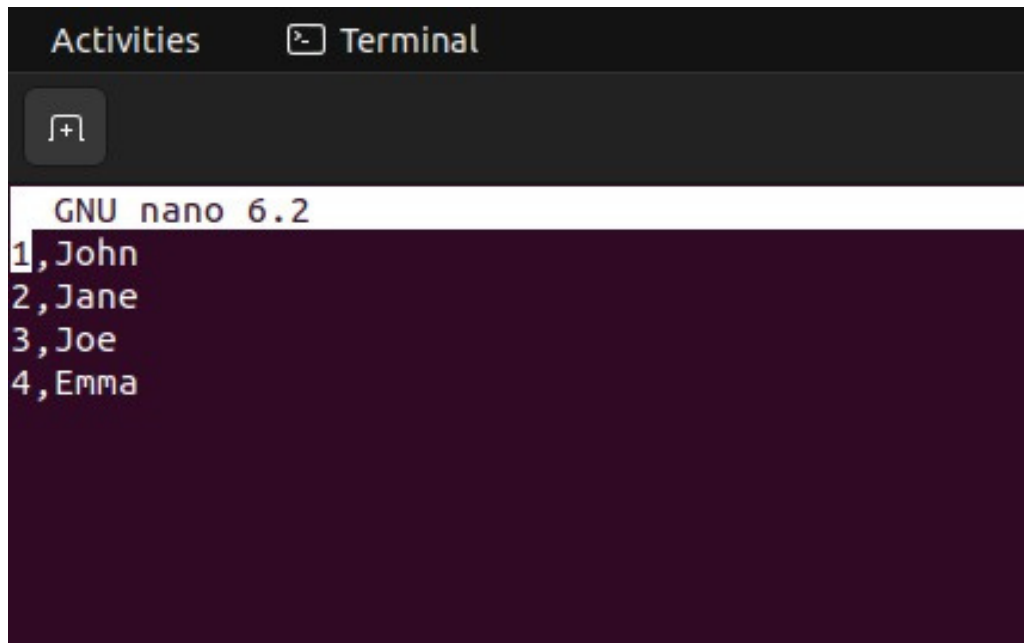


Exp. No : 4

## User Defined Function (UDF) in PIG

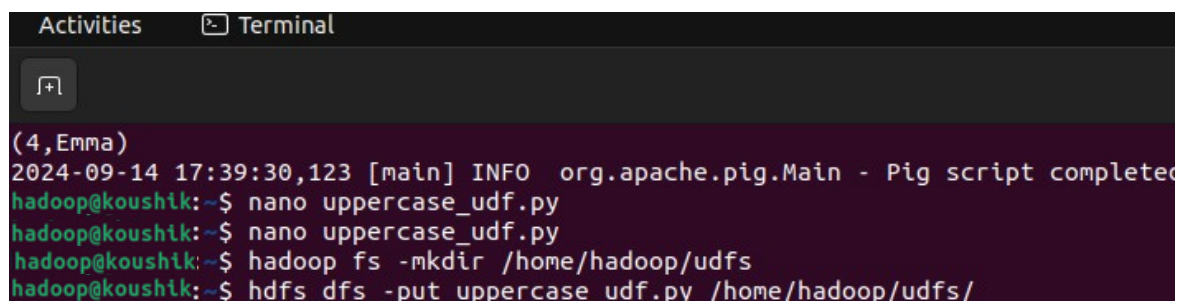
1. Create sample.txt



The screenshot shows a terminal window with the title 'Activities Terminal'. A new terminal session is open, displaying 'GNU nano 6.2'. The file 'sample.txt' is being edited, containing the following text:

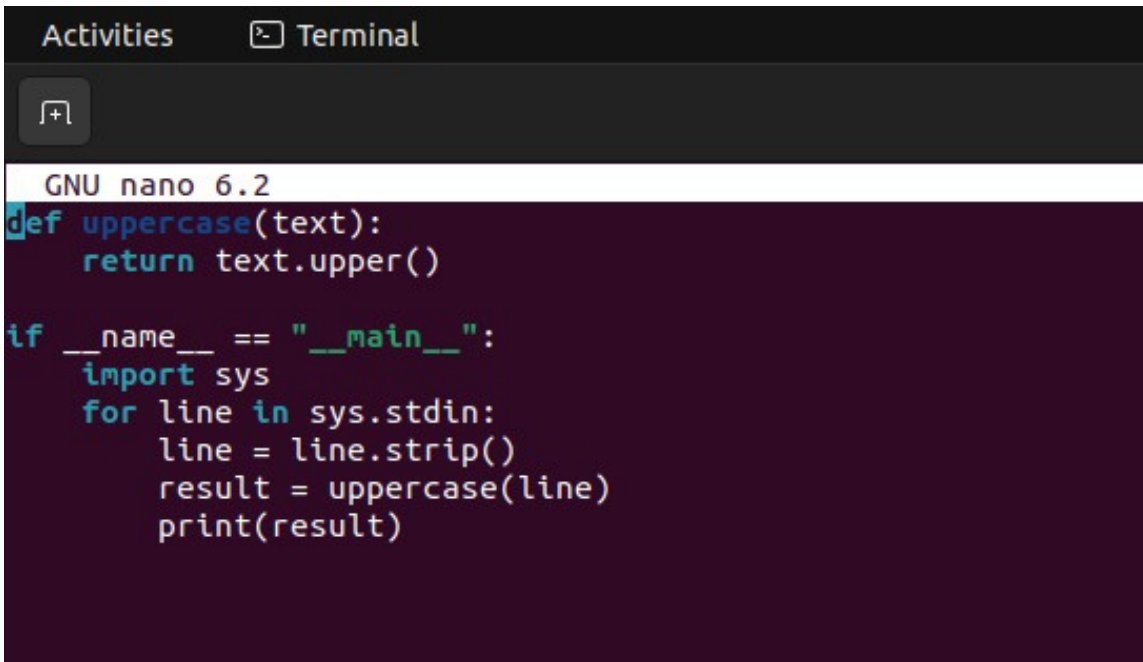
```
1,John
2,Jane
3,Joe
4,Emma
```

2. Create demo\_pig.pig file
3. Execute demo\_pig.pig
4. Create uppercase\_udf.py



The screenshot shows a terminal window with the title 'Activities Terminal'. The terminal output shows the execution of a Pig script, which completed successfully. The user then creates a file named 'uppercase\_udf.py' using nano, and uploads it to the Hadoop Distributed File System (HDFS) using the 'hdfs dfs -put' command.

```
(4, Emma)
2024-09-14 17:39:30,123 [main] INFO org.apache.pig.Main - Pig script completed
hadoop@koushik:~$ nano uppercase_udf.py
hadoop@koushik:~$ nano uppercase_udf.py
hadoop@koushik:~$ hadoop fs -mkdir /home/hadoop/udfs
hadoop@koushik:~$ hdfs dfs -put uppercase_udf.py /home/hadoop/udfs/
```



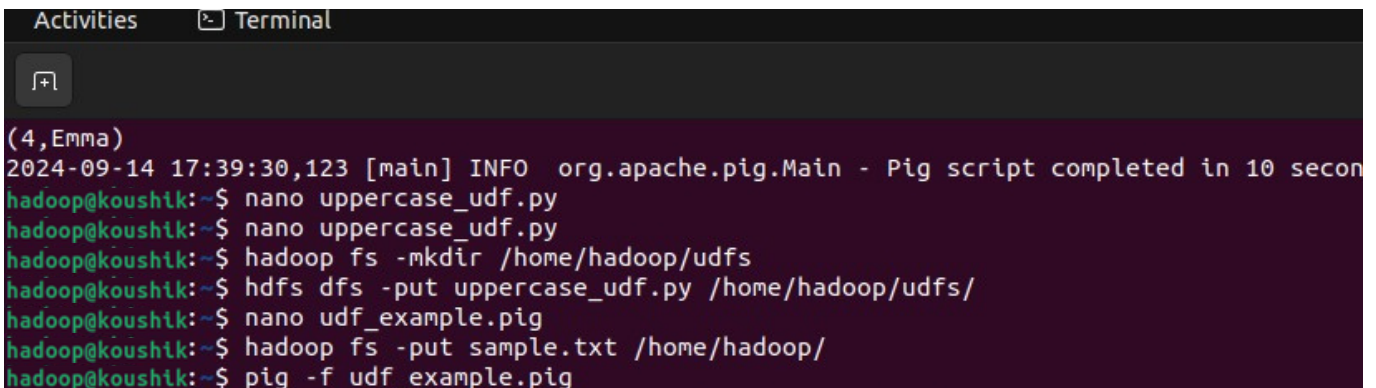
```

GNU nano 6.2
def uppercase(text):
    return text.upper()

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

```

5. Upload uppercase\_udf.py file to HDFS Storage.

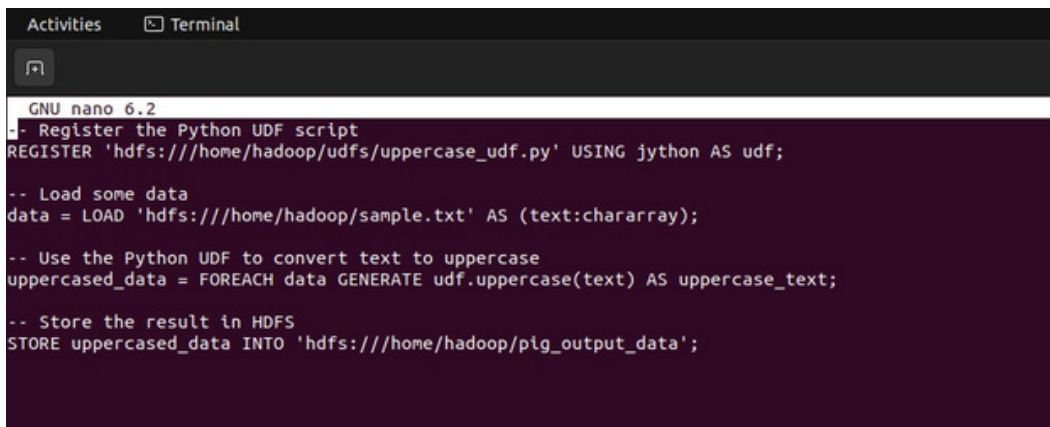


```

(4, Emma)
2024-09-14 17:39:30,123 [main] INFO  org.apache.pig.Main - Pig script completed in 10 seconds
hadoop@koushik:~$ nano uppercase_udf.py
hadoop@koushik:~$ nano uppercase_udf.py
hadoop@koushik:~$ hadoop fs -mkdir /home/hadoop/udfs
hadoop@koushik:~$ hdfs dfs -put uppercase_udf.py /home/hadoop/udfs/
hadoop@koushik:~$ nano udf_example.pig
hadoop@koushik:~$ hadoop fs -put sample.txt /home/hadoop/
hadoop@koushik:~$ pig -f udf_example.pig

```

6. Create udf\_example.pig



```

GNU nano 6.2
-- 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 to convert text to uppercase
uppercased_data = FOREACH data GENERATE udf.uppercase(text) AS uppercase_text;

-- Store the result in HDFS
STORE uppercased_data INTO 'hdfs:///home/hadoop/pig_output_data';

```

7. Execute udf\_example.pig

**Output :**

```
hadoop@koushik:~$ hdfs dfs -cat /home/hadoop/pig_output_data/part-m-00000  
1,JOHN  
2,JANE  
3,JOE  
4,EMMA  
hadoop@koushik:~$
```