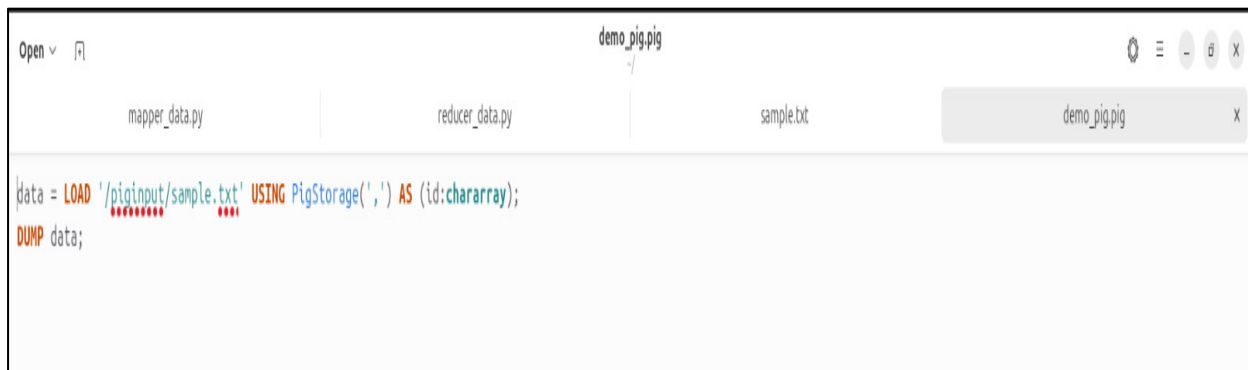


EXP NO: 4**CREATE UDF IN PIG****\$start-all.sh****\$ jps****\$wget <https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz> \$****tar xvzf pig-0.16.0.tar.gz****\$nano ~/.bashrc****\$mv pig-0.16.0 pig****\$pig****\$cd DA-Lab \$mkdir exp4 \$cd exp4 \$nano sample.txt****\$nano demo_pig.pig**

```
data = LOAD '/piginput/sample.txt' USING PigStorage(',') AS (id:chararray);
DUMP data;
```

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

\$pig demo_pig.pig

```

2024-09-13 14:43:17,945 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - Jo
bTracker metrics system already initialized!
2024-09-13 14:43:17,949 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - Jo
bTracker metrics system already initialized!
2024-09-13 14:43:17,953 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - Jo
bTracker metrics system already initialized!
2024-09-13 14:43:17,964 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduc
eLayer.MapReduceLauncher - Success!
2024-09-13 14:43:17,970 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.
default.name is deprecated. Instead, use fs.defaultFS
2024-09-13 14:43:17,971 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBa
ckend has already been initialized
2024-09-13 14:43:17,989 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat
- Total input files to process : 1
2024-09-13 14:43:17,990 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.Map
RedUtil - Total input paths to process : 1
(john)
(jane)
(joe)
(Emma)
2024-09-13 14:43:18,065 [main] INFO org.apache.pig.Main - Pig script completed in 10 secon
ds and 403 milliseconds (10403 ms)

```

\$nano uppercase_udf.py**\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/uppercase_udf.py /exp4****\$nano udf_example.pig**


```

REGISTER 'hdfs:///piginput/uppercase.py' USING jython AS udf;
data = LOAD 'hdfs:///piginput/sample.txt' AS (text:chararray);
uppercased_data = FOREACH data GENERATE udf.uppercase(text) AS uppercase_text;
STORE uppercased_data INTO 'hdfs:///piginput/pig_output_data';

```

\$pig -f udf_example.pig**\$hdfs dfs -cat /exp4/output/***

```

2024-09-13 14:51:40,000 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduc
eLayer.MapReduceLauncher - Success!
2024-09-13 14:51:40,040 [main] INFO org.apache.pig.Main - Pig script completed in 15 secon
ds and 570 milliseconds (15570 ms)
hadoop@ubuntu:~$ hdfs dfs -cat /piginput/pig_output_data/part-m-00000
JOHN
JANE
JOE
EMMA
hadoop@ubuntu:~$

```