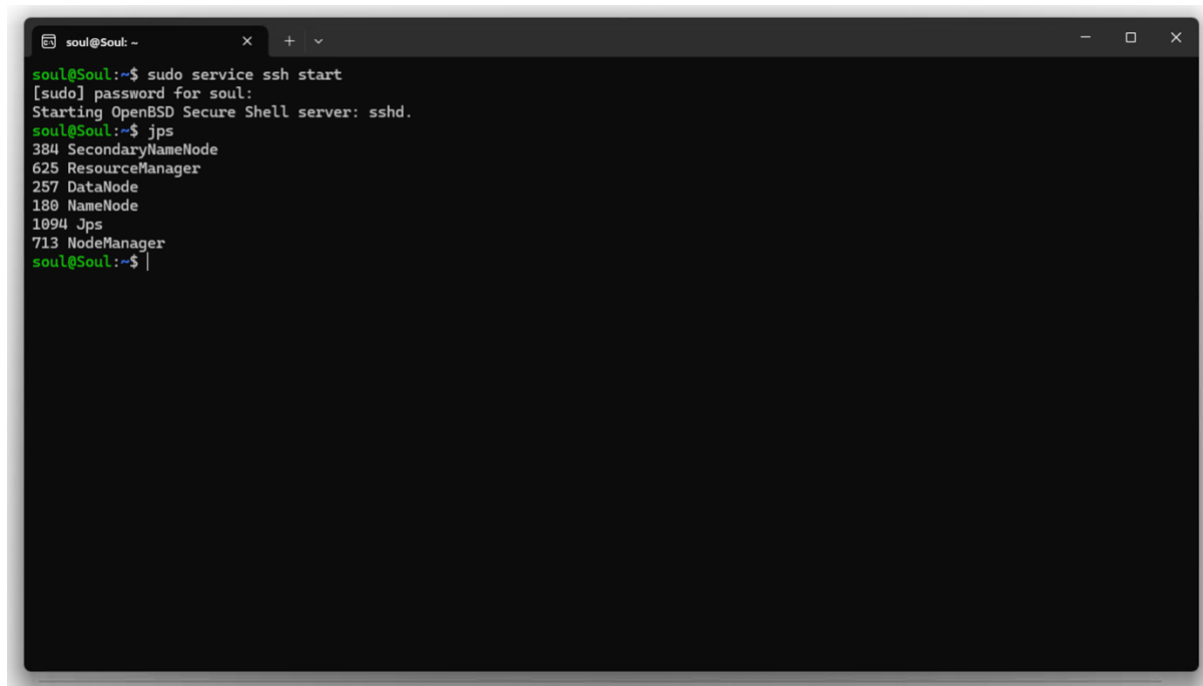


EXP NO: 4

## Create UDF In Pig

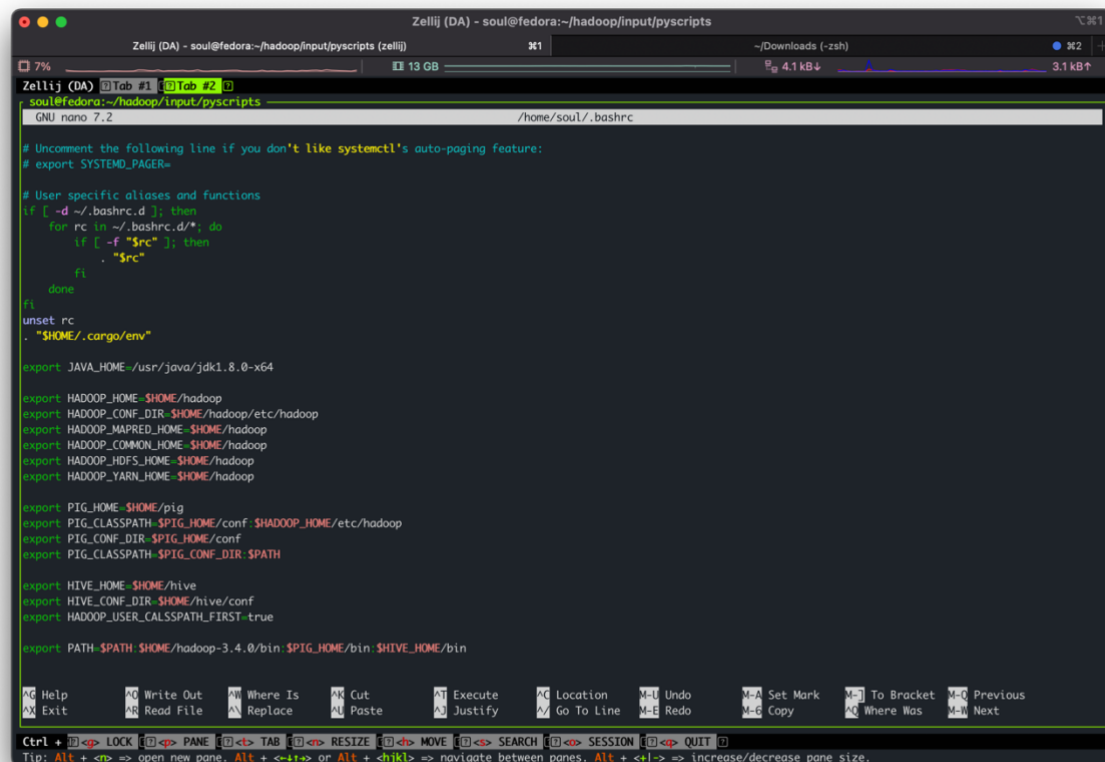
Starting Hadoop :



A terminal window titled 'soul@Soul: ~' showing the execution of 'sudo service ssh start' and 'jps'. The output of 'jps' lists several processes: 384 SecondaryNameNode, 625 ResourceManager, 257 DataNode, 180 NameNode, 1094 Jps, and 713 NodeManager.

```
soul@Soul:~$ sudo service ssh start
[sudo] password for soul:
Starting OpenBSD Secure Shell server: sshd.
soul@Soul:~$ jps
384 SecondaryNameNode
625 ResourceManager
257 DataNode
180 NameNode
1094 Jps
713 NodeManager
soul@Soul:~$
```

\$sudo nano ~/.bashrc



A terminal window titled 'Zellij (DA) - soul@fedora:~/hadoop/input/pyscripts' showing the nano editor editing the file '/home/soul/.bashrc'. The editor content includes comments about systemd, user-specific aliases, and environment variable exports for Hadoop, Pig, and Hive. The bottom of the screen shows a detailed command palette with various shortcuts like Ctrl+L for LOCK, Alt+O for Write Out, and Ctrl+Q for QUIT.

```
Zellij (DA) - soul@fedora:~/hadoop/input/pyscripts
Zellij (DA) - soul@fedora:~/hadoop/input/pyscripts (zellij)
GNU nano 7.2 /home/soul/.bashrc

# Uncomment the following line if you don't like systemd's auto-paging feature:
# export SYSTEMD_PAGER=

# User specific aliases and functions
if [ -d ~/.bashrc.d ]; then
  for rc in ~/.bashrc.d/*; do
    if [ -f "$rc" ]; then
      . "$rc"
    fi
  done
fi
unset rc
. "$HOME/.cargo/env"

export JAVA_HOME=/usr/java/jdk1.8.0-x64

export HADOOP_HOME=$HOME/hadoop
export HADOOP_CONF_DIR=$HOME/hadoop/etc/hadoop
export HADOOP_MAPRED_HOME=$HOME/hadoop
export HADOOP_COMMON_HOME=$HOME/hadoop
export HADOOP_HDFS_HOME=$HOME/hadoop
export HADOOP_YARN_HOME=$HOME/hadoop

export PIG_HOME=$HOME/pig
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_HOME/etc/hadoop
export PIG_CONF_DIR=$PIG_HOME/conf
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH

export HIVE_HOME=$HOME/hive
export HIVE_CONF_DIR=$HOME/hive/conf
export HADOOP_USER_CLASSPATH_FIRST=true

export PATH=$PATH:$HOME/hadoop-3.4.0/bin:$PIG_HOME/bin:$HIVE_HOME/bin

Zellij (DA) - soul@fedora:~/hadoop/input/pyscripts
  Help  Write Out  Where Is  Cut  Execute  Location  M-L  Set Mark  To Bracket  Previous
  Exit  Read File  Replace  Paste  Justify  Go To Line  M-B  Copy  Where Was  Next

Ctrl + L: LOCK, Ctrl + O: PANE, Ctrl + T: TAB, Ctrl + R: RESIZE, Ctrl + M: MOVE, Ctrl + S: SEARCH, Ctrl + Q: SESSION, Ctrl + W: QUIT
Tip: Alt + C: => open new pane, Alt + <-> or Alt + <hjkl> => navigate between panes, Alt + <+> => increase/decrease pane size.
```

Input Files (sample.txt, demo.pig, udf\_example.pig, uppercase\_udf.py) :

```

Zellij (BA) Tab #1
soul@fedora:~/hadoop/input/pig
soul@fedora:~/hadoop/input/pig$ cd ..
soul@fedora:~/hadoop/input$ ls
'DA Assignment 3' dataset.txt Experiments pig pyscripts shellscripts wordCount.txt
soul@fedora:~/hadoop/input$ cd pig/
soul@fedora:~/hadoop/input/pig$ ls
demo.pig sample.txt udf_example.pig uppercase_udf.py
soul@fedora:~/hadoop/input/pig$ cat demo.pig
-- Load the data from HDFS
data = LOAD '/input_dir/sample.txt' USING PigStorage(',') AS (id:int, name:chararray);
-- Dump the data to check if it was loaded correctly
DUMP data;
soul@fedora:~/hadoop/input/pig$ cat sample.txt
1,Jerin
2,Sheela
3,Bellavita
soul@fedora:~/hadoop/input/pig$ cat udf_example.pig
-- Register the Python UDF script
REGISTER 'hdfs:///uppercase_udf.py' USING jython AS udf;
-- Load some data
data = LOAD 'hdfs:///input_dir/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:///pig_output_data';
soul@fedora:~/hadoop/input/pig$
  
```

```

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)
  
```

Ctrl + <=> LOCK    <=> PANE    <=> TAB    <=> RESIZE    <=> MOVE    <=> SEARCH    <=> SESSION    <=> QUIT

Tip: Alt + <=> => open new pane, Alt + <=>+<=> or Alt + <=>+<=> => navigate between panes, Alt + <=>+<=> => increase/decrease pane size.

Executing Pig Script :

```

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime  Alias F
ature  Outputs
job_1724946859197_0006  1      0      11      11      11      11      0      0      0      0      data,uppercased_data  MAP_ONLY  hdfs:///pig_output_data,

Input(s):
Successfully read 3 records (398 bytes) from: "hdfs:///input_dir/sample.txt"

Output(s):
Successfully stored 3 records (29 bytes) in: "hdfs:///pig_output_data"

Counters:
Total records written : 3
Total bytes written : 29
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
  
```

Output :

```

soul@fedora:~/hadoop-3.4.0/input/pig$ hdfs dfs -cat /pig_output_data/*
1,JERIN
2,SHEELA
3,BELLAVITA
  
```

**Result :**

Thus the Pig Script is successfully executed and the desired output is achieved.