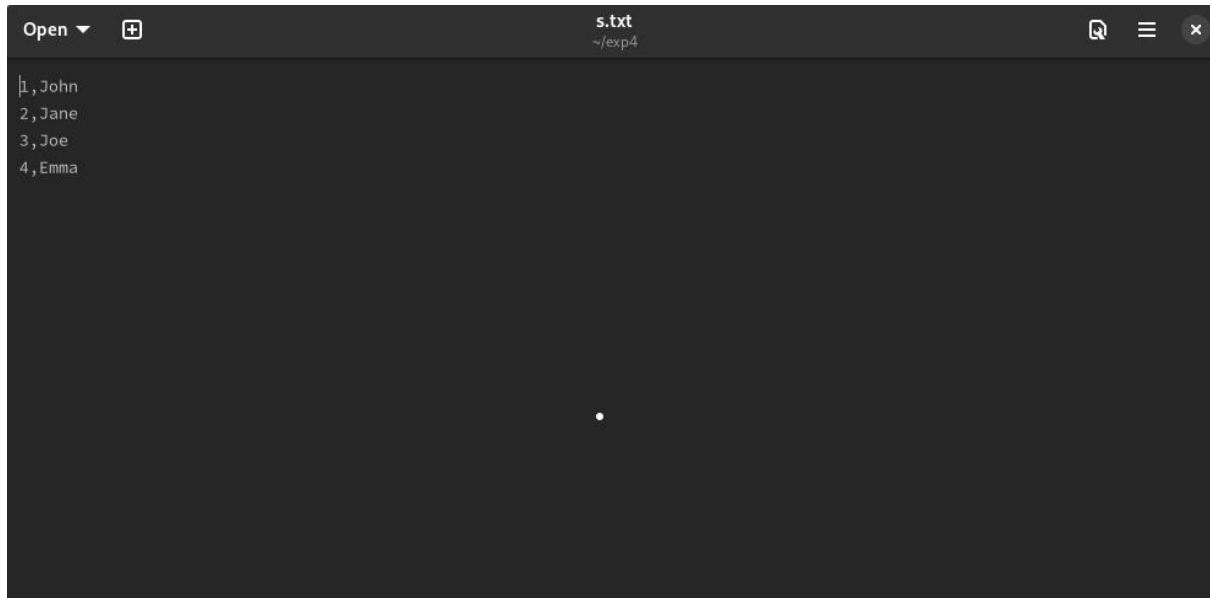
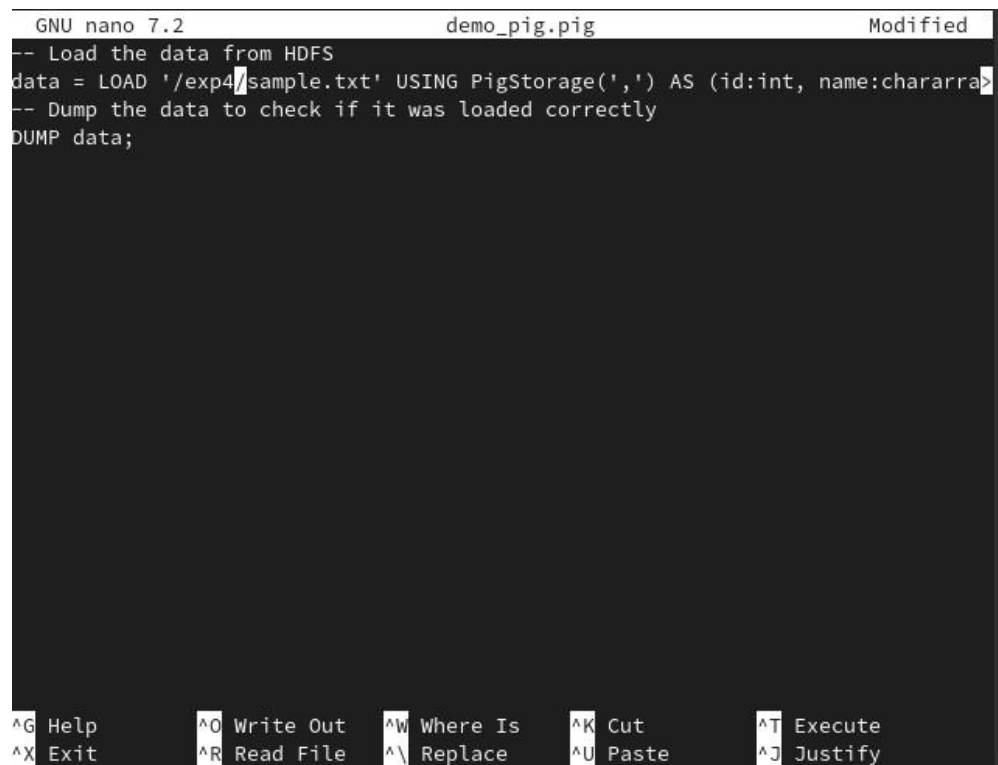


Exp. No : 4**User Defined Function (UDF) in PIG****1. Create sample.txt**A screenshot of a text editor window titled 's.txt' with a path of '~/.exp4'. The window contains four lines of text: '1, John', '2, Jane', '3, Joe', and '4, Emma'. The editor has a dark background and standard window controls at the top.**2. Create demo_pig.pig file**A screenshot of a GNU nano 7.2 editor window titled 'demo_pig.pig'. The window contains the following Pig script:

```
-- 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;
```

 The bottom of the window shows a status bar with various keyboard shortcuts like ^G Help, ^O Write Out, ^W Where Is, ^K Cut, ^T Execute, ^X Exit, ^R Read File, ^\ Replace, ^U Paste, and ^J Justify.

3. Execute demo_pig.pig

```

kaarokki@fedora:~/exp4$ pig demo_pig.pig
2024-10-20 12:03:59,586 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-20 12:03:59,588 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-10-20 12:03:59,589 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-10-20 12:03:59,691 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-10-20 12:03:59,691 [main] INFO org.apache.pig.Main - Logging error messages to: /home/kaarokki/exp4/pig_1729440239683.log
2024-10-20 12:04:00,114 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/kaarokki/.pigbootup not found
2024-10-20 12:04:00,202 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:04:00,202 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-10-20 12:04:00,202 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-10-20 12:04:01,077 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:04:01,077 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:9001
2024-10-20 12:04:01,078 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-10-20 12:04:01,116 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-demo_pig.pig-c5f049ba-f2cf-4d1b-b7f3-6b741bc3112c
2024-10-20 12:04:01,116 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-10-20 12:04:01,786 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:04:01,786 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

```

```

ried 9 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1000 MILLISECONDS)
2024-10-20 12:42:07,537 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2024-10-20 12:42:07,537 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-10-20 12:42:07,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-10-20 12:42:07,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:42:07,546 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-10-20 12:42:07,546 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2024-10-20 12:42:07,572 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-10-20 12:42:07,573 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,John)
(2,Jane)
(3,Joe)
(4,Emma)
2024-10-20 12:42:07,684 [main] INFO org.apache.pig.Main - Pig script completed in 2 minutes, 56 seconds and 38 milliseconds (176038 ms)

```

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

```

kaarokki@fedora:~/exp4$ pig udf_example.pig
2024-10-20 12:39:45,920 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-20 12:39:45,921 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-10-20 12:39:45,921 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-10-20 12:39:45,958 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2024-10-20 12:39:45,958 [main] INFO org.apache.pig.Main - Logging error messages to: /home/kaarokki/exp4/pig_1729442385953.log
2024-10-20 12:39:46,132 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/kaarokki/.pigbootup not found
2024-10-20 12:39:46,177 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:39:46,177 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-10-20 12:39:46,177 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-10-20 12:39:46,518 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:39:46,518 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:9001
2024-10-20 12:39:46,519 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-10-20 12:39:46,538 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf_example.pig-5852dde0-7b93-483d-b65d-4a72df1a9d06
2024-10-20 12:39:46,538 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2024-10-20 12:39:46,570 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-10-20 12:39:46,570 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

```

Output :

```

kaarokki@fedora:~/exp4$ hdfs dfs -ls /exp4/output
Found 2 items
-rw-r--r-- 1 kaarokki supergroup          0 2024-10-20 12:40 /exp4/output/_SUCCESS
-rw-r--r-- 1 kaarokki supergroup       27 2024-10-20 12:40 /exp4/output/part-m-00000
kaarokki@fedora:~/exp4$ hdfs dfs -cat /exp4/output
cat: `/exp4/output': Is a directory
kaarokki@fedora:~/exp4$ hdfs dfs -cat /exp4/output/part-m-50
cat: `/exp4/output/part-m-50': No such file or directory
kaarokki@fedora:~/exp4$ hdfs dfs -cat /exp4/output/part-m-00000
1,JOHN
2,JANE
3,JOE
4,EMMA

```