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

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

$ tar xvzf pig-0.16.0.tar.gz

```
hadoop@fedora:~/Downloads
bash: cd: downloads: No such file or directory
hadoop@fedora:~$ cd Downloads
hadoop@fedora:~/Downloads$ ls
apache-hive-4.0.0-bin.tar.gz  pig-0.17.0.tar.gz
hadoop@fedora:~/Downloads$ tar xvzf pig-0.17.0.tar.gz
pig-0.17.0/
pig-0.17.0/bin/
pig-0.17.0/conf/
pig-0.17.0/contrib/
pig-0.17.0/contrib/piggybank/
pig-0.17.0/contrib/piggybank/java/
pig-0.17.0/contrib/piggybank/java/build/
pig-0.17.0/contrib/piggybank/java/build/classes/
pig-0.17.0/contrib/piggybank/java/build/classes/org/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/convert/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/diff/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/truncate/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/decode/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/math/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/stats/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/string/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/apachelogparser/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/xml/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/allloader/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/apachelog/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/avro/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/hiverc/
pig-0.17.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/partition/
pig-0.17.0/contrib/piggybank/java/build/docs/
pig-0.17.0/contrib/piggybank/java/build/docs/api/
pig-0.17.0/contrib/piggybank/java/build/test/
pig-0.17.0/contrib/piggybank/java/build/test/classes/
pig-0.17.0/contrib/piggybank/java/lib/
pig-0.17.0/contrib/piggybank/java/src/
pig-0.17.0/contrib/piggybank/java/src/main/
pig-0.17.0/contrib/piggybank/java/src/main/java/
pig-0.17.0/contrib/piggybank/java/src/main/java/org/
pig-0.17.0/contrib/piggybank/java/src/main/java/org/apache/
pig-0.17.0/contrib/piggybank/java/src/main/java/org/apache/pig/
pig-0.17.0/contrib/piggybank/java/src/main/java/org/apache/pig/piggybank/
pig-0.17.0/contrib/piggybank/java/src/main/java/org/apache/pig/piggybank/evaluation/
```

$nano ~/.bashrc

```
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=$PIG_HOME/conf
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
export HIVE_HOME=/home/jeciyazhini/hive
export PATH=$PATH:$HIVE_HOME/bin
export HADOOP_USER_CLASSPATH_FIRST=true
```

$pig

```
                                                    Oct 20  10:54 PM
                                                    hadoop@fedora:~/Downloads
hadoop@fedora:~/Downloads$ nano ~/.bashrc
hadoop@fedora:~/Downloads$ pig
2024-10-20 22:54:10,327 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-20 22:54:10,328 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-10-20 22:54:10,328 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-10-20 22:54:10,375 [main] INFO  org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-10-20 22:54:10,375 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/hadoop/Downloads/pig_1729445050373.log
2024-10-20 22:54:10,411 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found
2024-10-20 22:54:10,826 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 22:54:10,826 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-10-20 22:54:11,347 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 22:54:11,347 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:9001
2024-10-20 22:54:11,365 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-298b7132-efae-4283-ba6a-e26cfaa59c3f
2024-10-20 22:54:11,365 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>
```

$nano sample.txt

```
hadoop@fedora:~/pig/experiments$ cat sample.txt
1,Jeci
2,Jerin
```

$nano uppercase_udf.py

```
hadoop@fedora:~/pig/experiments$ cat 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)
```

$nano udf_example.pig

```
hadoop@fedora:~/pig/experiments$ cat udf_example.pig
-- Register the Python UDF script
REGISTER 'hdfs:///uppercase_udf.py' USING jython AS udf;
 -- Load some data
data = LOAD 'hdfs:///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:///pig_output_data';
```

$pig -f udf_example.pig

```
HadoopVersion PigVersion    UserId StartedAt      FinishedAt      Features
3.4.0   0.17.0  hadoop 2024-09-01 22:07:31   2024-09-01 22:09:27   UNKNOWN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime     MinMapTime     AvgMapTime     MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime      Alias   Feature Outputs
job_1725208407393_0002  1       0       n/a     n/a     n/a     n/a     0       0       0       0       data,uppercased_data    MAP_ONLY      hdfs:///pig_output_data,

Input(s):
Successfully read 0 records from: "hdfs:///sample.txt"

Output(s):
Successfully stored 0 records in: "hdfs:///pig_output_data"

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

Job DAG:
job_1725208407393_0002
```

hadoop@fedora:~/pig/experiments$ hdfs dfs -cat /pig_output_data/*

```
hadoop@fedora:~/pig/experiments$ hdfs dfs -cat /pig_output_data/*
1,JECI
2,JERIN
```