Exercise : 04 210701134

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

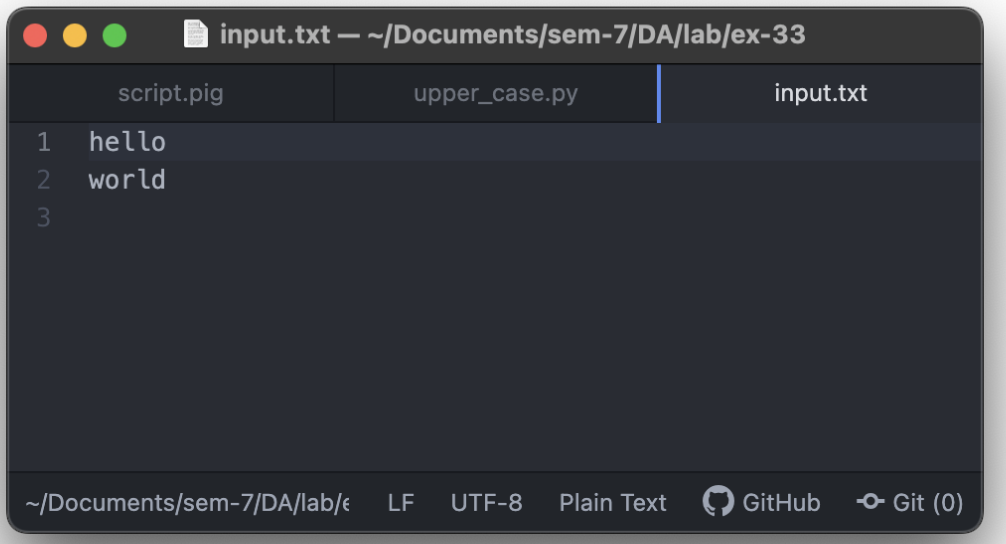
Aim:

To create user defined functions in Apache Pig and execute it in MapReduce / HDFS mode.

Procedure:

1. Write the Python UDF : Created a Python function that reverses strings.
2. Register the UDF in Pig : Registered the Python script in the Pig script using the REGISTER command.
3. Use the UDF in Pig Script : Applied the UDF to the data set using FOREACH ... GENERATE.
4. Execute the Pig Script : Ran the Pig script in Map Reduce mode to process the data on HDFS.

Output:



upper_case.py — ~/Documents/sem-7/DA/lab/ex-33

script.pig | upper_case.py | input.txt

```python
#!/usr/bin/python3
@outputSchema("word:chararray")
def to_upper(word):
    return word.upper()

```

~/Documents/sem-7/DA/lab/ex- LF UTF-8 Python GitHub Git (0)

script.pig — ~/Documents/sem-7/DA/lab/ex-33

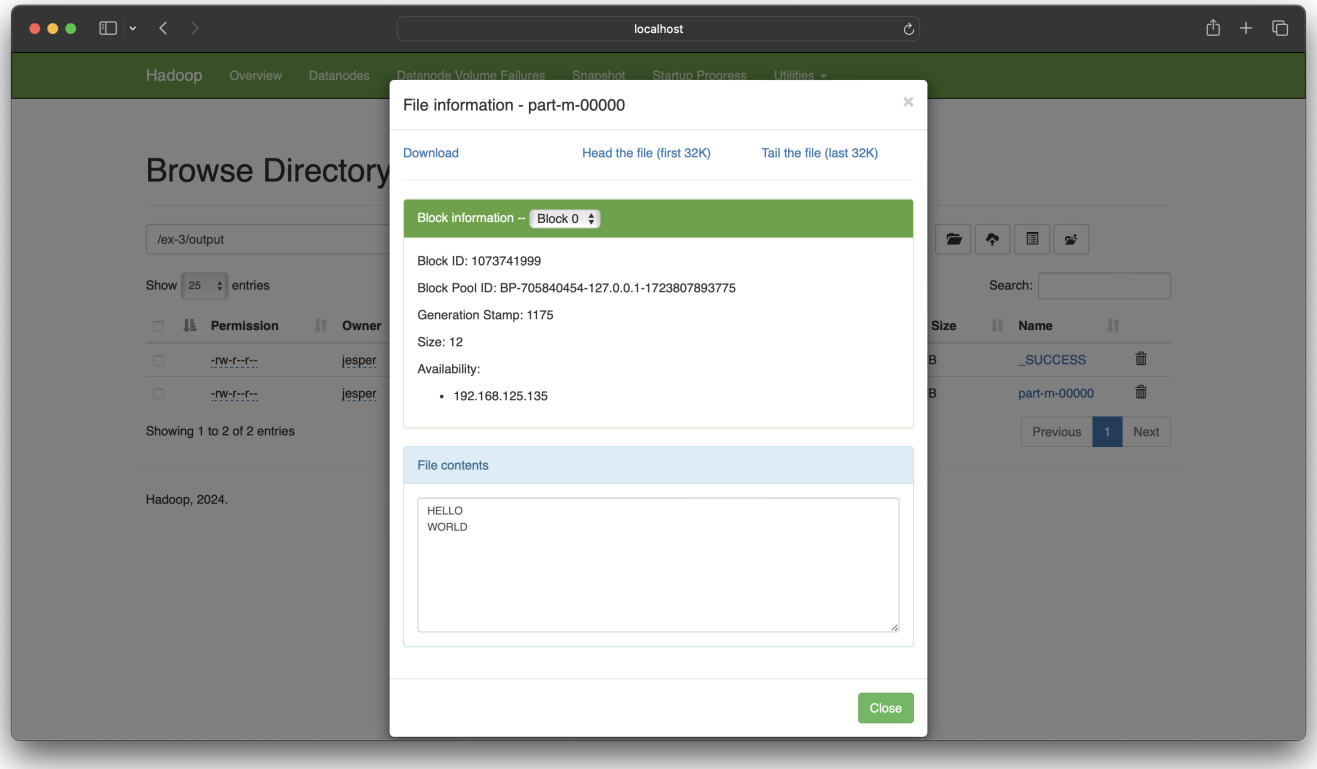
script.pig | upper_case.py | input.txt

```
REGISTER 'upper_case.py' USING jython as myfuncs;

data = LOAD '/ex-3/input.txt' AS (word:chararray);

upper_data = FOREACH data GENERATE myfuncs.to_upper(word);

STORE upper_data INTO '/ex-3/output';

```

~/Documents/sem-7/DA/lab/ex-33/scrip LF UTF-8 Plain Text GitHub Git (0)

ex-33 — -zsh — 166×27

```
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias
  Feature  Outputs
job_1725606640689_0002  1       0       n/a     n/a     n/a     n/a     0       0       0       0       data,upper_data MAP_ONLY        /ex-3/output,

Input(s):
Successfully read 0 records from: "/ex-3/input.txt"

Output(s):
Successfully stored 0 records in: "/ex-3/output"

Counters:
Total records written : 0
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1725606640689_0002


2024-09-06 12:51:15,957 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to get job related diagnostics
2024-09-06 12:51:46,467 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggreg
ation.
2024-09-06 12:51:46,467 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-09-06 12:51:46,497 [main] INFO  org.apache.pig.Main - Pig script completed in 2 minutes, 56 seconds and 321 milliseconds (176321 ms)
jesper@j ex-33 %
```



Result:

Thus the Installation, Configuration and run Hadoop and HDFS is successfully executed.