

## EXP NO: 4

## CREATE UDF IN PIG

## \$start-all.sh

```

(hadoop@kali)-[~]
└─$ 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 [kali]
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-11 04:59:16,429 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

```

## \$ jps

```

(hadoop@kali)-[~]
└─$ jps
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13597 DataNode
13471 NameNode

```

\$wget <https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz>

## \$ tar xvfz pig-0.16.0.tar.gz

```

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
File Actions Edit View Help
(hadoop@kali)-[~]
└─$ wget https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
--2024-08-29 10:55:32-- https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
Resolving d1cdn.apache.org (d1cdn.apache.org) ... 208.146.42.244, 151.101.2.132
Connecting to d1cdn.apache.org (d1cdn.apache.org)[208.146.42.244]:443 ... connected.
HTTP request sent, awaiting response... 200 OK
Length: 17729333 (169M) [application/x-gzip]
Saving to: 'pig-0.16.0.tar.gz'

pig-0.16.0.tar.gz 100%[=====] 169.07M 26.5MB/s in 6.5s

2024-08-29 10:55:41 (26.0 MB/s) - 'pig-0.16.0.tar.gz' saved [17729333/17729333]

(hadoop@kali)-[~]
└─$ tar xvfz pig-0.16.0.tar.gz
pig-0.16.0/
pig-0.16.0/bin/
pig-0.16.0/conf/
pig-0.16.0/contrib/
pig-0.16.0/contrib/piggybank/
pig-0.16.0/contrib/piggybank/java/
pig-0.16.0/contrib/piggybank/java/build/
pig-0.16.0/contrib/piggybank/java/build/classes/
pig-0.16.0/contrib/piggybank/java/build/classes/org/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/convert/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/diff/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/truncate/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/decode/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/math/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/stats/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/string/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/apacheLogParser/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/xml/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/allloader/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/apachelog/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/avro/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/hivert/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/partition/
pig-0.16.0/contrib/piggybank/java/build/docs/
pig-0.16.0/contrib/piggybank/java/build/docs/api/
pig-0.16.0/contrib/piggybank/java/build/test/
pig-0.16.0/contrib/piggybank/java/build/test/classes/
pig-0.16.0/contrib/piggybank/java/lib/
pig-0.16.0/contrib/piggybank/java/src/

```

## \$nano ~/.bashrc

```
#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

```
hadoop@kali: ~$ mv pig-0.16.0 pig $pig
hadoop@kali: ~$ pig
Picked up _JAVA_OPTIONS: -Dswt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-29 10:50:45.222 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-08-29 10:50:45.226 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-08-29 10:50:45.227 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-08-29 10:50:45.477 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746538) compiled Jun 01 2016, 23:10:49
2024-08-29 10:50:45.477 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig-1724942589897_0002.log
2024-08-29 10:50:45.563 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootstrap not found
2024-08-29 10:50:46.687 [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 10:50:46.721 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-29 10:50:46.723 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-29 10:50:46.776 [main] INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to Hadoop file system at: hdfs://localhost:9800
2024-08-29 10:50:46.872 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-29 10:50:46.978 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-82b2e2d-235b-475b-9b0b-c0da2f46rc2
2024-08-29 10:50:46.979 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt> quit
2024-08-29 10:50:48.426 [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

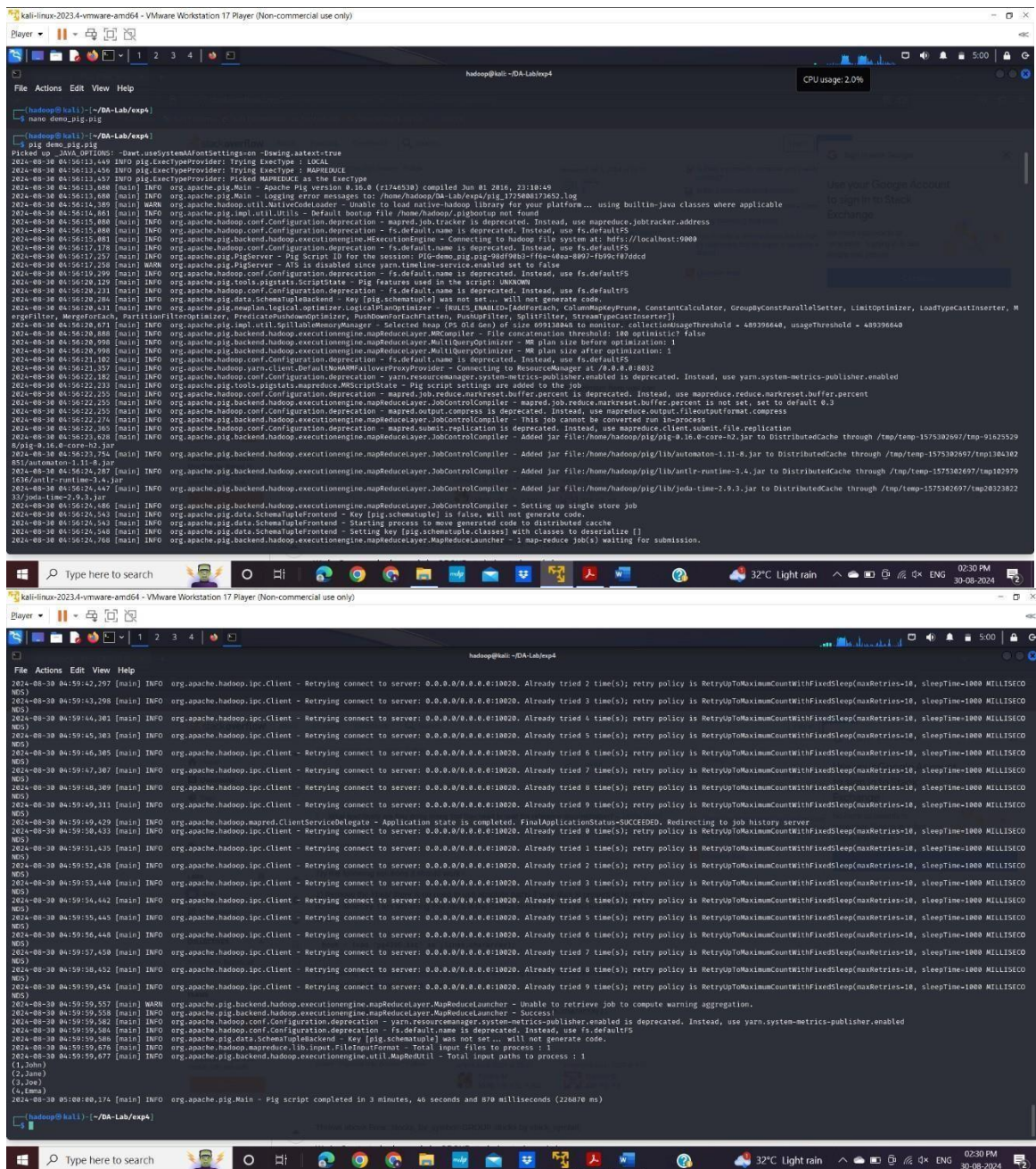
```
hadoop@kali: ~/DA-Lab/exp4
GNU nano 7.2 sample.txt
1,John
2,Jane
3,Joe
4,Emma
```

\$nano demo\_pig.pig

```
hadoop@kali: ~/DA-Lab/exp4
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
$ nano demo_pig.pig
(hadoop@kali) ~/DA-Lab/exp4
$ pig demo_pig.pig
Pig Demo
Pig Script ID for the session: pig-demo.pig-98d98b1-f6e4-40e4-8097-fb99c67ddcd
2024-08-30 04:59:42,297 [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 MILLISECONDS)
2024-08-30 04:59:43,298 [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 MILLISECONDS)
2024-08-30 04:59:44,301 [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 MILLISECONDS)
2024-08-30 04:59:45,303 [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 MILLISECONDS)
2024-08-30 04:59:46,305 [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 MILLISECONDS)
2024-08-30 04:59:47,307 [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 MILLISECONDS)
2024-08-30 04:59:48,309 [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 MILLISECONDS)
2024-08-30 04:59:49,311 [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 MILLISECONDS)
2024-08-30 04:59:49,429 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-08-30 04:59:50,433 [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 MILLISECONDS)
2024-08-30 04:59:51,435 [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 MILLISECONDS)
2024-08-30 04:59:52,438 [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 MILLISECONDS)
2024-08-30 04:59:53,440 [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 MILLISECONDS)
2024-08-30 04:59:54,442 [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 MILLISECONDS)
2024-08-30 04:59:55,445 [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 MILLISECONDS)
2024-08-30 04:59:56,448 [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 MILLISECONDS)
2024-08-30 04:59:57,450 [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 MILLISECONDS)
2024-08-30 04:59:58,452 [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 MILLISECONDS)
2024-08-30 04:59:59,454 [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 MILLISECONDS)
2024-08-30 04:59:59,557 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduce.launcher - Unable to retrieve job to compute warning aggregation.
2024-08-30 04:59:59,558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.launcher - Success
2024-08-30 04:59:59,560 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.launcher - Success
2024-08-30 04:59:59,562 [main] 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 04:59:59,564 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:59:59,566 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2024-08-30 04:59:59,576 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input files to process : 1
2024-08-30 04:59:59,577 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
2024-08-30 05:00:00,174 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 46 seconds and 870 milliseconds (226070 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**

```

hadoop@kali: ~/DA-Lab/exp4
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)
Player
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
(hadoop@kali)~[~/DA-Lab/exp4]
$ nano uppercase_udf.py
(hadoop@kali)~[~/DA-Lab/exp4]
$ cd ../..
(hadoop@kali)~[~/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 05:02:54,541 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(hadoop@kali)~[~/hadoop/bin]
$ cd ../..
(hadoop@kali)~[~/DA-Lab/exp4]
$ nano udf_example.pig
(hadoop@kali)~[~/DA-Lab/exp4]
$ pig -f udf_example.pig
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-30 05:06:10,591 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-08-30 05:06:10,601 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-08-30 05:06:10,601 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-08-30 05:06:10,606 INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746538) compiled Jun 01 2016, 23:10:49
2024-08-30 05:06:10,605 INFO org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-30 05:06:10,617 INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/DA-Lab/exp4/pig.172508077801e.log
2024-08-30 05:06:10,620 INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-30 05:06:10,633 INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-08-30 05:06:12,430 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:12,513 INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf_example.pig-1d6d0c3-kali-vbox-vm-501390638a1
2024-08-30 05:06:12,513 INFO org.apache.pig.PigServer - ATIS is disabled since yarn.timeline-service.enabled set to false
2024-08-30 05:06:12,725 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:12,725 INFO org.apache.pig.scripting.jython.JythonScriptEngine - Created tmp python.cachedir/tmp/pig.jython.378184265984473167
2024-08-30 05:06:13,018 INFO org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.remainers is empty. This is not expected unless on testing.
2024-08-30 05:06:13,044 INFO org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF: udf.uppercase
2024-08-30 05:06:13,784 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:13,807 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:14,407 INFO org.apache.pig.scripting.jython.JythonFunction - No schema defined for function 'uppercase' in /tmp/pig213131342629392933005tmp/uppercase_udf.py
2024-08-30 05:06:15,583 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:15,763 INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MCompiler - File concatenation threshold: 100 optimistic: false
2024-08-30 05:06:16,630 INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2024-08-30 05:06:16,659 INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2024-08-30 05:06:16,694 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:16,902 INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set ... will not generate code.
2024-08-30 05:06:19,132 INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED:AddressSch, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, M
2024-08-30 05:06:19,403 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:19,403 INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MCompiler - File concatenation threshold: 100 optimistic: false
2024-08-30 05:06:19,858 INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2024-08-30 05:06:19,859 INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2024-08-30 05:06:19,924 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

```

```

kali-linux-20234-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
hadoop@kali:~/hadoop/bin
2024-09-20 05:18:12,729 [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
2024-09-20 05:18:13,731 [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
2024-09-20 05:18:14,733 [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
2024-09-20 05:18:15,735 [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
2024-09-20 05:18:16,738 [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
2024-09-20 05:18:17,740 [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
2024-09-20 05:18:17,859 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-09-20 05:18:18,862 [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
2024-09-20 05:18:19,864 [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
2024-09-20 05:18:20,866 [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
2024-09-20 05:18:21,868 [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
2024-09-20 05:18:22,871 [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
2024-09-20 05:18:23,874 [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
2024-09-20 05:18:24,876 [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
2024-09-20 05:18:25,879 [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
2024-09-20 05:18:26,881 [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
2024-09-20 05:18:27,883 [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
2024-09-20 05:18:27,987 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2024-09-20 05:18:27,988 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-09-20 05:18:28,108 [main] INFO org.apache.pig.Main - Pig script completed in 4 minutes, 9 seconds and 747 milliseconds (249747 ms)

hadoop@kali:~/hadoop/bin
$ cd ../.. /hadoop/bin

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... 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

```