

```
#PIG settings
export PIG_HOME=/home/hadoop/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_HOME/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
```

\$mv pig-0.16.0 pig

\$pig

```
kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Beyer | 1 2 3 4 | 10:59
File Actions Edit View Help
pig 0.16.0/tutorial/src/org/apache/pig/tutorial/tutorialUtil.java
pig 0.16.0/bin/pig
pig 0.16.0/bin/pig.cmd
pig 0.16.0/bin/pig.py

(hadoop@kali:~) [-]
└─ ls
DA-Lab Desktop Documents Downloads hadoop hadoop-1.4.0.tar.gz hadoopdata Music Pictures pig-0.16.0 pig-0.16.0.tar.gz Public Templates Videos
(hadoop@kali:~) [-]
└─ mv pig-0.16.0 pig
└─ ls
DA-Lab Desktop Documents Downloads hadoop hadoop-1.4.0.tar.gz hadoopdata Music Pictures pig pig-0.16.0.tar.gz Public Templates Videos
(hadoop@kali:~) [-]
└─ nano ~/.bashrc
(hadoop@kali:~) [-]
└─ source ~/.bashrc
(hadoop@kali:~) [-]
└─ java
Picked up _JAVA_OPTIONS: -Dswt.useSystemAAFontSettings=on -Dswing.aatext=true
55824 ResourceManager
54824 SecondaryNameNode
55130 NodeManager
54393 DataNode
64953 JPS
54267 NameNode
(hadoop@kali:~) [-]
└─ pig
Picked up _JAVA_OPTIONS: -Dswt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-29 18:58:45.222 [INFO] pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-08-29 18:58:45.226 [INFO] pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-08-29 18:58:45.227 [INFO] pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-08-29 18:58:45.477 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1745538) compiled Jun 01 2016, 23:10:40
2024-08-29 18:58:45.477 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig-1724943525458.log
2024-08-29 18:58:45.563 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/pigbootstrap not found
2024-08-29 18:58:45.667 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-29 18:58:46.721 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2024-08-29 18:58:46.721 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-29 18:58:46.724 [main] INFO org.apache.pig.backend.hadoop.executionengine.HadoopExecutionEngine - Connecting to Hadoop file system at: hdfs://localhost:9000
2024-08-29 18:58:48.975 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-5c3dc2d3-235b-475c-ab06-cbd8d2f4bfc2
2024-08-29 18:58:48.975 [main] WARN org.apache.pig.PigServer - AJS is disabled since yarn.timeline-service.enabled set to false
gruntp> quit
2024-08-29 18:58:48.426 [main] INFO org.apache.pig.Main - Pig script completed in 15 seconds and 392 milliseconds (15392 ms)
(hadoop@kali:~) [-]
```

\$cd DA-Lab

\$mkdir exp4

\$cd exp4

\$nano sample.txt

```
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2 sample.txt
1, John
2, Jane
3, Joe
4, Emma
```

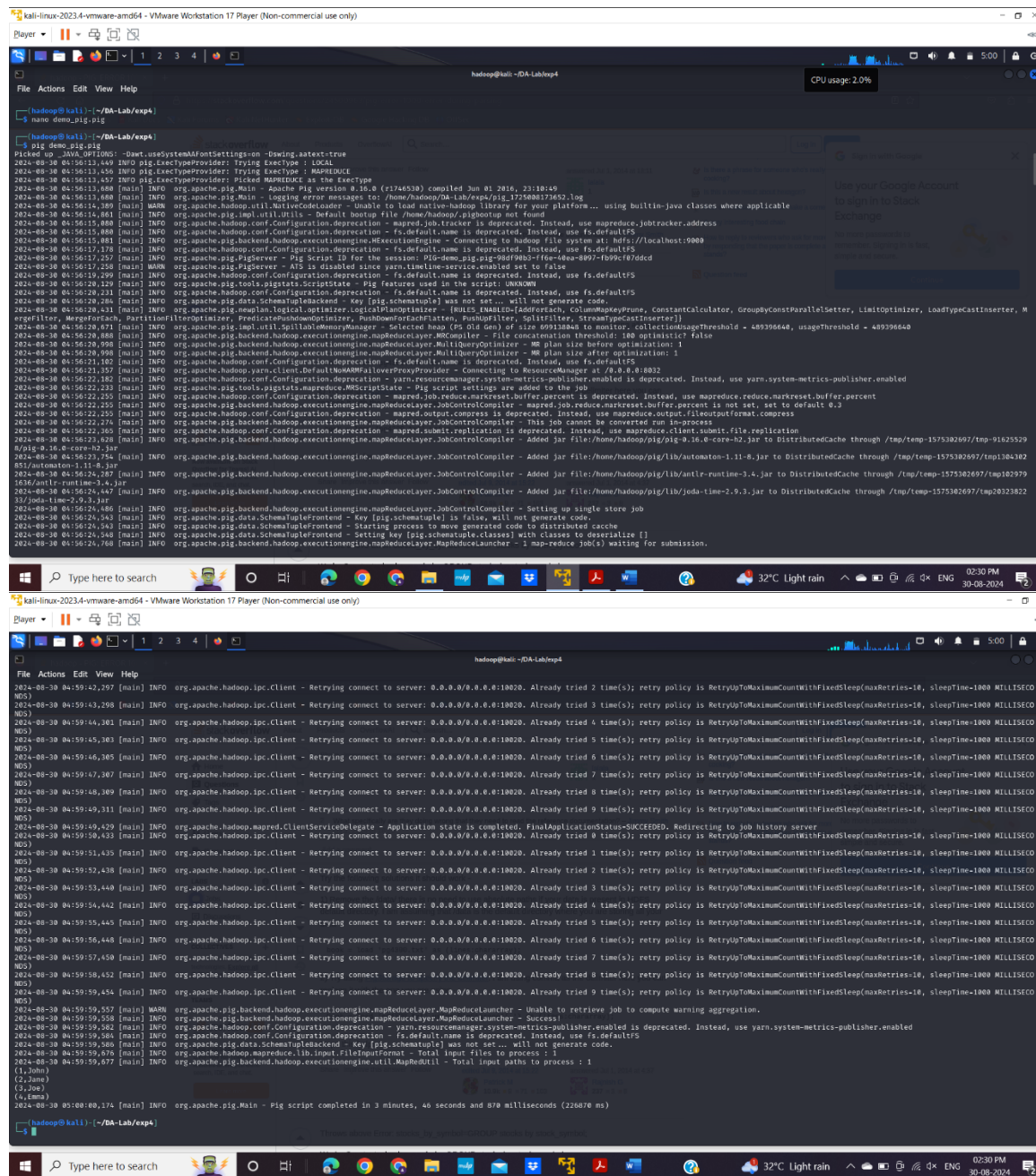
\$nano demo_pig.pig

```
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2 demo_pig.pig
-- Load the data from HDFS
data = LOAD '/exp4/sample.txt' USING PigStorage(',') AS (id:int, name:chararray);
-- Dump the data to check if it was loaded correctly
DUMP data;
```

\$hdfs dfs -mkdir /exp4

\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/sample.txt /exp4

\$pig demo_pig.pig



```

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
File Actions Edit View Help
hadoop@kali: ~/DA-Lab/exp4
CPU usage: 2.0%

hadoop@kali:~/DA-Lab/exp4$ nano demo_pig.pig
Picked up _JAVA_OPTIONS: -Dant.useSystemFontSettings=on -Dswing.aatext=true
2024-08-30 04:56:11,449 INFO org.apache.pig.ExecTypeProvider: Trying ExecType: LOCAL
2024-08-30 04:56:11,456 INFO org.apache.pig.ExecTypeProvider: Trying ExecType: MAPREDUCE
2024-08-30 04:56:11,457 INFO org.apache.pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-08-30 04:56:11,680 INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r748539) compiled Jun 01 2016, 23:10:48
2024-08-30 04:56:11,680 INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/DA-Lab/exp4/pig/375000173652.log
2024-08-30 04:56:14,389 INFO org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-30 04:56:14,661 INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/pig/bootstrap not found
2024-08-30 04:56:15,800 INFO org.apache.hadoop.conf.Configuration.deprecation - mapped.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-30 04:56:15,800 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:15,800 INFO org.apache.pig.backend.hadoop.executionengine.MixedExecutionEngine - Connecting to hadoop file system at hdfs://localhost:9000
2024-08-30 04:56:17,178 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:17,257 INFO org.apache.pig.PigServer - Pig Script ID for the session: Pig-demo_pig-pig-9d8f9bb5-ff6c-4be4-b097-fb9c47d0dd
2024-08-30 04:56:17,258 INFO org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-08-30 04:56:19,299 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:20,129 INFO org.apache.pig.toolkit.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2024-08-30 04:56:20,231 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:20,246 INFO org.apache.pig.data.SchemaTableBackend - Key [pig.schematuple] was not set... will not generate code
2024-08-30 04:56:20,431 INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - RULES_ENABLED[AdaptorEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstrParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeSortJoin, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForJoinFilter, PushDownFilter, SplitFilter, StreamTypeCastInserter]
2024-08-30 04:56:20,473 INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 60938848 to monitor. collectionUsageThreshold - 489396648, usageThreshold - 489396640
2024-08-30 04:56:20,680 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - File concatenation threshold: 100 optimistic? false
2024-08-30 04:56:20,720 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Pig features used in the script: UNKNOWN
2024-08-30 04:56:20,998 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - MR plan size before optimization: 1
2024-08-30 04:56:21,102 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - MR plan size after optimization: 1
2024-08-30 04:56:21,182 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:21,225 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Connecting to ResourceManager at /s:0.0.0.0:8032
2024-08-30 04:56:21,337 INFO org.apache.hadoop.conf.Configuration.deprecation - mapped.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2024-08-30 04:56:21,352 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Pig script settings are added to the job
2024-08-30 04:56:21,255 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - mapped.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
2024-08-30 04:56:21,255 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - mapped.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2024-08-30 04:56:21,255 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - mapped.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2024-08-30 04:56:21,274 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - This job cannot be converted run in-process
2024-08-30 04:56:21,274 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - mapped.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication
2024-08-30 04:56:21,628 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Added jar file:/home/hadoop/pig/pig-0.16.0-core-h2.jar to DistributedCache through /tmp/temp-1575302697/tmp-91625529
8/19/0.16.0-core-h2.jar
2024-08-30 04:56:21,736 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Added jar file:/home/hadoop/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-1575302697/tmp1364362
857/automaton-1.11-8.jar
2024-08-30 04:56:21,747 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Added jar file:/home/hadoop/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-1575302697/tmp102799
1636/antlr-runtime-3.4.jar
2024-08-30 04:56:21,747 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Added jar file:/home/hadoop/pig/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/temp-1575302697/tmp20323822
32/joda-time-2.9.3.jar
2024-08-30 04:56:21,686 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Setting up single store job
2024-08-30 04:56:21,694 INFO org.apache.pig.data.SchemaTableBackend - Key [pig.schematuple] is false, will not generate code.
2024-08-30 04:56:21,694 INFO org.apache.pig.data.SchemaTableBackend - Starting process to move generated code to distributed cache
2024-08-30 04:56:21,694 INFO org.apache.pig.data.SchemaTableBackend - Setting key [pig.schematuple.class] with classes to deserialize: []
2024-08-30 04:56:21,748 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - 1 map-reduce job(s) waiting for submission.

2024-08-30 04:59:42,257 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:43,298 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 3 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:44,361 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 4 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:45,383 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 5 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:46,365 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 6 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:47,367 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 7 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:48,389 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 8 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:49,311 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:49,429 INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-08-30 04:59:50,433 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:51,435 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:52,438 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:53,440 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 3 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:54,442 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 4 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:55,445 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 5 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:56,448 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 6 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:57,450 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 7 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:58,452 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 8 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:59,454 INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISEC)
2024-08-30 04:59:59,557 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Unable to retrieve job to compute warning aggregation.
2024-08-30 04:59:59,558 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Success!
2024-08-30 04:59:59,582 INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-08-30 04:59:59,584 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:59:59,586 INFO org.apache.pig.data.SchemaTableBackend - Key [pig.schematuple] was not set... will not generate code.
2024-08-30 04:59:59,676 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process: 1
2024-08-30 04:59:59,677 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaCompiler - Total input paths to process: 1
(1,John)
(2,James)
(3,Joe)
(4,Emma)
2024-08-30 05:00:00,174 INFO org.apache.pig.Main - pig script completed in 3 minutes, 46 seconds and 870 milliseconds (226078 ms)
hadoop@kali:~/DA-Lab/exp4$

```

\$nano uppercase_udf.py



```

hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2
uppercase_udf.py
def uppercase(text):
    return text.upper()
if __name__ == "__main__":
    import sys
    for line in sys.stdin:
        line = line.strip()
        result = uppercase(line)
        print(result)

```

\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/uppercase_udf.py /exp4

```
(hadoop@kali)-[~/hadoop/bin]
$ ./hdfs dfs -ls /exp4
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:26:01,736 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 3 items
drwxr-xr-x   - hadoop supergroup          0 2024-08-30 05:07 /exp4/output
-rw-r--r--   1 hadoop supergroup          27 2024-08-30 04:43 /exp4/sample.txt
-rw-r--r--   1 hadoop supergroup        172 2024-08-30 05:02 /exp4/uppercase_udf.py
```

\$nano udf_example.pig

```
File Actions Edit View Help
GNU nano 7.2 udf_example.pig
-- Register the Python UDF script
REGISTER 'hdfs:///exp4/uppercase_udf.py' USING jython AS udf;
-- Load some data
data = LOAD 'hdfs:///exp4/sample.txt' AS (text:chararray);
-- Use the Python UDF
uppercased_data = FOREACH data GENERATE udf.uppercase(text) AS uppercase_text;
-- Store the result
STORE uppercased_data INTO 'hdfs:///exp4/output';
```

\$pig -f udf_example.pig

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)

hadoop@kali: ~/DA-Lab/exp4

```
$ nano uppercase_udf.py
$ cd ../..
$ cd ~/hadoop/bin
$ ./hdfs dfs -put ~/DA-Lab/exp4/uppercase_udf.py /exp4
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-30 05:07:39,341 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
$ cd ../..
$ cd ~/DA-Lab/exp4
$ nano udf_example.pig
$ pig -f udf_example.pig
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-30 05:08:18,591 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-08-30 05:08:18,601 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-08-30 05:08:18,601 INFO pig.ExecTypeProvider: Picked MapReduce as the ExecType
2024-08-30 05:08:18,836 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-08-30 05:08:18,836 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/DA-Lab/exp4/pig-175088778818.log
2024-08-30 05:08:19,085 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-30 05:08:20,127 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/pigbootstrap not found
2024-08-30 05:08:20,382 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-30 05:08:20,383 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:20,383 [main] INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-08-30 05:08:22,438 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:22,513 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf_example.pig-1dcd0cc-jell-4763-a9e8-561b5b6b3da1
2024-08-30 05:08:22,513 [main] WARN org.apache.pig.PigServer - AFS is disabled since yarn.timeline-service.enabled set to false
2024-08-30 05:08:22,725 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:24,290 [main] INFO org.apache.pig.scripthing.jython.JythonScriptEngine - created tmp python.cachedir=/tmp/pig_jython.3781842655894475147
2024-08-30 05:08:26,810 [main] WARN org.apache.pig.scripthing.jython.JythonScriptEngine - pig.cmd.pys.remainders is empty. This is not expected unless on testing.
2024-08-30 05:08:26,844 [main] INFO org.apache.pig.scripthing.jython.JythonScriptEngine - Register Scripting UDF: udf.uppercase
2024-08-30 05:08:27,784 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:27,967 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:38,407 [main] INFO org.apache.pig.scripthing.jython.JythonFunction - No schema defined for function 'uppercase' in /tmp/pig313342629392933005tmp/uppercase_udf.py
2024-08-30 05:08:38,583 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:39,743 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2024-08-30 05:08:38,830 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2024-08-30 05:08:39,858 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:39,982 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schema tuple] was not set... Will not generate code.
2024-08-30 05:08:39,132 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES ENABLED: AddOrderBy, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCaster, M
easureFilter, MergeFilterCatcher, PartialFilterOptimizer, PredicatePushdownOptimizer, PushDownForCastPattern, PushDownFilter, SplitFilter, StreamTypeCaster, S
2024-08-30 05:08:39,403 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699138848 to monitor. collectionUsageThreshold = 489396640, usageThreshold = 489396640
2024-08-30 05:08:39,484 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MCCompiler - File concatenation threshold: 168 optimistic? false
2024-08-30 05:08:39,558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MQ plan size before optimization: 1
2024-08-30 05:08:39,855 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MQ plan size after optimization: 1
2024-08-30 05:08:39,994 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
File Actions Edit View Help
2024-08-30 05:10:12,729 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 4 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:13,731 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 5 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:14,733 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 6 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:15,735 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 7 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:16,738 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 8 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:17,740 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:17,859 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed, finalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-08-30 05:10:18,862 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:19,864 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:20,866 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:21,868 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 3 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:22,871 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 4 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:23,874 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 5 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:24,876 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 6 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:25,879 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 7 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:26,881 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 8 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:27,883 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NDS)
2024-08-30 05:10:27,987 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2024-08-30 05:10:27,988 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-08-30 05:10:28,108 [main] INFO org.apache.pig.Main - Pig script completed in 4 minutes, 9 seconds and 747 milliseconds (249747 ms)

(hadoop@kali) [~/BA-Lab/exp4]
$ cd ../hadoop/bin

(hadoop@kali) [~/hadoop/bin]
$ ./hdfs dfs -cat /exp4/output/part-m-00000
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-30 05:12:25,908 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,JOHN
2,JANE
3,JOE
4,EMMA

```

\$hdfs dfs -cat /exp4/output/*

```

(hadoop@kali) [~/hadoop/bin]
$ ./hdfs dfs -cat /exp4/output/*
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:33:32,731 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
1,JOHN
2,JANE
3,JOE
4,EMMA

```