hadoop@ab4cseibm133: ~

```
hadoop@ab4cseibm133:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ab4cseibm133]
Starting resourcemanager
Starting nodemanagers
hadoop@ab4cseibm133:~$ jps
22513 DataNode
22996 ResourceManager
23238 NodeManager
23591 Jps
22364 NameNode
22717 SecondaryNameNode
hadoop@ab4cseibm133:~$ hadoop fs -mkdir hdfs://localhost:9000/ex10
mkdir: `hdfs://localhost:9000/ex10': File exists
hadoop@ab4cseibm133:~$ hadoop fs -put '/home/hadoop/student_details.txt' /ex10
hadoop@ab4cseibm133:~$ pig -x local
2022-10-06 10:05:13,220 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2022-10-06 10:05:13,220 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2022-10-06 10:05:13,277 [main] INFO  org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-10-06 10:05:13,277 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1665030913276.log
2022-10-06 10:05:13,328 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found
2022-10-06 10:05:13,441 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-10-06 10:05:13,443 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2022-10-06 10:05:13,502 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-10-06 10:05:13,514 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-192573b8-a43f-465a-b4a8-8e07b0d819fb
2022-10-06 10:05:13,514 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt> student_details = LOAD 'hdfs://localhost:9000/pig_data/student_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray,age:int, phone:chararray, city:chararray);
2022-10-06 10:06:14,346 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
```

hadoop@ab4cseibm133: ~

```
Counters:
Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1838784688_0003        ->      job_local773754855_0004,
job_local773754855_0004 ->      job_local2015787209_0005,
job_local2015787209_0005

2022-10-06 10:13:52,689 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,691 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,692 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,696 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,698 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,699 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,703 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,705 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,706 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:13:52,708 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-10-06 10:13:52,709 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-10-06 10:13:52,709 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-10-06 10:13:52,711 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-10-06 10:13:52,711 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(8,Bharathi,Nambiayar,24,9848022333,Chennai)
(7,Komal,Nayak,24,9848022334,trivendram)
(6,Archana,Mishra,23,9848022335,Chennai)
(5,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar)
(3,Rajesh,Khanna,22,9848022339,Delhi)
(2,siddarth,Battacharya,22,9848022338,Kolkata)
(4,Preethi,Agarwal,21,9848022330,Pune)
(1,Rajiv,Reddy,21,9848022337,Hyderabad)
grunt>
```

hadoop@ab4cseibm133: ~

```
Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   Median
ReduceTime      Alias   Feature Outputs
job_local664789670_0006 1       1       n/a     n/a     n/a     n/a     n/a     n/a     n/a     n/a     group_data,student_details      GROUP_
BY      file:/tmp/temp1257080311/tmp-1137725553,

Input(s):
Successfully read 8 records (1356 bytes) from: "hdfs://localhost:9000/ex10/student_details.txt"

Output(s):
Successfully stored 4 records in: "file:/tmp/temp1257080311/tmp-1137725553"

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

Job DAG:
job_local664789670_0006


2022-10-06 10:35:26,077 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:35:26,078 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:35:26,080 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 10:35:26,082 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-10-06 10:35:26,083 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-10-06 10:35:26,083 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-10-06 10:35:26,085 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-10-06 10:35:26,085 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(21,{(4,Preethi,Agarwal,21,9848022330,Pune),(1,Rajiv,Reddy,21,9848022337,Hyderabad)})
(22,{(3,Rajesh,Khanna,22,9848022339,Delhi),(2,siddarth,Battacharya,22,9848022338,Kolkata)})
(23,{(6,Archana,Mishra,23,9848022335,Chennai),(5,Trupthi,Mohanthy,23,9848022336,Bhuwaneshwar)})
(24,{(8,Bharathi,Nambiayar,24,9848022333,Chennai),(7,Komal,Nayak,24,9848022334,trivendram)})
grunt>
```

hadoop@ab4cseibm133: ~

```
Input(s):
Successfully read 8 records from: "hdfs://localhost:9000/ex10/customers.txt"
Successfully read 8 records from: "hdfs://localhost:9000/ex10/customers.txt"

Output(s):
Successfully stored 7 records in: "file:/tmp/temp-1019090238/tmp-1326185156"

Counters:
Total records written : 7
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1957753093_0001

2022-10-06 11:16:22,833 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:16:22,835 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:16:22,836 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:16:22,841 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning ACCE
SSING_NON_EXISTENT_FIELD 8 time(s).
2022-10-06 11:16:22,841 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-10-06 11:16:22,843 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-10-06 11:16:22,843 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-10-06 11:16:22,847 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-10-06 11:16:22,847 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Ramesh,32,Ahmedabad,2000,1,Ramesh,32,Ahmedabad,2000)
(2,Khilan,25,Delhi,1500,2,Khilan,25,Delhi,1500)
(3,kaushik,23,Kota,2000,3,kaushik,23,Kota,2000)
(4,Chaitali,25,Mumbai,6500,4,Chaitali,25,Mumbai,6500)
(5,Hardik,27,Bhopal,8500,5,Hardik,27,Bhopal,8500)
(6,Komal,22,MP,4500,6,Komal,22,MP,4500)
(7,Muffy,24,Indore,10000,7,Muffy,24,Indore,10000)
grunt>
```

hadoop@ab4cseibm133: ~

hadoop@ab4cseibm133: ~   ×   hadoop@ab4cseibm133: ~   ×

```
Job Stats (time in seconds):
JobId       Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   Median
Reducetime      Alias   Feature Outputs
job_local875847837_0002 1       0       n/a     n/a     n/a     n/a     0       0       0       0       filter_data,student_details     MAP_ON
LY      file:/tmp/temp-1019090238/tmp-66183060,

Input(s):
Successfully read 8 records (717 bytes) from: "hdfs://localhost:9000/ex10/student_details.txt"

Output(s):
Successfully stored 2 records in: "file:/tmp/temp-1019090238/tmp-66183060"

Counters:
Total records written : 2
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local875847837_0002

2022-10-06 11:29:13,638 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:29:13,640 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:29:13,641 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:29:13,644 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-10-06 11:29:13,645 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-10-06 11:29:13,645 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-10-06 11:29:13,646 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-10-06 11:29:13,647 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(6,Archana,Mishra,23,9848022335,Chennai)
(8,Bharathi,Nambiayar,24,9848022333,Chennai)
grunt>
```

hadoop@ab4cseibm133: ~

hadoop@ab4cseibm133: ~   ×   hadoop@ab4cseibm133: ~   ×

```
Input(s):
Successfully read 8 records (1056 bytes) from: "hdfs://localhost:9000/ex10/student_details.txt"

Output(s):
Successfully stored 8 records in: "file:/tmp/temp-1019090238/tmp-845785205"

Counters:
Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local684544580_0003

2022-10-06 11:31:04,864 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:31:04,866 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:31:04,867 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-10-06 11:31:04,870 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-10-06 11:31:04,870 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-10-06 11:31:04,871 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-10-06 11:31:04,872 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-10-06 11:31:04,872 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,21,Hyderabad)
(2,22,Kolkata)
(3,22,Delhi)
(4,21,Pune)
(5,23,Bhuwaneshwar)
(6,23,Chennai)
(7,24,trivendram)
(8,24,Chennai)
grunt>
```

Join:

customers = LOAD 'hdfs://localhost:9000/ex10/customers.txt' USING PigStorage(',') as (id:int, name:chararray, age:int, address:chararray, salary:int);

orders = LOAD 'hdfs://localhost:9000/ex10/orders.txt' USING PigStorage(',') as (oid:int, date:chararray, customer_id:int, amount:int);

customers1 = LOAD 'hdfs://localhost:9000/ex10/customers.txt' USING PigStorage(',') as (id:int, name:chararray, age:int, address:chararray, salary:int);

customers2 = LOAD 'hdfs://localhost:9000/ex10/customers.txt' USING PigStorage(',') as (id:int, name:chararray, age:int, address:chararray, salary:int);

Group:

group_data = GROUP student_details by age;
Dump group_data;

Project:

foreach_data = FOREACH student_details GENERATE id,age,city;
Dump foreach_data;

Filter:

filter_data = FILTER student_details BY city == 'Chennai';
Dump filter_data;

Sort:

order_by_data = ORDER student_details BY age DESC;
Dump order_by_data;