

EXP NO: 4**CREATE UDF IN PIG****\$start-all.sh****\$ jps****\$wget <https://dlcdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz>****\$ tar xvfz pig-0.16.0.tar.gz****\$nano ~/.bashrc**

```

hadoop@fedora:~/da/exp4
Job DAG:
job_local1466974387_0001

2024-09-11 14:36:03,944 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:36:03,957 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:36:03,972 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:36:04,007 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Success!
2024-09-11 14:36:04,022 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:36:04,026 [main] WARN org.apache.pig.data.SchemaToolBackend - SchemaToolBackend has already been initialized
2024-09-11 14:36:04,073 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-09-11 14:36:04,074 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,John)
(2,Jane)
(3,Joe)
(4,Emma)
2024-09-11 14:36:04,276 [main] INFO org.apache.pig.Main - Pig script completed in 15 seconds and 101 milliseconds (15101 ms)
[hadoop@fedora exp4] $ cd ..
[hadoop@fedora da] $ ls
demo pig exp4
[hadoop@fedora da] $ cd exp4
[hadoop@fedora exp4] $ ls
demo pig pig pig_1726045169605.log pig_1726045407699.log pig_1726045457396.log sample.txt
[hadoop@fedora exp4] $ nano uppercase_udf.py
[hadoop@fedora exp4] $ nano uppercase_udf.py
[hadoop@fedora exp4] $ hdfs dfs -copyFromLocal uppercase_udf.py /pig
[hadoop@fedora exp4] $ nano udf_example.pig
[hadoop@fedora exp4] $ pig -f udf_example.pig
2024-09-11 14:48:12,588 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-11 14:48:12,591 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-11 14:48:12,591 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-11 14:48:12,710 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-11 14:48:12,710 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/da/exp4/pig_1726040292686.log
2024-09-11 14:48:13,510 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/pigbootstrap not found
2024-09-11 14:48:13,675 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-09-11 14:48:13,675 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:48:13,675 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-09-11 14:48:14,254 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:48:14,513 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf-example.pig-708c3640-fe1f-4903-bcce-a08c300b6851
2024-09-11 14:48:14,514 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-09-11 14:48:15,000 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:48:15,799 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - created tmp python.cachedir=/tmp/pig_jython_12909388127781137576
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.python.core.PySystemState (file:/home/hadoop/pig/lib/jython-standalone-2.7.0.jar) to method java.io.Console.encoding()
WARNING: Please consider reporting this to the maintainers of org.python.core.PySystemState
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
2024-09-11 14:48:22,513 [main] WARN org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.remaining is empty. This is not expected unless on testing.
2024-09-11 14:48:22,530 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF: udf_uppercase
2024-09-11 14:48:23,198 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:48:23,244 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:48:23,401 [main] INFO org.apache.pig.scripting.jython.JythonFunction - No schema defined for function 'uppercase' in /tmp/pig224194580207460181tmp/uppercase_udf.py
2024-09-11 14:48:23,507 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:48:23,602 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

```

\$mv pig-0.16.0 pig

\$pig

\$cd DA-Lab

\$mkdir exp4

\$cd exp4

\$nano sample.txt

\$nano demo_pig.pig

\$hdfs dfs -mkdir /exp4

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

\$pig demo_pig.pig

\$nano uppercase_udf.py

```

hadoop@fedora:~/da/exp4
$ bash
hadoop@fedora:~/da/exp4$ nano sample.txt
hadoop@fedora:~/da/exp4$ nano demo_pig.pig
hadoop@fedora:~/da/exp4$ pig demo_pig.pig
2024-09-11 14:29:29,555 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-11 14:29:29,559 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-11 14:29:29,609 INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:40
2024-09-11 14:29:29,609 (main) INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/da/exp4/pig.1726045109665.log
2024-09-11 14:29:30,489 (main) INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootstrap not found
2024-09-11 14:29:30,636 (main) INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-09-11 14:29:30,636 (main) INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:29:30,636 (main) INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-09-11 14:29:31,649 (main) INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:29:31,906 (main) INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-demo_pig.pig-ef66083-9018-4c76-0457-fb2aa97789ec
2024-09-11 14:29:31,906 (main) WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-09-11 14:29:32,310 (main) ERROR org.apache.pig.PigServer - exception during parsing: Error during parsing. <file demo_pig.pig, line 5, column 0> mismatched input 'DUMP' expecting SEMI_COLON
Failed to parse: <file demo_pig.pig, line 5, column 0> mismatched input 'DUMP' expecting SEMI_COLON
    at org.apache.pig.parser.QueryParserDriver.parse(QueryParserDriver.java:244)
    at org.apache.pig.parser.QueryParserDriver.parse(QueryParserDriver.java:182)
    at org.apache.pig.PigServer$Graph.parseQuery(PigServer.java:1819)
    at org.apache.pig.PigServer$Graph.access$800(PigServer.java:1527)
    at org.apache.pig.PigServer.parseAndBuild(PigServer.java:480)
    at org.apache.pig.PigServer.executeBatch(PigServer.java:483)
    at org.apache.pig.PigServer.executeBatch(PigServer.java:471)
    at org.apache.pig.tools.grunt.GruntParser.executeBatch(GruntParser.java:172)
    at org.apache.pig.tools.grunt.GruntParser.parseStopOnError(GruntParser.java:235)
    at org.apache.pig.tools.grunt.GruntParser.parseStopOnError(GruntParser.java:206)
    at org.apache.pig.tools.grunt.Grunt.exec(Grunt.java:81)
    at org.apache.pig.Main.run(Main.java:130)
    at org.apache.pig.Main.main(Main.java:176)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(Native Method)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.base/java.lang.reflect.Method.invoke(Method.java:566)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:328)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:243)
2024-09-11 14:29:32,317 (main) ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1208: <file demo_pig.pig, line 5, column 0> mismatched input 'DUMP' expecting SEMI_COLON
Details at logfile: /home/hadoop/da/exp4/pig.1726045109665.log
2024-09-11 14:29:32,381 (main) INFO org.apache.pig.Main - Pig script completed in 2 seconds and 966 milliseconds (2966 ms)
hadoop@fedora:~/da/exp4$ hdfs dfs -mkdir /exp4
hadoop@fedora:~/da/exp4$ nano demo_pig.pig
hadoop@fedora:~/da/exp4$ pig demo_pig.pig
2024-09-11 14:33:27,591 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-11 14:33:27,594 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-11 14:33:27,724 (main) INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:40
2024-09-11 14:33:27,724 (main) INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/da/exp4/pig.1726045487699.log
2024-09-11 14:33:28,532 (main) INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootstrap not found
2024-09-11 14:33:28,690 (main) INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-09-11 14:33:28,690 (main) INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:33:28,690 (main) INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-09-11 14:33:29,916 (main) INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-11 14:33:29,969 (main) INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-demo_pig.pig-ef66083-9018-4c76-0457-fb2aa97789ec
2024-09-11 14:33:29,969 (main) WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false

```

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

```
$nano udf_example.pig
```

```
$pig -f udf_example.pig
```

```
hadoop@fedora:~/da/exp4
at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.base/java.lang.reflect.Method.invoke(Method.java:566)
at org.apache.hadoop.util.RunJar.run(RunJar.java:328)
at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
2024-09-11 14:33:30,316 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1200: <file demo_pig.pig, line 5, column
0> mismatched input 'DUMP' expecting SEMI_COLON
Details at logfile: /home/hadoop/da/exp4/pig_1726045407699.log
2024-09-11 14:33:30,380 [main] INFO org.apache.pig.Main - Pig script completed in 2 seconds and 918 milliseconds (291
8 ms)
[hadoop@fedora exp4]$ nano demo_pig.pig
[hadoop@fedora exp4]$ pig demo_pig.pig
2024-09-11 14:34:17,296 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-11 14:34:17,299 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-11 14:34:17,299 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-11 14:34:17,420 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016,
23:10:49
2024-09-11 14:34:17,420 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/da/exp4/pig_1726045
457396.log
2024-09-11 14:34:18,225 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/.pigbootstrap not
found
2024-09-11 14:34:18,369 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecate
ed. Instead, use mapreduce.jobtracker.address
2024-09-11 14:34:18,369 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-11 14:34:18,370 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to ha
doo file system at: hdfs://localhost:9000
2024-09-11 14:34:19,515 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-11 14:34:19,575 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-demo_pig.pig-114288
6c-f5d8-4f00-9076-ce9b4317a5b1
2024-09-11 14:34:19,576 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled se
t to false
2024-09-11 14:34:20,618 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-11 14:34:21,209 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKN
OWN
2024-09-11 14:34:21,271 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-11 14:34:21,294 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... wil
l not generate code.
2024-09-11 14:34:21,378 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[A
ddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, Me
rgeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, S
plitFilter, StreamTypeCastInserter]}
2024-09-11 14:34:21,553 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (G1 Old Gen) of s
ize 1048576000 to monitor. collectionUsageThreshold = 734003200, usageThreshold = 734003200
2024-09-11 14:34:21,671 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File co
ncatenation threshold: 100 optimistic? false
2024-09-11 14:34:21,727 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
- MR plan size before optimization: 1
2024-09-11 14:34:21,728 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
- MR plan size after optimization: 1
2024-09-11 14:34:21,793 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-11 14:34:22,033 [main] INFO org.apache.hadoop.metrics2.impl.MetricsConfig - Loaded properties from hadoop-met
```

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

```
hadoop@fedora:~/da/exp4
GC time elapsed (ms)=0
Total committed heap usage (bytes)=204472320
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
2024-09-11 14:48:31,941 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local125569218_0001_e_000000_0
2024-09-11 14:48:31,942 [Thread-24] INFO org.apache.hadoop.mapred.LocalJobRunner - map task executor complete.
2024-09-11 14:48:32,245 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 58% complete
2024-09-11 14:48:32,245 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_local125569218_0001]
2024-09-11 14:48:35,761 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:48:35,775 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:48:35,777 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2024-09-11 14:48:35,777 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2024-09-11 14:48:35,781 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:48:35,850 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2024-09-11 14:48:35,858 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
3.3.6 0.16.0 hadoop 2024-09-11 14:48:24 2024-09-11 14:48:35 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTimeAvgReduceTime MedianReduceTime Alias Feature Outputs
job_local125569218_0001 1 0 n/a n/a n/a n/a 0 0 0 0 data,uppercased_data MAP_ONLY hdfs:///pig/pig_output_data,

Input(s):
Successfully read 4 records (42400493 bytes) from: "hdfs:///pig/sample.txt"

Output(s):
Successfully stored 4 records (42400294 bytes) in: "hdfs:///pig/pig_output_data"

Counters:
Total records written : 4
Total bytes written : 42400294
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job Bags:
job_local125569218_0001

2024-09-11 14:48:35,869 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:48:35,877 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:48:35,884 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-11 14:48:35,942 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-09-11 14:48:36,488 [main] INFO org.apache.pig.Main - Pig script completed in 23 seconds and 530 milliseconds (23530 ms)
[hadoop@fedora exp4]$ hdfs dfs -cat /pig/pig_output_data/*
1,JOHN
2,JANE
3,JOE
4,EMMA
[hadoop@fedora exp4]$
```