EX NO: 4 Roll No: 210701150

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.

EX NO: 4 Roll No: 210701150

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

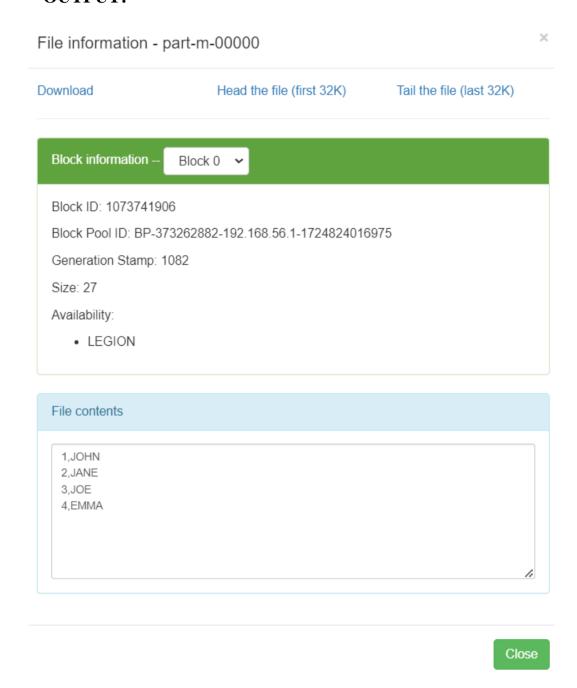
```
C:\hadoop_pigex4>pig -x mapreduce
2024-09-02 14:14:08,735 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-02 14:14:08,737 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-02 14:14:08,737 INFO pig.ExecTypeProvider: Picked MAPREDUCE s the ExecType
2024-09-02 14:14:09,200 [main] INFO org.apache.pig.Main - Apache Pig version 0:17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-09-02 14:14:09,200 [main] INFO org.apache.pig.Main - Logging error messages 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 14:14:09,658 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker
address
2024-09-02 14:14:09,658 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost
9000
2024-09-00 14:14:10,287 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-6da9ba70-8037-4d0a-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.72755694816i24080851
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
esting.
2024-09-02 14:14:27,476 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF: myudf.to_upper
2024-09-02 14:14:27,476 [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.JythonScriptEngine - Register scripting UDF: myudf.to_upper
2024-09-02 14:14:28,267 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapred
uce.output.textoutputformat.separator
2024-09-02 14:14:12:28,393 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2024-09-02 14:14:12:393 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2024-09-02 14:14:12:39,144 [main] INFO org.apache.pig.nephan.logical.potmizer.pimizer - Regueen_Independent of the complex of the c
```

EX NO: 4 Roll No: 210701150

OUTPUT:



RESULT:

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