EX NO: 4 Roll No: 210701153

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).

- 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.

```
C: > hadoop_pigex4 > \equiv \text{script.pig}
1   -- Register the Python UDF script
2   REGISTER 'uppercase.py' USING jython AS myudf;
3
4   -- Load the input file from HDFS
5   data = LOAD 'hdfs:///pigex4/wordeg.txt' USING PigStorage(',') AS (line: chararray);
6
7   -- Apply the UDF to convert each line to uppercase
8   uppercased_data = FOREACH data GENERATE myudf.to_upper(line);
9
10   -- Store the result in HDFS
11   STORE uppercased_data INTO 'hdfs:///pigex4/output' USING PigStorage(',');
```

EX NO: 4 Roll No: 210701153

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

```
C:\Windows\System32>hadoop fs -mkdir /pigex4
C:\Windows\System32>hadoop fs -put C:/hadoop_pigex4/wordeg.txt /pigex4
```

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

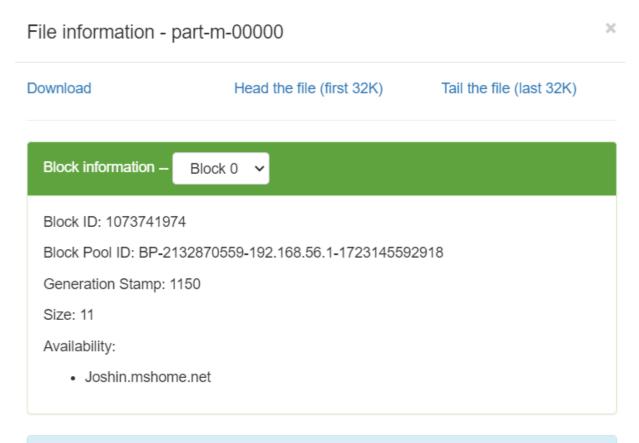
```
C:\hadoop_pigex4>pig -x mapreduce
2024-09-02 14:14:88,735 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-02 14:14:88,737 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-02 14:14:98,737 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-02 14:14:99,200 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-09-02 14:14:99,200 [main] INFO org.apache.pig.Main - Logging error messages to: C:\hadoop\logs\pig_1725266649194.log
2024-09-02 14:14:99,224 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\asus/.pigbootup not found
2024-09-02 14:14:99,658 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\asus/.pigbootup not found
2024-09-02 14:14:99,658 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file C:\Users\asus/.pigbootup not found
2024-09-02 14:14:19,9658 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:
9000
2024-09-02 14:14:10,287 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-6da9ba70-0837-4408-8211-7b5a5c9fc189
2024-09-02 14:14:10,288 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
```

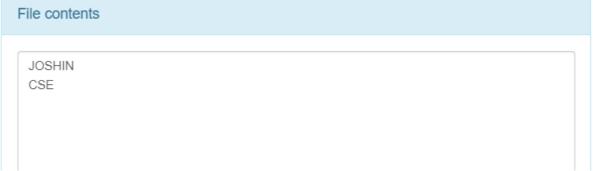
7. After executing the above command you will enter the grunt shell. Here we can execute the script.pig

```
grunt> exec script.pig
2024-09-02 14:14:17,379 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - created tmp python.cachedir=C:\Users\asus\AppData\Local\Temp\pig_j
ython_7275569481612488851
2024-09-02 14:14:25,922 [main] WARN org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.remainders is empty. This is not expected unless on t
exting.
2024-09-02 14:14:27,476 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF: myudf.to_upper
2024-09-02 14:14:28,267 [main] INFO org.apache.pig.scripting.jython.JythonFurnction - Schema 'word:chararray' defined for func to_upper
2024-09-02 14:14:28,267 [main] INFO org.apache.pig.scripting.jython.JythonFurnction - Schema 'word:chararray' defined for func to_upper
2024-09-02 14:14:28,293 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2024-09-02 14:14:28,293 [main] INFO org.apache.pig.scripting.jython.JythonFurnction - Schematuple] was not set... will not generate code.
2024-09-02 14:14:28,374 [main] INFO org.apache.pig.newlpan.logical.optimizer.cojicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, Constant
Calculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter]}
2024-09-02 14:14:28,644 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - readed tmp python.cached the post of size 699400102 to monitor. collectionUs
2024-09-02 14:14:28,644 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - pig.scripting.jethon.JythonScriptEngine - readed tmp python.cached tmp python.cached the python.cached the python cache is not expected unless org.apache.pig.scripting.jython.JythonScriptEngine - pig.scripting.jython.JythonScriptEngine - pig.scripting.jython.JythonScriptEngin
```

EX NO: 4 Roll No: 210701153

OUTPUT:





RESULT:

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