

(2) Calculate the number of females (age between 18-30 years) who were victims in different crimes in different states.

CODE:

```
cage = LOAD '/user/training/Somaiya/crimeage' using PigStorage(',') AS
(state:chararray,crime:chararray,year:long,mblw18:long,fblw18:long,mbtw18_30:long,fbtw18_30:long,
mbtw30_45:long,fbtw30_45:long,mbtw45_60:long,fbtw45_60:long,mabv60:long,fabv60:long,tmale:long,tfemale:long);
```

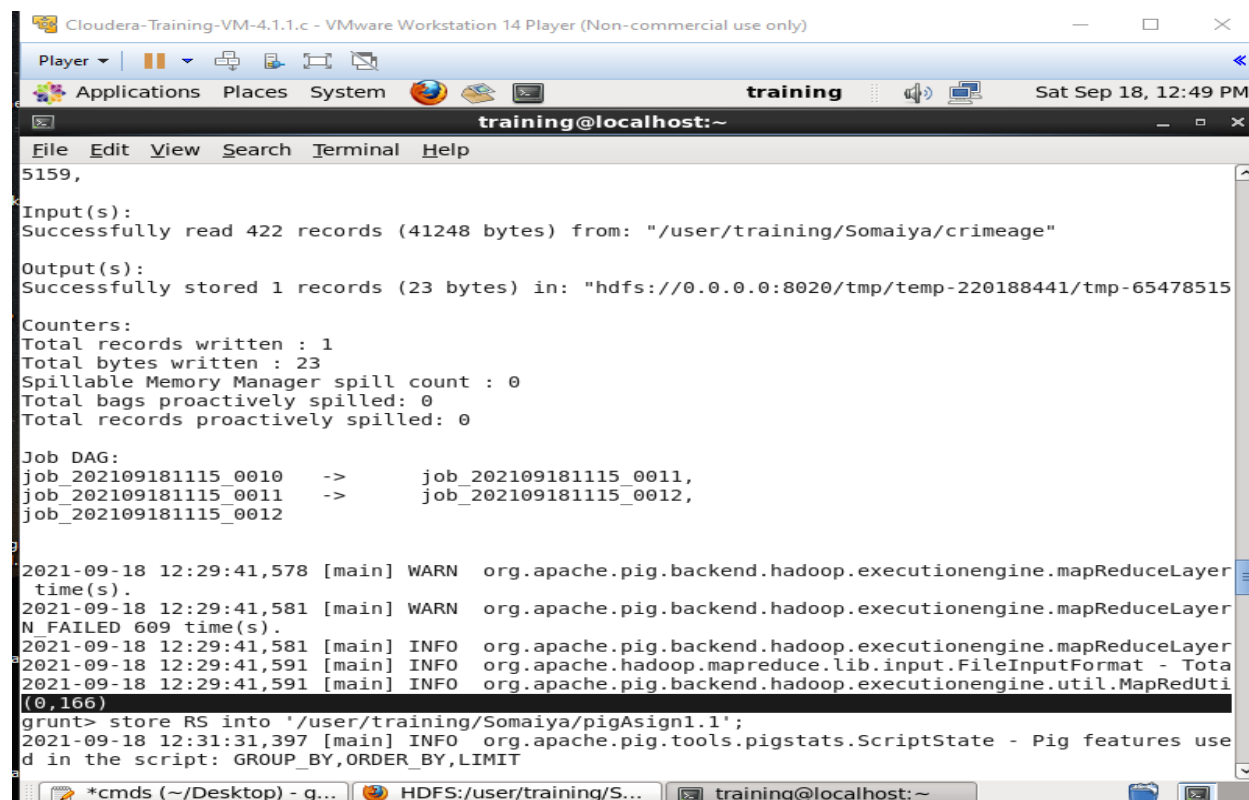
```
X = group cage by fbtw18_30;
```

```
Y = FOREACH X generate group, COUNT(cage);
```

```
RSLT = LIMIT Y 1;
```

```
dump RSLT;
```

OUTPUT:



The screenshot shows a terminal window titled "training@localhost:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output displays the execution of a Pig script. It starts with "Input(s):" showing 422 records read from "/user/training/Somaiya/crimeage". This is followed by "Output(s):" showing 1 record stored in HDFS. A "Counters:" section lists statistics like "Total records written : 1" and "Total bytes written : 23". A "Job DAG:" section shows a sequence of jobs. The bottom of the terminal shows log messages from the Pig execution engine, including warnings and info messages, and ends with the command "grunt> store RS into '/user/training/Somaiya/pigAssign1.1';".

```
5159,
Input(s):
Successfully read 422 records (41248 bytes) from: "/user/training/Somaiya/crimeage"

Output(s):
Successfully stored 1 records (23 bytes) in: "hdfs://0.0.0.0:8020/tmp/temp-220188441/tmp-65478515"

Counters:
Total records written : 1
Total bytes written : 23
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_202109181115_0010  ->      job_202109181115_0011,
job_202109181115_0011  ->      job_202109181115_0012,
job_202109181115_0012

2021-09-18 12:29:41,578 [main] WARN    org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
time(s).
2021-09-18 12:29:41,581 [main] WARN    org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
N FAILED 609 time(s).
2021-09-18 12:29:41,581 [main] INFO    org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
2021-09-18 12:29:41,591 [main] INFO    org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tota
2021-09-18 12:29:41,591 [main] INFO    org.apache.pig.backend.hadoop.executionengine.util.MapRedUti
(0,166)
grunt> store RS into '/user/training/Somaiya/pigAssign1.1';
2021-09-18 12:31:31,397 [main] INFO    org.apache.pig.tools.pigstats.ScriptState - Pig features use
d in the script: GROUP_BY,ORDER_BY,LIMIT
```

STORING OUTPUT ON HDFS:

Command: store RSLT into '/user/training/Somaiya/pigAssign1.1';

The screenshot shows a VMware Workstation 14 Player window titled "Cloudera-Training-VM-4.1.1.c - VMware Workstation 14 Player (Non-commercial use only)". Inside the VM, a Mozilla Firefox browser window is open with the title "HDFS:/user/training/Somaiya/pigAssign1.1/part-r-00000 - Mozilla Firefox". The browser's address bar shows the URL "localhost.localdomain:50075/browseBlock.jsp?blockId=223565540". The page content displays the file path "File: /user/training/Somaiya/pigAssign1.1/part-r-00000" and a "Goto:" field with the same path and a "go" button. Below this, there are links for "Go back to dir listing" and "Advanced view/download options". The main content area is mostly empty, with a small table header visible at the top left showing "0" and "166". The browser's toolbar includes "File", "Edit", "View", "History", "Bookmarks", "Tools", and "Help". The page's "Most Visited" section lists "Hadoop NameNode", "Hadoop JobTracker", "Hadoop YARN", "Hue", and "HBase Master". The bottom of the screen shows the taskbar with three open windows: "*cmds (~/Desktop) - q...", "HDFS:/user/training/S...", and "training@localhost: ~".

(4) Find crime wise which state is having highest count.

CODE:

```
cage = LOAD '/user/training/Somaiya/crimeage' using PigStorage(',') AS
(state:chararray,crime:chararray,year:long,mblw18:long,fblw18:long,mbtw18_30:long,fbtw18_30
:long,mbtw30_45:long,fbtw30_45:long,mbtw45_60:long,fbtw45_60:long,mabv60:long,fabv60:lon
g,tmale:long,tfemale:long);
```

A = group cage by state;

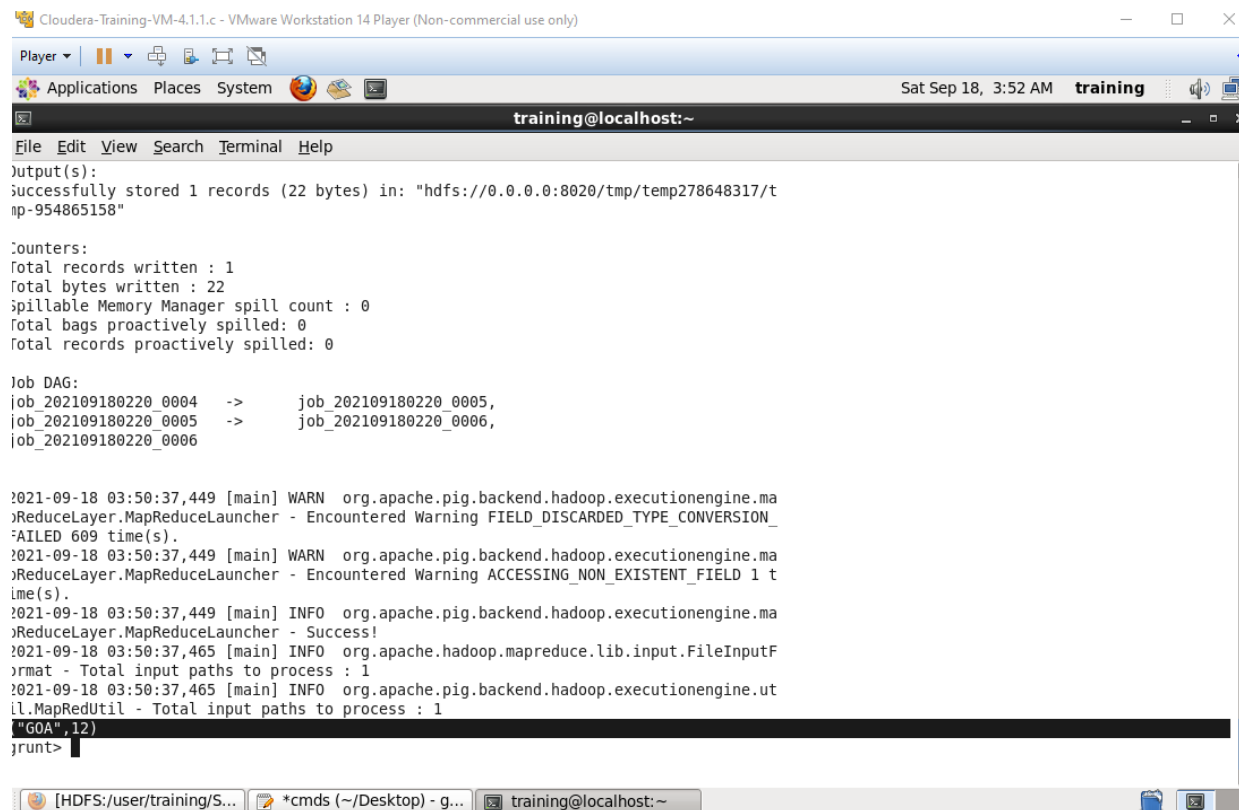
B = FOREACH A generate group, COUNT(cage);

C = order B by \$1 desc;

RS = LIMIT C

dump RS;

OUTPUT:



The screenshot shows a terminal window titled "training@localhost:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The output of a Pig script is displayed, showing that 1 record (22 bytes) was successfully stored in HDFS. It also shows job DAG information and various log messages from the Pig backend and Hadoop MapReduce. The logs include warnings about field conversion and accessing non-existent fields, as well as successful completion messages. The terminal ends with a prompt "jrunt>".

```
Output(s):
successfully stored 1 records (22 bytes) in: "hdfs://0.0.0.0:8020/tmp/temp278648317/t
mp-954865158"

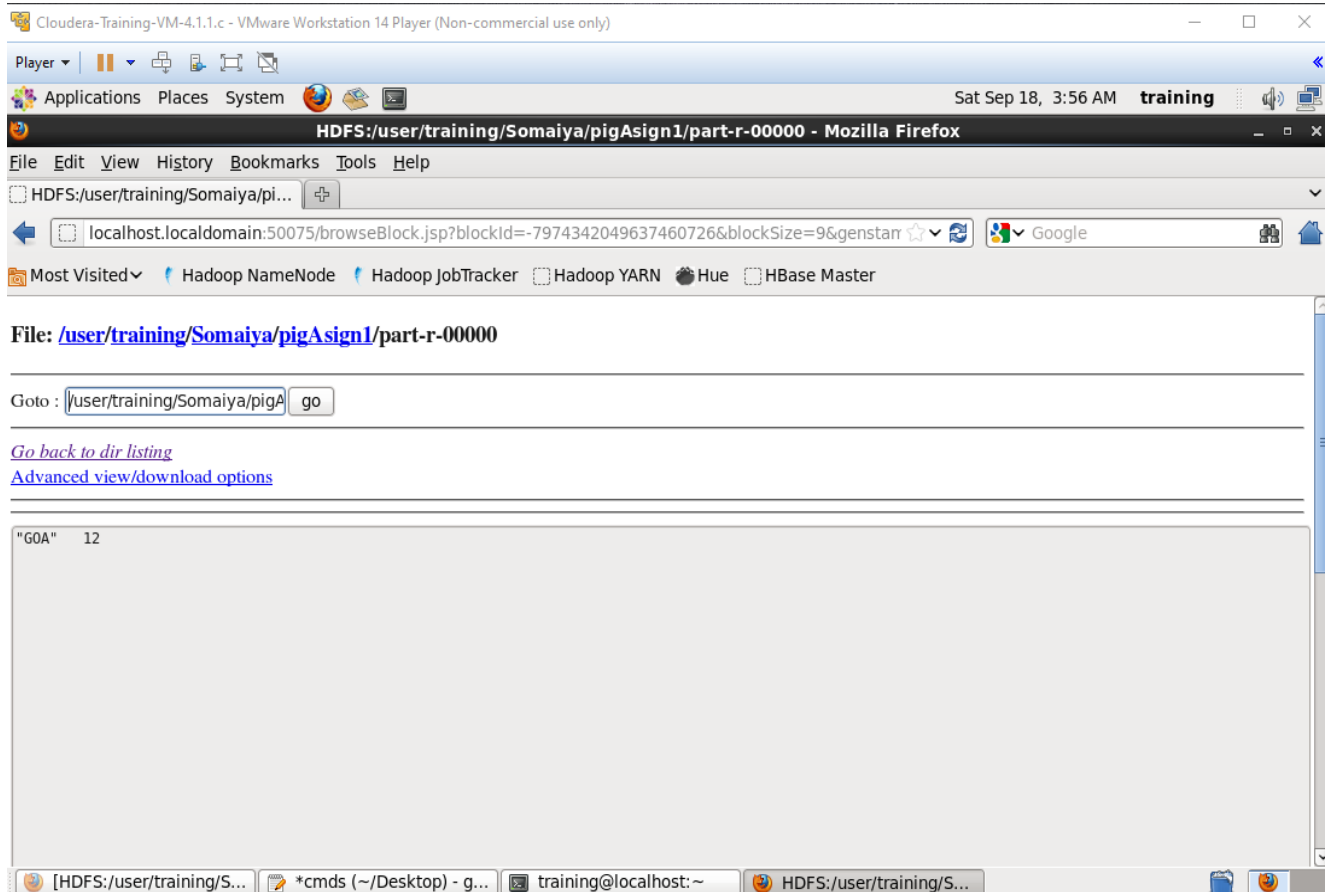
Counters:
Total records written : 1
Total bytes written : 22
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_202109180220_0004 -> job_202109180220_0005,
job_202109180220_0005 -> job_202109180220_0006,
job_202109180220_0006

2021-09-18 03:50:37,449 [main] WARN org.apache.pig.backend.hadoop.executionengine.ma
pReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_
FAILED 609 time(s).
2021-09-18 03:50:37,449 [main] WARN org.apache.pig.backend.hadoop.executionengine.ma
pReduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 1 t
ime(s).
2021-09-18 03:50:37,449 [main] INFO org.apache.pig.backend.hadoop.executionengine.ma
pReduceLayer.MapReduceLauncher - Success!
2021-09-18 03:50:37,465 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputF
ormat - Total input paths to process : 1
2021-09-18 03:50:37,465 [main] INFO org.apache.pig.backend.hadoop.executionengine.ut
il.MapRedUtil - Total input paths to process : 1
"G0A",12)
jrunt>
```

STORING OUTPUT ON HDFS:

Command: store RS into '/user/training/Somaiya/pigAssign1';



(1) To find the total number of crimes which occurred in all the states in the year 2006.

CODE:

```
cyear = LOAD '/user/training/Somaiya/crimeyear' using PigStorage(',') AS
(state:chararray,crime:chararray,y2001:int,y2002:int,y2003:int,y2004:int,y2005:int,y2006:int,y2007:int,y
2008:int,y2009:int,y2010:int,y2011:int,y2012:int);
```

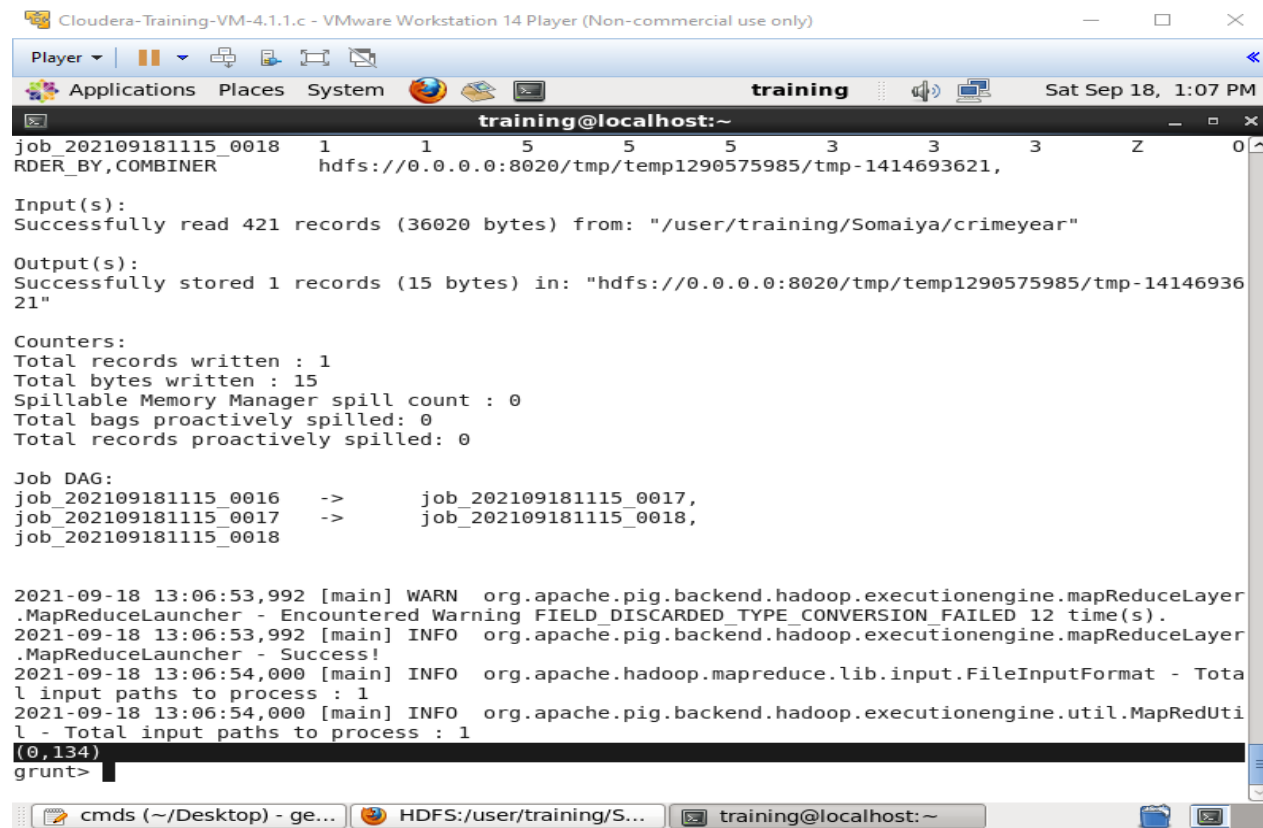
```
X = group cyear by y2006;
```

```
Y = FOREACH X generate group,COUNT(cage);
```

```
R = LIMIT Y 1;
```

```
dump R;
```

OUTPUT:



```
Cloudera-Training-VM-4.1.1.c - VMware Workstation 14 Player (Non-commercial use only)
Player | Applications Places System training Sat Sep 18, 1:07 PM
training@localhost:~
job_202109181115_0018 1 1 5 5 5 3 3 3 Z 0
RDER_BY,COMBINER hdfs://0.0.0.0:8020/tmp/temp1290575985/tmp-1414693621,

Input(s):
Successfully read 421 records (36020 bytes) from: "/user/training/Somaiya/crimeyear"

Output(s):
Successfully stored 1 records (15 bytes) in: "hdfs://0.0.0.0:8020/tmp/temp1290575985/tmp-1414693621"

Counters:
Total records written : 1
Total bytes written : 15
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_202109181115_0016 -> job_202109181115_0017,
job_202109181115_0017 -> job_202109181115_0018,
job_202109181115_0018

2021-09-18 13:06:53,992 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 12 time(s).
2021-09-18 13:06:53,992 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer
.MapReduceLauncher - Success!
2021-09-18 13:06:54,000 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total
input paths to process : 1
2021-09-18 13:06:54,000 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil
l - Total input paths to process : 1
(0,134)
grunt>
```

STORING OUTPUT ON HDFS:

Command: store R into '/user/training/Somaiya/pigAssign1.2';

