**EXP 4: Create UDF (User Defined Functions) in Apache Pig and execute it in MapReduce / HDFS mode**

# AIM:

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

# PROCEDURE:

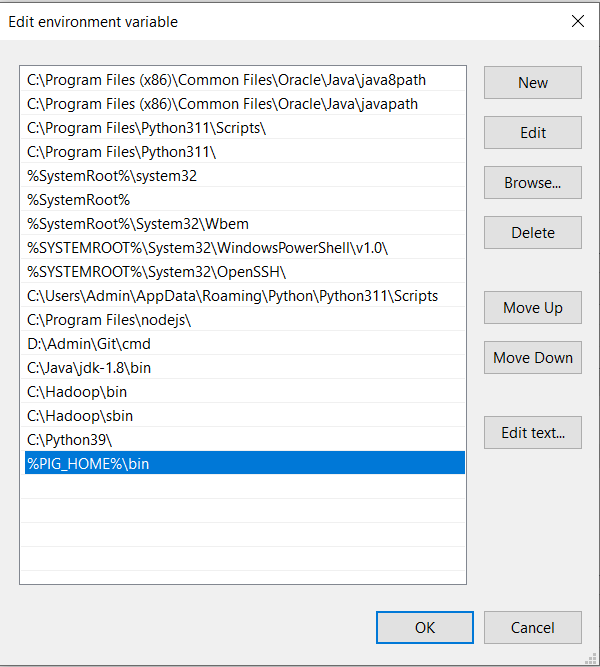
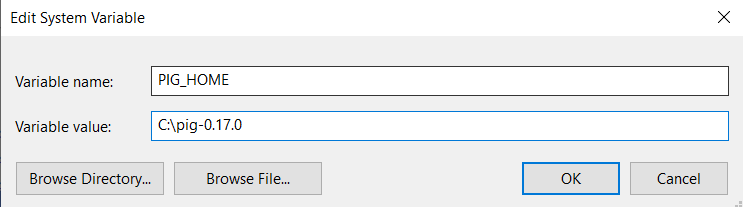
**Pig Download and installation:**

1. **Download Pig:**

Download Pig from “https://downloads.apache.org/pig/pig-0.17.0/”



1. **Add the environment variable for Pig:**



1. Go to C:\pig-0.17.0\bin and open pig (Windows Command Script)



1. Open Windows Powershell and type “pig –x local” and check whether pig grunt appears.

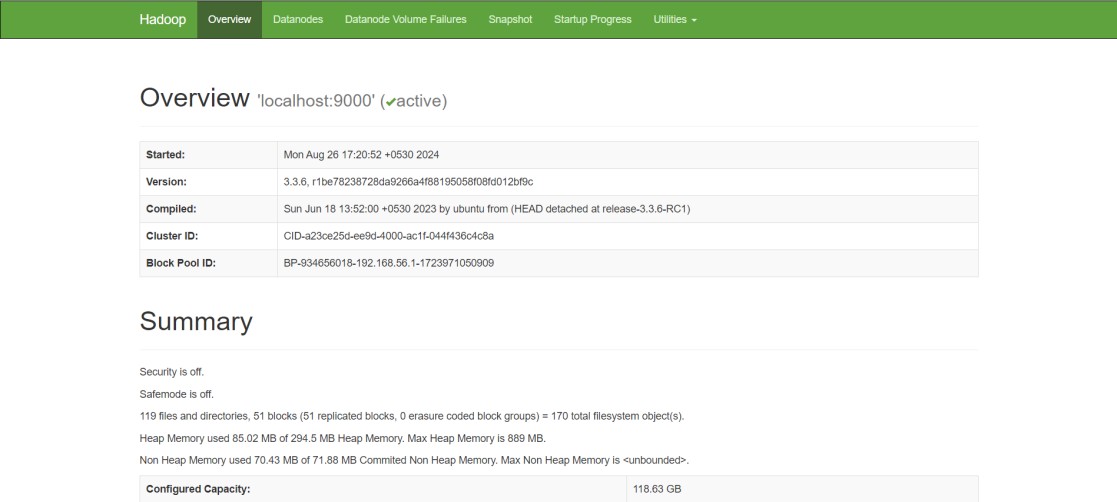
# Pig is successfully installed.

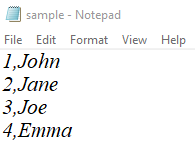
**Create UDF:**

# Start Hadoop services:

Open command prompt as an administrator start-dfs.cmd

start-yarn.cmd

1. Open the browser and go to the URL “localhost:9870”
2. Create a text file “sample.txt”:



1. Create a Directory in HDFS and copy the Input File to HDFS

hdfs dfs –mkdir /UDF

hadoop fs -put C:/Users/user/Documents/Pig/sample.txt /UDF





1. Create a Python file “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)

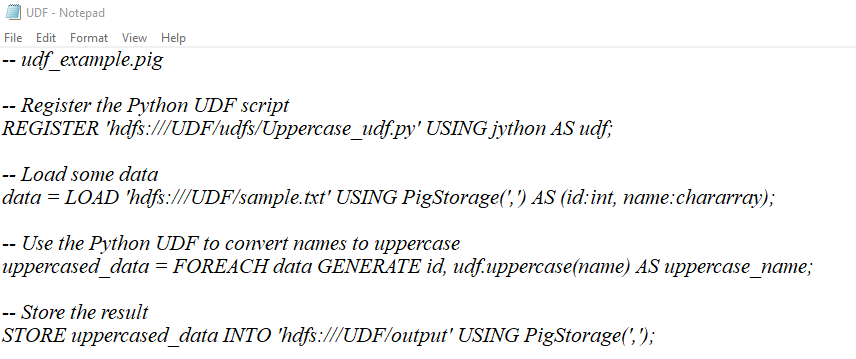
1. Create a Directory in HDFS and copy the Input File to HDFS hdfs dfs –mkdir /UDF/udfs

hadoop fs -put C:/Users/user/Documents/Pig/Uppercase\_udf.py /UDF/udfs

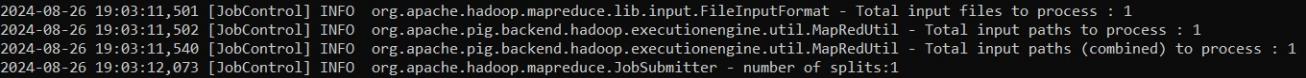


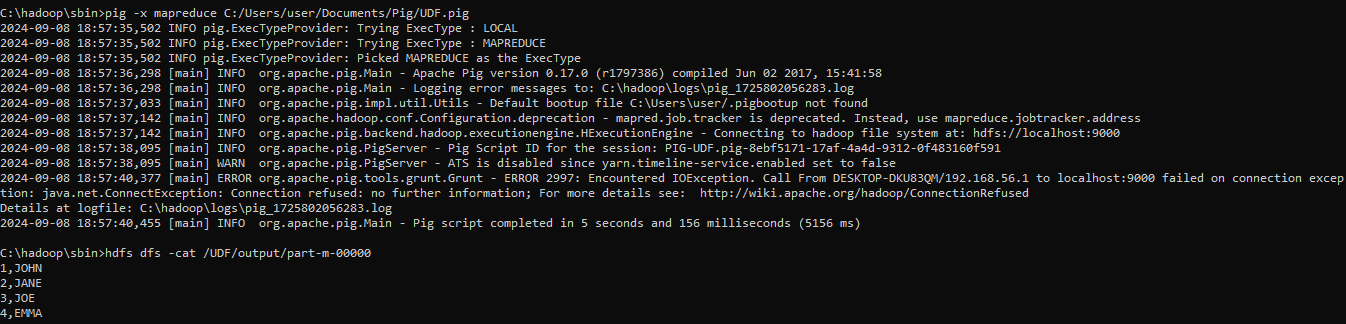


1. Create pig file “UDF.pig”:



1. Execute Pig file

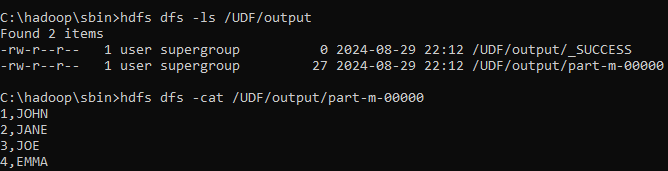
pig -x mapreduce C:/Users/user/Documents/Pig/UDF.pig



1. View the Output

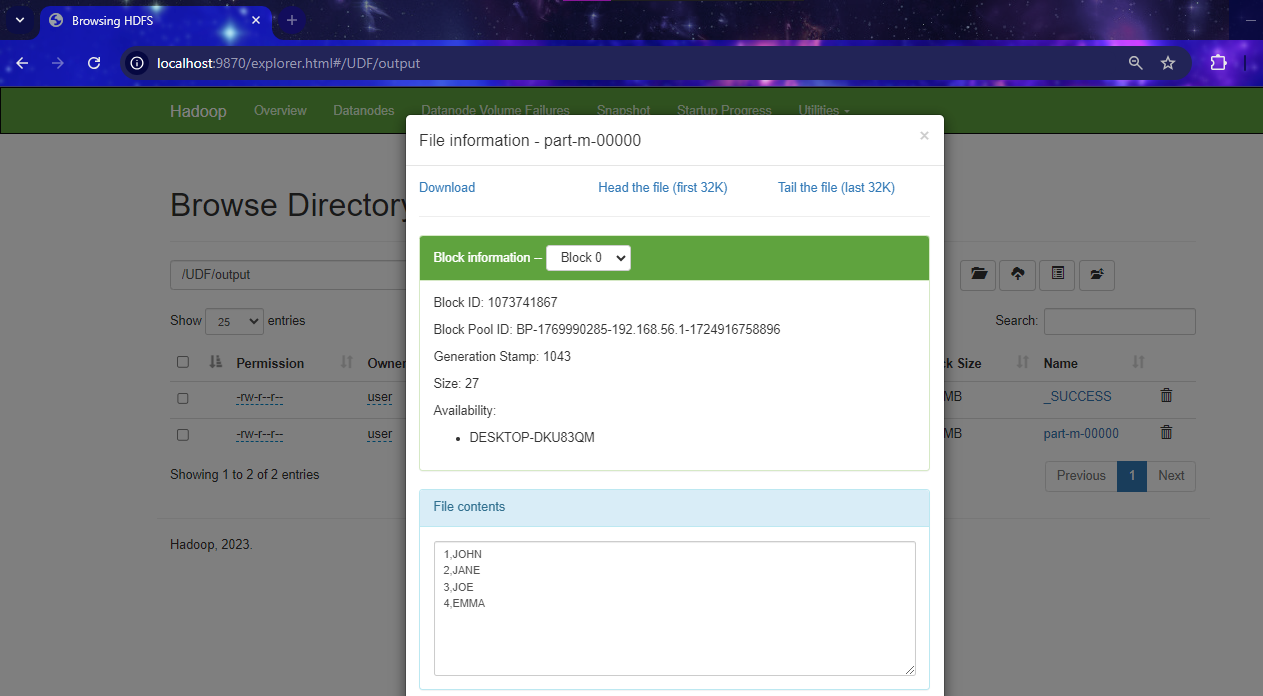
hdfs dfs -ls /UDF/output

hdfs dfs -cat /UDF/output/part-m-00000



1. Once the map reduce operations are performed successfully, the output will be present in the specified directory.

“/UDF/output/part-m-00000”



# RESULT:

Thus, UDF in Apache Pig has been created and executed in MapReduce/HDFS mode successfully.