ROLL NUMBER: 210701102

EXP NO: 4 **CREATE UDF IN PIG**

$start-all.sh

```
kamal@kamal-VirtualBox: ~
kamal@kamal-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as kamal in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 4790.  Stop it first and ensure /tmp/hadoop-kamal-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 4916.  Stop it first and ensure /tmp/hadoop-kamal-datanode.pid file is empty before retry.
Starting secondary namenodes [kamal-VirtualBox]
kamal-VirtualBox: secondarynamenode is running as process 5112.  Stop it first and ensure /tmp/hadoop-kamal-secondarynamenode.pid fi
le is empty before retry.
Starting resourcemanager
resourcemanager is running as process 5333.  Stop it first and ensure /tmp/hadoop-kamal-resourcemanager.pid file is empty before ret
ry.
Starting nodemanagers
localhost: nodemanager is running as process 5455.  Stop it first and ensure /tmp/hadoop-kamal-nodemanager.pid file is empty before
retry.
```

$ jps

```
kamal@kamal-VirtualBox:~$ jps
4916 DataNode
5333 ResourceManager
4790 NameNode
6919 Jps
5112 SecondaryNameNode
5455 NodeManager
```

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

$ tar xvzf pig-0.16.0.tar.gz

```
kamal@kamal-VirtualBox: ~
kamal@kamal-VirtualBox:~$ tar xvzf pig-0.16.0.tar.gz
pig-0.16.0/
pig-0.16.0/bin/
pig-0.16.0/conf/
pig-0.16.0/contrib/
pig-0.16.0/contrib/piggybank/
pig-0.16.0/contrib/piggybank/java/
pig-0.16.0/contrib/piggybank/java/build/
pig-0.16.0/contrib/piggybank/java/build/classes/
pig-0.16.0/contrib/piggybank/java/build/classes/org/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/convert/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/diff/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/truncate/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/decode/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/math/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/stats/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/string/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/apachelogparser/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/xml/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/allloader/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/apachelog/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/avro/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/hiverc/
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/
pig-0.16.0/contrib/piggybank/java/src/main/
pig-0.16.0/contrib/piggybank/java/src/main/java/
pig-0.16.0/contrib/piggybank/java/src/main/java/org/
```

$nano ~/.bashrc

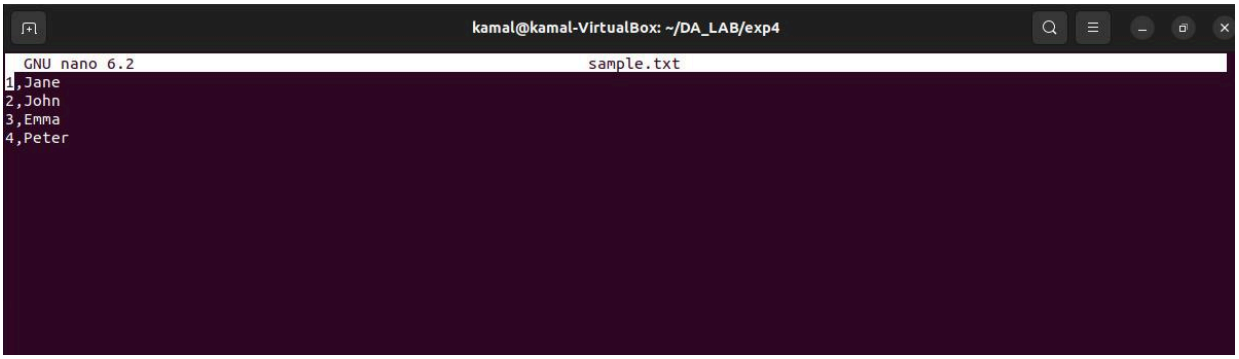
```
#PIG settings
export PIG_HOME=~/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_INSTALL/etc/hadoop/
export PIG_CONF_DIR=$PIG_HOME/conf
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
```

$mv pig-0.16.0 pig

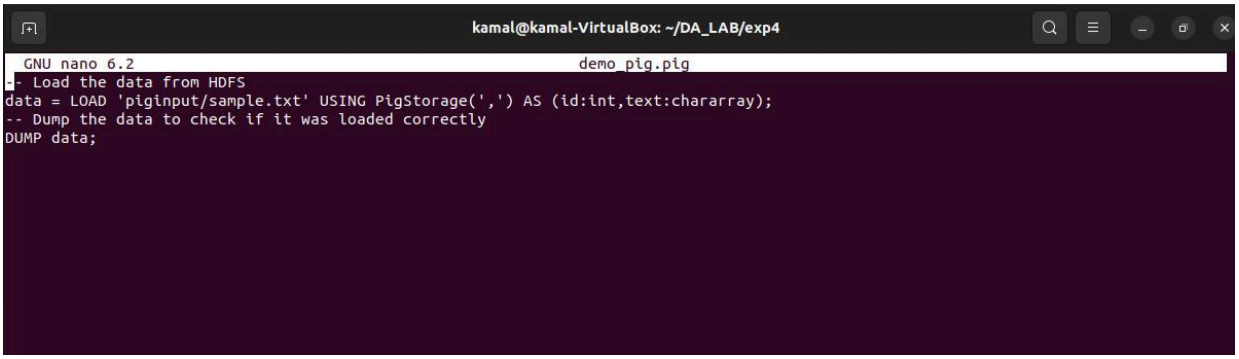
$pig

```
kamal@kamal-VirtualBox:~$ pig
2024-10-12 13:01:39,155 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-12 13:01:39,163 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-10-12 13:01:39,164 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-10-12 13:01:39,283 [main] INFO  org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-10-12 13:01:39,283 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/kamal/pig_1728718299270.log
2024-10-12 13:01:39,352 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/kamal/.pigbootup not found
2024-10-12 13:01:39,836 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, u
se mapreduce.jobtracker.address
2024-10-12 13:01:39,836 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-10-12 13:01:39,836 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file syst
em at: hdfs://localhost:9000
2024-10-12 13:01:40,659 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-10-12 13:01:40,683 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-a878a9a9-2f4d-4742-850e-6
f70e04790cb
2024-10-12 13:01:40,683 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>
```

$cd DA-Lab
$mkdir exp4
$cd exp4
$nano sample.txt



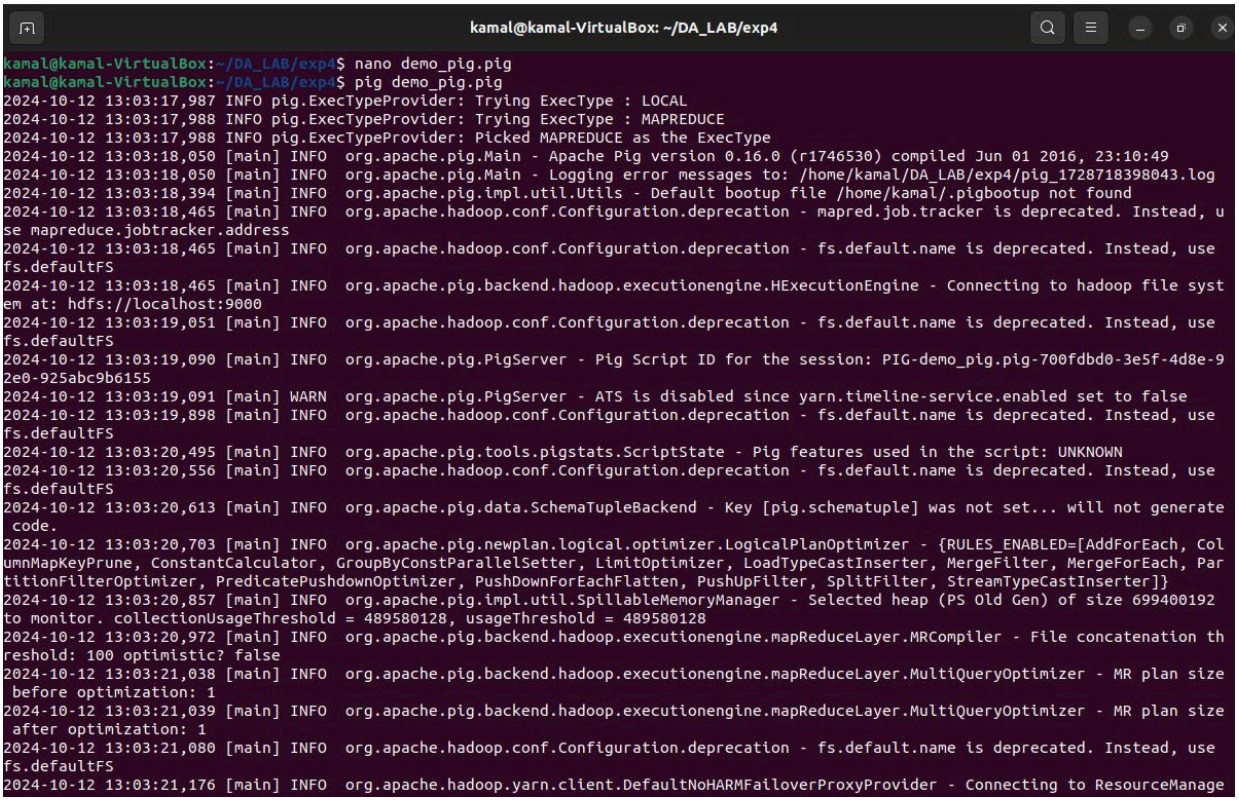
$nano demo_pig.pig



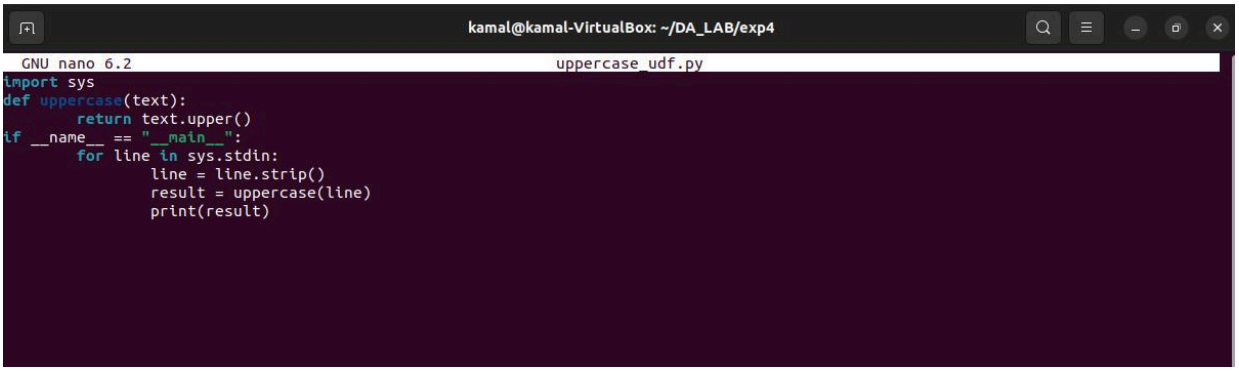
$hdfs dfs -mkdir /exp4

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

$pig demo_pig.pig

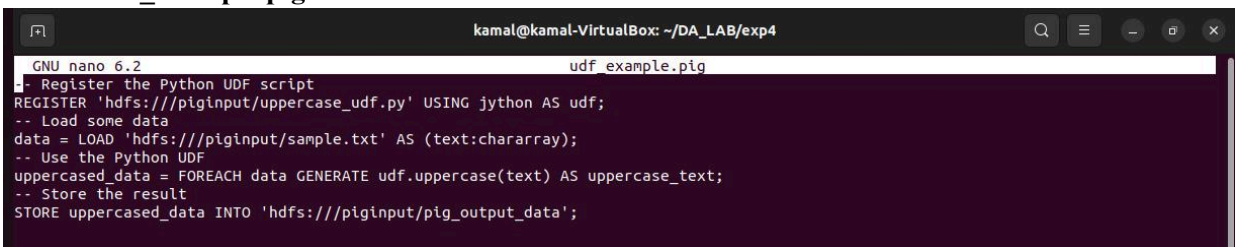


$nano uppercase_udf.py

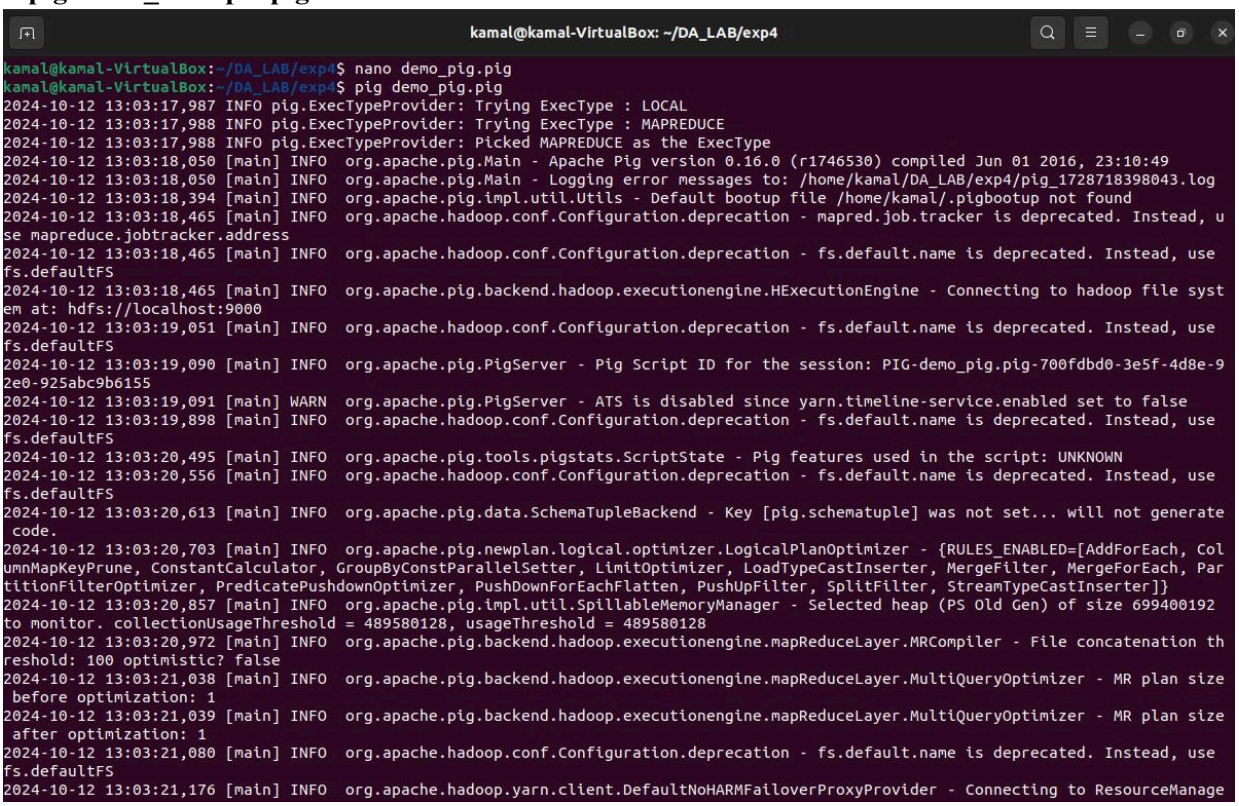


$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/uppercase_udf.py /exp4

$nano udf_example.pig



$pig -f udf_example.pig



ROLL NUMBER: 210701102

$hdfs dfs -cat /exp4/output/*

```
2024-09-04 12:42:06,312 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-04 12:42:06,318 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-04 12:42:06,336 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2024-09-04 12:42:06,448 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-09-04 12:42:06,724 [main] INFO  org.apache.pig.Main - Pig script completed in 43 seconds and 342 milliseconds (43342 ms)
karthickragav@fedora:~/dalab/exp4$ hdfs dfs -ls /
Found 7 items
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:41 /home
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:28 /piginput
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:42 /tmp
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:37 /udfs
drwxr-xr-x   - karthickragav supergroup          0 2024-09-02 11:02 /weatherdata
drwxr-xr-x   - karthickragav supergroup          0 2024-09-01 21:13 /word_count_in_py
drwxr-xr-x   - karthickragav supergroup          0 2024-09-01 20:59 /word_count_in_python
karthickragav@fedora:~/dalab/exp4$ hdfs dfs -ls /tmp
karthickragav@fedora:~/dalab/exp4$ hdfs dfs -ls /home
Found 1 items
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:41 /home/hadoop
karthickragav@fedora:~/dalab/exp4$ ls
demo_pig.pig          pig_1725290479440.log  pig_1725290702550.log  pig_1725432772737.log  udf_example.pig
pig_1725289163833.log  pig_1725290532876.log  pig_1725432499143.log  sample.txt             uppercase_udf.py
karthickragav@fedora:~/dalab/exp4$ cd
karthickragav@fedora:~$ hdfs dfs -ls /home
Found 1 items
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:41 /home/hadoop
karthickragav@fedora:~$ hdfs dfs -ls /home/hadoop
Found 1 items
drwxr-xr-x   - karthickragav supergroup          0 2024-09-04 12:42 /home/hadoop/pig_output_data
karthickragav@fedora:~$ hdfs dfs -ls /home/hadoop/pig_output_data
Found 2 items
-rw-r--r--   1 karthickragav supergroup          0 2024-09-04 12:42 /home/hadoop/pig_output_data/_SUCCESS
-rw-r--r--   1 karthickragav supergroup         27 2024-09-04 12:42 /home/hadoop/pig_output_data/part-m-00000
karthickragav@fedora:~$ hdfs dfs -ls /home/hadoop/pig_output_data/part-m-00000
-rw-r--r--   1 karthickragav supergroup         27 2024-09-04 12:42 /home/hadoop/pig_output_data/part-m-00000
karthickragav@fedora:~$ hdfs dfs -cat /home/hadoop/pig_output_data/part-m-00000
1,JHON
2,JANE
3,JOE
4,EMMA
karthickragav@fedora:~$
```