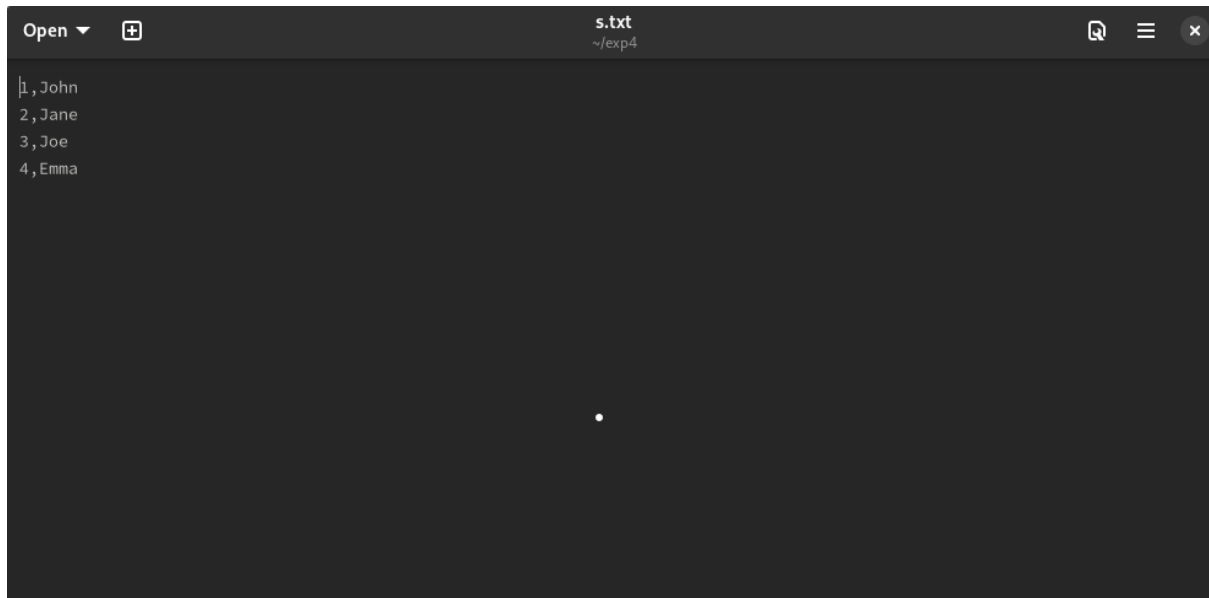
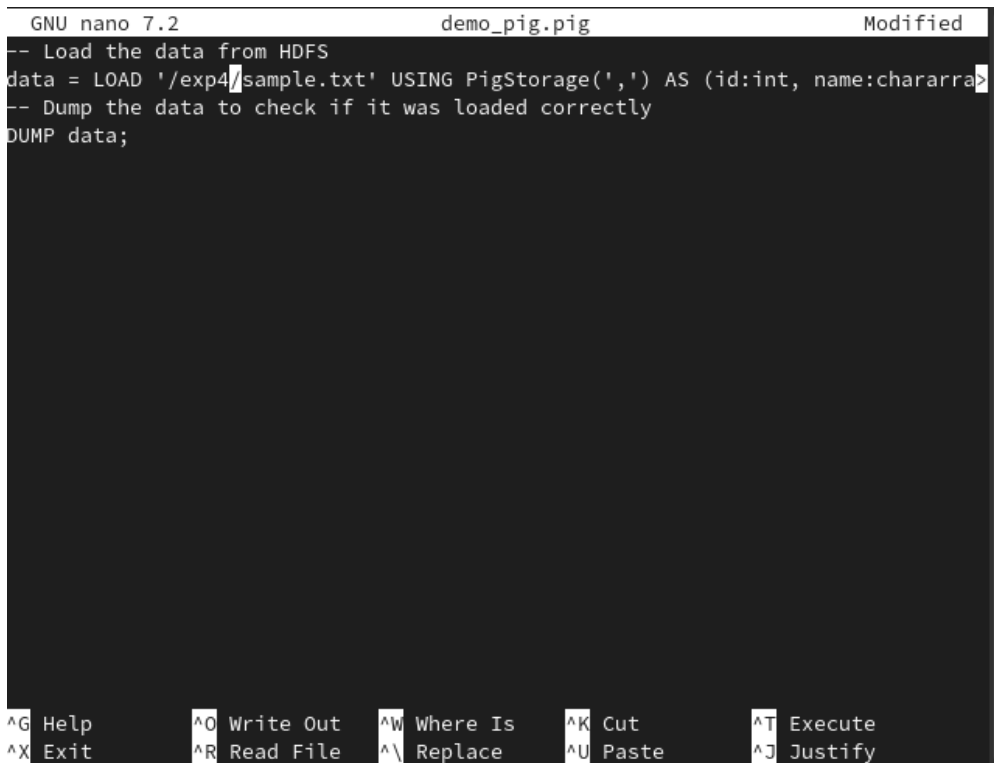


Exp. No : 4**User Defined Function (UDF) in PIG****1. Create sample.txt**A screenshot of a text editor window titled 's.txt' with a path '~ / exp4'. The editor contains four lines of text: '1, John', '2, Jane', '3, Joe', and '4, Emma'. The cursor is positioned at the end of the first line.

```
Open ▾ + s.txt ~ / exp4
1, John
2, Jane
3, Joe
4, Emma
```

2. Create demo_pig.pig fileA screenshot of a nano text editor window titled 'demo_pig.pig' with a status 'Modified'. The editor contains Pig Latin code to load data from 'sample.txt' and dump it. The bottom of the window shows nano editor shortcuts.

```
GNU nano 7.2 demo_pig.pig Modified
-- Load the data from HDFS
data = LOAD '/exp4/sample.txt' USING PigStorage(',') AS (id:int, name:chararra>
-- Dump the data to check if it was loaded correctly
DUMP data;
```

^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify

3. Execute demo_pig.pig

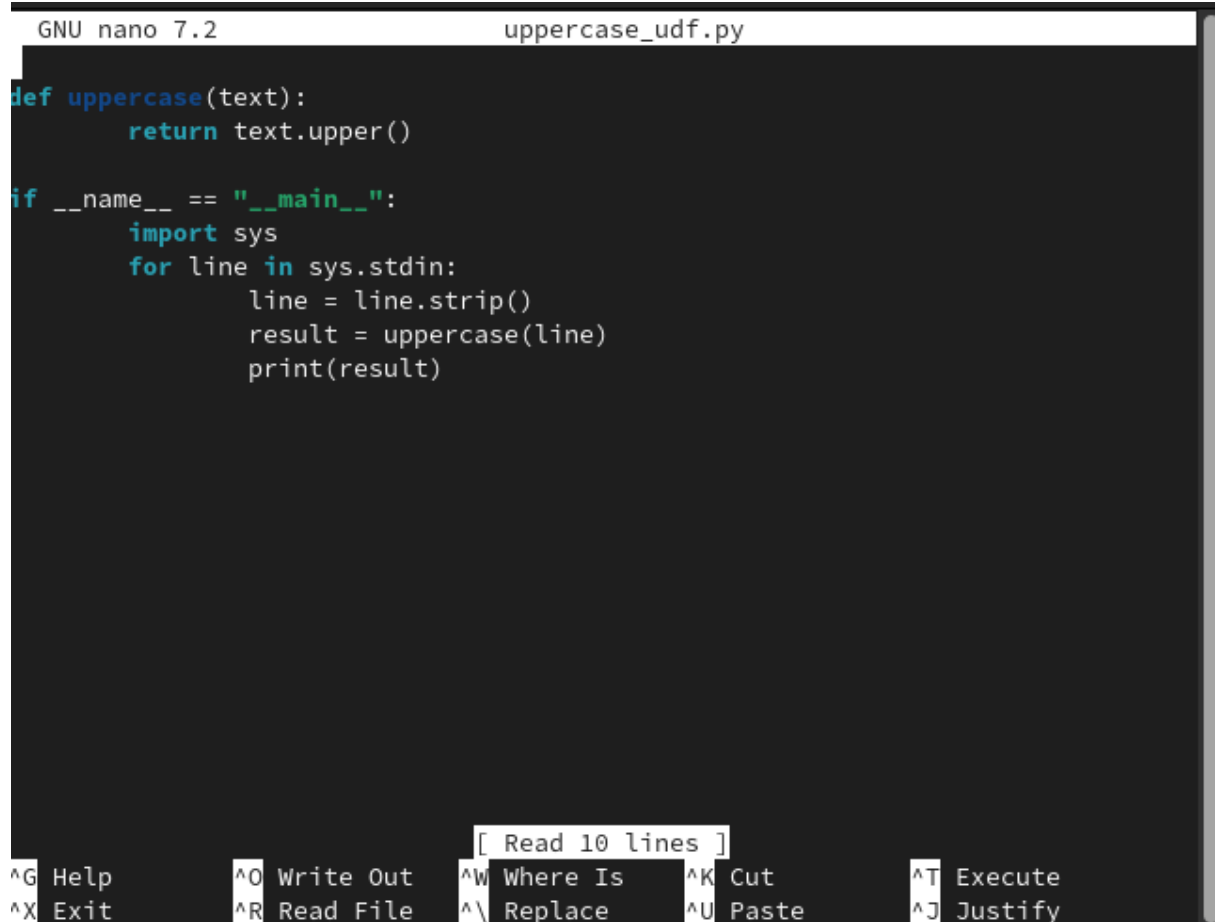
```

harini@fedora:~/exp4$ pig demo.pig
2024-10-12 05:51:59,843 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-12 05:51:59,845 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-10-12 05:51:59,846 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-10-12 05:51:59,905 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-10-12 05:51:59,906 [main] INFO org.apache.pig.Main - Logging error messages to: /home/harini/exp4/pig_1728726719904.log
2024-10-12 05:52:00,304 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/harini/.pigbootup not found
2024-10-12 05:52:00,345 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
mapreduce.jobtracker.address
2024-10-12 05:52:00,345 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs
.defaultFS
2024-10-12 05:52:00,345 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system
at: hdfs://localhost:9000
2024-10-12 05:52:00,910 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
mapreduce.jobtracker.address
2024-10-12 05:52:00,910 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tra
cker at: localhost:9001
2024-10-12 05:52:00,912 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs
.defaultFS
2024-10-12 05:52:00,978 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-demo.pig-f52af5f5-70cf-4007-8edb-35
e08b917837
2024-10-12 05:52:00,978 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-10-12 05:52:01,356 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use

2024-10-12 05:55:07,493 [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)
2024-10-12 05:55:08,593 [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)
2024-10-12 05:55:08,702 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrie
ve job to compute warning aggregation.
2024-10-12 05:55:08,703 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-10-12 05:55:08,708 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
mapreduce.jobtracker.address
2024-10-12 05:55:08,709 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs
.defaultFS
2024-10-12 05:55:08,710 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.
enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-10-12 05:55:08,710 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate c
ode.
2024-10-12 05:55:08,789 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-10-12 05:55:08,791 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,John)
(2,Jane)
(3,Joe)
(4,Emma)
2024-10-12 05:55:08,948 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 9 seconds and 132 milliseconds (189132 m
s)

```

4. Create uppercase_udf.py



```
GNU nano 7.2                                uppercase_udf.py

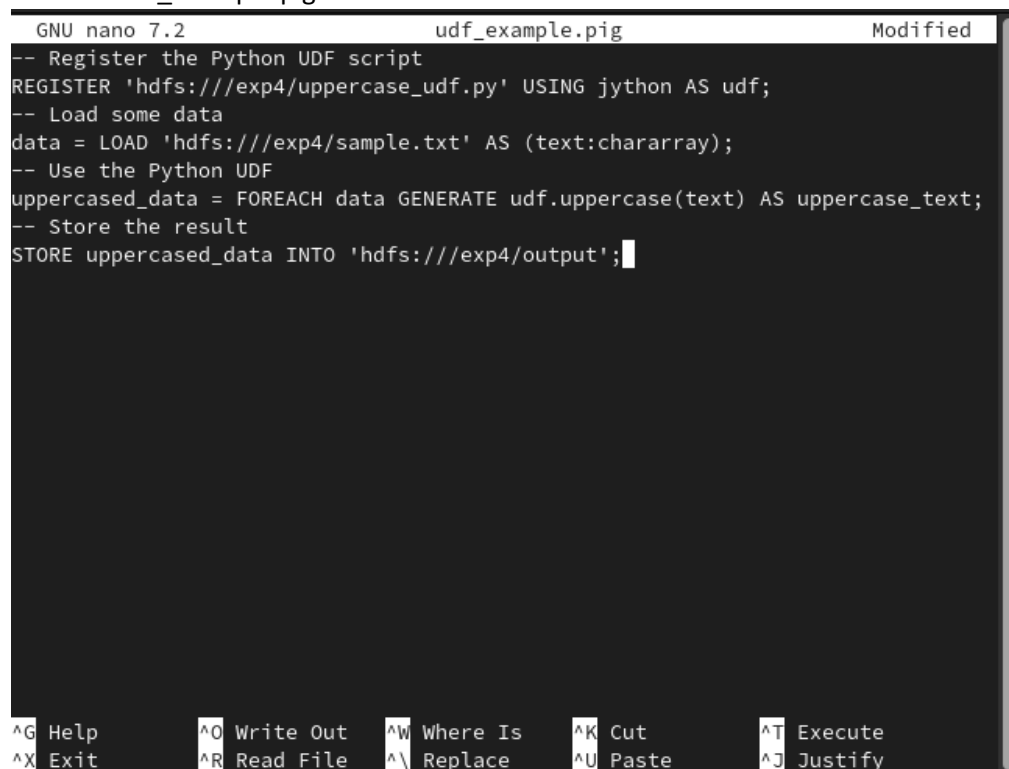
def uppercase(text):
    return text.upper()

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

[Read 10 lines]

^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify

5. Create udf_example.pig



```
GNU nano 7.2                                udf_example.pig                                Modified

-- 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';
```

^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify

5. Execute udf_example.pig

```
2024-10-10 21:01:20,882 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-10 21:01:20,884 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-10-10 21:01:20,885 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-10-10 21:01:20,980 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0
2024-10-10 21:01:20,981 [main] INFO org.apache.pig.Main - Logging error messages to:
2024-10-10 21:01:21,555 [main] INFO org.apache.pig.impl.util.Utils - Default bootup
2024-10-10 21:01:21,626 [main] INFO org.apache.hadoop.conf.Configuration.deprecation
2024-10-10 21:01:21,626 [main] INFO org.apache.hadoop.conf.Configuration.deprecation
2024-10-10 21:01:21,626 [main] INFO org.apache.pig.backend.hadoop.executionengine.HE
2024-10-10 21:01:22,461 [main] INFO org.apache.hadoop.conf.Configuration.deprecation
2024-10-10 21:01:22,482 [main] INFO org.apache.pig.PigServer - Pig Script ID for the
2024-10-10 21:01:22,483 [main] WARN org.apache.pig.PigServer - ATS is disabled since
2024-10-10 21:01:22,562 [main] INFO org.apache.hadoop.conf.Configuration.deprecation
```

Output :

```
1,JOHN
2,JANE
3,JOE
4,EMMA
```