

EXP NO: 4

CREATE UDF IN PIG

\$start-all.sh

\$ jps

```

krish@krish-VirtualBox: ~
Starting namenodes on [localhost]
Starting datanodes
localhost: datanode is running as process 4172. Stop it first and ensure /tmp/hadoop-krish-datanode.pid file is empty before retry.
Starting secondary namenodes [krish-VirtualBox]
krish-VirtualBox: secondarynamenode is running as process 4366. Stop it first and ensure /tmp/hadoop-krish-secondarynamenode.pid file is empty before retry.
Starting resourcemanager
resourcemanager is running as process 4573. Stop it first and ensure /tmp/hadoop-krish-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 4715. Stop it first and ensure /tmp/hadoop-krish-nodemanager.pid file is empty before retry.
krish@krish-VirtualBox:~$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as krish in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [krish-VirtualBox]
Stopping nodemanagers
Stopping resourcemanager
krish@krish-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as krish 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 [krish-VirtualBox]
Starting resourcemanager
Starting nodemanagers
krish@krish-VirtualBox:~$ jps
8721 Jps
8009 SecondaryNameNode
7626 NameNode
7773 DataNode
8351 NodeManager
8207 ResourceManager
krish@krish-VirtualBox:~$ nano mapper.py
krish@krish-VirtualBox:~$ nano mapper.py
krish@krish-VirtualBox:~$ nano reducer.py
krish@krish-VirtualBox:~$

```

\$wget <https://dclcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz>

```

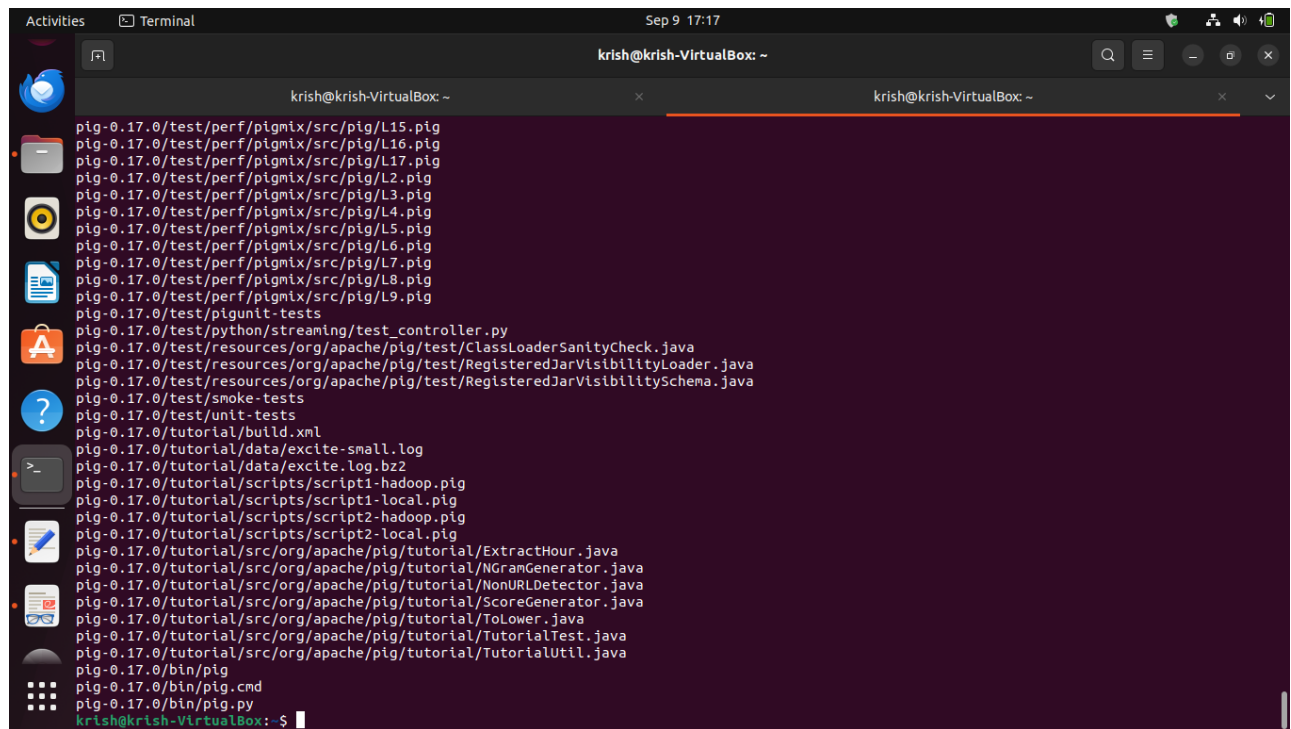
krish@krish-VirtualBox: ~
krish@krish-VirtualBox:~$ wget https://dclcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz
--2024-09-09 17:15:38-- https://dclcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz
Resolving dclcdn.apache.org (dclcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dclcdn.apache.org (dclcdn.apache.org)|151.101.2.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 230606579 (220M) [application/x-gzip]
Saving to: 'pig-0.17.0.tar.gz'

pig-0.17.0.tar.gz      100%[=====] 219.92M  5.86MB/s   in 41s
2024-09-09 17:16:28 (5.36 MB/s) - 'pig-0.17.0.tar.gz' saved [230606579/230606579]

krish@krish-VirtualBox:~$

```

\$ tar xvfz pig-0.16.0.tar.gz

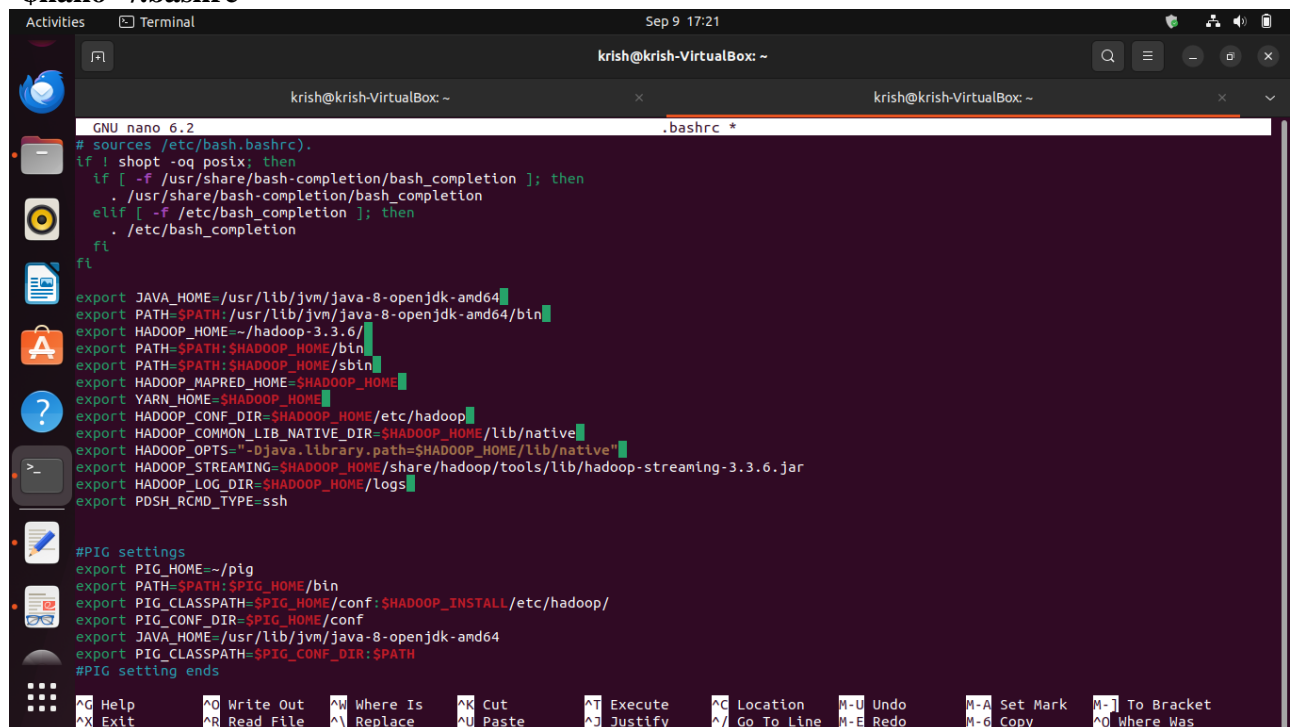


```

krish@krish-VirtualBox: ~
krish@krish-VirtualBox: ~
pig-0.17.0/test/perf/pigmix/src/pig/L15.pig
pig-0.17.0/test/perf/pigmix/src/pig/L16.pig
pig-0.17.0/test/perf/pigmix/src/pig/L17.pig
pig-0.17.0/test/perf/pigmix/src/pig/L2.pig
pig-0.17.0/test/perf/pigmix/src/pig/L3.pig
pig-0.17.0/test/perf/pigmix/src/pig/L4.pig
pig-0.17.0/test/perf/pigmix/src/pig/L5.pig
pig-0.17.0/test/perf/pigmix/src/pig/L6.pig
pig-0.17.0/test/perf/pigmix/src/pig/L7.pig
pig-0.17.0/test/perf/pigmix/src/pig/L8.pig
pig-0.17.0/test/perf/pigmix/src/pig/L9.pig
pig-0.17.0/test/pigunit-tests
pig-0.17.0/test/python/streaming/test_controller.py
pig-0.17.0/test/resources/org/apache/pig/test/ClassLoaderSanityCheck.java
pig-0.17.0/test/resources/org/apache/pig/test/RegisteredJarVisibilityLoader.java
pig-0.17.0/test/resources/org/apache/pig/test/RegisteredJarVisibilitySchema.java
pig-0.17.0/test/smoke-tests
pig-0.17.0/test/unit-tests
pig-0.17.0/tutorial/build.xml
pig-0.17.0/tutorial/data/excite-small.log
pig-0.17.0/tutorial/data/excite.log.bz2
pig-0.17.0/tutorial/scripts/script1-hadoop.pig
pig-0.17.0/tutorial/scripts/script1-local.pig
pig-0.17.0/tutorial/scripts/script2-hadoop.pig
pig-0.17.0/tutorial/scripts/script2-local.pig
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/ExtractHour.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/MGramGenerator.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/NonURLDetector.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/ScoreGenerator.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/ToLower.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/TutorialTest.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/TutorialUtil.java
pig-0.17.0/bin/pig
pig-0.17.0/bin/pig.cmd
pig-0.17.0/bin/pig.py
krish@krish-VirtualBox: ~$

```

\$nano ~/.bashrc



```

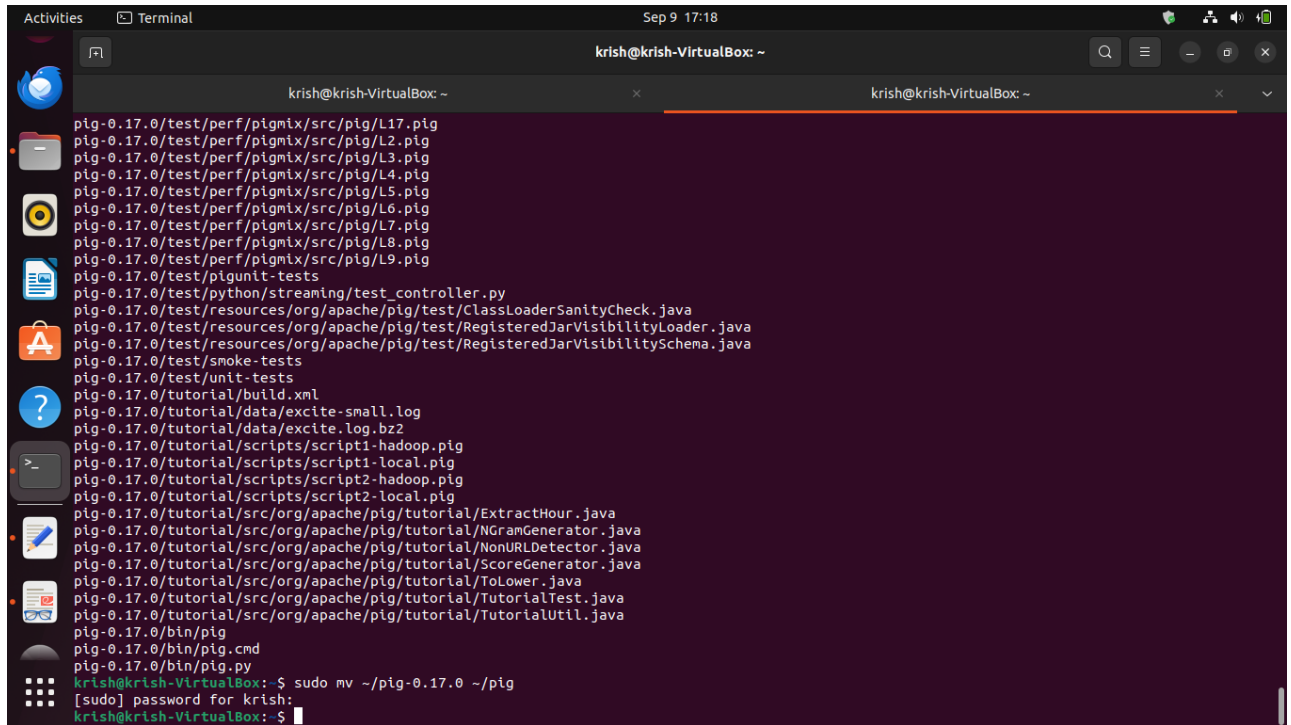
GNU nano 6.2                                .bashrc *
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin
export HADOOP_HOME=/hadoop-3.3.6/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh

#PIG settings
export PIG_HOME=/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_INSTALL/etc/hadoop/
export PIG_CONF_DIR=$PIG_HOME/conf
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
#PIG setting ends

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^J Execute    ^C Location   ^U Undo       ^M-A Set Mark ^-] To Bracket
^X Exit      ^R Read File  ^_ Replace    ^U Paste      ^D Justify    ^_ Go To Line ^-E Redo      ^M-6 Copy     ^_ Where Was

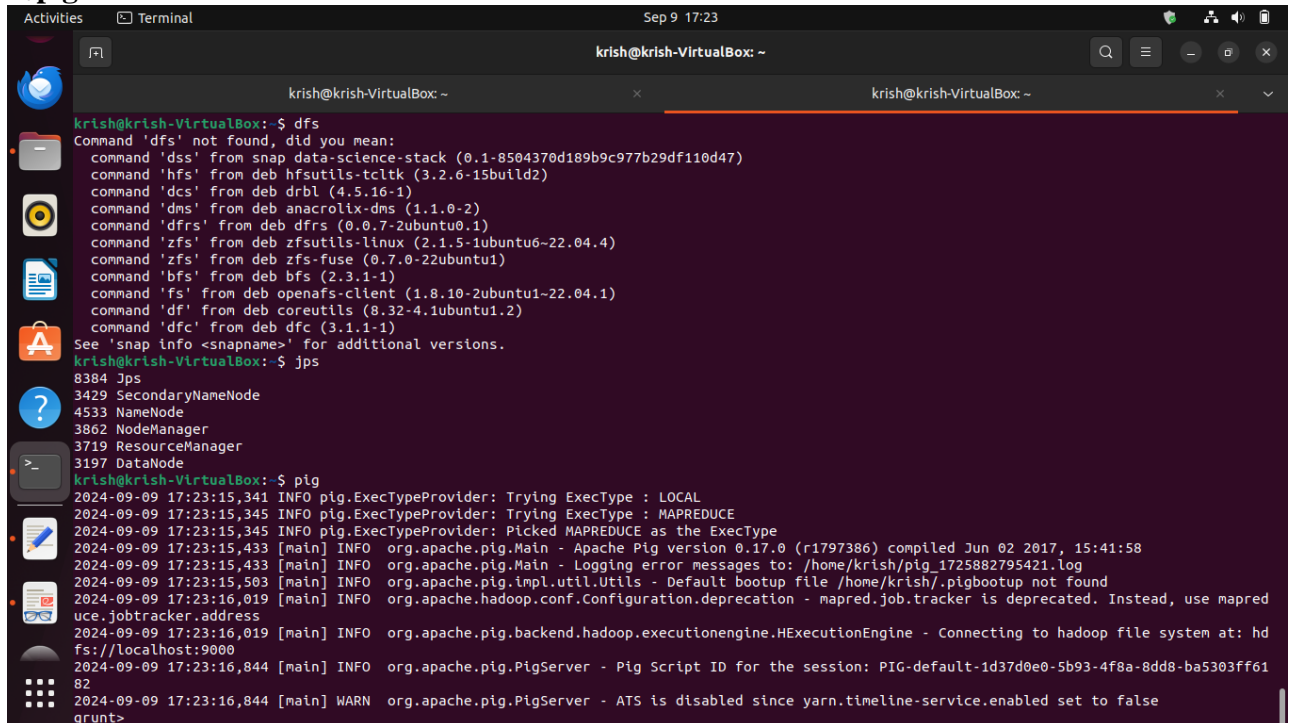
```

\$mv pig-0.16.0 pig


```

krish@krish-VirtualBox: ~
krish@krish-VirtualBox: ~
pig-0.17.0/test/perf/pigmix/src/pig/L17.pig
pig-0.17.0/test/perf/pigmix/src/pig/L2.pig
pig-0.17.0/test/perf/pigmix/src/pig/L3.pig
pig-0.17.0/test/perf/pigmix/src/pig/L4.pig
pig-0.17.0/test/perf/pigmix/src/pig/L5.pig
pig-0.17.0/test/perf/pigmix/src/pig/L6.pig
pig-0.17.0/test/perf/pigmix/src/pig/L7.pig
pig-0.17.0/test/perf/pigmix/src/pig/L8.pig
pig-0.17.0/test/perf/pigmix/src/pig/L9.pig
pig-0.17.0/test/pigunit-tests
pig-0.17.0/test/python/streaming/test_controller.py
pig-0.17.0/test/resources/org/apache/pig/test/ClassLoaderSanityCheck.java
pig-0.17.0/test/resources/org/apache/pig/test/RegisteredJarVisibilityLoader.java
pig-0.17.0/test/resources/org/apache/pig/test/RegisteredJarVisibilitySchema.java
pig-0.17.0/test/smoke-tests
pig-0.17.0/test/unit-tests
pig-0.17.0/tutorial/build.xml
pig-0.17.0/tutorial/data/excite-small.log
pig-0.17.0/tutorial/data/excite.log.bz2
pig-0.17.0/tutorial/scripts/script1-hadoop.pig
pig-0.17.0/tutorial/scripts/script1-local.pig
pig-0.17.0/tutorial/scripts/script2-hadoop.pig
pig-0.17.0/tutorial/scripts/script2-local.pig
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/ExtractHour.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/NGramGenerator.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/NonURLDetector.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/ScoreGenerator.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/ToLower.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/TutorialTest.java
pig-0.17.0/tutorial/src/org/apache/pig/tutorial/TutorialUtil.java
pig-0.17.0/bin/pig
pig-0.17.0/bin/pig.cmd
pig-0.17.0/bin/pig.py
krish@krish-VirtualBox:~$ sudo mv ~/pig-0.17.0 ~/pig
[sudo] password for krish:
krish@krish-VirtualBox:~$

```

\$pig


```

krish@krish-VirtualBox:~$ dfs
Command 'dfs' not found, did you mean:
command 'dss' from snap data-science-stack (0.1-8504370d189b9c977b29df110d47)
command 'hfs' from deb hfsutils-tcltk (3.2.6-15build2)
command 'dcs' from deb drbl (4.5.16-1)
command 'dms' from deb anacrolix-dms (1.1.0-2)
command 'dfrs' from deb dfrs (0.0.7-2ubuntu0.1)
command 'zfs' from deb zfsutils-linux (2.1.5-1ubuntu6-22.04.4)
command 'zfs' from deb zfs-fuse (0.7.0-22ubuntu1)
command 'bfs' from deb bfs (2.3.1-1)
command 'fs' from deb openafs-client (1.8.10-2ubuntu1-22.04.1)
command 'df' from deb coreutils (8.32-4.1ubuntu1.2)
command 'dfc' from deb dfc (3.1.1-1)
See 'snap info <snapname>' for additional versions.
krish@krish-VirtualBox:~$ jps
8384 Jps
3429 SecondaryNameNode
4533 NameNode
3862 NodeManager
3719 ResourceManager
3197 DataNode
krish@krish-VirtualBox:~$ pig
2024-09-09 17:23:15,341 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-09 17:23:15,345 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-09 17:23:15,345 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-09 17:23:15,433 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-09-09 17:23:15,433 [main] INFO org.apache.pig.Main - Logging error messages to: /home/krish/pig_1725882795421.log
2024-09-09 17:23:15,503 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/krish/.pigbootstrap not found
2024-09-09 17:23:16,019 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
uce.jobtracker.address
2024-09-09 17:23:16,019 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hd
fs://localhost:9000
2024-09-09 17:23:16,844 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-1d37d0e0-5b93-4f8a-8dd8-ba5303ff61
82
2024-09-09 17:23:16,844 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>

```