EX NO: 4   **Create UDF in PIG**

Aim:

To create User Defined Function in Apache Pig and execute it on map reduce.

Procedure:

Step 1:

Create a file named "pig_udf_text.txt" and populate it with the data.

1,hello

2,apache

3,pig

4,user

Step 2:

Create program **uppercase_udf.py**

```python
def uppercase(text):

    return text.upper()

if __name__ == "__main__":

    import sys

    for line in sys.stdin:

        line = line.strip()

        result = uppercase(line)

        print(result)
```

Create a pig file

```
REGISTER 'hdfs:///user/upperr/udf.jar';

DEFINE ToUpperCase udf();

data = LOAD 'hdfs:///user/upperr/data.txt' AS (name:chararray);

uppercased_data = FOREACH data GENERATE ToUpperCase(name) AS uppercased_name;

STORE uppercased_data INTO 'hdfs:///user/upperr/output_data';
```

Step 3:

Put python file and text file inside the hadoop directory.

/user/Admin/home/hadoop/pig_input_data

Step 4:

Run the pig file

pig –f example.pig

After executing the pig file, success message will be displayed.

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

```
C:\Users\hp>jps
11040 ResourceManager
20176 DataNode
5304 NameNode
8904 NodeManager
11244 Jps
```

Step 5:

Check the output in the output directory created.

```
C:\Users\hp>hadoop fs -cat /user/Admin/home/hadoop/pig_output_data/part-m-00000
1,HELLO
2,APACHE
3,PIG
4,USER
```

Result:

Thus the user defined program in pig was executed successfully.