**Exp.No: 4**

**Create User Defined Function (UDF) in Apache Pig and execute it in MapReduce**

**AIM:**

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

**PROCEDURE:**

**Step-1:** 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 comment to download Apache Pig in Ubuntu:

**wget** <https://dlcdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz>

**Step-2:** To untar pig-0.16.0.tar.gz file run the following command:

**tar xvzf pig-0.16.0.tar.gz**

**Step 3:** To create a pig folder and move pig-0.16.0 to the pig folder, execute the following command:

**sudo mv /home/hdoop/pig-0.16.0 /home/hdoop/pig**

**Step 4:** 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 settings**

**export PIG\_HOME=/home/hdoop/pig**

**export PATH=$PATH:$PIG\_HOME/bin**

**export PIG\_CLASSPATH=$PIG\_HOME/conf:$HADOOP\_INSTALL/etc/hadoop/**

**export PIG\_CONF\_DIR=$PIG\_HOME/conf**

**export JAVA\_HOME=/usr/lib/jvm/java-8-openjdkamd64**

**export PIG\_CLASSPATH=$PIG\_CONF\_DIR:$PATH**

**#PIG setting ends**

**Step 5:** Run the following command to make the changes effective in the .bashrc file:

**source .bashrc**

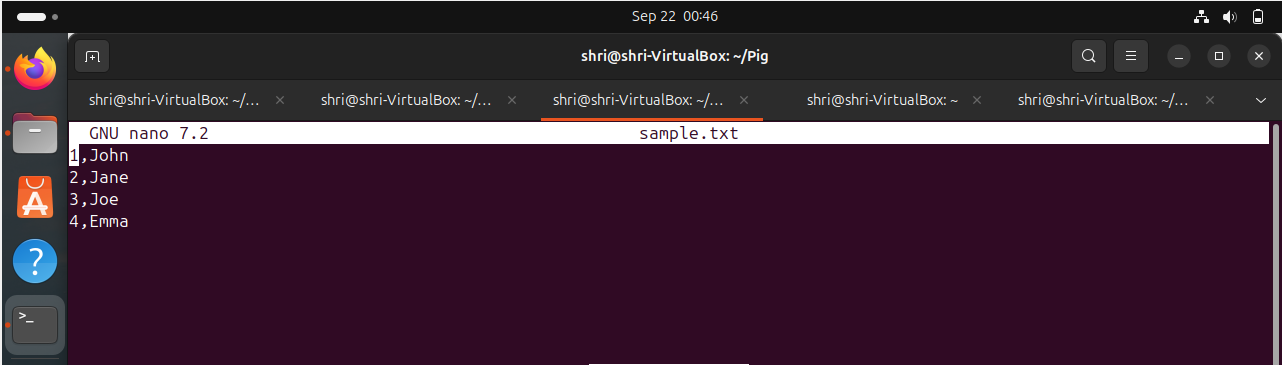
**Step 6:** To start all Hadoop daemons, navigate to the hadoop-3.2.1/sbin folder and run the following commands:

**./start-dfs.sh**

**./start-yarn.sh**

**Step 7:** Create a sample text file

**nano sample.txt**

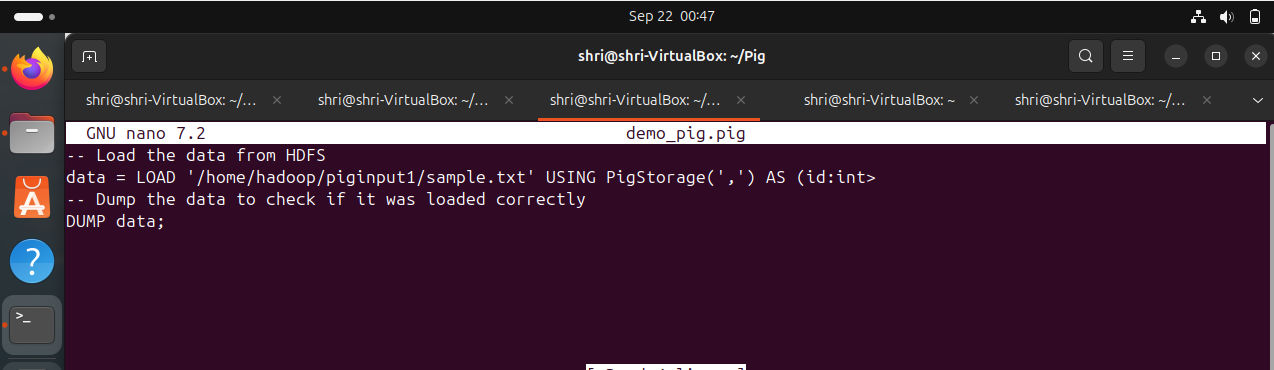
****

**Step 8:** Add the text file to the Hadoop environment.

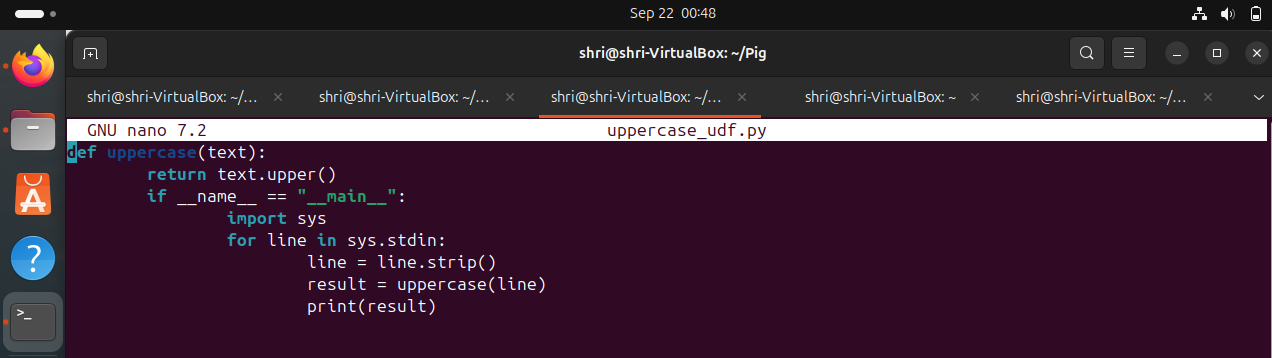
**hadoop fs -put sample.txt /home/hadoop/piginput/**

**Step 9:** Create PIG File

**nano demo\_pig.pig**

****

**Step 10:** Create udf file and save as uppercase\_udf.py



**Step 11:** Create the udfs folder on hadoop

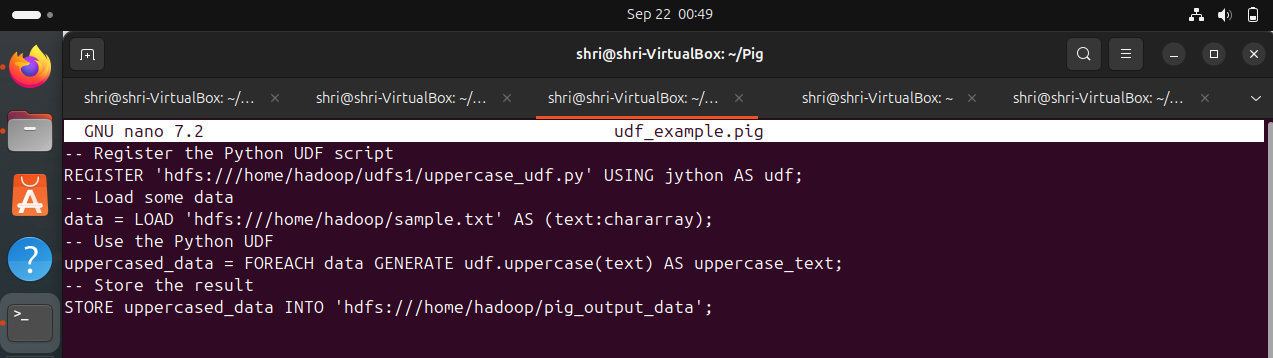
**hadoop fs -mkdir /home/hadoop/udfs**

**Step 12:** Put the upppercase\_udf.py in to the above folder

**hdfs dfs -put uppercase\_udf.py /home/hadoop/udfs/**

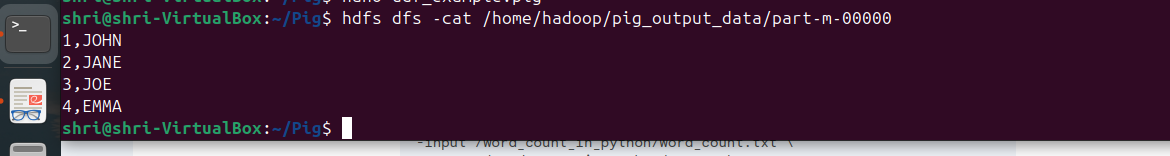
**Step 13:** Create a file named udf\_example.pig

**nano udf\_example.pig**

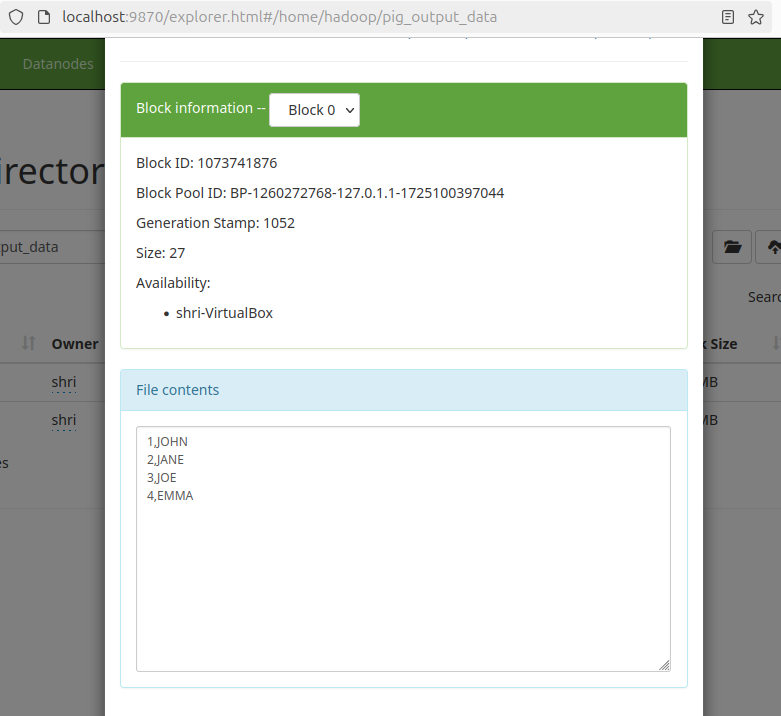
****

**Step 14:** To view the output use the command below

**hdfs dfs -cat /home/hadoop/pig\_output\_data/part-m-00000**

****

**Step 15:** The result in the Namenode is as follows:



**RESULT:**

Thus the program is executed successfully and output is verified.