ONLY hdf://quickstart.cloudera:8020/tmp/temp262147121/tmp-385917375"

Input(s):
Successfully read 6 records (614 bytes) from: "/user/cloudera/Training/pig/STUDENT1.txt"

Output(s):
Successfully stored 6 records (194 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp262147121/tmp-385917375"

Counters:
Total records written : 6
Total bytes written : 194
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1644548343526_0039

2022-03-29 21:29:00,458 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-29 21:29:00,461 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:29:00,461 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-29 21:29:00,463 [main] INFO org.apache.pig.data.SchemaTupleBackend - key [pig.schemaTuple] was not set... will not generate code.
2022-03-29 21:29:00,479 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-29 21:29:00,479 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,2019-01-10T00:00:00.000-08:00,29000,22,NEW YORK)
(2,2018-01-11T00:00:00.000-08:00,15000,21,FLORIDA)
(3,2020-01-10T00:00:00.000-08:00,17000,23,BOSTON)
(4,2021-01-30T00:00:00.000-08:00,25000,24,CALIFORNIA)
(5,2022-01-15T00:00:00.000-08:00,21000,23,CHICAGO)
(6,2018-01-14T00:00:00.000-08:00,14000,22,SEATTLE)
2022-03-29 21:29:00,626 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-29 21:29:00,626 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

Display the patricular data:

Step 9:

Mention the data which we want to check. It will be display in listing form.

```
(2,RISHI,21,FLORIDA,15000,2018-09-11)
2022-03-29 21:35:36,565 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 21:35:36,575 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 21:35:36,575 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 21:35:36,576 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 21:35:36,581 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 21:35:36,620 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 21:35:36,632 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Aliases being processed per job phase (AliasName[line,offset]): M: listing[1,10],listing[-1,-1],listing[3,10] C: R:
2022-03-29 21:35:36,633 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 21:35:36,637 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 21:35:36,638 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 21:35:36,639 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 21:35:36,639 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 21:35:36,671 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 21:35:36,674 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Aliases being processed per job phase (AliasName[line,offset]): M: listing[1,10],listing[-1,-1],listing[3,10] C: R:
2022-03-29 21:35:36,680 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-29 21:35:36,687 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-29 21:35:36,687 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-29 21:35:36,688 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-29 21:35:36,688 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-29 21:35:36,707 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-03-29 21:35:36,725 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Aliases being processed per job phase (AliasName[line,offset]): M: listing[1,10],listing[-1,-1],listing[3,10] C: R:

-----
| listing | sid:int | name:chararray | age:int | address:chararray | salary:long | date:chararray | |
|---|---|---|---|---|---|---|---|
| | 2 | | RISHI | 21 | FLORIDA | 15000 | 2018-09-11 |
|-----|-----|-----|-----|-----|-----|
| listing | sid:int | name:datetime | date:chararray | salary:long | age:int | address:chararray |
|-----|-----|-----|-----|-----|-----|
| | 2 | 2018-01-11T00:00:00.000-08:00 | 2018-09-11 | 15000 | 21 | FLORIDA |
|-----|-----|-----|-----|-----|-----|
```