EX NO: 4 Roll No: 210701138

# Create UDF (User Defined Functions) in Apache Pig and execute it in MapReduce/HDFS mode

### AIM:

To create UDF (User Defined Functions) in Apache Pig and execute it in MapReduce/HDFS mode.

### **PROCEDURE:**

- 1. Install and Configure Apache Pig.
- 2. Create a python UDF (User Defined Functions).

```
C: > hadoop_pigex4 >  uppercase.py

1  @outputSchema("word:chararray")

2  def to_upper(word):

3  return word.upper()
```

- 3. Install Jython because Pig will use it to interpret the Python UDFs.
- 4. Create a Pig script that registers and uses the Python UDF.

EX NO: 4 Roll No: 210701138

5. Create Directory pigex4 and put the input files inside the created directory.

```
C:\Windows\System32>hdfs dfs -mkdir /pigudfs
C:\Windows\System32>hdfs dfs -put C:\Users\Manoj\Desktop\DA\Ex4PIG\sample.txt /pigudfs
```

6. Use the command pig -x mapreduce is used to run Apache Pig scripts in MapReduce mode

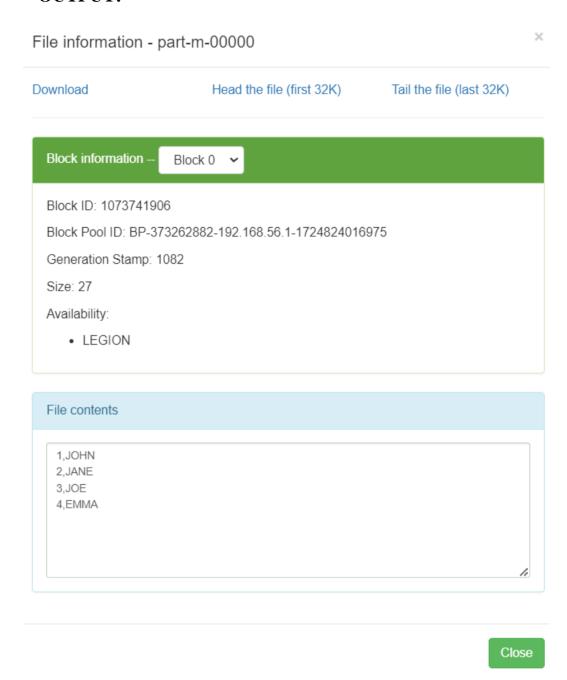
```
C:\hadoop_pigex4>pig -x mapreduce
2074-09-02 14:14:08,735 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2074-09-02 14:14:08,737 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-02 14:14:08,737 INFO pig.ExecTypeProvider: Picked MAPREDUCE
2024-09-02 14:14:09,730 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-09-02 14:14:09,200 [main] INFO org.apache.pig.Main - Logging error measages to: C:\hadoop\logs\pig_1725266649194.log
2024-09-02 14:14:09,224 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\asus.pigbootup not found
2024-09-02 14:14:09,657 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\asus.pigbootup not found
2024-09-02 14:14:09,658 [main] INFO org.apache.pig.backend.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.
2024-09-02 14:14:09,658 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:
9080
2024-09-02 14:14:10,287 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-6da9ba70-0a37-4d0a-8211-7b5a5c9fc189
2024-09-02 14:14:10,288 [main] WARN org.apache.pig.PigServer - ATS is disabled since varn.timeline-service.enabled set to false
```

7. After executing the above command you will enter the grunt shell. Here we can execute the script.pig

```
grunt> exec script.pig
3294-99-21 14:14:77,379 [main] INFO
3294-99-22 14:14:17,379 [main] INFO
3294-99-22 14:14:17,379 [main] INFO
3294-99-22 14:14:25,922 [main] INFO
3294-99-23 14:14:25,922 [main] INFO
3294-99-23 14:14:25,922 [main] INFO
3294-99-23 14:14:25,726 [main] INFO
3294-99-23 14:14:27,476 [main] INFO
3294-99-23 14:14:27,476 [main] INFO
3294-99-23 14:14:28,276 [main] INFO
3294-99-24 [main] INFO
3294-99-24 [main] INFO
3294-99-24 [main] INFO
3294-99-24 [main] INFO
3294-99-25 [main] INFO
3294-99-25 [main] INFO
3294-99-25 [main] INFO
3294-99-25 [main] INFO
3294-99-26 [main] INFO
3294-99
```

EX NO: 4 Roll No: 210701138

## **OUTPUT:**



### **RESULT:**

Thus, to create a UDF in Apache Pig and execute in MapReduce mode has been executed successfully.