EXP NO: 4

CREATE UDF IN PIG

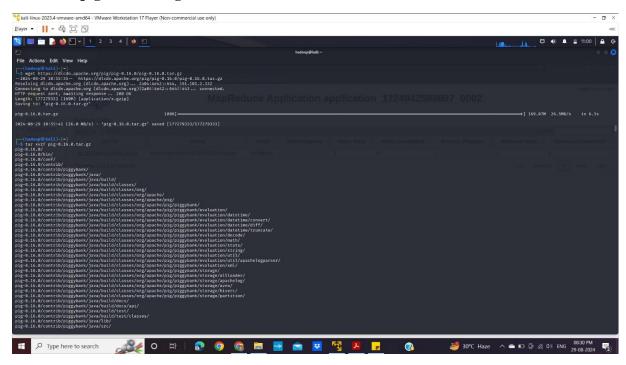
\$start-all.sh

\$ jps

```
(hadoop® kali)-[~]
$ jps
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13397 DataNode
13471 NameNode
```

\$wget https://dlcdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz

\$ tar xvzf pig-0.16.0.tar.gz

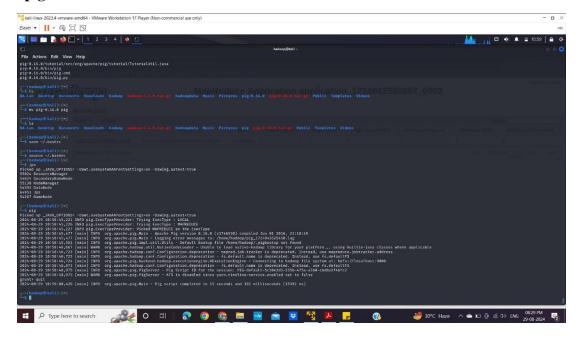


\$nano ~/.bashrc

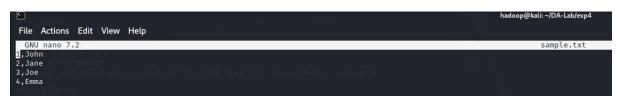
```
#PIG settings
export PIG_HOME=/home/hadoop/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_HOME/etc/hadoop/
export PIG_CONF_DIR=$PIG_HOME/conf
#export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
#PIG setting ends
```

\$mv pig-0.16.0 pig

\$pig



\$cd DA-Lab \$mkdir exp4 \$cd exp4 \$nano sample.txt



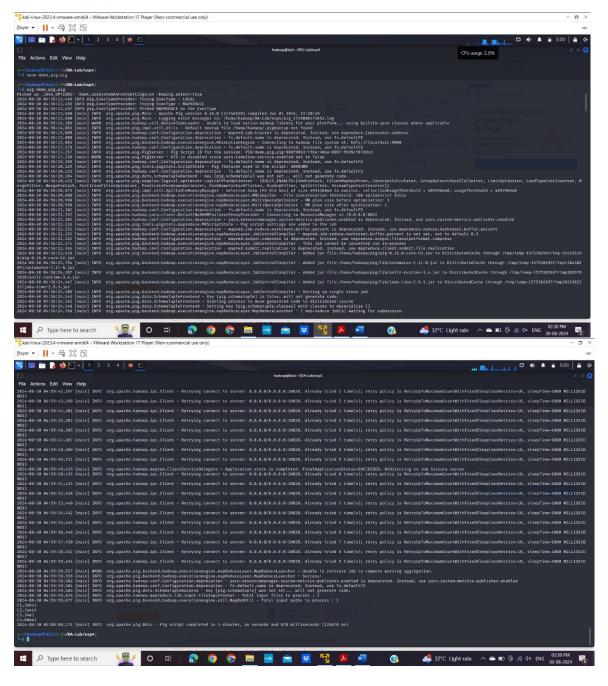
\$nano demo_pig.pig



\$hdfs dfs -mkdir /exp4

\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/sample.txt /exp4

\$pig demo_pig.pig



\$nano uppercase_udf.py

```
File Actions Edit View Help

GNU nano 7.2

def uppercase(text):
    return text.upper()
    if __nane__ = "__main__":
        import sys
    for line in sys.stdin:
        line = line.strip()
        result = uppercase(line)
        print(result)
```

\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/uppercase_udf.py /exp4

```
(hadoop® kali)-[~/hadoop/bin]
$ ./hdfs dfs -ls /exp4
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:26:01,736 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform ...
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2024-08-30 05:07 /exp4/output
-rw-r--r- 1 hadoop supergroup 27 2024-08-30 04:43 /exp4/sample.txt
-rw-r--r- 1 hadoop supergroup 172 2024-08-30 05:02 /exp4/uppercase_udf.py
```

\$nano udf_example.pig

```
File Actions Edit View Help

GNU nano 7.2

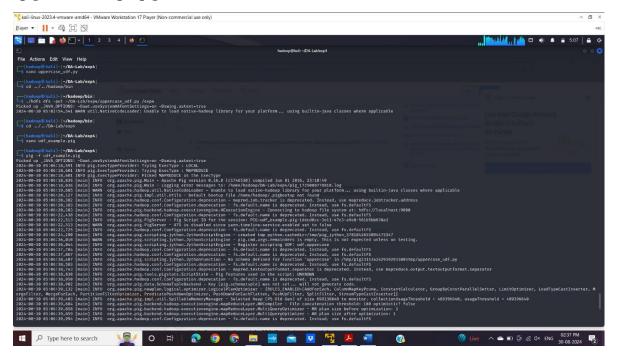
-- Register the Python UDF script
REGISTER 'hdfs:///exp4/uppercase_udf.py' USING jython AS udf;

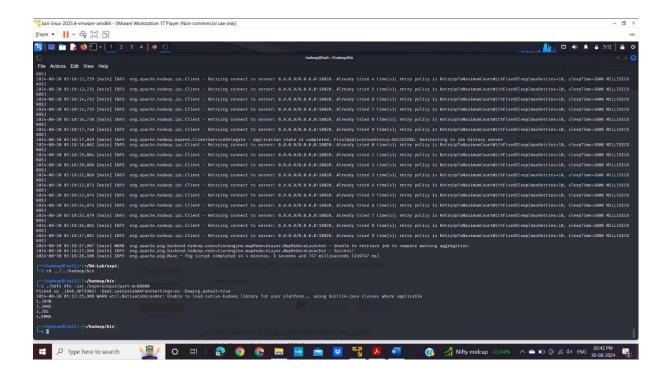
-- Load some data
data = LOAD 'hdfs:///exp4/sample.txt' AS (text:chararray);

-- Use the Python UDF
uppercased_data = FOREACH data GENERATE udf.uppercase(text) AS uppercase_text;

-- Store the result
STORE uppercased_data INTO 'hdfs:///exp4/output';
```

\$pig -f udf_example.pig





\$hdfs dfs -cat /exp4/output/*

```
(hadoop® kali)-[~/hadoop/bin]

$ ./hdfs dfs -cat /exp4/output/*
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:33:32,731 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform ...
1,JOHN
2,JANE
3,JOE
4,EMMA
```