

## EXP 4: Create UDF in PIG

### Step-by-step installation of Apache Pig on Hadoop cluster on Ubuntu

Step 1: Login into Ubuntu

Step 2: Go to <https://pig.apache.org/releases.html> and copy the path of the latest version of pig that you want to install. Run the following command to download Apache Pig in Ubuntu:

```
$ wget https://dlcdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
```

Step 3: To untar pig-0.16.0.tar.gz file run the following command:

```
$ tar xvzf pig-0.16.0.tar.gz
```

Step 4: To create a pig folder and move pig-0.16.0 to the pig folder, execute the following command:

```
$ sudo mv /home/hadoop/pig-0.16.0 /home/hadoop/pig
```

Step 5: Now open the .bashrc file to edit the path and variables/settings for pig. Run the following command:

```
$ sudo nano .bashrc
```

Add the below given to .bashrc file at the end and save the file.

```
#PIG settings
```

```
export PIG_HOME=/home/hadoop/pig
```

```
Export PATH=$PATH:$PIG_HOME/bin
```

```
Export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_INSTALL/etc/hadoop/
```

```
Export PIG_CONF_DIR=$PIG_HOME/confexport JAVA_HOME=/usr/lib/jvm/
```

```
java-8-openjdk- amd64
```

```
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
```

```
#PIG setting ends
```

Step 6: Run the following command to make the changes effective in the .bashrc file:

```
$ source .bashrc
```

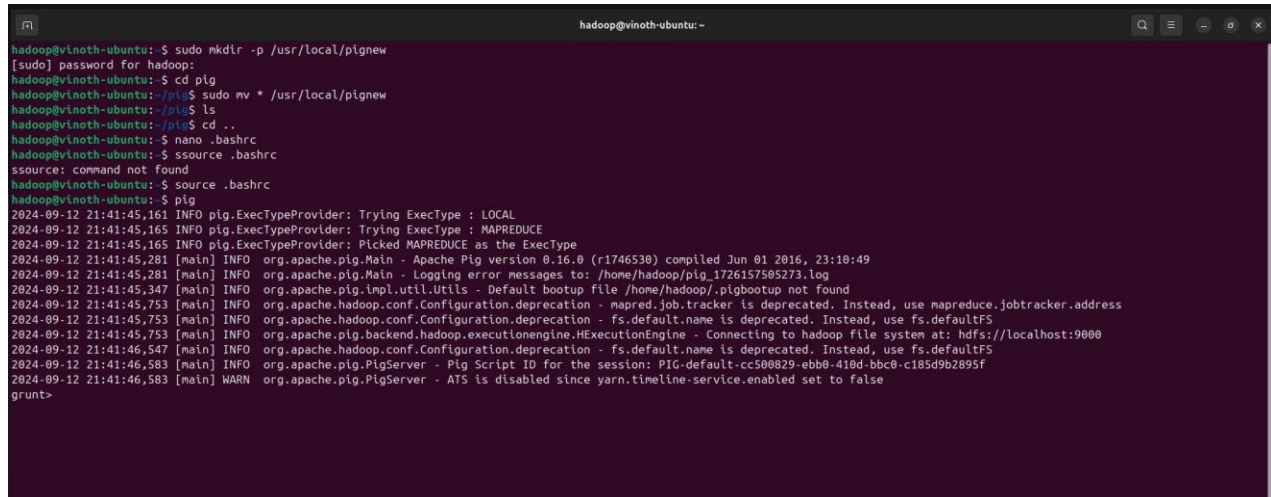
Step 7: To start all Hadoop daemons, navigate to the hadoop-3.2.1/sbin folder and run the following commands:

```
$ ./start-dfs.sh$ ./start-yarn$ jps
```

Step 8: Now you can launch pig by executing the following command:

```
$ pig
```

Step 9: Now you are in pig and can perform your desired tasks on pig. You can come out of the pig by the quit command:  
> quit;

A terminal window titled 'hadoop@vinoth-ubuntu: -' showing the installation and execution of Apache Pig. The user runs 'sudo mkdir -p /usr/local/pignew', 'cd pig', 'sudo mv \* /usr/local/pignew', 'ls', 'cd ..', 'nano .bashrc', 'ssource .bashrc' (which fails with 'command not found'), and 'source .bashrc'. Then they run 'pig', which outputs a series of log messages including 'org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49' and 'org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-cc500829-ebb0-410d-bbc8-c185d9b2895f'. The prompt returns to 'grunt>'.

```
hadoop@vinoth-ubuntu:~$ sudo mkdir -p /usr/local/pignew
[sudo] password for hadoop:
hadoop@vinoth-ubuntu:~$ cd pig
hadoop@vinoth-ubuntu:~/pig$ sudo mv * /usr/local/pignew
hadoop@vinoth-ubuntu:~/pig$ ls
hadoop@vinoth-ubuntu:~/pig$ cd ..
hadoop@vinoth-ubuntu:~$ nano .bashrc
hadoop@vinoth-ubuntu:~$ ssource .bashrc
ssource: command not found
hadoop@vinoth-ubuntu:~$ source .bashrc
hadoop@vinoth-ubuntu:~$ pig
2024-09-12 21:41:45,161 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-12 21:41:45,165 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-12 21:41:45,165 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-12 21:41:45,281 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-12 21:41:45,281 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1726157585273.log
2024-09-12 21:41:45,347 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found
2024-09-12 21:41:45,753 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-09-12 21:41:45,753 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-12 21:41:45,753 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-09-12 21:41:46,547 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-09-12 21:41:46,583 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-cc500829-ebb0-410d-bbc8-c185d9b2895f
2024-09-12 21:41:46,583 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline.service.enabled set to false
grunt>
```

## CREATE USER DEFINED FUNCTION(UDF)

```
hadoop@ubuntu:~/Documents$ hadoop fs -mkdir /home
hadoop@ubuntu:~/Documents$ hadoop fs -mkdir /home/hadoop
hadoop@ubuntu:~/Documents$ hadoop fs -mkdir /home/hadoop/piginput
hadoop@ubuntu:~/Documents$ hadoop fs -put sample.txt
/home/hadoop/piginput
hadoop@ubuntu:~/Documents$ nano demo_pig.pig
hadoop@ubuntu:~/Documents$ pig demo_pig.pig
2024-09-12 22:14:17,196 INFO pig.ExecTypeProvider: Trying ExecType :
LOCAL
2024-09-12 22:14:17,198 INFO pig.ExecTypeProvider: Trying ExecType :
MAPREDUCE
2024-09-12 22:14:17,198 INFO pig.ExecTypeProvider: Picked MAPREDUCE as
the ExecType
2024-09-12 22:14:17,262 [main] INFO org.apache.pig.Main - Apache Pig version
0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-12 22:14:17,262 [main] INFO org.apache.pig.Main - Logging error
messages to: /home/hadoop/Documents/pig_1726159457255.log
2024-09-12 22:14:17,503 [main] INFO org.apache.pig.impl.util.Utils - Default
bootup file /home/hadoop/.pigbootup not found
2024-09-12 22:14:17,578 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is
deprecated. Instead, use mapreduce.jobtracker.address
```

2024-09-12 22:14:17,578 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS

2024-09-12 22:14:17,578 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting  
to hadoop file system at: hdfs://localhost:9000

2024-09-12 22:14:17,948 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS

2024-09-12 22:14:17,960 [main] INFO org.apache.pig.PigServer - Pig Script ID  
for the session: PIG-demo\_pig.pig-161613a1-cbde-4eff-a1fa-3e0e980e9988

2024-09-12 22:14:17,961 [main] WARN org.apache.pig.PigServer - ATS is  
disabled since yarn.timeline-service.enabled set to false

2024-09-12 22:14:18,369 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS

2024-09-12 22:14:18,578 [main] INFO org.apache.pig.tools.pigstats.ScriptState -  
Pig features used in the script: UNKNOWN

2024-09-12 22:14:18,637 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS

2024-09-12 22:14:18,664 [main] INFO org.apache.pig.data.SchemaTupleBackend  
- Key [pig.schematuple] was not set... will not generate code.

2024-09-12 22:14:18,731 [main] INFO  
org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer -  
{RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator,  
GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter,  
MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer,  
PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

2024-09-12 22:14:18,794 [main] INFO  
org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen)  
of size 699400192 to monitor. collectionUsageThreshold = 489580128,  
usageThreshold = 489580128

2024-09-12 22:14:18,837 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler -  
File concatenation threshold: 100 optimistic? false

2024-09-12 22:14:18,862 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOpti  
mizer - MR plan size before optimization: 1

2024-09-12 22:14:18,862 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOpti  
mizer - MR plan size after optimization: 1

2024-09-12 22:14:18,886 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS

2024-09-12 22:14:18,977 [main] INFO  
org.apache.hadoop.metrics2.impl.MetricsConfig - Loaded properties from  
hadoop-metrics2.properties

2024-09-12 22:14:19,032 [main] INFO  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - Scheduled Metric snapshot  
period at 10 second(s).

2024-09-12 22:14:19,032 [main] INFO  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
started

2024-09-12 22:14:19,057 [main] INFO  
org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are  
added to the job

2024-09-12 22:14:19,063 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation -  
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use  
mapreduce.reduce.markreset.buffer.percent

2024-09-12 22:14:19,063 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2024-09-12 22:14:19,063 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is  
deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2024-09-12 22:14:19,072 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - This job cannot be converted run in-process

2024-09-12 22:14:19,127 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is  
deprecated. Instead, use mapreduce.client.submit.file.replication

2024-09-12 22:14:19,822 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/pig-0.16.0-core-h2.jar to DistributedCache  
through /tmp/temp-1322036369/tmp391406508/pig-0.16.0-core-h2.jar

2024-09-12 22:14:20,313 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/automaton-1.11-8.jar to  
DistributedCache through  
/tmp/temp-1322036369/tmp-998973096/automaton-1.11-8.jar

2024-09-12 22:14:20,816 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/antlr-runtime-3.4.jar to

DistributedCache through  
/tmp/temp-1322036369/tmp1544990091/antlr-runtime-3.4.jar  
2024-09-12 22:14:20,896 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/joda-time-2.9.3.jar to DistributedCache  
through /tmp/temp-1322036369/tmp-1874402139/joda-time-2.9.3.jar  
2024-09-12 22:14:20,910 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Setting up single store job  
2024-09-12 22:14:20,923 [main] INFO org.apache.pig.data.SchemaTupleFrontend  
- Key [pig.schematuple] is false, will not generate code.  
2024-09-12 22:14:20,923 [main] INFO org.apache.pig.data.SchemaTupleFrontend  
- Starting process to move generated code to distributed cacche  
2024-09-12 22:14:20,923 [main] INFO org.apache.pig.data.SchemaTupleFrontend  
- Setting key [pig.schematuple.classes] with classes to deserialize []  
2024-09-12 22:14:21,014 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - 1 map-reduce job(s) waiting for submission.  
2024-09-12 22:14:21,082 [JobControl] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:21,128 [JobControl] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.task.id is deprecated.  
Instead, use mapreduce.task.attempt.id  
2024-09-12 22:14:21,207 [JobControl] WARN  
org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User  
classes may not be found. See Job or Job#setJar(String).  
2024-09-12 22:14:21,328 [JobControl] INFO org.apache.pig.builtin.PigStorage -  
Using PigTextInputFormat  
2024-09-12 22:14:21,361 [JobControl] INFO  
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to  
process : 1  
2024-09-12 22:14:21,361 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths to process : 1  
2024-09-12 22:14:21,411 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths (combined) to process : 1  
2024-09-12 22:14:21,452 [JobControl] INFO  
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2024-09-12 22:14:21,568 [JobControl] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS

2024-09-12 22:14:21,647 [JobControl] INFO  
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job:  
job\_local971728215\_0001  
2024-09-12 22:14:21,647 [JobControl] INFO  
org.apache.hadoop.mapreduce.JobSubmitter - Executing with tokens: []  
2024-09-12 22:14:22,258 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_30f6cc3d-0397-4c1  
a-94ef-110800f1d87e/pig-0.16.0-core-h2.jar <-  
/home/hadoop/Documents/pig-0.16.0-core-h2.jar  
2024-09-12 22:14:22,314 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp-1322036369/tmp391406508/pig-0.16.0-core-h2.jar  
as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_30f6cc3d-0397-  
4c1a-94ef-110800f1d87e/pig-0.16.0-core-h2.jar  
2024-09-12 22:14:22,315 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_4117f54e-6d4f-4cb  
1-8e00-def71b7405a4/automaton-1.11-8.jar <-  
/home/hadoop/Documents/automaton-1.11-8.jar  
2024-09-12 22:14:22,392 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp-1322036369/tmp-998973096/automaton-1.11-8.jar  
as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_4117f54e-6d4f-  
4cb1-8e00-def71b7405a4/automaton-1.11-8.jar  
2024-09-12 22:14:22,392 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_da5ad78e-5f2f-44e0  
-8fed-6e10ac5424bb/antlr-runtime-3.4.jar <-  
/home/hadoop/Documents/antlr-runtime-3.4.jar  
2024-09-12 22:14:22,520 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp-1322036369/tmp1544990091/antlr-runtime-3.4.jar  
as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_da5ad78e-5f2f-  
44e0-8fed-6e10ac5424bb/antlr-runtime-3.4.jar  
2024-09-12 22:14:22,520 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_78615d63-c453-418  
9-bad0-6fc2d5ce98e1/joda-time-2.9.3.jar <-  
/home/hadoop/Documents/joda-time-2.9.3.jar

2024-09-12 22:14:22,607 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp-1322036369/tmp-1874402139/joda-time-2.9.3.jar  
as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_78615d63-c453-4189-bad0-6fc2d5ce98e1/joda-time-2.9.3.jar  
2024-09-12 22:14:22,689 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_30f6cc3d-0397-4c1a-94ef-110800f1d87e/pig-0.16.0-core-h2.jar  
2024-09-12 22:14:22,690 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_4117f54e-6d4f-4cb1-8e00-def71b7405a4/automaton-1.11-8.jar  
2024-09-12 22:14:22,690 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_da5ad78e-5f2f-44e0-8fed-6e10ac5424bb/antlr-runtime-3.4.jar  
2024-09-12 22:14:22,690 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local971728215\_0001\_78615d63-c453-4189-bad0-6fc2d5ce98e1/joda-time-2.9.3.jar  
2024-09-12 22:14:22,703 [JobControl] INFO org.apache.hadoop.mapreduce.Job -  
The url to track the job: http://localhost:8080/  
2024-09-12 22:14:22,708 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_local971728215\_0001  
2024-09-12 22:14:22,730 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases data  
2024-09-12 22:14:22,730 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: data[2,7],data[-1,-1] C: R:  
2024-09-12 22:14:22,762 [Thread-19] INFO  
org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter set in config null  
2024-09-12 22:14:22,812 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2024-09-12 22:14:22,812 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_local971728215\_0001]  
2024-09-12 22:14:22,892 [Thread-19] INFO  
org.apache.hadoop.conf.Configuration.deprecation -

mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use  
mapreduce.reduce.markreset.buffer.percent  
2024-09-12 22:14:22,892 [Thread-19] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:14:22,901 [Thread-19] INFO  
org.apache.hadoop.mapreduce.lib.output.PathOutputCommitterFactory - No output  
committer factory defined, defaulting to FileOutputCommitterFactory  
2024-09-12 22:14:22,924 [Thread-19] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output  
Committer Algorithm version is 2  
2024-09-12 22:14:22,924 [Thread-19] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter -  
FileOutputCommitter skip cleanup \_temporary folders under output directory:false,  
ignore cleanup failures: false  
2024-09-12 22:14:22,927 [Thread-19] INFO  
org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter is  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigOutputCom  
mitter  
2024-09-12 22:14:23,099 [Thread-19] INFO  
org.apache.hadoop.mapred.LocalJobRunner - Waiting for map tasks  
2024-09-12 22:14:23,100 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner - Starting task:  
attempt\_local971728215\_0001\_m\_000000\_0  
2024-09-12 22:14:23,184 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.PathOutputCommitterFactory - No output  
committer factory defined, defaulting to FileOutputCommitterFactory  
2024-09-12 22:14:23,185 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output  
Committer Algorithm version is 2  
2024-09-12 22:14:23,186 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter -  
FileOutputCommitter skip cleanup \_temporary folders under output directory:false,  
ignore cleanup failures: false  
2024-09-12 22:14:23,226 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Using ResourceCalculatorProcessTree : [ ]  
2024-09-12 22:14:23,263 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.MapTask - Processing split: Number of splits :1  
Total Length = 27  
Input split[0]:  
Length = 27  
ClassName: org.apache.hadoop.mapreduce.lib.input.FileSplit  
Locations:



-----

2024-09-12 22:14:23,315 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.builtin.PigStorage - Using PigTextInputFormat

2024-09-12 22:14:23,318 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigRecordReader - Current split being processed  
hdfs://localhost:9000/home/hadoop/piginput/sample.txt:0+27

2024-09-12 22:14:23,332 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.PathOutputCommitterFactory - No output  
committer factory defined, defaulting to FileOutputCommitterFactory

2024-09-12 22:14:23,332 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output  
Committer Algorithm version is 2

2024-09-12 22:14:23,335 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter -  
FileOutputCommitter skip cleanup \_temporary folders under output directory:false,  
ignore cleanup failures: false

2024-09-12 22:14:23,426 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen)  
of size 699400192 to monitor. collectionUsageThreshold = 489580128,  
usageThreshold = 489580128

2024-09-12 22:14:23,427 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set...  
will not generate code.

2024-09-12 22:14:23,451 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly\$  
Map - Aliases being processed per job phase (AliasName[line,offset]): M:  
data[2,7],data[-1,-1] C: R:

2024-09-12 22:14:23,495 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner -

2024-09-12 22:14:23,967 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task -  
Task:attempt\_local971728215\_0001\_m\_000000\_0 is done. And is in the process  
of committing

2024-09-12 22:14:23,990 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner -

2024-09-12 22:14:23,993 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Task  
attempt\_local971728215\_0001\_m\_000000\_0 is allowed to commit now

2024-09-12 22:14:24,043 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of  
task 'attempt\_local971728215\_0001\_m\_000000\_0' to  
hdfs://localhost:9000/tmp/temp-1322036369/tmp926395784

2024-09-12 22:14:24,046 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner - map

2024-09-12 22:14:24,046 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Task

'attempt\_local971728215\_0001\_m\_000000\_0' done.

2024-09-12 22:14:24,065 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Final Counters for

attempt\_local971728215\_0001\_m\_000000\_0: Counters: 21

File System Counters

FILE: Number of bytes read=438

FILE: Number of bytes written=6174142

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=5378234

HDFS: Number of bytes written=5378257

HDFS: Number of read operations=54

HDFS: Number of large read operations=0

HDFS: Number of write operations=13

HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework

Map input records=4

Map output records=4

Input split bytes=380

Spilled Records=0

Failed Shuffles=0

Merged Map outputs=0

GC time elapsed (ms)=0

Total committed heap usage (bytes)=418381824

File Input Format Counters

Bytes Read=0

File Output Format Counters

Bytes Written=0

2024-09-12 22:14:24,068 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner - Finishing task:

attempt\_local971728215\_0001\_m\_000000\_0

2024-09-12 22:14:24,068 [Thread-19] INFO

org.apache.hadoop.mapred.LocalJobRunner - map task executor complete.

2024-09-12 22:14:24,318 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - 50% complete

2024-09-12 22:14:24,318 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - Running jobs are [job\_local971728215\_0001]

2024-09-12 22:14:27,851 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:27,856 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:27,861 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.map.tasks is  
deprecated. Instead, use mapreduce.job.maps  
2024-09-12 22:14:27,862 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is  
deprecated. Instead, use mapreduce.job.reduces  
2024-09-12 22:14:27,862 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:27,960 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - 100% complete  
2024-09-12 22:14:27,963 [main] INFO  
org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
3.3.6	0.16.0hadoop		2024-09-12 22:14:19	2024-09-12 22:14:27	
UNKNOWN					

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
job_local971728215_0001											1	0	n/a n/a n/a n/a 0 00
0	data	MAP_ONLY											

hdfs://localhost:9000/tmp/temp-1322036369/tmp926395784,

Input(s):

Successfully read 4 records (5378234 bytes) from:  
"/home/hadoop/piginput/sample.txt"

Output(s):

Successfully stored 4 records (5378257 bytes) in:  
"hdfs://localhost:9000/tmp/temp-1322036369/tmp926395784"

Counters:

Total records written : 4  
Total bytes written : 5378257  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0

Job DAG:  
job\_local971728215\_0001

2024-09-12 22:14:27,964 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:27,966 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:27,967 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:14:27,980 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - Success!  
2024-09-12 22:14:27,993 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:14:27,995 [main] WARN  
org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already  
been initialized  
2024-09-12 22:14:28,040 [main] INFO  
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to  
process : 1  
2024-09-12 22:14:28,043 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths to process : 1  
(1,John)  
(2,Jane)  
(3,Joe)  
(4,Emma)  
2024-09-12 22:14:28,181 [main] INFO org.apache.pig.Main - Pig script  
completed in 11 seconds and 16 milliseconds (11016 ms)  
hadoop@ubuntu:~/Documents\$ ^C  
hadoop@ubuntu:~/Documents\$

```
hadoop@ubuntu:~/Documents$ nano uppercase_udf.py
hadoop@ubuntu:~/Documents$ hadoop fs -mkdir /home/hadoop/udfs/
hadoop@ubuntu:~/Documents$ nano udf_example.pig
hadoop@ubuntu:~/Documents$ pig -f udf_example.pig
hadoop@ubuntu:~/Documents$ hadoop fs -put uppercase_udf.py
/home/hadoop/udfs/
hadoop@ubuntu:~/Documents$ pig -f udf_example.pig
2024-09-12 22:25:53,446 INFO pig.ExecTypeProvider: Trying ExecType :
LOCAL
2024-09-12 22:25:53,448 INFO pig.ExecTypeProvider: Trying ExecType :
MAPREDUCE
2024-09-12 22:25:53,448 INFO pig.ExecTypeProvider: Picked MAPREDUCE as
the ExecType
2024-09-12 22:25:53,491 [main] INFO org.apache.pig.Main - Apache Pig version
0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-12 22:25:53,491 [main] INFO org.apache.pig.Main - Logging error
messages to: /home/hadoop/Documents/pig_1726160153487.log
2024-09-12 22:25:53,699 [main] INFO org.apache.pig.impl.util.Utils - Default
bootup file /home/hadoop/.pigbootup not found
2024-09-12 22:25:53,751 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is
deprecated. Instead, use mapreduce.jobtracker.address
2024-09-12 22:25:53,752 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-12 22:25:53,752 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting
to hadoop file system at: hdfs://localhost:9000
2024-09-12 22:25:54,119 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-12 22:25:54,143 [main] INFO org.apache.pig.PigServer - Pig Script ID
for the session: PIG-udf_example.pig-3a0a6c91-935b-4e72-b01c-21789c73b10d
2024-09-12 22:25:54,143 [main] WARN org.apache.pig.PigServer - ATS is
disabled since yarn.timeline-service.enabled set to false
2024-09-12 22:25:54,177 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.
Instead, use fs.defaultFS
2024-09-12 22:25:54,430 [main] INFO
org.apache.pig.scripting.jython.JythonScriptEngine - created tmp
python.cachedir=/tmp/pig_jython_2186509906898523982
2024-09-12 22:25:57,333 [main] WARN
org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.remainers is
empty. This is not expected unless on testing.
```

2024-09-12 22:25:57,339 [main] INFO  
org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF:  
udf.uppercase  
2024-09-12 22:25:57,694 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:25:57,703 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:25:57,884 [main] INFO  
org.apache.pig.scripting.jython.JythonFunction - No schema defined for function  
'uppercase' in /tmp/pig2391808375011456533tmp/uppercase\_udf.py  
2024-09-12 22:25:57,926 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:25:58,101 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation -  
mapred.textoutputformat.separator is deprecated. Instead, use  
mapreduce.output.textoutputformat.separator  
2024-09-12 22:25:58,202 [main] INFO org.apache.pig.tools.pigstats.ScriptState -  
Pig features used in the script: UNKNOWN  
2024-09-12 22:25:58,268 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:25:58,278 [main] INFO org.apache.pig.data.SchemaTupleBackend  
- Key [pig.schematuple] was not set... will not generate code.  
2024-09-12 22:25:58,371 [main] INFO  
org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer -  
{RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator,  
GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter,  
MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer,  
PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

2024-09-12 22:25:58,445 [main] INFO  
org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen)  
of size 699400192 to monitor. collectionUsageThreshold = 489580128,  
usageThreshold = 489580128  
2024-09-12 22:25:58,520 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler -  
File concatenation threshold: 100 optimistic? false  
2024-09-12 22:25:58,559 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOpti  
mizer - MR plan size before optimization: 1

2024-09-12 22:25:58,559 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOpti  
mizer - MR plan size after optimization: 1  
2024-09-12 22:25:58,589 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:25:58,718 [main] INFO  
org.apache.hadoop.metrics2.impl.MetricsConfig - Loaded properties from  
hadoop-metrics2.properties  
2024-09-12 22:25:58,808 [main] INFO  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - Scheduled Metric snapshot  
period at 10 second(s).  
2024-09-12 22:25:58,808 [main] INFO  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
started  
2024-09-12 22:25:58,828 [main] INFO  
org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are  
added to the job  
2024-09-12 22:25:58,833 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation -  
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use  
mapreduce.reduce.markreset.buffer.percent  
2024-09-12 22:25:58,834 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2024-09-12 22:25:58,834 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is  
deprecated. Instead, use mapreduce.output.fileoutputformat.compress  
2024-09-12 22:25:58,840 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - This job cannot be converted run in-process  
2024-09-12 22:25:58,853 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is  
deprecated. Instead, use mapreduce.client.submit.file.replication  
2024-09-12 22:25:59,522 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/pig-0.16.0-core-h2.jar to DistributedCache  
through /tmp/temp956332242/tmp1195535656/pig-0.16.0-core-h2.jar  
2024-09-12 22:26:00,717 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/jython-standalone-2.7.0.jar to  
DistributedCache through  
/tmp/temp956332242/tmp-275867342/jython-standalone-2.7.0.jar

2024-09-12 22:26:01,219 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/automaton-1.11-8.jar to  
DistributedCache through  
/tmp/temp956332242/tmp568970725/automaton-1.11-8.jar  
2024-09-12 22:26:01,286 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/antlr-runtime-3.4.jar to  
DistributedCache through  
/tmp/temp956332242/tmp-2133106477/antlr-runtime-3.4.jar  
2024-09-12 22:26:01,790 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/usr/local/pignew/lib/joda-time-2.9.3.jar to DistributedCache  
through /tmp/temp956332242/tmp-877263801/joda-time-2.9.3.jar  
2024-09-12 22:26:02,299 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Added jar file:/tmp/PigScriptUDF-0f01b5a3f0424cbfa9184a6cd1311214.jar  
to DistributedCache through  
/tmp/temp956332242/tmp-1279289345/PigScriptUDF-0f01b5a3f0424cbfa9184a6c  
d1311214.jar  
2024-09-12 22:26:02,308 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCom  
piler - Setting up single store job  
2024-09-12 22:26:02,332 [main] INFO org.apache.pig.data.SchemaTupleFrontend  
- Key [pig.schematuple] is false, will not generate code.  
2024-09-12 22:26:02,333 [main] INFO org.apache.pig.data.SchemaTupleFrontend  
- Starting process to move generated code to distributed cacche  
2024-09-12 22:26:02,333 [main] INFO org.apache.pig.data.SchemaTupleFrontend  
- Setting key [pig.schematuple.classes] with classes to deserialize []  
2024-09-12 22:26:02,403 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - 1 map-reduce job(s) waiting for submission.  
2024-09-12 22:26:02,426 [JobControl] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:02,461 [JobControl] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.task.id is deprecated.  
Instead, use mapreduce.task.attempt.id  
2024-09-12 22:26:02,506 [JobControl] WARN  
org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set. User  
classes may not be found. See Job or Job#setJar(String).  
2024-09-12 22:26:02,585 [JobControl] INFO org.apache.pig.builtin.PigStorage -  
Using PigTextInputFormat



2024-09-12 22:26:02,597 [JobControl] INFO  
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to  
process : 1  
2024-09-12 22:26:02,597 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths to process : 1  
2024-09-12 22:26:02,632 [JobControl] INFO  
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input  
paths (combined) to process : 1  
2024-09-12 22:26:02,658 [JobControl] INFO  
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1  
2024-09-12 22:26:02,742 [JobControl] INFO  
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job:  
job\_local1189089322\_0001  
2024-09-12 22:26:02,742 [JobControl] INFO  
org.apache.hadoop.mapreduce.JobSubmitter - Executing with tokens: []  
2024-09-12 22:26:03,217 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_6aab6a5b-3fb7-4cf  
5-9ff3-3b48bf77897b/pig-0.16.0-core-h2.jar <-  
/home/hadoop/Documents/pig-0.16.0-core-h2.jar  
2024-09-12 22:26:03,320 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp956332242/tmp1195535656/pig-0.16.0-core-h2.jar  
as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_6aab6a5b-3fb7  
-4cf5-9ff3-3b48bf77897b/pig-0.16.0-core-h2.jar  
2024-09-12 22:26:03,583 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_1d4925b8-b3e0-45  
33-b051-52472be3f91b/jython-standalone-2.7.0.jar <-  
/home/hadoop/Documents/jython-standalone-2.7.0.jar  
2024-09-12 22:26:03,588 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp956332242/tmp-275867342/jython-standalone-2.7.  
0.jar as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_1d4925b8-b3e  
0-4533-b051-52472be3f91b/jython-standalone-2.7.0.jar  
2024-09-12 22:26:03,588 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_9c6e353a-126f-46  
df-9035-21b47930e952/automaton-1.11-8.jar <-  
/home/hadoop/Documents/automaton-1.11-8.jar

2024-09-12 22:26:03,606 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp956332242/tmp568970725/automaton-1.11-8.jar as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_9c6e353a-126f  
-46df-9035-21b47930e952/automaton-1.11-8.jar  
2024-09-12 22:26:03,606 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_2af9dbdd-38f3-45  
4e-ad31-a1f19c160bad/antlr-runtime-3.4.jar <-  
/home/hadoop/Documents/antlr-runtime-3.4.jar  
2024-09-12 22:26:03,633 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp956332242/tmp-2133106477/antlr-runtime-3.4.jar  
as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_2af9dbdd-38f3  
-454e-ad31-a1f19c160bad/antlr-runtime-3.4.jar  
2024-09-12 22:26:03,640 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_5320506f-c665-4a  
06-b080-3fe64525044c/joda-time-2.9.3.jar <-  
/home/hadoop/Documents/joda-time-2.9.3.jar  
2024-09-12 22:26:03,697 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp956332242/tmp-877263801/joda-time-2.9.3.jar as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_5320506f-c66  
5-4a06-b080-3fe64525044c/joda-time-2.9.3.jar  
2024-09-12 22:26:03,697 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Creating symlink:  
/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_481971b7-5f9f-47  
b4-9cca-d2ac02ad23ab/PigScriptUDF-0f01b5a3f0424cbfa9184a6cd1311214.jar <-  
/home/hadoop/Documents/PigScriptUDF-0f01b5a3f0424cbfa9184a6cd1311214.jar  
2024-09-12 22:26:03,726 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager - Localized  
hdfs://localhost:9000/tmp/temp956332242/tmp-1279289345/PigScriptUDF-0f01b5  
a3f0424cbfa9184a6cd1311214.jar as  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_481971b7-5f9f  
-47b4-9cca-d2ac02ad23ab/PigScriptUDF-0f01b5a3f0424cbfa9184a6cd1311214.jar  
2024-09-12 22:26:03,894 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_6aab6a5b-3fb7  
-4cf5-9ff3-3b48bf77897b/pig-0.16.0-core-h2.jar  
2024-09-12 22:26:03,894 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -

file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_1d4925b8-b3e0-4533-b051-52472be3f91b/jython-standalone-2.7.0.jar  
2024-09-12 22:26:03,894 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_9c6e353a-126f-46df-9035-21b47930e952/automaton-1.11-8.jar  
2024-09-12 22:26:03,894 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_2af9dbdd-38f3-454e-ad31-a1f19c160bad/antlr-runtime-3.4.jar  
2024-09-12 22:26:03,894 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_5320506f-c665-4a06-b080-3fe64525044c/joda-time-2.9.3.jar  
2024-09-12 22:26:03,894 [JobControl] INFO  
org.apache.hadoop.mapred.LocalDistributedCacheManager -  
file:/tmp/hadoop-hadoop/mapred/local/job\_local1189089322\_0001\_481971b7-5f9f-47b4-9cca-d2ac02ad23ab/PigScriptUDF-0f01b5a3f0424cbfa9184a6cd1311214.jar  
2024-09-12 22:26:03,899 [JobControl] INFO org.apache.hadoop.mapreduce.Job -  
The url to track the job: http://localhost:8080/  
2024-09-12 22:26:03,908 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_local1189089322\_0001  
2024-09-12 22:26:03,908 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases data,uppercased\_data  
2024-09-12 22:26:03,908 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: data[5,7],uppercased\_data[-1,-1] C: R:  
2024-09-12 22:26:03,960 [Thread-25] INFO  
org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter set in config null  
2024-09-12 22:26:03,960 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete  
2024-09-12 22:26:03,960 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_local1189089322\_0001]  
2024-09-12 22:26:04,053 [Thread-25] INFO  
org.apache.hadoop.conf.Configuration.deprecation -  
mapred.textoutputformat.separator is deprecated. Instead, use  
mapreduce.output.textoutputformat.separator  
2024-09-12 22:26:04,056 [Thread-25] INFO  
org.apache.hadoop.conf.Configuration.deprecation -

mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use  
mapreduce.reduce.markreset.buffer.percent  
2024-09-12 22:26:04,057 [Thread-25] INFO  
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated.  
Instead, use fs.defaultFS  
2024-09-12 22:26:04,061 [Thread-25] INFO  
org.apache.hadoop.mapreduce.lib.output.PathOutputCommitterFactory - No output  
committer factory defined, defaulting to FileOutputCommitterFactory  
2024-09-12 22:26:04,062 [Thread-25] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output  
Committer Algorithm version is 2  
2024-09-12 22:26:04,062 [Thread-25] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter -  
FileOutputCommitter skip cleanup \_temporary folders under output directory:false,  
ignore cleanup failures: false  
2024-09-12 22:26:04,063 [Thread-25] INFO  
org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter is  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigOutputCom  
mitter  
2024-09-12 22:26:04,185 [Thread-25] INFO  
org.apache.hadoop.mapred.LocalJobRunner - Waiting for map tasks  
2024-09-12 22:26:04,186 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner - Starting task:  
attempt\_local1189089322\_0001\_m\_000000\_0  
2024-09-12 22:26:04,336 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.PathOutputCommitterFactory - No output  
committer factory defined, defaulting to FileOutputCommitterFactory  
2024-09-12 22:26:04,336 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output  
Committer Algorithm version is 2  
2024-09-12 22:26:04,336 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter -  
FileOutputCommitter skip cleanup \_temporary folders under output directory:false,  
ignore cleanup failures: false  
2024-09-12 22:26:04,379 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Using ResourceCalculatorProcessTree : [ ]  
2024-09-12 22:26:04,400 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.MapTask - Processing split: Number of splits :1  
Total Length = 27  
Input split[0]:  
Length = 27  
ClassName: org.apache.hadoop.mapreduce.lib.input.FileSplit  
Locations:

-----

2024-09-12 22:26:04,481 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.builtin.PigStorage - Using PigTextInputFormat

2024-09-12 22:26:04,513 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigRecordReader - Current split being processed  
hdfs://localhost:9000/home/hadoop/piginput/sample.txt:0+27

2024-09-12 22:26:04,519 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.PathOutputCommitterFactory - No output  
committer factory defined, defaulting to FileOutputCommitterFactory

2024-09-12 22:26:04,519 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - File Output  
Committer Algorithm version is 2

2024-09-12 22:26:04,519 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter -  
FileOutputCommitter skip cleanup \_temporary folders under output directory:false,  
ignore cleanup failures: false

2024-09-12 22:26:04,605 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen)  
of size 699400192 to monitor. collectionUsageThreshold = 489580128,  
usageThreshold = 489580128

2024-09-12 22:26:04,607 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set...  
will not generate code.

2024-09-12 22:26:04,648 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly\$  
Map - Aliases being processed per job phase (AliasName[line,offset]): M:  
data[5,7],uppercased\_data[-1,-1] C: R:

2024-09-12 22:26:04,677 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner -

2024-09-12 22:26:04,832 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task -  
Task:attempt\_local1189089322\_0001\_m\_000000\_0 is done. And is in the process  
of committing

2024-09-12 22:26:04,838 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner -

2024-09-12 22:26:04,838 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Task  
attempt\_local1189089322\_0001\_m\_000000\_0 is allowed to commit now

2024-09-12 22:26:04,902 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of  
task 'attempt\_local1189089322\_0001\_m\_000000\_0' to  
hdfs://localhost:9000/home/hadoop/pig\_output\_data

2024-09-12 22:26:04,903 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner - map  
2024-09-12 22:26:04,903 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Task  
'attempt\_local1189089322\_0001\_m\_000000\_0' done.  
2024-09-12 22:26:04,910 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.Task - Final Counters for  
attempt\_local1189089322\_0001\_m\_000000\_0: Counters: 21  
File System Counters  
FILE: Number of bytes read=438  
FILE: Number of bytes written=43492306  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=42400493  
HDFS: Number of bytes written=42400295  
HDFS: Number of read operations=63  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=17  
HDFS: Number of bytes read erasure-coded=0  
Map-Reduce Framework  
Map input records=4  
Map output records=4  
Input split bytes=380  
Spilled Records=0  
Failed Shuffles=0  
Merged Map outputs=0  
GC time elapsed (ms)=0  
Total committed heap usage (bytes)=431489024  
File Input Format Counters  
Bytes Read=0  
File Output Format Counters  
Bytes Written=0  
2024-09-12 22:26:04,910 [LocalJobRunner Map Task Executor #0] INFO  
org.apache.hadoop.mapred.LocalJobRunner - Finishing task:  
attempt\_local1189089322\_0001\_m\_000000\_0  
2024-09-12 22:26:04,915 [Thread-25] INFO  
org.apache.hadoop.mapred.LocalJobRunner - map task executor complete.  
2024-09-12 22:26:04,991 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - 50% complete  
2024-09-12 22:26:04,996 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - Running jobs are [job\_local1189089322\_0001]

2024-09-12 22:26:09,014 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:09,027 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:09,032 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.map.tasks is  
deprecated. Instead, use mapreduce.job.maps  
2024-09-12 22:26:09,032 [main] INFO  
org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is  
deprecated. Instead, use mapreduce.job.reduces  
2024-09-12 22:26:09,032 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:09,081 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - 100% complete  
2024-09-12 22:26:09,084 [main] INFO  
org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
3.3.6	0.16.0hadoop		2024-09-12 22:25:58	2024-09-12 22:26:09	
UNKNOWN					

## Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
job_local1189089322_0001	1	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	00	0	0	data,uppercased_data
MAP_ONLY													
hdfs:///home/hadoop/pig_output_data,													

Input(s):

Successfully read 4 records (42400493 bytes) from:  
"hdfs:///home/hadoop/piginput/sample.txt"

Output(s):

Successfully stored 4 records (42400295 bytes) in:  
"hdfs:///home/hadoop/pig\_output\_data"

Counters:

Total records written : 4  
Total bytes written : 42400295  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0

Job DAG:  
job\_local1189089322\_0001

2024-09-12 22:26:09,086 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:09,090 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:09,091 [main] WARN  
org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system  
already initialized!  
2024-09-12 22:26:09,105 [main] INFO  
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau  
ncher - **Success!**  
2024-09-12 22:26:09,185 [main] INFO org.apache.pig.Main - Pig script  
completed in 15 seconds and 709 milliseconds (15709 ms)  
hadoop@vinoth-ubuntu:~/Documents\$

**hadoop@ubuntu:~/Documents\$ hadoop fs -ls**  
**/home/hadoop/pig\_output\_data**  
Found 2 items  
-rw-r--r-- 1 hadoop supergroup 0 2024-09-12 22:26  
/home/hadoop/pig\_output\_data/\_SUCCESS  
-rw-r--r-- 1 hadoop supergroup 27 2024-09-12 22:26  
/home/hadoop/pig\_output\_data/part-m-00000  
**hadoop@ubuntu:~/Documents\$ hadoop fs -cat**  
**/home/hadoop/pig\_output\_data/part-m-00000**

## **OUTPUT:**

1,JOHN  
2,JANE  
3,JOE  
4,EMMA  
hadoop@ubuntu:~/Documents\$



