# BIG DATA ANALYTICS LAB

**(Course Code: 22UPCSC1C11)**

**A programming laboratory record submitted to Periyar University, Salem**

**In partial fulfillment of the requirements for the degree of**

**MASTER OF COMPUTER APPLICATIONS**

**By**

**ELANCHEZHIAN M.**

**[Reg. No.: U22PG507CAP006]**



**DEPARTMENT OF COMPUTER SCIENCE**

**PERIYAR UNIVERSITY**

**(NAAC `A++` Grade with CGPA 3.61) – NIRF RANK 59 – ARIIA RANK 10**

**PERIYAR PALKALAI NAGAR,**

**SALEM – 636 011.**

**(APRIL - 2023)**

# CERTIFICATE

    This is to certify that the Programming Laboratory entitled **"BIG DATA ANALYTICS - LAB (22UPCSC1C11)"** is a bonafide record work done by Mr. /Ms. _____

Register No: _____as partial fulfillment of the requirements for the degree of Master of Computer Applications, in the Department of Computer Science, Periyar University, Salem, during the Academic Year 2023-2024.

    Staff In-charge              Head of the Department

  Submitted for the practical examination held on……...………………

    Internal Examiner            External Examiner

# CONTENTS

## SOURCE CODE:

**hduser@elan-VirtualBox:/home/elan$ start-dfs.sh**
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [elan-VirtualBox]
2023-03-04 22:24:49,572 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hduser@elan-VirtualBox:/home/elan$ start-yarn.sh**
Starting resourcemanager
resourcemanager is running as process 8979.  Stop it first and ensure
/tmp/hadoop-hduser-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 9110.  Stop it first and ensure
/tmp/hadoop-hduser-nodemanager.pid file is empty before retry.

**hduser@elan-VirtualBox:/home/elan$ jps**
8979 ResourceManager
9110 NodeManager
12888 DataNode
13098 SecondaryNameNode
13535 Jps
12751 NameNode

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /**
2023-03-04 22:32:15,190 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Found 1 items
drwx------   - hduser supergroup        0 2023-03-04 22:17 /user

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -mkdir /elan**
2023-03-04 22:33:30,869 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /**
2023-03-04 22:33:37,423 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hduser supergroup          0 2023-03-04 22:33 /**elan**
drwx------   - hduser supergroup          0 2023-03-04 22:17 /user

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -touchz  /elan/magu.txt**
2023-03-04 22:34:15,934 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /elan/**
2023-03-04 22:34:26,146 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 hduser supergroup          0 2023-03-04 22:34 /elan/**magu.txt**

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -put AI.txt /elan/**
2023-03-04 22:36:16,386 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /elan/**
2023-03-04 22:36:25,867 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 hduser supergroup          7 2023-03-04 22:36 /elan/**AI.txt**
-rw-r--r--   1 hduser supergroup          0 2023-03-04 22:34 /elan/magu.txt

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -cat /elan/AI.txt**
2023-03-04 22:38:21,461 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Hellow

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -du /elan/AI.txt**
2023-03-04 22:39:50,335 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
7  7  /elan/AI.txt

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -dus /user**
2023-03-04 22:40:51,214 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
dus: DEPRECATED: Please use 'du -s' instead.
**0  0  /user**

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -stat /elan/**
2023-03-04 22:41:17,430 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
2023-03-04 17:06:18

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -setrep -R 4 /elan**
2023-03-04 22:42:40,864 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Replication 4 set: /elan/AI.txt
Replication 4 set: /elan/magu.txt

**hduser@elan-VirtualBox:/home/elan$ echo $?**
0

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -test /elan/magu.txt&echo
$?**
[1] 15182
0

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -count -q /elan/AI.txt**
2023-03-04 22:46:43,775 WARN util.NativeCodeLoader: Unable to
load native-hadoop library for your platform... using builtin-java classes
where applicable
 none          inf       none        inf       0        1        7 /elan/AI.txt


**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -mkdir /temp**
2023-03-04 22:49:01,972 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable


**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -cp /elan /temp**
2023-03-04 22:49:35,176 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable


**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /temp**
2023-03-04 22:49:50,868 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Found 1 items
drwxr-xr-x   - hduser supergroup          0 2023-03-04 22:49 /temp/elan


**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /**
2023-03-04 22:50:08,561 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Found 3 items
drwxr-xr-x   - hduser supergroup          0 2023-03-04 22:36 /**elan**
drwxr-xr-x   - hduser supergroup          0 2023-03-04 22:49 /**temp**
drwx------   - hduser supergroup          0 2023-03-04 22:17 **/user**


**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -mv /elan/AI.txt /temp/AI.txt**
2023-03-04 22:53:23,022 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /temp/**
2023-03-04 22:53:40,368 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   4 hduser supergroup         7 2023-03-04 22:36 /temp/**AI.txt**
drwxr-xr-x   - hduser supergroup         0 2023-03-04 22:49 /temp/**elan**

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -rmr /temp**
2023-03-04 22:54:30,679 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
rmr: DEPRECATED: Please use '-rm -r' instead.
2023-03-04 22:54:32,095 INFO fs.TrashPolicyDefault: Moved: 'hdfs://localhost:54310/temp' to trash at: hdfs://localhost:54310/user/hduser/.Trash/Current/temp

**hduser@elan-VirtualBox:/home/elan$ hdfs dfs -ls /**
2023-03-04 22:54:50,505 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hduser supergroup         0 2023-03-04 22:53 /**elan**
drwx------   - hduser supergroup         0 2023-03-04 22:17 /**user**

**hduser@elan-VirtualBox:/home/elan$ stop-all.sh**
WARNING: Stopping all Apache Hadoop daemons as hduser in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [elan-VirtualBox]
2023-03-04 23:33:57,243 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Stopping nodemanagers
Stopping resourcemanage

## SOURCE CODE:

**hadoop@ubuntu:~$ start-all.sh**
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10
seconds.
WARNING: This is not a recommended production deployment
configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [ubuntu]
2023-03-04 21:29:38,334 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Starting resourcemanager
Starting nodemanagers

**hadoop@ubuntu:~$ mkdir Bda**
**hadoop@ubuntu:~$ cd Bda**
**hadoop@ubuntu:~/Bda$ nano minMaxMapper.py**

```
import sys
for line in sys.stdin:
    val = line.strip()
    year = val[0:4]
    temp = val[5:9]
    if temp != '9999':
        print('%s\t%s' % (year, temp))
```

**hadoop@ubuntu:~/Bda$ nano minMaxReducer.py**

```
from operator import itemgetter
import sys

(last_key, max_val, min_val) = (None, 0, 0)

for line in sys.stdin:
    (key, val) = line.strip().split("\t")

    if last_key and last_key != key:
        print('%s\t%s\t%s' % (last_key, max_val, min_val))
        (last_key, max_val, min_val) = (key, float(val), float(val)
```

```
              else:
                  (last_key, max_val, min_val) = (key, max(max_val,
                  float(val)), min(min_val, float(val)))
                  if min_val == 0.0:
                      min_val = max_val
              if last_key:
                  print('%s\t%s\t%s' % (last_key, max_val, min_val))
```

**hadoop@ubuntu:~/Bda$ nano temp.txt**

```
                        2011    41
                        2011    72
                        2012    55
                        2012    43
                        2013    51
                        2013    62
                        2014    23
                        2015    45
                        2016    47
                        2016    65
                        2017    43
                        2017    23
                        2018    60
                        2019    65
                        2022    78
                        2022    45
                        2023    22
```

**hadoop@ubuntu:~/Bda$ ls**
minMaxMapper.py  minMaxReducer.py  temp.txt

**hadoop@ubuntu:~/Bda$ cat temp.txt | python minMaxMapper.py |**
**python minMaxReducer.py**

| | | |
|---|---|---|
| 2011 | 72.0 | 41.0 |
| 2012 | 55.0 | 43.0 |
| 2013 | 62.0 | 51.0 |
| 2014 | 23.0 | 23.0 |
| 2015 | 45.0 | 45.0 |
| 2016 | 65.0 | 47.0 |

| 2017 | 43.0 | 23.0 |
| 2018 | 60.0 | 60.0 |
| 2019 | 65.0 | 65.0 |
| 2022 | 78.0 | 45.0 |
| 2023 | 22.0 | 22.0 |

**hadoop@ubuntu:~/Bda$ hdfs dfs -mkdir /BigData**
2023-03-04 21:44:12,439 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~/Bda$ hdfs dfs -ls /**
2023-03-04 21:44:25,881 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hadoop supergroup          0 2023-03-04 21:44 /**BigData**
drwxr-xr-x   - hadoop supergroup          0 2023-03-04 21:43 /elan

**hadoop@ubuntu:~/Bda$ hdfs dfs -put minMaxReducer.py /BigData**
2023-03-04 21:45:16,642 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~/Bda$ hdfs dfs -put minMaxMapper.py /BigData**
2023-03-04 21:45:40,287 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~/Bda$ hdfs dfs -put temp.txt /BigData**
2023-03-04 21:46:10,971 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~/Bda$ hdfs dfs -ls /BigData/**
2023-03-04 21:46:26,374 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 3 items

-rw-r--r--   1 hadoop supergroup        161 2023-03-04 21:45 /BigData/**minMaxMapper.py**

-rw-r--r--   1 hadoop supergroup        554 2023-03-04 21:45 /BigData/**minMaxReducer.py**

-rw-r--r--   1 hadoop supergroup        136 2023-03-04 21:46 /BigData/**temp.txt**

**hadoop@ubuntu:~/Bda$ hadoop jar /usr/local/hadoop/share/hadoop/ tools/lib/hadoop-streaming-*.jar -file minMaxMapper.py -mapper "python minMaxMapper.py" -file minMaxReducer.py -reducer "python minMaxReducer.py" -input /BigData/temp.txt -output /BigData/ P02RESULT**

**hadoop@ubuntu:~/Bda$ hdfs dfs -ls /BigData/**
2023-03-04 22:01:01,336 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items

drwxr-xr-x   - hadoop supergroup          0 2023-03-04 21:58 /BigData/**P02RESULT**

-rw-r--r--   1 hadoop supergroup        161 2023-03-04 21:45 /BigData/**minMaxMapper.py**

-rw-r--r--   1 hadoop supergroup        554 2023-03-04 21:45 /BigData/**minMaxReducer.py**

-rw-r--r--   1 hadoop supergroup        136 2023-03-04 21:46 /BigData/**temp.txt**

**hadoop@ubuntu:~/Bda$ hdfs dfs -ls /BigData/P02RESULT/**
2023-03-04 22:02:41,070 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items

-rw-r--r--   1 hadoop supergroup          0 2023-03-04 21:58 /BigData/**P02RESULT/_SUCCESS**

-rw-r--r--   1 hadoop supergroup        165 2023-03-04 21:58 /BigData/**P02RESULT/part-00000**

**hadoop@ubuntu:~/Bda$ hdfs dfs -cat /BigData/P02RESULT/part-***
2023-03-04 22:03:05,017 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

| | | |
|------|------|------|
| 2011 | 72.0 | 41.0 |
| 2012 | 55.0 | 43.0 |
| 2013 | 62.0 | 51.0 |
| 2014 | 23.0 | 23.0 |
| 2015 | 45.0 | 45.0 |
| 2016 | 65.0 | 47.0 |
| 2017 | 43.0 | 23.0 |
| 2018 | 60.0 | 60.0 |
| 2019 | 65.0 | 65.0 |
| 2022 | 78.0 | 45.0 |
| 2023 | 22.0 | 22.0 |

**hadoop@ubuntu:~/Bda$ stop-all.sh**
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [0.0.0.0]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
2023-03-04 22:10:56,169 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Stopping nodemanagers
localhost: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
Stopping resourcemanager

# SOURCE CODE:

**hadoop@ubuntu:~$ start-all.sh**
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [ubuntu]
2023-03-05 04:56:47,216 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

**hadoop@ubuntu:~$ nano words.txt**
Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes

**hadoop@ubuntu:~$ nano wordMapper.py**
```
import sys
for line in sys.stdin:
    line=line.strip()
    words=line.split()
    for word in words:
        print('%s\t%s'%(word,"1"))
```

**hadoop@ubuntu:~$ nano wordReducer.py**
```
import sys

wordcount = { }

for line in sys.stdin:
    line = line.strip()
    word, count = line.split()
```

```
                    try:
                        wordcount[word] = wordcount[word] + int(count)
                    except:
                        wordcount[word] = int(count)

                for word in wordcount.keys():
                    print('%s\t%s' % (word, wordcount[word]))
```

**hadoop@ubuntu:~$ hdfs dfs -mkdir -p /Pro3**
2023-03-05 05:19:47,272 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hadoop@ubuntu:~$ hdfs dfs -ls /**
2023-03-05 05:20:00,556 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Found 4 items
drwxr-xr-x   - hadoop supergroup          0 2023-03-04 21:57 /BigData
drwxr-xr-x   - hadoop supergroup          0 2023-03-05 05:19 /**Pro3**
drwxr-xr-x   - hadoop supergroup          0 2023-03-04 21:43 /elan
drwx------   - hadoop supergroup          0 2023-03-04 21:52 /tmp

**hadoop@ubuntu:~$ hdfs dfs -chmod 700 /Pro3**
2023-03-05 05:20:26,829 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hadoop@ubuntu:~$ hdfs dfs -put words.txt /Pro3**
2023-03-05 05:21:14,052 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hadoop@ubuntu:~$ hdfs dfs -put wordMapper.py /Pro3**
2023-03-05 05:21:33,014 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable

**hadoop@ubuntu:~$ hdfs dfs -put wordReducer.py /Pro3**

2023-03-05 05:21:47,709 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -ls /Pro3**

2023-03-05 05:22:00,937 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 3 items

-rw-r--r--   1 hadoop supergroup      159 2023-03-05 05:21 /Pro3/**wordMapper.py**

-rw-r--r--   1 hadoop supergroup      342 2023-03-05 05:21 /Pro3/**wordReducer.py**

-rw-r--r--   1 hadoop supergroup      242 2023-03-05 05:21 /Pro3/**words.txt**

**hadoop@ubuntu: ~$ cat words.txt | python wordMapper.py | sort | python wordReducer.py**

| | |
|---|---|
| and | 2 |
| structured, | 1 |
| semi-structured | 1 |
| is | 1 |
| terabytes | 1 |
| analytics | 1 |
| diverse | 1 |
| in | 1 |
| techniques | 1 |
| data, | 1 |
| different | 2 |
| from | 2 |
| to | 1 |
| Big | 1 |
| analytic | 1 |
| large, | 1 |
| include | 1 |
| that | 1 |
| very | 1 |

| | |
|---|---|
| use | 1 |
| unstructured | 1 |
| data | 2 |
| advanced | 1 |
| zettabytes | 1 |
| sources, | 1 |
| sizes | 1 |
| of | 1 |
| against | 1 |
| sets | 1 |
| the | 1 |

**hadoop@ubuntu: ~$ hadoop jar /usr/local/hadoop/share/hadoop/ tools/lib/hadoop-streaming-\*.jar -D mapred.reduce.tasks=2 -file wordMapper.py -mapper "python wordMapper.py" -file wordReducer.py -reducer "python wordReducer.py" -input /Pro3/ words.txt -output /Pro3/WordCoutResult**

**hadoop@ubuntu: ~$ hdfs dfs -ls /Pro3/WordCoutResult**
2023-03-05 05:30:11,919 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
-rw-r--r--   1 hadoop supergroup        0 2023-03-05 05:29 /Pro3/**WordCoutResult/_SUCCESS**
-rw-r--r--   1 hadoop supergroup      141 2023-03-05 05:29 /Pro3/**WordCoutResult/part-00000**
-rw-r--r--   1 hadoop supergroup      135 2023-03-05 05:29 /Pro3/**WordCoutResult/part-00001**

**hadoop@ubuntu:~$ hdfs dfs -cat /Pro3/WordCoutResult/part\***
2023-03-05 05:31:28,786 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

| | |
|---|---|
| and | 2 |
| different | 2 |
| zettabytes | 1 |
| structured, | 1 |

| | |
|---|---|
| to | 1 |
| that | 1 |
| of | 1 |
| terabytes | 1 |
| sets | 1 |
| use | 1 |
| against | 1 |
| analytic | 1 |
| large, | 1 |
| in | 1 |
| the | 1 |
| techniques | 1 |
| data, | 1 |
| from | 2 |
| sources, | 1 |
| sizes | 1 |
| very | 1 |
| Big | 1 |
| is | 1 |
| unstructured | 1 |
| analytics | 1 |
| diverse | 1 |
| semi-structured | 1 |
| include | 1 |
| data | 2 |
| advanced | 1 |

**hadoop@ubuntu:~$ stop-all.sh**
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [0.0.0.0]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
2023-03-05 05:32:11,114 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Stopping nodemanagers
Stopping resourcemanager

## SOURCE CODE:

```
test> use BDA

BDA> db.stuInfo.insertOne(
{"_id" : 1, Name:'Elanchezhian', Age:'21', Course:'MCA'}
)

BDA> db.stuInfo.insertMany([
{"_id" : 2,Name : 'Dhanush', Age : '22', Course:'MSC CS'},
{"_id" : 3,Name : 'Hariharan', Age : '20', Course: 'MSC DS'},
{"_id" : 4,Name : 'Lachu', Age : '24', Course: 'MSc IT'}
])

BDA> db.stuInfo.find()

BDA> db.stuInfo.updateOne({ "_id":3 }, { $set: { Course: 'MSc IT' } })

BDA> db.stuInfo.updateMany({},{$set:{Batch:'2022'}})

BDA> db.stuInfo.find()

BDA> db.stuInfo.deleteOne({Name:'Lachu'})

BDA> db.stuinfo.deleteMany({Batch:'2022'})

BDA> db.stuInfo.find()
```

## OUTPUT:

- Insert Single Documents:

```
{ acknowledged: true, insertedId: 1 }
```

- Insert Many Documents:

```
{ acknowledged: true, insertedIds: { '0': 2, '1': 3, '2': 4 } }
```

- Database:

```
[
  { _id: 1, Name: 'Elanchezhian', Age: '21', Course: 'MCA' },
  { _id: 2, Name: 'Dhanush', Age: '22', Course: 'MSC CS' },
  { _id: 3, Name: 'Hariharan', Age: '20', Course: 'MSC DS' },
  { _id: 4, Name: 'Lachu', Age: '24', Course: 'MSc IT' }
]
```

- Update Single Document:

```
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
```

- Update Multiple Document:

```
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 4,
  modifiedCount: 4,
  upsertedCount: 0
}
```

➤ Database:

```
[
  {
    _id: 1,
    Name: 'Elanchezhian',
    Age: '21',
    Course: 'MCA',
    Batch: '2022'
  },
  {
    _id: 2,
    Name: 'Dhanush',
    Age: '22',
    Course: 'MSC CS',
    Batch: '2022'
  },
  {
    _id: 3,
    Name: 'Hariharan',
    Age: '20',
    Course: 'MSc IT',
    Batch: '2022'
  },
  { _id: 4, Name: 'Lachu', Age: '24', Course: 'MSc IT', Batch: '2022' }
]
```

➤ Delete Single Documents:

```
{ acknowledged: true, deletedCount: 1 }
```

➤ Delete Many Documents:

```
{ acknowledged: true, deletedCount: 3 }
```

## SOURCE CODE:

```
BDA > db.stuinfo.insertMany([
{ "_id": 1, Name: 'Divya', Age: '22', Course: 'MSC CS' },
{ "_id": 2, Name: 'Shalini', Age: '20', Course: 'MSC DS' },
{ "_id": 3, Name: 'Hariharan', Age: '24', Course: 'MSc IT' }
])


BDA > db.stuinfo.find()

BDA > db.stuinfo.find({Name:'Shalini'})

BDA > db.stuinfo.find({$or:[{Name:'Divya'},{Age:'20'}]})

BDA > db.stuinfo.find({Name:{$in:["Hariharan","Divya"]}})

BDA > db.stuinfo.find({ Age: { $gt:'21'}})

BDA > db.stuinfo.find({ Age: { $lt:'21'}})

BDA > db.stuinfo.find({ Name: { $eq:'Shalini'}})

BDA > db.stuinfo.find({ Name: { $nin:['Shalini']}})
```

## **OUTPUT:**

➤ Finding Documents:

```
[
  { _id: 1, Name: 'Divya', Age: '22', Course: 'MSC CS' },
  { _id: 2, Name: 'Shalini', Age: '20', Course: 'MSC DS' },
  { _id: 3, Name: 'Hariharan', Age: '24', Course: 'MSc IT' }
]
```

➤ AND Operator:

:

```
[ { _id: 2, Name: 'Shalini', Age: '20', Course: 'MSC DS' } ]
```

➤ OR Operator:

```
[
  { _id: 1, Name: 'Divya', Age: '22', Course: 'MSC CS' },
  { _id: 2, Name: 'Shalini', Age: '20', Course: 'MSC DS' }
]
```

➤ IN Operator:

```
[
  { _id: 1, Name: 'Divya', Age: '22', Course: 'MSC CS' },
  { _id: 3, Name: 'Hariharan', Age: '24', Course: 'MSc IT' }
]
```

➤ Greater than Operator:

```
[
  { _id: 1, Name: 'Divya', Age: '22', Course: 'MSC CS' },
  { _id: 3, Name: 'Hariharan', Age: '24', Course: 'MSc IT' }
]
```

➤ Less than Operator:

```
[ { _id: 2, Name: 'Shalini', Age: '20', Course: 'MSC DS' } ]
```

➢ Equal to Operator:

`

```
[ { _id: 2, Name: 'Shalini', Age: '20', Course: 'MSC DS' } ]
```

➢ NIN Operator:

```
[
  { _id: 1, Name: 'Divya', Age: '22', Course: 'MSC CS' },
  { _id: 3, Name: 'Hariharan', Age: '24', Course: 'MSc IT' }
]
```

## SOURCE CODE:

**rr@ubuntu:~$ start-all.sh**
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [ubuntu]
2023-03-04 21:29:38,334 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

**rr@ubuntu:~$ cd BDA**
**rr@ubuntu:~/BDA$ gedit emp.txt**

```
Elanchezhian,20,100000,CEO
Sivakumar,21,50000,Software_Devloper
Magudeshwaren,23,10000,Designer
Dhasush,22,80000,HR
Hariharan,23,4000,Cleaner
Sanjai,44,70000,Recptionist
Lachu,24,10000,Security
```

**rr@ubuntu:~/BDA$ hive**

**hive> CREATE DATABASE IF NOT EXISTS ELAN;**
OK
Time taken: 0.361 seconds

**hive> SHOW DATABASES;**
OK
default
elan
kera_company
Time taken: 0.571 seconds, Fetched: 3 row(s)

**hive> ALTER DATABASE ELAN SET DBPROPERTIES('CREATED BY'='ELANCHEZHIAN');**
OK
Time taken: 0.295 seconds

**hive> CREATE TABLE EMP_CAT(EMP_NAME STRING, AGE INT, SALARY FLOAT,EMP_ROLL STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE;**
OK
Time taken: 0.138 seconds

**hive> LOAD DATA LOCAL INPATH '/home/rr/BDA/emp.txt' INTO TABLE EMP_CAT;**
Loading data to table default.emp_cat
Table default.emp_cat stats: [numFiles=1, totalSize=193]
OK
Time taken: 0.424 seconds

**hive> SELECT * FROM EMP_CAT;**
OK
Elanchezhian      20    100000.0    CEO
Sivakumar  21     50000.0     Software_Devloper
Magudeshwaren  23     10000.0     Designer
Dhasush    22     80000.0     HR
Hariharan  23     4000.0      Cleaner
Sanjai     44     70000.0     Recptionist
Lachu      24     10000.0     Security
Time taken: 0.186 seconds, Fetched: 7 row(s)

**hive> CREATE TABLE EMP_CAT2 LIKE EMP_CAT;**
OK
Time taken: 0.177 seconds

**hive> INSERT INTO TABLE EMP_CAT2 SELECT * FROM EMP_CAT;**
Query ID = rr_20230327001239_f8973b2d-95eb-4d6d-9b68-473c3c68bd1e
Total jobs = 3

Launching Job 1 out of 3

.

.

OK

Time taken: 27.428 seconds

**hive> SELECT * FROM EMP_CAT2;**

OK

| | | | |
|---|---|---|---|
| Elanchezhian | 20 | 100000.0 | CEO |
| Sivakumar | 21 | 50000.0 | Software_Devloper |
| Magudeshwaren | 23 | 10000.0 | Designer |
| Dhasush | 22 | 80000.0 | HR |
| Hariharan | 23 | 4000.0 | Cleaner |
| Sanjai | 44 | 70000.0 | Recptionist |
| Lachu | 24 | 10000.0 | Security |

Time taken: 0.171 seconds, Fetched: 7 row(s)

**hive> SELECT * FROM EMP_CAT2 WHERE AGE > 21;**

OK

| | | | |
|---|---|---|---|
| Magudeshwaren | 23 | 10000.0 | Designer |
| Dhasush | 22 | 80000.0 | HR |
| Hariharan | 23 | 4000.0 | Cleaner |
| Sanjai | 44 | 70000.0 | Recptionist |
| Lachu | 24 | 10000.0 | Security |

Time taken: 0.295 seconds, Fetched: 5 row(s)

**hive> SELECT * FROM EMP_CAT2 WHERE SALARY >80000;**

OK

Elanchezhian     20     100000.0     CEO

Time taken: 0.159 seconds, Fetched: 1 row(s)

**rr@ubuntu:~$ stop-all.sh**
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [0.0.0.0]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
2023-03-05 05:32:11,114 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Stopping nodemanagers
Stopping resourcemanager

# SOURCE CODE:

**hadoop@ubuntu:~$ start-all.sh**

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [ubuntu]
2023-04-05 07:39:45,309 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

**hadoop@ubuntu:~$ gedit wCount.txt**

> Pig Represents Big Data as data flows.
> Pig is a high-level platform or tool which is used to process the large datasets.
> It provides a high-level of abstraction for processing over the MapReduce.
> It provides a high-level scripting language, known as Pig Latin which is used to develop the data analysis codes.

**hadoop@ubuntu:~$ hdfs dfs -put wCount.txt /elan**

2023-04-05 07:44:11,347 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -ls /elan**

2023-04-05 07:44:25,000 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 1 items
-rw-r--r--   1 hadoop supergroup        313 2023-04-05 07:44 /elan/**wCount.txt**

**hadoop@ubuntu:~$ gedit wCount.pig**

```
Input_Data = LOAD '/elan/wCount.txt' AS (line:chararray);
tozkenized_data = FOREACH Input_Data GENERATE
FLATTEN(TOKENIZE(line)) AS word;
grouped_data = GROUP tozkenized_data BY word;
word_count = FOREACH grouped_data GENERATE group AS word,
COUNT(tozkenized_data) AS count;
DUMP ordered_data;
```

**hadoop@ubuntu:~$ pig wCount.pig**

```
                (a,3)
                (It,2)
                (as,2)
                (is,3)
                (of,1)
                (or,1)
                (to,2)
                (Big,1)
                (Pig,3)
                (for,1)
                (the,3)
                (Data,1)
                (data,2)
                (over,1)
                (tool,1)
                (used,2)
                (Latin,1)
                (known,1)
                (large,1)
                (which,2)
                (codes.,1)
                (flows.,1)
                (develop,1)
                (process,1)
```

(analysis,1)
(language,1)
(platform,1)
(provides,2)
(datasets.,1)
(scripting,1)
(MapReduce.,1)
(Represents,1)
(high-level,3)
(processing,1)
(abstraction,1)

**hadoop@ubuntu:~$ stop-all.sh**
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [0.0.0.0]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
2023-04-05 07:39:45,309 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Stopping nodemanagers
Stopping resourcemanager

# SOURCE CODE:

**hadoop@ubuntu:~$ start-all.sh**

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [0.0.0.0]
Starting datanodes
Starting secondary namenodes [ubuntu]
2023-04-05 07:39:45,309 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

**hadoop@ubuntu:~$ gedit minMax.txt**

```
2000 45
2000 15
2001 26
2002 24
2003 21
2003 36
2004 26
```

**hadoop@ubuntu:~$ hdfs dfs -put minMax.txt /BigData**

2023-04-05 08:20:55,357 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -ls /BigData**

2023-04-05 08:28:38,199 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 6 items
drwxr-xr-x   - hadoop supergroup        0 2023-03-04 21:58 /BigData/P02RESULT
-rw-r--r--   1 hadoop supergroup       65 2023-04-05 08:20 /BigData/**minMax.txt**
-rw-r--r--   1 hadoop supergroup      161 2023-03-04 21:45 /BigData/minMaxMapper.py
-rw-r--r--   1 hadoop supergroup      554 2023-03-04 21:45 /BigData/minMaxReducer.py
-rw-r--r--   1 hadoop supergroup      136 2023-03-04 21:46 /BigData/temp.txt
-rw-r--r--   1 hadoop supergroup      131 2023-04-05 07:11 /BigData/wordCount.txt

**hadoop@ubuntu:~$ gedit temp.pig**

```
A = LOAD '/BigData/minMax.txt' using PigStorage (' ') AS
(year:int,temp:int);
B = GROUP A BY year;
C = FOREACH B GENERATE group, MAX(A.temp), MIN(A.temp);
DUMP C;
```

**hadoop@ubuntu:~$ pig temp.pig**

```
(2000,45,15)
(2001,26,26)
(2002,24,24)
(2003,36,21)
(2004,26,26)
(2005,37,37)
```

**hadoop@ubuntu:~$ stop-all.sh**
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [0.0.0.0]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
2023-04-05 07:39:45,309 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using builtin-java classes where
applicable
Stopping nodemanagers
Stopping resourcemanager

## SOURCE CODE:

**hadoop@ubuntu:~$ start-all.sh**

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [0.0.0.0]

Starting datanodes

Starting secondary namenodes [ubuntu]

2023-04-12 06:37:01,144 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Starting resourcemanager

Starting nodemanagers

**hadoop@ubuntu:~$ gedit employee.txt**

```
10012,Santhosh,25,35000

10021,Elanchezhian,26,45250

12045,Suresh,31,85000

14502,Karthi,23,65000

18545,Selva,21,15000
```

**hadoop@ubuntu:~$ gedit udf.py**

```
@outputSchema('num:long')

def get_length(data):

  str_data = ''.join([chr(x) for x in data])

  return len(str_data)
```

**hadoop@ubuntu:~$ hdfs dfs -put employee.txt /BDA**

2023-04-12 06:43:15,455 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -put udf.py /BDA**

2023-04-12 06:43:26,792 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -ls /BDA**

2023-04-12 06:44:01,322 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 2 items

-rw-r--r--   1 hadoop supergroup        117 2023-04-12 06:43 /BDA/**employee.txt**

-rw-r--r--   1 hadoop supergroup        120 2023-04-12 06:43 /BDA/**udf.py**

**hadoop@ubuntu:~$ pig**

 **grunt> REGISTER 'udf.py' USING jython as pyudf**

 **grunt> A = LOAD '/BDA/employee.txt' USING PigStorage();**

 **grunt> B = FOREACH A GENERATE $0,pyudf.get_length($0);**

 **grunt> DUMP B;**

  (10012,Santhosh,25,35000,23)

  (10021,Elanchezhian,26,45250,27)

  (12045,Suresh,31,85000,21)

  (14502,Karthi,23,65000,21)

  (18545,Selva,21,15000,20)

 **grunt> quit**

```
hadoop@ubuntu:~$ gedit emp_Del.txt

                    10012,Salem,MCA,Developer

                    10021,Erode,MCA,CEO

                    12045,Trichy,MCA,Manager

                    14502,Rahul,B.E,GM

                    18545,Rehan,M.E,MD

hadoop@ubuntu:~$ hdfs dfs -put emp_Del.txt  /BDA

hadoop@ubuntu:~$ pig

    grunt> A = LOAD '/BDA/employee.txt' USING PigStorage(',') AS
    (id:int,name:chararray,age:int,Salary:int);

    grunt> B = LOAD '/BDA/emp_Del.txt ' USING PigStorage(',') AS
    (id:int,location:chararray,Qualification:chararray,Designation:cha
    rarray);

    grunt> join_data = JOIN A by id,B by id;

        (10012,Santhosh,25,35000,10012,Salem,MCA,Developer)

        (10021,Elanchezhian,26,45250,10021,Erode,MCA,CEO)

        (12045,Suresh,31,85000,12045,Trichy,MCA,Manager)

        (14502,Karthi,23,65000,14502,Rahul,B.E,GM)

        (18545,Selva,21,15000,18545,Rehan,M.E,MD)

    grunt> SPLIT A INTO X IF age==25, Y IF age<=22;

    grunt> dump X;

        (10012,Santhosh,25,35000)

    grunt> dump Y;

        (18545,Selva,21,15000)

    grunt> grade_grp = GROUP A BY name;
```

```
grunt> grade_avg = FOREACH grade_grp GENERATE A.name,
A.age, AVG(A.Salary);

grunt> dump grade_avg;

        ({(Selva)},{(21)},15000.0)

        ({(Karthi)},{(23)},65000.0)

        ({(Suresh)},{(31)},85000.0)

        ({(Santhosh)},{(25)},35000.0)

        ({(Elanchezhian)},{(26)},45250.0)

grunt> cogroup_data = COGROUP A BY id,B BY id;

grunt> dump cogroup_data;

        (10012,{(10012,Santhosh,25,35000)},{(10012,Salem,MCA,Developer)})

        (10021,{(10021,Elanchezhian,26,45250)},{(10021,Erode,MCA,CEO)})

        (12045,{(12045,Suresh,31,85000)},{(12045,Trichy,MCA,Manager)})

        (14502,{(14502,Karthi,23,65000)},{(14502,Rahul,B.E,GM)})

        (18545,{(18545,Selva,21,15000)},{(18545,Rehan,M.E,MD)})

grunt> M = FOREACH A GENERATE TOMAP(name, age);

grunt> dump M;

                ([Santhosh#25])

                ([Elanchezhian#26])

                ([Suresh#31])

                ([Karthi#23])

                ([Selva#21])

grunt> quit;
```

**hadoop@ubuntu:~$ stop-all.sh**

WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: Use CTRL-C to abort.

Stopping namenodes on [0.0.0.0]

Stopping datanodes

Stopping secondary namenodes [ubuntu]

2023-04-12 07:24:45,390 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Stopping nodemanagers

Stopping resourcemanager

## SOURCE CODE:

**hadoop@ubuntu:~$ start-all.sh**

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [0.0.0.0]

Starting datanodes

Starting secondary namenodes [ubuntu]

2023-04-12 07:26:04,750 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Starting resourcemanager

Starting nodemanagers

**hadoop@ubuntu:~$ gedit dels.txt**

```
1,Elan,25,A

2,Dhanush,22,B

3,Lachu,30,C

4,Sanjai,27,B

5,Akil,24,A
```

**hadoop@ubuntu:~$ hdfs dfs -mkdir /BDA**

2023-04-12 06:42:59,540 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -chmod 755 /BDA**

2023-04-12 06:43:37,776 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -put dels.txt /BDA**

2023-04-12 07:34:08,843 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

**hadoop@ubuntu:~$ hdfs dfs -ls /BDA**

2023-04-12 07:34:30,977 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 4 items

-rw-r--r--   1 hadoop supergroup        66 2023-04-12 07:24 /BDA**/dels.txt**

**hadoop@ubuntu:~$ pig**

    **grunt> A = LOAD '/BDA/dels.txt ' USING PigStorage(',') as (id:int,name:chararray,age:int,grade:chararray);**

    **grunt> DUMP A;**

        (1,Elan,25,A)

        (2,Dhanush,22,B)

        (3,Lachu,30,C)

        (4,Sanjai,27,B)

        (5,Akil,24,A)

    **grunt> fewrecords = LIMIT A 2;**

    **grunt> DUMP fewrecords;**

        (1,Elan,25,A)

        (2,Dhanush,22,B)

**grunt> agelimit = FILTER A BY age>25;**

**grunt> DUMP agelimit;**

(3,Lachu,30,C)

(4,Sanjai,27,B)

**grunt> DESCRIBE A;**

A: {id: int,name: chararray,age: int,grade: chararray}

**grunt> groupbyage = GROUP A BY age;**

**grunt> DUMP groupbyage;**

(22,{(2,Dhanush,22,B)})

(24,{(5,Akil,24,A)})

(25,{(1,Elan,25,A)})

(27,{(4,Sanjai,27,B)})

(30,{(3,Lachu,30,C)})

**grunt> maxage = FOREACH groupbyage GENERATE(A.id, A.age), MAX(A.age);**

**grunt> DUMP maxage;**

(({(2)},{(22)}),22)

(({(5)},{(24)}),24)

(({(1)},{(25)}),25)

(({(4)},{(27)}),27)

(({(3)},{(30)}),30)

**grunt> allagevalues = GROUP A ALL;**

**grunt> DUMP allagevalues;**

(all,{(5,Akil,24,A),(4,Sanjai,27,B),(3,Lachu,30,C),(2,Dhanush,22,B),(1,Elan,25,A)})

**grunt> sumofage = FOREACH allagevalues GENERATE group,SUM(A.age);**

**grunt> dump sumofage;**

    (all,128)

**grunt> dist = DISTINCT A;**

**grunt> DUMP dist;**

    (1,Elan,25,A)

    (2,Dhanush,22,B)

    (3,Lachu,30,C)

    (4,Sanjai,27,B)

    (5,Akil,24,A)

**grunt> quit**

**hadoop@ubuntu:~$ stop-all.sh**

WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: Use CTRL-C to abort.

Stopping namenodes on [0.0.0.0]

Stopping datanodes

Stopping secondary namenodes [ubuntu]

2023-04-12 08:11:41,223 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Stopping nodemanagers

Stopping resourcemanager