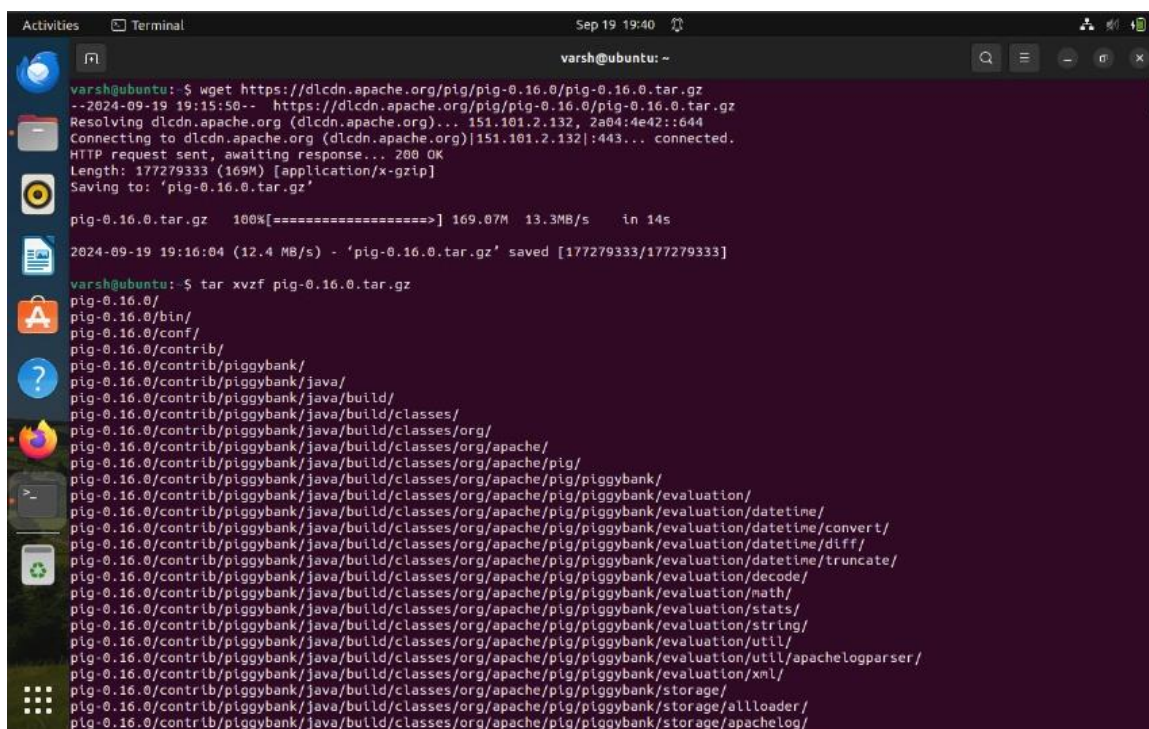


**Exp.No.: 4****Create UDF in PIG****Step-by-step installation of Apache Pig on Hadoop cluster on Ubuntu Pre-requisite:**

1. Ubuntu 16.04 or higher version running (I have installed Ubuntu on Oracle VM (Virtual Machine) VirtualBox),
2. Run Hadoop on ubuntu (I have installed Hadoop 3.2.1 on Ubuntu 16.04). You may refer to my blog “How to install Hadoop installation” click [here](#) for Hadoop installation).

**Pig installation steps****Step 1: Login into Ubuntu**


```

varsh@ubuntu: ~
--2024-09-19 19:15:50-- https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
Resolving d1cdn.apache.org (d1cdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to d1cdn.apache.org (d1cdn.apache.org)[151.101.2.132]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 177279333 (169M) [application/x-gzip]
Saving to: 'pig-0.16.0.tar.gz'

pig-0.16.0.tar.gz 100%[=====] 169.07M 13.3MB/s in 14s

2024-09-19 19:16:04 (12.4 MB/s) - 'pig-0.16.0.tar.gz' saved [177279333/177279333]

varsh@ubuntu: ~$ tar xvfz pig-0.16.0.tar.gz
pig-0.16.0/
pig-0.16.0/bin/
pig-0.16.0/conf/
pig-0.16.0/contrib/
pig-0.16.0/contrib/piggybank/
pig-0.16.0/contrib/piggybank/java/
pig-0.16.0/contrib/piggybank/java/build/
pig-0.16.0/contrib/piggybank/java/build/classes/
pig-0.16.0/contrib/piggybank/java/build/classes/org/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/convert/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/diff/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/truncate/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/decode/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/math/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/stats/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/string/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/uttl/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/uttl/apache/logparser/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/xml/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/allloader/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/apache/log/

```

**Step 2:** Go to <https://pig.apache.org/releases.html> and copy the path of the latest version of pig that you want to install. Run the following command to download Apache Pig in Ubuntu:

```
$ wget https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
```

**Step 3:** To untar pig-0.16.0.tar.gz file run the following command:

```
$ tar xvfz pig-0.16.0.tar.gz
```

**Step 4:** To create a pig folder and move pig-0.16.0 to the pig folder, execute the following command:

```
$ sudo mv /home/hadoop/pig-0.16.0 /home/hadoop/pig
```

**Step 5:** Now open the .bashrc file to edit the path and variables/settings for pig. Run the following command:

```
$ sudo nano .bashrc
```

Add the below given to .bashrc file at the end and save the file.

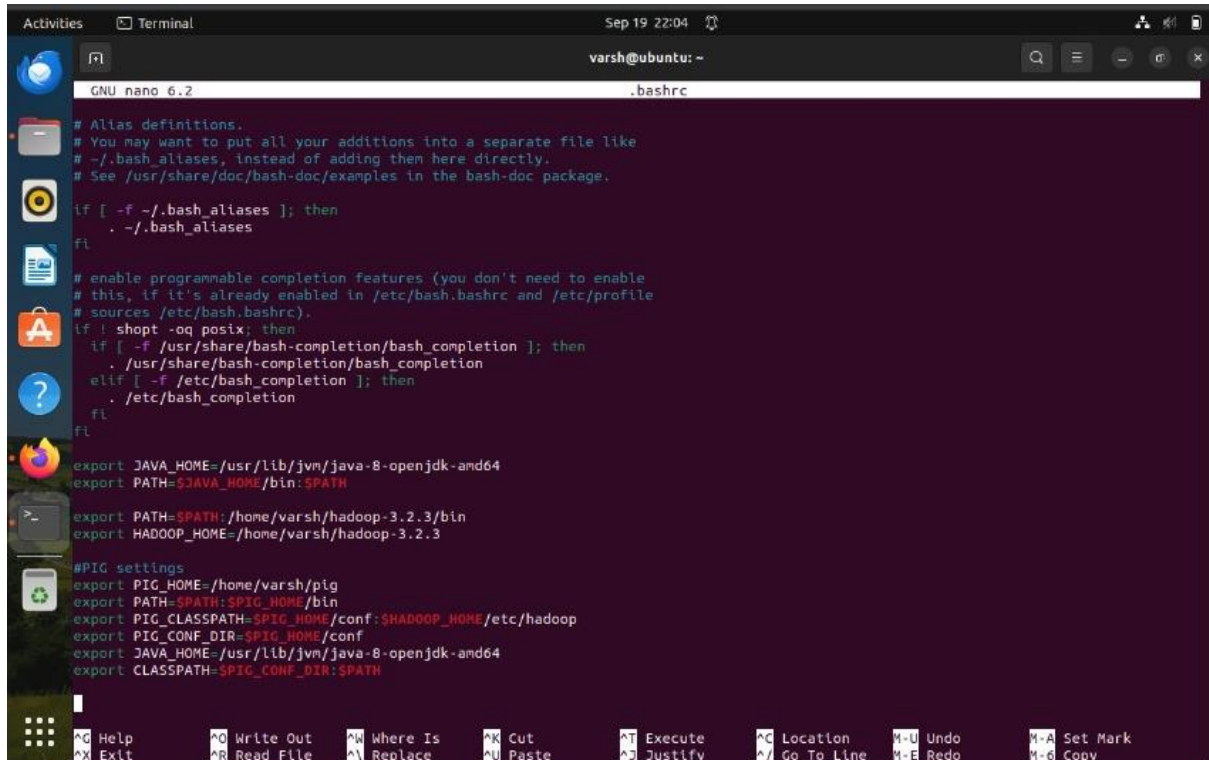
```
#PIG settingsexport PIG_HOME=/home/hadoop/pigexport
```

```
PATH=$PATH:$PIG_HOME/binexport
```

```
PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_INSTALL/etc/hadoop/export
```

```
PIG_CONF_DIR=$PIG_HOME/confexport JAVA_HOME=/usr/lib/jvm/java-8-
```

```
openjdkamd64export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH#PIG setting ends
```



```

GNU nano 6.2 .bashrc
# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    elif [ -f /etc/bash_completion ]; then
        . /etc/bash_completion
    fi
fi

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$JAVA_HOME/bin:$PATH

export PATH=$PATH:/home/varsh/hadoop-3.2.3/bin
export HADOOP_HOME=/home/varsh/hadoop-3.2.3

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

**Step 6:** Run the following command to make the changes effective in the .bashrc file:

```
$ source .bashrc
```

**Step 7:** To start all Hadoop daemons, navigate to the hadoop-3.2.1/sbin folder and run the following commands:

```
$ ./start-dfs.sh$ ./start-yarn$ jps
```

```

Activities  Terminal  Sep 19 22:03
varsh@ubuntu: ~
WARNING: Stopping all Apache Hadoop daemons as varsh in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
Stopping nodemanagers
Stopping resourcemanager
varsh@ubuntu: ~$ /home/varsh/hadoop-3.2.3/sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as varsh 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 [ubuntu]
Starting resourcemanager
Starting nodemanagers
varsh@ubuntu: ~$ jps
6656 NameNode
7464 Jps
7177 ResourceManager
6972 SecondaryNameNode
6780 DataNode
7295 NodeManager

```

**Step 8:** Now you can launch pig by executing the following command: \$ pig

```

varsh@ubuntu: ~$ pig
2024-09-19 22:01:59,329 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-19 22:01:59,338 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-19 22:01:59,330 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-19 22:01:59,500 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-19 22:01:59,500 [main] INFO org.apache.pig.Main - Logging error messages to: /home/varsh/pig_1726763519421.log
2024-09-19 22:01:59,624 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/varsh/.pigbootup not found
2024-09-19 22:02:00,881 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
se mapreduce.jobtracker.address
2024-09-19 22:02:00,883 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-19 22:02:00,884 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file syst
em at: hdfs://localhost:9000
2024-09-19 22:02:04,542 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-702d16fd-0dce-41b9-b887-9
8c3dab11d1a
2024-09-19 22:02:04,546 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt>

```

**Step 9:** Now you are in pig and can perform your desired tasks on pig. You can come out of the pig by the quit command:

>quit;

## CREATE USER DEFINED FUNCTION(UDF)

**Aim :**

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

### PROCEDURE:

#### Create a sample text file

hadoop@Ubuntu:~/Documents\$ nano sample.txt

Paste the below content to sample.txt

- 1,John
- 2,Jane
- 3,Joe
- 4,Emma

hadoop@Ubuntu:~/Documents\$ hadoop fs -put sample.txt /home/hadoop/piginput/

#### Create PIG File

hadoop@Ubuntu:~/Documents\$ nano demo\_pig.pig

paste the below the content to demo\_pig.pig

-- Load the data from HDFS



data = LOAD '/home/hadoop/piginput/sample.txt' USING PigStorage(',') AS (id:int>

-- Dump the data to check if it was loaded correctly

DUMP data;

----- Run the above file

hadoop@Ubuntu:~/Documents\$ pig demo\_pig.pig

```

varsh@ubuntu:~$ pig demo_pig.pig
2024-09-20 00:04:56,649 [main] INFO org.apache.pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-20 00:04:56,666 [main] INFO org.apache.pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-20 00:04:56,668 [main] INFO org.apache.pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-20 00:04:56,762 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-20 00:04:56,771 [main] INFO org.apache.pig.Main - Logging error messages to: /home/varsh/pig_1726770896732.log
2024-09-20 00:04:57,514 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/varsh/.pigbootstrap not found
2024-09-20 00:04:57,647 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
se mapreduce.jobtracker.address
2024-09-20 00:04:57,658 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:04:57,679 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file syst
em at: hdfs://localhost:9000
2024-09-20 00:04:58,611 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-demo_pig.pig-6da2bf29-1628-4cc6-a
f94-e1451d6fef77
2024-09-20 00:04:58,613 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-09-20 00:04:59,556 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:04:59,958 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2024-09-20 00:05:00,031 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:05:00,090 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate
code.
2024-09-20 00:05:00,237 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, Col
umnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, Par
titionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2024-09-20 00:05:00,350 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192
to monitor. collectionUsageThreshold = 489580128, usageThreshold = 489580128
2024-09-20 00:05:00,504 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MRCompiler - File concatenation th
reshold: 100 optimistic? false
2024-09-20 00:05:00,549 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MultiQueryOptimizer - MR plan size
before optimization: 1
2024-09-20 00:05:00,550 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MultiQueryOptimizer - MR plan size
after optimization: 1
2024-09-20 00:05:00,655 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:05:00,762 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2024-09-20 00:05:01,074 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publishe
r.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-09-20 00:05:01,163 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the jo

```

Create udf file and save as uppercase\_udf.py

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)
```

----- Create the udfs folder on hadoop

hadoop@Ubuntu:~/Documents\$ hadoop fs -mkdir /home/hadoop/udfs

put the uppercase\_udf.py in to the abv folder

hadoop@Ubuntu:~/Documents\$ hdfs dfs -put uppercase\_udf.py /home/hadoop/udfs/

hadoop@Ubuntu:~/Documents\$ nano udf\_example.pig copy and paste the below content on udf\_example.pig

-- Register the Python UDF script

REGISTER 'hdfs:///home/hadoop/udfs/uppercase\_udf.py' USING jython AS udf;

```
-- Load some data
data = LOAD 'hdfs:///home/hadoop/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:///home/hadoop/pig_output_data';
```

### place sample.txt file on hadoop

```
hadoop@Ubuntu:~/Documents$ hadoop fs -put sample.txt /home/hadoop/
```

### To Run the pig file

```
hadoop@Ubuntu:~/Documents$ pig -f udf_example.pig
```

```
onstatus-SUCCESS. Redirecting to job history server
2024-09-20 00:08:10,946 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:11,957 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:12,960 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:13,963 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
3 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:14,971 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
4 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:15,975 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
5 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:16,976 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
6 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:17,979 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
7 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:18,988 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
8 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:19,991 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried
9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-09-20 00:08:20,095 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retr
ieve job to compute warning aggregation.
2024-09-20 00:08:20,096 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-09-20 00:08:20,106 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publishe
r.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-09-20 00:08:20,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:08:20,113 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate
code.
2024-09-20 00:08:20,214 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-09-20 00:08:20,215 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process :
1
(1,John)
(2,Jane)
(3,Joe)
(4,Emma)
2024-09-20 00:08:20,730 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 24 seconds and 41 milliseconds (204041
ms)
varsh@ubuntu: ~
```

### To check the output file is created

```
hadoop@Ubuntu:~/Documents$ hdfs dfs -ls /home/hadoop/pig_output_data
Found 2 items
```

If you need to examine the files in the output folder, use:

To view the output

```
hadoop@Ubuntu:~/Documents$ hdfs dfs -cat /home/hadoop/pig_output_data/part-m000000
```

```

Activities Terminal Sep 20 00:37
varsh@ubuntu: ~
(3,Joe)
(4,Emma)
2024-09-20 00:00:20,730 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 24 seconds and 41 milliseconds (204041ms)
varsh@ubuntu: $ hdfs dfs -mkdir /home/varsh/hadoop/udfs
varsh@ubuntu: $ touch uppercase_udf.py
varsh@ubuntu: $ hdfs dfs -put uppercase_udf.py /home/varsh/hadoop/udfs/
varsh@ubuntu: $ nano udf_example.pig
varsh@ubuntu: $ hdfs dfs -put udf.txt /home/varsh/hadoop/
varsh@ubuntu: $ pig -f udf_example.pig
2024-09-20 00:29:01,952 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-09-20 00:29:01,955 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-09-20 00:29:01,955 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-09-20 00:29:02,059 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-09-20 00:29:02,059 [main] INFO org.apache.pig.Main - Logging error messages to: /home/varsh/pig_1726772342046.log
2024-09-20 00:29:02,809 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/varsh/.pigbootup not found
2024-09-20 00:29:02,809 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
se.mapreduce.job.tracker.address
2024-09-20 00:29:02,809 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:29:02,809 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file syst
em at: hdfs://localhost:9000
2024-09-20 00:29:03,783 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf_example.pig-a171f53e-ad36-445
1-a605-b2ba714a20e
2024-09-20 00:29:03,784 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-09-20 00:29:03,909 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:29:04,515 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - created tmp python.cachedir=/tmp/pig_jytho
n_2660556237302379217
2024-09-20 00:29:15,780 [main] WARN org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.remaining is empty. This is
not expected unless on testing.
2024-09-20 00:29:15,795 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - Register scripting UDF: udf.uppercase
2024-09-20 00:29:16,463 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:29:16,482 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS
2024-09-20 00:29:16,953 [main] INFO org.apache.pig.scripting.jython.JythonFunction - No schema defined for function 'uppercase' in
/tmp/pig1021296038017057373tmp/uppercase_udf.py
2024-09-20 00:29:17,135 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use
fs.defaultFS

```

```

varsh@ubuntu: $ hdfs dfs -ls /home/varsh/hadoop/pig_output_data
Found 2 items
-rw-r--r-- 3 varsh supergroup 0 2024-09-20 00:30 /home/varsh/hadoop/pig_output_data/_SUCCESS
-rw-r--r-- 3 varsh supergroup 27 2024-09-20 00:30 /home/varsh/hadoop/pig_output_data/part-m-00000
varsh@ubuntu: $ hdfs dfs -cat /home/varsh/hadoop/pig_output_data/part-m-00000
1,JOHN
2,JANE
3,JOE
4,EMMA
varsh@ubuntu: $

```

## Result:

Thus the program to create User Define Function in Apache Pig and execute it on map reduce has been done successfully.