Ex:4   Create UDF (User Defined Functions) in Apache Pig and execute it inMapReduce/HDFS mode

PROCEDURE:

1. Ensure that Apache Pig is installed and configured.

```
C:\hadoop\sbin>start-all
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\hadoop\sbin>pig --version
Apache Pig version 0.17.0 (r1797386)
compiled Jun 02 2017, 15:41:58
```

2. Create a text file for input that contains "data analytics pig experiment".

3. Create a python UDF (User Defined Functions).

```python
@outputSchema("word:chararray")
def to_upper(word):
    if word is None:
        return None
    return word.upper()
```

4.Jython should be installed as Pig will use it to interpret the Python UDFs.

5.Create directory in hdfs for executing udf function using pig.

```
C:\hadoop\sbin>hadoop fs -mkdir /samplePig

C:\hadoop\sbin>hadoop fs -put D:\Data_Analytics\apachePig\input.txt /samplePig

C:\hadoop\sbin>hadoop fs -put D:\Data_Analytics\apachePig\convert_to_upper.pig /samplePig
```

6.Create a Pig script that registers and uses the Python UDF.

```
-- Register the Jython standalone JAR using the correct URI format
REGISTER 'file:///C:/jython-standalone-2.7.4.jar';

-- Register the Python UDF script
REGISTER 'D:/Data_Analytics/apachePig/my_udf.py' USING jython AS myudfs;

-- Load the input file from HDFS
data = LOAD 'hdfs://localhost:9000/samplePig/input.txt' AS (line: chararray);

-- Apply the UDF to convert each line to uppercase
uppercased_data = FOREACH data GENERATE myudfs.to_upper(line);

-- Store the result in HDFS
STORE uppercased_data INTO 'hdfs://localhost:9000/samplePig/output';
```

7.Execute the Pig Script in MapReduce Mode using the command:

```
C:\hadoop>pig -x mapreduce hdfs://localhost:9000/samplePig/convert_to_upper.pig
2024-08-30 17:32:35,942 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-08-30 17:32:35,943 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-08-30 17:32:35,943 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
```

OUTPUT:

File contents

DATA ANALYTICS PIG EXPERIMENT

RESULT:

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