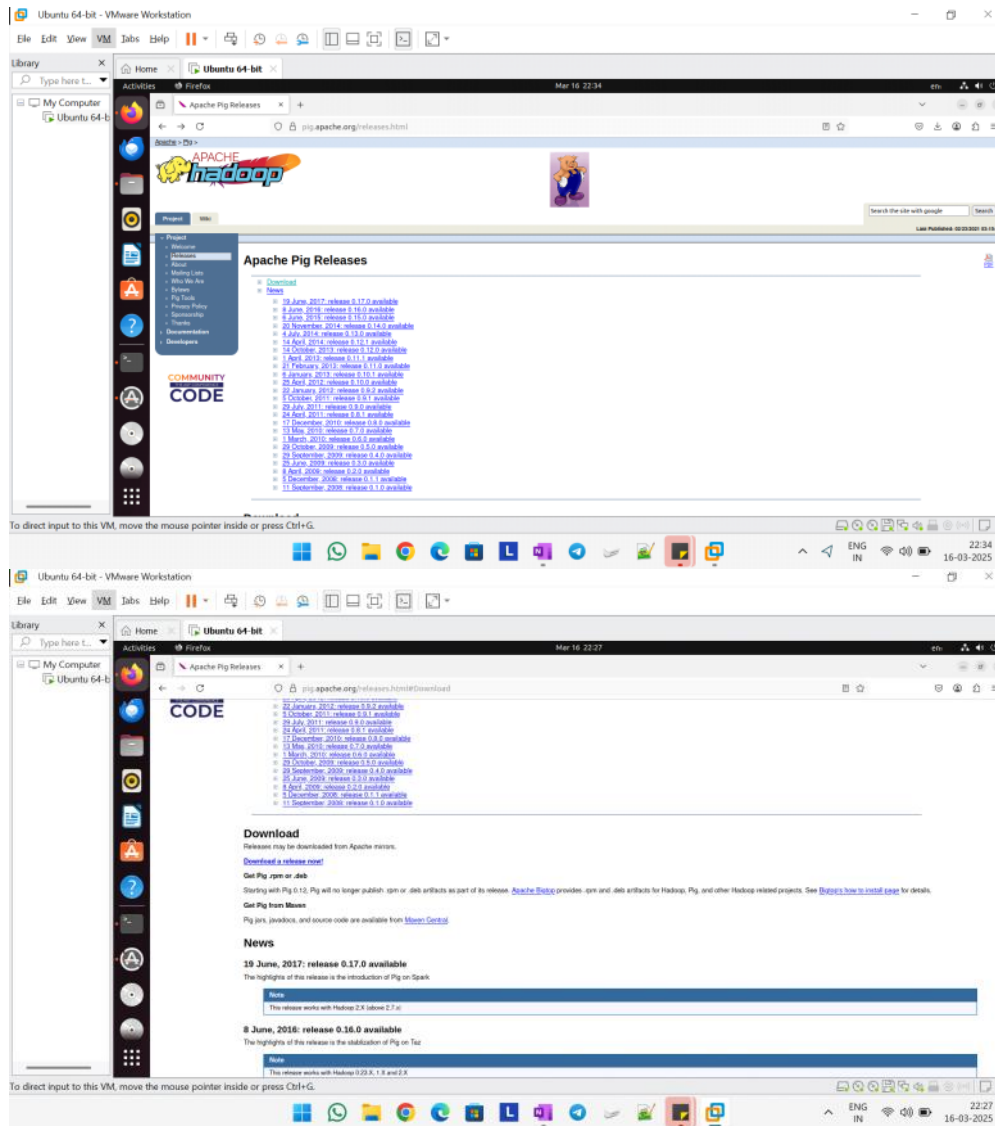
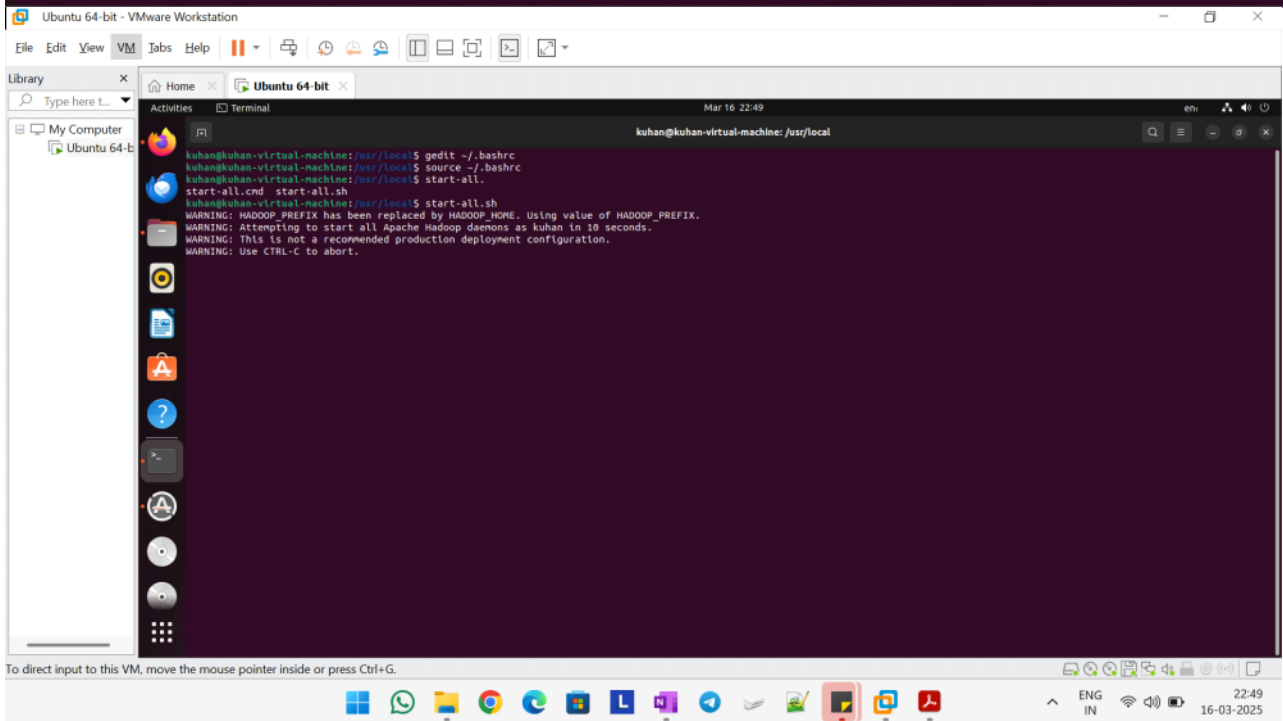
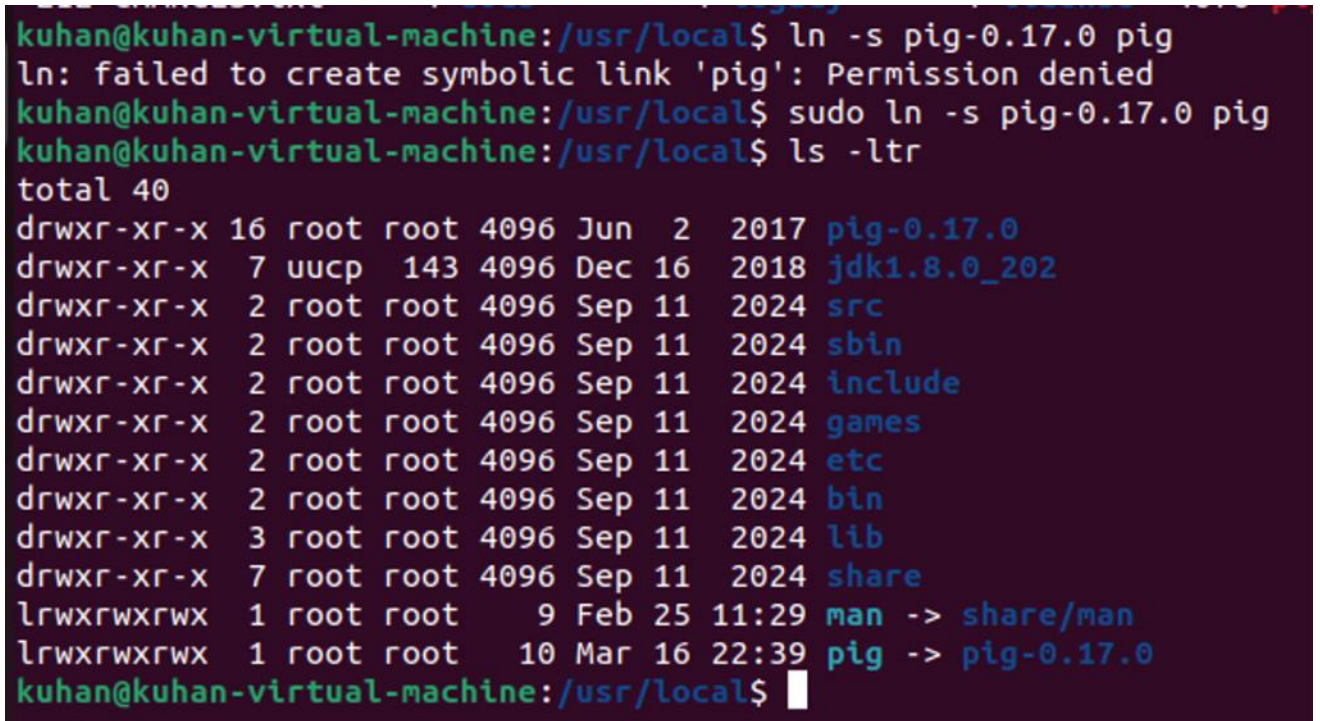
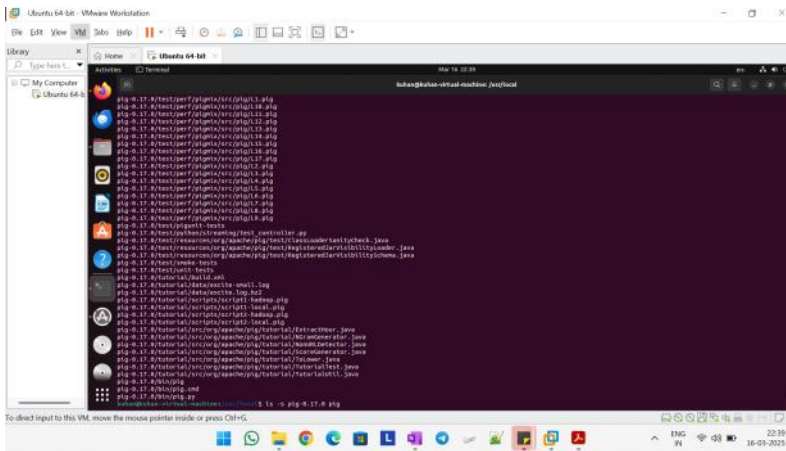


Pig Installation

16 March 2025 22:26

Submitted by **Kuhan.K**
Roll No. **2024MCS110028**





Activities Text Editor Mar 17 00:58

Open [?] .bashrc ~/

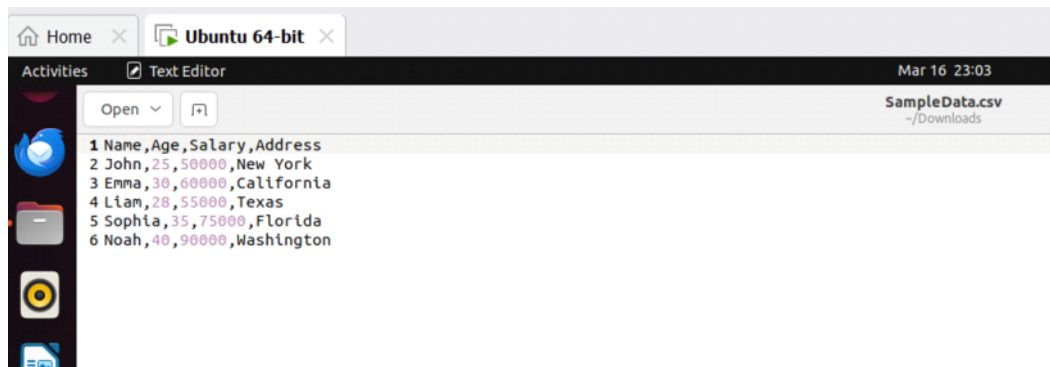
```
102 # See /usr/share/doc/bash-doc/examples in the bash-doc package.
103
104 if [ -f ~/.bash_aliases ]; then
105     . ~/.bash_aliases
106 fi
107
108 # enable programmable completion features (you don't need to enable
109 # this, if it's already enabled in /etc/bash.bashrc and /etc/profile
110 # sources /etc/bash.bashrc).
111 if ! shopt -oq posix; then
112     if [ -f /usr/share/bash-completion/bash_completion ]; then
113         . /usr/share/bash-completion/bash_completion
114     elif [ -f /etc/bash_completion ]; then
115         . /etc/bash_completion
116     fi
117 fi
118
119 # Set Java Home
120 export JAVA_HOME=/usr/local/jdk1.8.0_202
121
122 # Set Hadoop Variables
123 export HADOOP_HOME=/opt/hadoop-3.4.0
124 export HADOOP_PREFIX=$HADOOP_HOME
125 export HADOOP_MAPRED_HOME=$HADOOP_HOME
126 export HADOOP_COMMON_HOME=$HADOOP_HOME
127 export HADOOP_HDFS_HOME=$HADOOP_HOME
128 export YARN_HOME=$HADOOP_HOME
129 export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
130 export YARN_CONF_DIR=$HADOOP_HOME/etc/hadoop
131
132 # Set Hadoop Classpath
133 export HADOOP_CLASSPATH='$HADOOP_HOME/bin/hadoop.classpath'
134
135 # Fix Library Path
136 export JAVA_LIBRARY_PATH=$HADOOP_HOME/lib/native
137
138 # Set PATH
139 export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$JAVA_LIBRARY_PATH
140
141 # Set Pig Variables
142 export PIG_HOME=/usr/local/pig-0.17.0
```

```
kuhan@kuhan-virtual-machine:/usr/local$ pig -version
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58
kuhan@kuhan-virtual-machine:/usr/local$
```

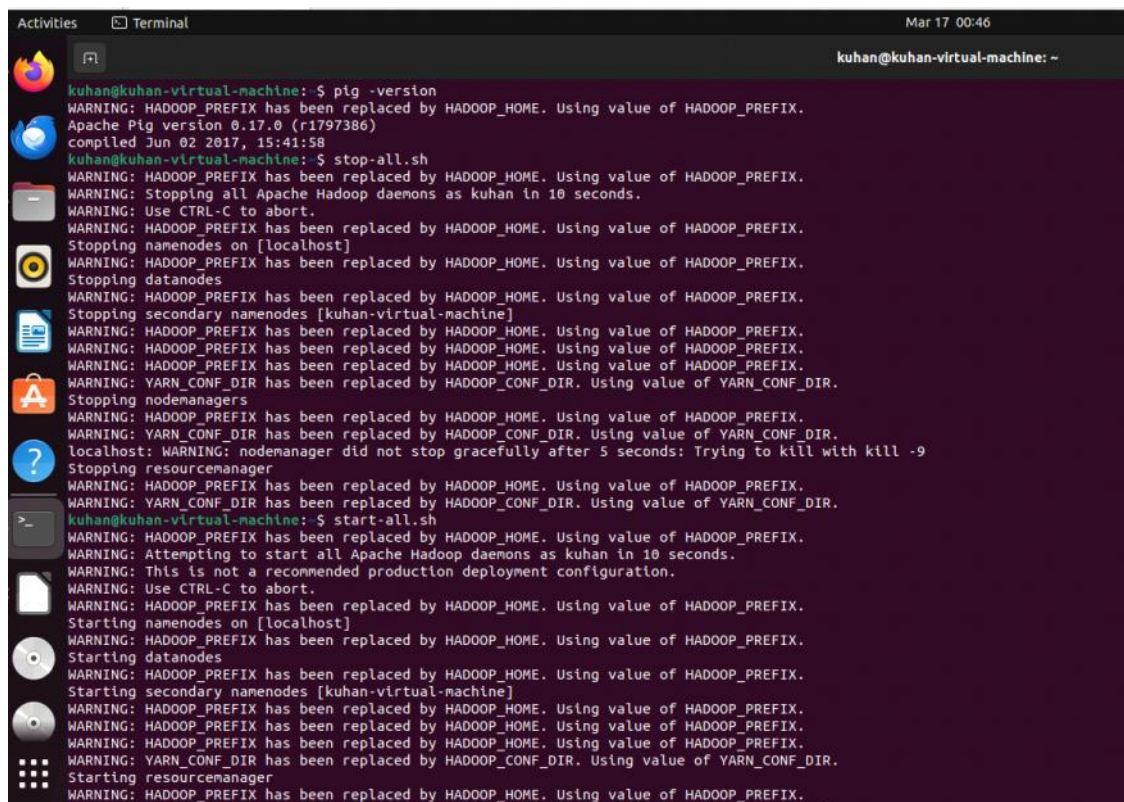

Pig commands

17 March 2025 00:45

create a csv file



check pig version start all the services



```
Activities Terminal Mar 17 00:46
kuhan@kuhan-virtual-machine: ~
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Stopping nodemanagers
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
localhost: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
Stopping resourcemanager
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
kuhan@kuhan-virtual-machine:~$ start-all.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Attempting to start all Apache Hadoop daemons as kuhan in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting namenodes on [localhost]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting datanodes
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting secondary namenodes [kuhan-virtual-machine]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Starting resourcemanager
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Starting nodemanagers
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
kuhan@kuhan-virtual-machine:~$ start-yarn.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Starting resourcemanager
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
resourcemanager is running as process 34444. Stop it first and ensure /tmp/hadoop-kuhan-resourcemanager.pid file is empty before retry.
Starting nodemanagers
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
localhost: nodemanager is running as process 34576. Stop it first and ensure /tmp/hadoop-kuhan-nodemanager.pid file is empty before retry.
kuhan@kuhan-virtual-machine:~$
```

```
Home Ubuntu 64-bit Mar 17 00:47
kuhan@kuhan-virtual-machine: ~
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
kuhan@kuhan-virtual-machine:~$ start-all.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Attempting to start all Apache Hadoop daemons as kuhan in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting namenodes on [localhost]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting datanodes
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Starting secondary namenodes [kuhan-virtual-machine]
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Starting resourcemanager
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Starting nodemanagers
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
kuhan@kuhan-virtual-machine:~$ start-yarn.sh
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
Starting resourcemanager
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
resourcemanager is running as process 34444. Stop it first and ensure /tmp/hadoop-kuhan-resourcemanager.pid file is empty before retry.
Starting nodemanagers
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
localhost: nodemanager is running as process 34576. Stop it first and ensure /tmp/hadoop-kuhan-nodemanager.pid file is empty before retry.
kuhan@kuhan-virtual-machine:~$ mr-jobhistory-daemon.sh start historyserver
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.
WARNING: Use of this script to start the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon start" instead.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
historyserver is running as process 28902. Stop it first and ensure /tmp/hadoop-kuhan-historyserver.pid file is empty before retry.
kuhan@kuhan-virtual-machine:~$
```

```
historyserver is running as process 28902.
kuhan@kuhan-virtual-machine:~$ jps
34576 NodeManager
28902 JobHistoryServer
35223 Jps
34026 DataNode
33900 NameNode
34444 ResourceManager
34254 SecondaryNameNode
kuhan@kuhan-virtual-machine:~$
```

Make directory
upload the csv file


```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -mkdir /pig_data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hadoop fs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 2 items
drwxr-xr-x - kuhan supergroup          0 2025-03-08 23:30 /data
drwxr-xr-x - kuhan supergroup          0 2025-03-16 23:11 /pig_data
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hdfs dfs -put ~/Downloads/SampleData.csv /pig_data
```

```
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$ hdfs dfs -ls /pig_data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rw-r--r-- 1 kuhan supergroup          141 2025-03-16 23:13 /pig_data/SampleData.csv
kuhan@kuhan-virtual-machine:/opt/hadoop-3.4.0$
```

```
kuhan@kuhan-virtual-machine:~$ hdfs dfs -rm -r /pig_data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Deleted /pig_data
kuhan@kuhan-virtual-machine:~$ hdfs dfs -rm -r /output
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Deleted /output
kuhan@kuhan-virtual-machine:~$ hdfs dfs -rm -r /tmp/hadoop-yarn/staging/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Deleted /tmp/hadoop-yarn/staging
kuhan@kuhan-virtual-machine:~$ hdfs dfs -ls /
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 2 items
drwxr-xr-x - kuhan supergroup          0 2025-03-08 23:30 /data
drwxrwxrwx - kuhan supergroup          0 2025-03-17 00:02 /tmp
kuhan@kuhan-virtual-machine:~$ hdfs dfs -mkdir -p /pig_data
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:~$ hdfs dfs -put ./Downloads/SampleData.csv /pig_data/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
kuhan@kuhan-virtual-machine:~$ hdfs dfs -ls /pig_data/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rw-r--r-- 1 kuhan supergroup          141 2025-03-17 00:25 /pig_data/SampleData.csv
```

```
kuhan@kuhan-virtual-machine:~$ hdfs dfs -ls /pig_data/
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Found 1 items
-rwxrwxrwx 1 kuhan supergroup          141 2025-03-17 00:25 /pig_data/SampleData.csv
kuhan@kuhan-virtual-machine:~$ hdfs dfs -cat /pig_data/SampleData.csv
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
Name,Age,Salary,Address
John,25,50000,New York
Emma,30,60000,California
Liam,28,55000,Texas
Sophia,35,75000,Florida
Noah,40,90000,Washington
kuhan@kuhan-virtual-machine:~$
```

Start pig in mapreduce mode
Inside the Pig shell (grunt>), run commands
Check if data is loaded properly
Dump data to check contents:


```
Home x Ubuntu 64-bit x
Activities Terminal Mar 17 00:50 enu
kuhan@kuhan-virtual-machine: ~
3.4.0 0.17.0 kuhan 2025-03-17 00:50:36 2025-03-17 00:50:54 FILTER
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime Alias Feature Outputs
Job_1742152579100_0002 1 0 2 2 2 0 0 0 0 data,filtered_data MAP_ONLY hdfs://localhost:9000/tmp/tempt1841971064/tmp1785592046,
Input(s):
Successfully read 6 records (513 bytes) from: "/pig_data/SampleData.csv"
Output(s):
Successfully stored 3 records (92 bytes) in: "hdfs://localhost:9000/tmp/tempt1841971064/tmp1785592046"
Counters:
Total records written : 3
Total bytes written : 92
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
Job_1742152579100_0002
2025-03-17 00:50:54,405 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFallOverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:50:54,410 [main] INFO org.apache.hadoop.mapred.clientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:50:54,437 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFallOverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:50:54,441 [main] INFO org.apache.hadoop.mapred.clientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:50:54,460 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFallOverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:50:54,464 [main] INFO org.apache.hadoop.mapred.clientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:50:54,481 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s).
2025-03-17 00:50:54,482 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2025-03-17 00:50:54,482 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2025-03-17 00:50:54,483 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2025-03-17 00:50:54,493 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2025-03-17 00:50:54,493 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Emma,30,60000,California)
(Sophia,35,75000,Florida)
(Noah,40,90000,Washington)
grunt>
```

Limit to first 3 records

```
(Sophia,35,75000,Florida)
(Noah,40,90000,Washington)
grunt> limited_data = LIMIT data 3;
grunt> DUMP limited_data;
```

```
(Noah,40,90000,Washington)
grunt> limited_data = LIMIT data 3;
grunt> DUMP limited_data;
2025-03-17 00:51:31,815 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2025-03-17 00:51:31,825 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2025-03-17 00:51:31,826 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2025-03-17 00:51:31,826 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LogicalPlanOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInspector], LogicalPlanOptimizer, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInspector}
2025-03-17 00:51:31,845 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File OutputCommitter Algorithm version is 2
2025-03-17 00:51:31,846 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures : false
2025-03-17 00:51:31,854 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2025-03-17 00:51:31,865 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2025-03-17 00:51:31,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input files to process : 1
2025-03-17 00:51:31,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2025-03-17 00:51:32,322 [main] INFO org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2025-03-17 00:51:32,334 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2025-03-17 00:51:32,337 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input files to process : 1
2025-03-17 00:51:32,337 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Name,,Address)
(John,25,50000,New York)
(Emma,30,60000,California)
grunt>
```

Order by Salary (Descending)

```
grunt> ordered_data = ORDER data BY Salary DESC;
grunt> DUMP ordered_data;
2025-03-17 00:51:51,598 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER_BY
2025-03-17 00:51:51,606 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2025-03-17 00:51:51,607 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2025-03-17 00:51:51,607 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LogicalPlanOptimizer, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInspector], LogicalPlanOptimizer, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInspector}
2025-03-17 00:51:51,610 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimize? false
2025-03-17 00:51:51,627 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope=72
2025-03-17 00:51:51,629 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 3
2025-03-17 00:51:51,629 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 3
2025-03-17 00:51:51,640 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2025-03-17 00:51:51,642 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFallOverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:51:51,647 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2025-03-17 00:51:51,648 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2025-03-17 00:51:51,648 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2025-03-17 00:51:51,711 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig-0.17.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/tempt1841971064/tmp-2072860633/pig-0.17.0-core-h2.jar
2025-03-17 00:51:52,137 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig-0.17.0/lib/automaton-1.11-8.jar to DistributedCache through /tmp/tempt1841971064/tmp-2072860633/pig-0.17.0-core-h2.jar
2025-03-17 00:51:52,564 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig-0.17.0/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/tempt1841971064/tmp-799550604/antlr-runtime-3.4.jar
2025-03-17 00:51:52,986 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/local/pig-0.17.0/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/tempt1841971064/tmp-1056958321/joda-time-2.9.3.jar
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
```



```
Activities Terminal Mar 17 00:52 eni kuhan@kuhan-virtual-machine: -

bled
2025-03-17 00:51:51,642 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:51:51,647 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2025-03-17 00:51:51,648 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - mapreduce.markreset.buffer.percent is not set, set to default 0.3
2025-03-17 00:51:51,648 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - This job cannot be converted run in-process
2025-03-17 00:51:51,711 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Added jar file:/usr/local/pig-0.17.0/pig-0.17.0-core-h2.jar to DistributedCache thro
ugh /tmp/tenp1841971064/tmp-2072860633/pig-0.17.0-core-h2.jar
2025-03-17 00:51:52,137 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Added jar file:/usr/local/pig-0.17.0/lib/automaton-1.11-8.jar to DistributedCache th
rough /tmp/tenp1841971064/tmp-2072860633/pig-0.17.0-core-h2.jar
2025-03-17 00:51:52,564 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Added jar file:/usr/local/pig-0.17.0/lib/antlr-runtime-3.4.jar to DistributedCache t
hrough /tmp/tenp1841971064/tmp-2072860633/pig-0.17.0-core-h2.jar
2025-03-17 00:51:52,586 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Added jar file:/usr/local/pig-0.17.0/lib/joda-time-2.9.3.jar to DistributedCache thr
ough /tmp/tenp1841971064/tmp-2072860633/pig-0.17.0-core-h2.jar
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Setting up single store job
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2025-03-17 00:51:52,987 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2025-03-17 00:51:52,997 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - 1 map-reduce job(s) waiting for submission.
2025-03-17 00:51:52,999 [JobControl] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:51:53,022 [JobControl] INFO org.apache.hadoop.mapreduce.JobResourceUploader - Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kuhan/.staging/job_1742152579100_0003
2025-03-17 00:51:53,024 [JobControl] WARN org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User classes may not be found. See Job or JobsetJar(String).
2025-03-17 00:51:53,034 [JobControl] INFO org.apache.pig.builtin.PigStorage - Using PigTextInputFormat
2025-03-17 00:51:53,035 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2025-03-17 00:51:53,035 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2025-03-17 00:51:53,037 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2025-03-17 00:51:53,082 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2025-03-17 00:51:54,309 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1742152579100_0003
2025-03-17 00:51:54,309 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Executing with tokens: []
2025-03-17 00:51:54,312 [JobControl] INFO org.apache.hadoop.mapreduce.YARNRunner - Job jar is not present. Not adding any jar to the list of resources.
2025-03-17 00:51:54,344 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1742152579100_0003
2025-03-17 00:51:54,349 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://kuhan-virtual-machine:8088/proxy/application_1742152579100_0003/
2025-03-17 00:51:54,349 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - HadoopJobId: job_1742152579100_0003
2025-03-17 00:51:54,349 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Processing aliases data
2025-03-17 00:51:54,349 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Detailed locations: M: data[1,7],data[-1,1] C: R:
2025-03-17 00:51:54,352 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - 0% complete
2025-03-17 00:51:54,352 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Running jobs are [job_1742152579100_0003]
2025-03-17 00:52:04,610 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - 16% complete
2025-03-17 00:52:04,610 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Running jobs are [job_1742152579100_0003]
2025-03-17 00:52:09,621 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - 33% complete
2025-03-17 00:52:09,624 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:09,633 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
```

```
Home x Ubuntu 64-bit x Activities Terminal Mar 17 00:52 eni kuhan@kuhan-virtual-machine: -

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

Job DAG:
job_1742152579100_0003 -> job_1742152579100_0004,
job_1742152579100_0004 -> job_1742152579100_0005,
job_1742152579100_0005

2025-03-17 00:52:55,370 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,385 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,434 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,444 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,481 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,488 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,547 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,556 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,589 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,595 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,620 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,628 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,676 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,717 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,727 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,777 [main] INFO org.apache.hadoop.yarn.client.DefaultNOHARMFalloverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2025-03-17 00:52:55,788 [main] INFO org.apache.hadoop.mapreduce.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2025-03-17 00:52:55,837 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s).
2025-03-17 00:52:55,837 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Success!
2025-03-17 00:52:55,838 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
bled
2025-03-17 00:52:55,839 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2025-03-17 00:52:55,853 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2025-03-17 00:52:55,853 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Noah,40,90000,Washington)
(Sophia,35,75000,Florida)
(Emma,30,60000,California)
(Llan,28,55000,Texas)
(John,25,50000,New York)
(Name,,,Address)
grunt>
```