

EXP NO: 4

CREATE UDF IN PIG

\$start-all.sh

```
File Actions Edit View Help
hadoop@kali: ~
hadoop@kali ~$ sudo start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [kali]
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-11 04:59:16,479 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers
```

\$ jps

```

[ubuntu@kali:~]$ java
Picked up _JAVA_OPTIONS: -Dwt.useSystemAAFontSettings-on -Dswing.aatext=true
14436 NodeManager
14772 Jps
13636 SecondaryNameNode
14311 ResourceManager
13597 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
```

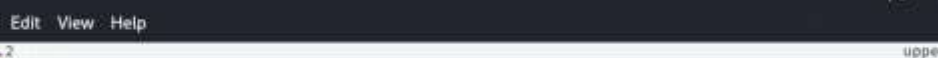
[illegible]

\$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
```


\$pig demo_pig.pig

\$nano uppercase udf.py



```
File Actions Edit View Help
GNU nano 7.2                                     uppercase_udf.py
def uppercase(text):
    return text.upper()
if __name__ == "__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

```

-- 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

```

WSL: Ubuntu-22.04
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2 udf_example.pig
-- 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';

```

The screenshot shows a Windows 10 desktop with a web browser window open. The browser address bar displays 'http://localhost:3000'. The main content area of the browser shows a log file viewer interface. The log file viewer has a search bar at the top and a list of log entries below. The log entries are organized into columns: 'Time', 'Level', 'Message', and 'Source'. The log entries show various messages from the 'http://localhost:3000' source, including 'GET / HTTP/1.1 200 OK' and 'POST / HTTP/1.1 200 OK'. The log viewer also shows a search bar and a list of log files.

```
$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
```