

**Ex No 4****CreateUDF(UserDefinedFunctions)inApachePigandexecuteitin  
MapReduce / HDFS mode****AIM:**

To create UDF in Apache Pig and execute it in MapReduce/HDFS mode.

**PROCEDURE:****1. Install Apache Pig****Download Pig:**

1. Download Pig from the Apache Pig download page:

Link: [Apache Pig 0.17.0 Download](#)

Extract the downloaded file (assuming you downloaded `pig-0.17.0.tar.gz`): `tar -`

```
xzf pig-0.17.0.tar.gz
```

Movetheextractedfolderto adirectory, such as `/usr/local/`: `sudo`

```
mv pig-0.17.0 /usr/local/pig
```

**2. Set Up Environment Variables for Pig**

Edit your `~/.profile` or `~/.zshrc` to include Pig in the `PATH`.

```
nano ~/.zshrc
```

Add the following lines:

```
export PIG_HOME=/usr/local/pig
export PATH=$PIG_HOME/bin:$PATH
```

Apply the changes:

```
source ~/.zshrc
```

**3. Verify Pig Installation**

Run the following command to check if Pig is installed correctly:

```
pig-x local
```

You should see the Pig Grunt shell prompt:

```
grunt>
```

Type `quit` to exit the shell.

## 4. Start Hadoop Services

Make sure your Hadoop is up and running. Start the required services:

```
cd /usr/local/hadoop/sbin  
./start-dfs.sh  
./start-yarn.sh
```

## 5. Prepare Input Data(ex4.txt)

Create a sample text file for testing the UDF, named `ex4.txt`: `nano`

```
ex4.txt
```

Example content:

```
1,John  
2,Soniya  
3,Vijay  
4,Sonu
```

Upload the file to HDFS:

```
hdfsdfs-mkdir /UDF  
hdfsdfs-put ex4.txt/UDF/
```

## 6. Create UDF in Python

Now, you need to write your Python UDF.

Create a Python file `uppercase_udf.py`:

```
nano uppercase_udf.py
```

Add the following code to `uppercase_udf.py`:

```
#!/usr/bin/python3
defuppercase(text):
    returntext.upper()

if __name__== "__main__":
    import sys
    for line in sys.stdin:
        line=line.strip()
        result=uppercase(line)
        print(result)
```

UploadthePythonUDFtoHDFS:

```
hdfsdfs-mkdir/UDF/udfs
hdfsdfs-putuppercase_udf.py/UDF/udfs/
```

MakesurethefileisinthecorrectHDFSdirectorybyrunning:

```
hdfsdfs-ls/UDF/udfs
```

## 7. WritePigScript(UDF.pig)

CreateaPigscripttoapplyyourUDF. Create

UDF.pig:

```
nanoUDF.pig
```

AddthefollowingPigscripttoUDF.pig:

```
--RegistertheUDF
REGISTERhdfs:///UDF/udfs/uppercase_udf.pyUSINGjythonASmyudfs;

--Loadtheex4.txtfilefrom HDFS
data=LOAD'hdfs:///UDF/ex4.txt'USINGPigStorage(',')AS
(id:int,name:chararray);

--ApplytheUDFtoeachline
uppercase_data=FOREACHdataGENERATEmyudfs.uppercase(name)AS upper_line;
```

```
--StoretheresultinHDFS
STOREuppercase_dataINTO'hdfs:///UDF/output'USINGPigStorage(',');
```

Savethefileandexit.

## 8. RunthePigScriptinMapReduce Mode

Nowthateverythingissetup,executethePigscriptinMapReduce mode:

```
hdfsdfs-chmod755/UDF/udfs/uppercase_udf.py hdfs
dfs -chmod 755 /UDF
hdfsdfs-chmod755/UDF/ex4.txt
```

```
pig-xmapreduceUDF.pig
```

## 9. ChecktheOutput

Afterthejobfinishes,youcanviewtheoutputinHDFS. List the

output directory:

```
hdfsdfs-ls/UDF/output
```

You should see something like:

```
Found 1 items
-rw-r--r--  3 user group      1232024-09-11 12:00
/UDF/output/part-m-00000
```

View the output file:

```
hdfsdfs-cat/UDF/output/part-m-00000
```

You should see the content in uppercase

**OUTPUT:**

```

Last login: Tue Sep 10 20:08:42 on ttys002
nativewit@Nativewits-MacBook-Air ~ %
cd /usr/local/Cellar/hadoop/3.4.0/libexec/sbin
nativewit@Nativewits-MacBook-Air sbin % ./start-dfs.sh

Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Nativewits-MacBook-Air.local]
2024-09-10 20:35:20,894 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % ./start-yarn.sh

Starting ResourceManager
Starting NodeManagers
nativewit@Nativewits-MacBook-Air sbin % nano ex4.txt

nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -mkdir /UDF

2024-09-10 20:36:25,288 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -put ex4.txt /UDF/

2024-09-10 20:36:31,380 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % nano uppercase_udf.py

nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -mkdir /UDF/udfs

2024-09-10 20:37:00,810 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -put uppercase_udf.py /UDF/udfs/

2024-09-10 20:37:06,897 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -ls /UDF/udfs

2024-09-10 20:37:12,402 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 nativewit supergroup 219 2024-09-10 20:37 /UDF/udfs/uppercase_udf.py
nativewit@Nativewits-MacBook-Air sbin % nano UDF.pig

nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -chmod 755 /UDF/udfs/uppercase_udf.py

2024-09-10 20:38:16,221 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -chmod 755 /UDF

2024-09-10 20:38:20,614 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % hdfs dfs -chmod 755 /UDF/ex4.txt

2024-09-10 20:38:25,425 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewit@Nativewits-MacBook-Air sbin % pig -x mapreduce UDF.pig

2024-09-10 20:38:30,793 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-10 20:38:36,795 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-10 20:38:36,795 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType

```

**RESULT:**

Thus,UDFinApachePighasbeencreatedandexecutedinMapReduce/HDFSmode successfully.