

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://dlcdn.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)
(hadoop@kali)~$ wget https://dlcdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
--2024-08-29 10:55:35-- https://dlcdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org) ... 204.144.2.244, 151.101.2.132
Connecting to dlcdn.apache.org (dlcdn.apache.org)|204.144.2.244|: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/state/
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/hiwerc/
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

[illegible]

\$cd DA-Lab

\$mkdir exp4

\$cd exp4

\$nano sample.txt

```
File Actions Edit View Help
GNU nano 7.2 sample.txt
1,John
2,Jane
3,Joe
4,Emma
```

\$nano demo_pig.pig

```

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-vm64 - 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: -Dwt.useSystemFontSettings=on -Dswing.text=true
2024-08-30 04:56:12.649 INFO org.apache.pig.ExecTypeProvider: Trying ExecType = LOCAL
2024-08-30 04:56:13.456 INFO org.apache.pig.ExecTypeProvider: Trying ExecType = MAPREDUCE
2024-08-30 04:56:13.657 INFO org.apache.pig.Main: Logging error messages to: /home/hadoop/DA-Lab/exp4/pig-175808173652.log
2024-08-30 04:56:15.269 INFO org.apache.hadoop.util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-30 04:56:16.861 INFO org.apache.pig.impl.util.Utils: Default bootstrap file /home/hadoop/pigbootstrap not found
2024-08-30 04:56:15.880 INFO org.apache.hadoop.conf.Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-30 04:56:15.881 INFO org.apache.hadoop.conf.Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:15.881 INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine: Connecting to hadoop file system at: hdfs://localhost:9000
2024-08-30 04:56:17.178 INFO org.apache.pig.PigServer: Pig Script ID for the session: PIG-demo-pig-pig-98d9f903-f6e4-4099-f699-f07ddc
2024-08-30 04:56:17.257 INFO org.apache.pig.PigServer: AJS is disabled since yarn.timeline-service.enabled set to false
2024-08-30 04:56:20.129 INFO org.apache.pig.tools.pigstats.ScriptState: Pig features used in the script: UNKNOWN
2024-08-30 04:56:20.221 INFO org.apache.hadoop.conf.Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:20.284 INFO org.apache.pig.data.SchemaTableBackend: Key [pig.schema.table] was not set... will not generate code.
2024-08-30 04:56:20.431 INFO org.apache.pig.mapplan.logical.optimizer.LogicalPlanOptimizer: [RULES [Enable[AddressCache, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, M
2024-08-30 04:56:20.671 INFO org.apache.pig.impl.util.SpillableMemoryManager: Selected heap (PS Old Gen) of size 899128048 to monitor, collectionUsageThreshold = 489396648, usageThreshold = 489396648
2024-08-30 04:56:20.690 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.MultiQueryOptimizer: MR plan size before optimization: 1
2024-08-30 04:56:21.302 INFO org.apache.hadoop.conf.Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 04:56:21.357 INFO org.apache.hadoop.yarn.client.DefaultHadoopWALinuxProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-08-30 04:56:22.233 INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState: Pig script settings are added to the job
2024-08-30 04:56:22.255 INFO org.apache.hadoop.conf.Configuration.deprecation: mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
2024-08-30 04:56:22.255 INFO org.apache.pig.data.SchemaTableFrontend: Starting process to move generated code to distributed cache
2024-08-30 04:56:22.274 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.JobControlCompiler: This job cannot be converted run in-process
2024-08-30 04:56:22.365 INFO org.apache.hadoop.conf.Configuration.deprecation: mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication
2024-08-30 04:56:22.628 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.JobControlCompiler: Added jar file:/home/hadoop/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/tmp-1575382697/tmp-91625529
2024-08-30 04:56:22.754 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.JobControlCompiler: Added jar file:/home/hadoop/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/tmp-1575382697/tmp1361482
2024-08-30 04:56:22.767 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.JobControlCompiler: Added jar file:/home/hadoop/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/tmp-1575382697/tmp1029799
2024-08-30 04:56:22.847 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.JobControlCompiler: Added jar file:/home/hadoop/pig/lib/joda-time-2.9.3.jar to DistributedCache through /tmp/tmp-1575382697/tmp20223822
2024-08-30 04:56:22.933 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.JobControlCompiler: Setting up single store job
2024-08-30 04:56:22.943 INFO org.apache.pig.data.SchemaTableFrontend: Key [pig.schema.table] is false, will not generate code.
2024-08-30 04:56:22.943 INFO org.apache.pig.data.SchemaTableFrontend: Starting process to move generated code to distributed cache
2024-08-30 04:56:23.548 INFO org.apache.pig.data.SchemaTableFrontend: Setting key [pig.schema.table.classes] with classes to deSerialize []
2024-08-30 04:56:24.768 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.Launcher: 1 map-reduce job(s) waiting for submission.

2024-08-30 04:59:42.297 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
2024-08-30 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 MILLISECO
2024-08-30 04:59:44.301 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
2024-08-30 04:59:45.303 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
2024-08-30 04:59:46.305 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
2024-08-30 04:59:47.307 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
2024-08-30 04:59:48.309 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
2024-08-30 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 MILLISECO
2024-08-30 04:59:50.429 INFO org.apache.hadoop.mapred.ClientServiceDelegate: Application state is completed. FinalApplicationStatus=SUCCEEDED, Redirecting to job history server
2024-08-30 04:59:50.533 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
2024-08-30 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 MILLISECO
2024-08-30 04:59:52.442 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
2024-08-30 04:59:53.448 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
2024-08-30 04:59:54.448 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
2024-08-30 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 MILLISECO
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 MILLISECO
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 MILLISECO
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 MILLISECO
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 MILLISECO
2024-08-30 04:59:59.557 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.Launcher: Unable to retrieve job to compute warning aggregation.
2024-08-30 04:59:59.558 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.mapreduce.Launcher: Success!
2024-08-30 04:59:59.566 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.data.SchemaTableBackend: Key [pig.schema.table] 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.util.MapReduceUtil: Total input paths to process: 1
(1,John)
(2,June)
(3,Joe)
(4,Anna)
2024-08-30 05:00:00.174 INFO org.apache.pig.Main: Pig script completed in 3 minutes, 46 seconds and 870 milliseconds (226076 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)

```

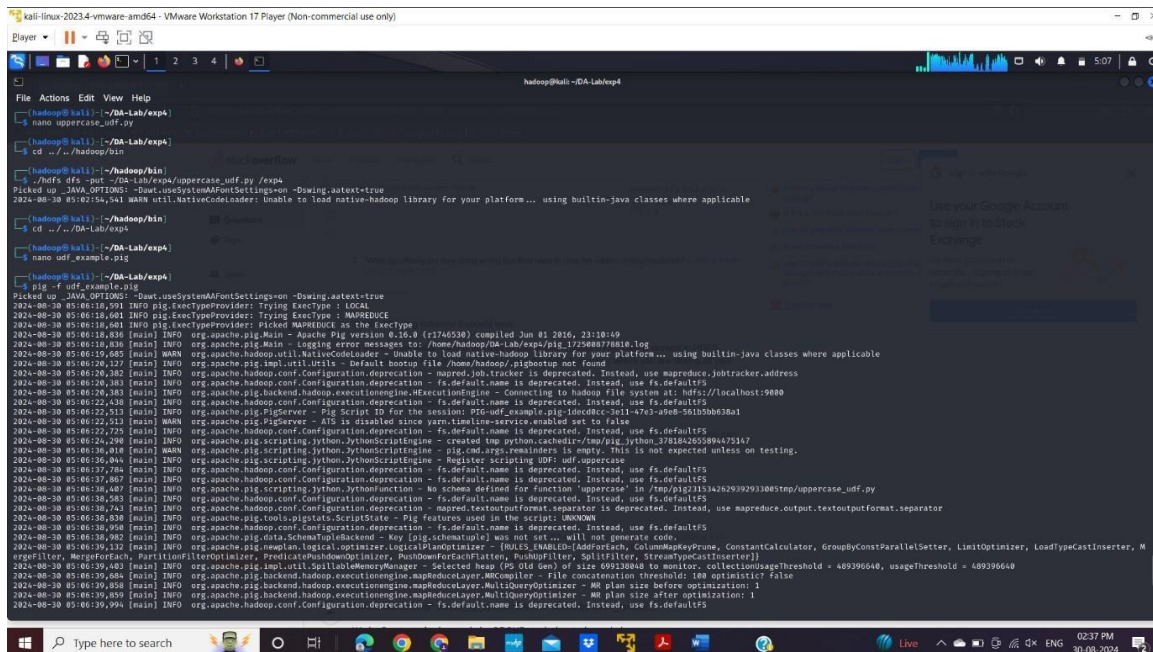

\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/uppercase_udf.py /exp4

```
(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

```
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
$ 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 05:06:10,636 INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746538) compiled Jun 01 2016, 23:10:49
2024-08-30 05:06:10,685 INFO org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-30 05:06:10,727 INFO org.apache.pig.impl.util.Utils - Default hadoop file /home/hadoop/.pigbootup not found
2024-08-30 05:06:10,730 INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-30 05:06:10,733 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:10,733 INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to Hadoop file system at: hdfs://localhost:9000
2024-08-30 05:06:10,733 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:10,733 INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf-example.pig-1dcd8cc3-1e11-47e3-a9e0-5610506838a1
2024-08-30 05:06:10,733 INFO org.apache.pig.PigServer - AFS is disabled since yarn.timeline-service.enabled set to false
2024-08-30 05:06:10,733 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:10,733 INFO org.apache.pig.scripting.jython.JythonScriptEngine - created tmp python.cachedir/tmp/pig_jython.378484403894475147
2024-08-30 05:06:10,733 INFO org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.reminders is empty. This is not expected unless on testing.
2024-08-30 05:06:10,733 INFO org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF: udf_uppercase
2024-08-30 05:06:10,733 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:10,733 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaDriver - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2024-08-30 05:06:10,733 INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:06:10,733 INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.cdmamap] was not set... will not generate code.
2024-08-30 05:06:10,733 INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED:[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, M
MergeForEach, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFilter, PushDownFilter, SplitFilter, StreamTypeCastInserter]]
2024-08-30 05:06:10,733 INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - Selected heap (PS Old Gen) of size 692538048 to monitor. collectionPageThreshold = 489396648, usageThreshold = 489396648
2024-08-30 05:06:10,733 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaDriver - File concatenation threshold: 100 optimistic: false
2024-08-30 05:06:10,733 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaDriver - MR plan size before optimization: 1
2024-08-30 05:06:10,733 INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.LambdaDriver - MR plan size after optimization: 1
2024-08-30 05:06:10,733 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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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 10 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECO
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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
NOS)
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

(hadoop@kali) [~/hadoop/bin]

```

\$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

```