

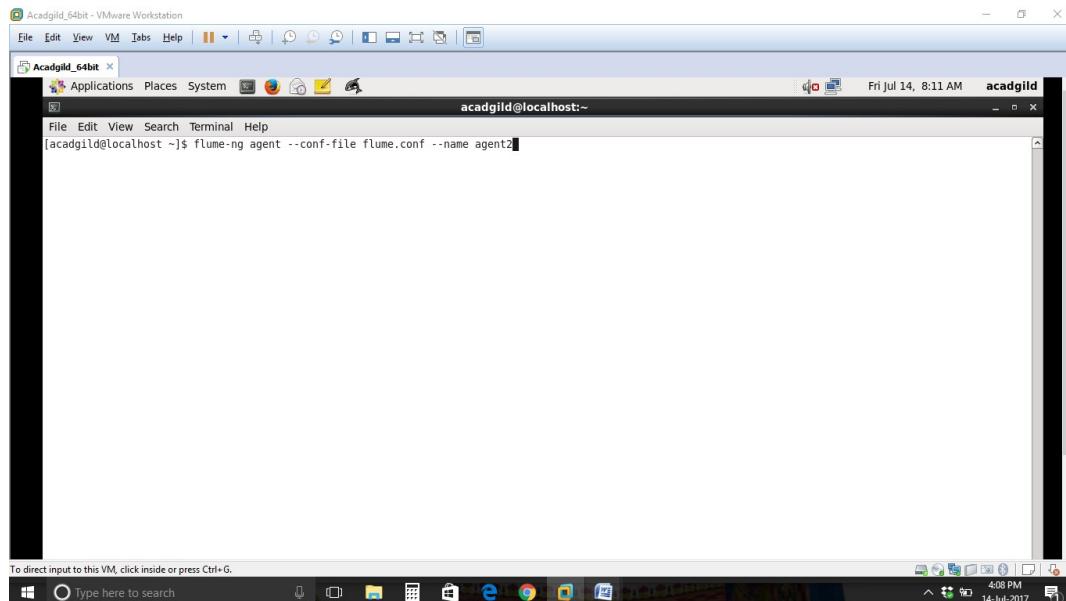
Project 2.2

USA CONSUMER FORUM DATA ANALYSIS

Task 1:

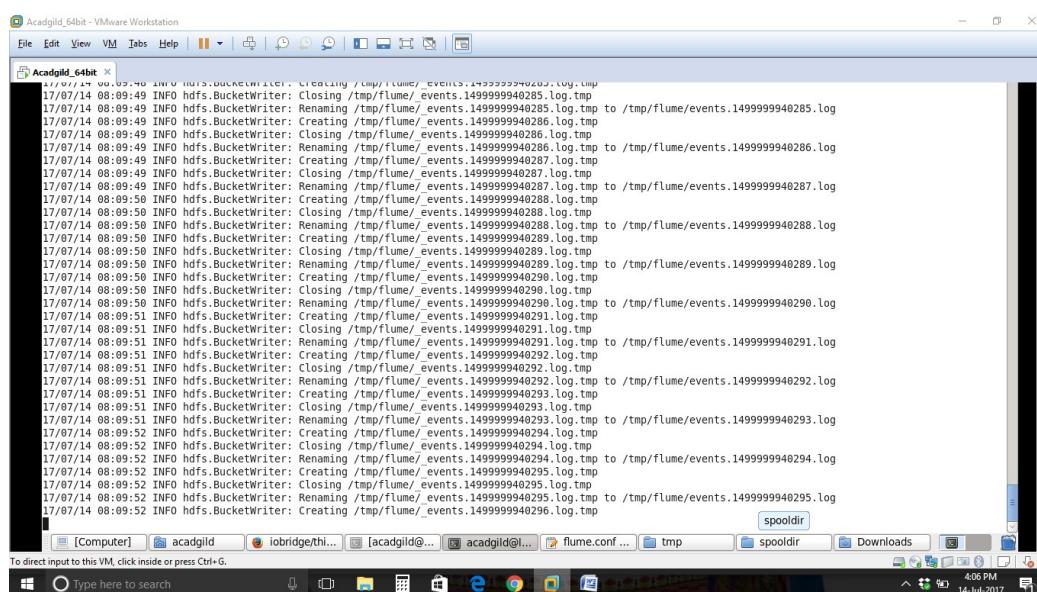
To copy the dataset into HDFS using Flume

Flume-ng agent --conf-file flume.conf --name agent2



```
[acadgild@localhost ~]$ flume-ng agent --conf-file flume.conf --name agent2
```

Flume agent



```
17/07/14 08:09:40 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940280.log.tmp
17/07/14 08:09:49 INFO hdfs.BucketWriter: Closing /tmp/flume/_events.149999940285.log.tmp
17/07/14 08:09:49 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940285.log.tmp to /tmp/flume/_events.149999940285.log
17/07/14 08:09:49 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940286.log.tmp
17/07/14 08:09:49 INFO hdfs.BucketWriter: Closing /tmp/flume/_events.149999940286.log.tmp to /tmp/flume/_events.149999940286.log
17/07/14 08:09:49 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940287.log.tmp
17/07/14 08:09:49 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940287.log.tmp to /tmp/flume/_events.149999940287.log
17/07/14 08:09:49 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940288.log.tmp
17/07/14 08:09:49 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940288.log.tmp to /tmp/flume/_events.149999940288.log
17/07/14 08:09:50 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940289.log.tmp
17/07/14 08:09:50 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940289.log.tmp to /tmp/flume/_events.149999940289.log
17/07/14 08:09:50 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940290.log.tmp
17/07/14 08:09:50 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940290.log.tmp to /tmp/flume/_events.149999940290.log
17/07/14 08:09:51 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940291.log.tmp
17/07/14 08:09:51 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940291.log.tmp to /tmp/flume/_events.149999940291.log
17/07/14 08:09:51 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940292.log.tmp
17/07/14 08:09:51 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940292.log.tmp to /tmp/flume/_events.149999940292.log
17/07/14 08:09:51 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940293.log.tmp
17/07/14 08:09:51 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940293.log.tmp to /tmp/flume/_events.149999940293.log
17/07/14 08:09:52 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940294.log.tmp
17/07/14 08:09:52 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940294.log.tmp to /tmp/flume/_events.149999940294.log
17/07/14 08:09:52 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940295.log.tmp
17/07/14 08:09:52 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940295.log.tmp to /tmp/flume/_events.149999940295.log
17/07/14 08:09:52 INFO hdfs.BucketWriter: Creating /tmp/flume/_events.149999940296.log.tmp
17/07/14 08:09:52 INFO hdfs.BucketWriter: Renaming /tmp/flume/_events.149999940296.log.tmp to /tmp/flume/_events.149999940296.log
```

CSV file copied into spool directory

The screenshot shows a VMware Workstation window titled "Acgild_64bit" running on a host machine named "Acgild_64bit - VMware Workstation". The window displays the Apache Flume log output from the command line, showing multiple instances of the `BucketWriter` class performing renaming and creation operations on log files. The log entries are timestamped from 08:10:23 to 08:10:27, with file names like `1499999940363.log` and `1499999940367.log`. In the foreground, a file browser window titled "spooldir" is open, showing a list of log files: `1499999940367.log`, `1499999940368.log`, `1499999940369.log`, `1499999940370.log`, and `1499999940371.log`. A blue box highlights the file "Consumer_Complaints.csv" which is selected and has a size of 106.4 MB. The status bar at the bottom of the file browser indicates "File / Edit / View / Places / Help" and "1499999940366.log". The taskbar at the bottom of the screen shows various application icons, and the system tray indicates the date as "14-Jul-2017" and the time as "4:06 PM".

File copying to HDFS done

Task 2

Write a pig script to find no of complaints which got timely response and copy the results into mysql using sqoop export

```
REGISTER 'home/acadgild/pig-0.14.0/lib/piggybank.jar';
```

```
DEFINE CSVLoader org.apache.pig.piggybank.storage.CSVLoader();
```

```
A = load 'user/acadgild/proj2/Consumer_Complaints.csv' using CSVLoader(',');
```

```
Dump A;
```

```
(12/01/2011,Credit card,,Credit card,Other fee,,,Capital One,HI,96813,Web,12/19/2011,Closed without relief,Yes,No,2219)
(12/01/2011,Credit card,,Credit card,Other,,,Capital One,CA,98632,Referral,12/01/2011,Closed without relief,Yes,No,2259)
(12/01/2011,Credit card,,Credit card,Identity theft / Fraud / Embezzlement,,,Capital One,NY,14224,Referral,12/01/2011,Closed without relief,Yes,No,2070)
(12/01/2011,Credit card,,Balance transfer / Fraud / Embezzlement,,,U.S. Bancorp,WA,98011,Referral,12/01/2011,Closed with relief,Yes,Yes,2225)
(12/01/2011,Mortgage,Other mortgage,Credit decision / Underwriting,,,JPMorgan Chase,NJ,08824,Referral,12/02/2011,Closed without relief,Yes,Yes,2404)
(12/01/2011,Credit card,,Balance transfer,,,U.S. Bancorp,WA,98011,Referral,12/01/2011,Closed with relief,Yes,No,2154)
(12/01/2011,Credit card,,Credit reporting,,,Bank of America,VA,20147,Referral,12/01/2011,Closed with relief,Yes,No,2422)
(12/01/2011,Mortgage,Conventional fixed mortgage,Loan modification,collection,foreclosure,,,Bank of America,OH,45248,Web,12/06/2011,Closed without relief,No,2323)
(12/01/2011,Credit card,,Other fee,,,Capital One,WA,99016,Phone,12/05/2011,Closed with relief,Yes,No,2278)
(12/01/2011,Credit card,,Closing/Cancelling account,,,Bank of America,CA,92011,Referral,12/07/2011,Untimely response,No,2260)
(12/01/2011,Credit card,,Identity theft / Fraud / Embezzlement,,,M&T Bank,PA,18692,Referral,12/06/2011,Closed with relief,Yes,No,3410)
(12/01/2011,Mortgage,FHA mortgage,Loan servicing, payments, escrow account,,,Bank of America,MD,20774,Web,12/02/2011,Closed without relief,Yes,No,2241)
(12/01/2011,Credit card,,Billing disputes,,,Bank of America,CA,90065,Referral,12/01/2011,Closed without relief,Yes,Yes,2298)
(12/01/2011,Credit card,,Credit determination,,,JPMorgan Chase,AZ,92262,Postal mail,12/02/2011,Closed without relief,Yes,No,2264)
(12/01/2011,Credit card,,Credit determination,,,Citibank,FL,33442,Referral,12/01/2011,Closed with relief,Yes,No,2088)
(12/01/2011,Mortgage,Conventional adjustable mortgage (ARM),Loan modification,collection,foreclosure,,,U.S. Bancorp,FL,33498,Web,12/05/2011,Closed without relief,Yes,No,2417)
(12/01/2011,Credit card,,Billing statement,,,Citibank,CA,92629,Web,12/05/2011,Closed with relief,Yes,No,2400)
(12/01/2011,Credit card,,Rewards,,,Barclays,IL,60022,Referral,12/01/2011,Closed with relief,Yes,No,2052)
(12/01/2011,Mortgage,Conventional fixed mortgage,Loan modification,collection,foreclosure,,,Bank of America,OR,97055,Phone,12/05/2011,Closed without relief,No,2328)
(12/01/2011,Credit card,,Other fee,,,Citibank,IL,60502,Referral,12/01/2011,Closed with relief,Yes,No,2200)
(12/01/2011,Mortgage,Conventional fixed mortgage,Application, originator, mortgage broker,,,Bank of America,NY,10023,Web,12/06/2011,Closed without relief,Yes,No,2275)
(12/01/2011,Credit card,,APR or interest rate,,,Citibank,CO,80503,Web,12/09/2011,Closed with relief,Yes,No,2327)
(12/01/2011,Credit card,,Overlimit fee,,,Wells Fargo,CA,94118,Referral,12/01/2011,Closed without relief,Yes,Yes,2065)
(12/01/2011,Credit card,,Balance transfer,,,Citibank,TX,75228,Web,12/02/2011,Closed with relief,Yes,No,2099)
(12/01/2011,Credit card,,Billing disputes,,,Citibank,VA,22304,Referral,12/01/2011,Closed without relief,Yes,Yes,2090)
(12/01/2011,Mortgage,Other mortgage,Loan modification,collection,foreclosure,,,SunTrust Bank,AR,85395,Referral,12/01/2011,Closed without relief,Yes,No,2092)
(12/01/2011,Credit card,,Billing disputes,,,Bank of America,MI,48081,Referral,12/01/2011,Closed without relief,Yes,Yes,2057)
(12/01/2011,Credit card,,Credit determination,,,JPMorgan Chase,VA,23889,Referral,12/01/2011,Closed without relief,Yes,No,2062)
(12/01/2011,Credit card,,Other fee,,,Citibank,CA,92037,Referral,12/01/2011,Closed with relief,Yes,No,2196)
grunt> [flume.conf (~) - edit]
```

B = Foreach A generate \$15 as id, \$13 as timely;

C= Filter B by timely='Yes';

D = Foreach C generate id,timely;

E = group D by timely;

F = Foreach E generate group as timely , COUNT(D.id) as numofcomplaints;

Dump F;

No of complaints resolved timely

```
Acadgild_64bit - VMware Workstation
File Edit View VMs Help ||| 
[Acadgild_64bit] successfully stored 0 TECLOGS 21.  HDFS://localhost:9000/clip/clip*190755550/clip214002/0

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

Job DAG:
job_1499928453184_0012

2017-07-14 09:11:51,942 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 09:11:51,995 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 09:11:52,160 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 09:11:52,160 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 09:11:52,377 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 09:11:52,419 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 09:11:52,646 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-07-14 09:11:52,647 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-07-14 09:11:52,652 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-07-14 09:11:52,652 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters
2017-07-14 09:11:52,652 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-14 09:11:52,652 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-07-14 09:11:52,681 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-07-14 09:11:52,681 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
[Yes, 45583]
grunt> [Click to switch to "Workspace 2"]
[Computer] [acadgild] [acadgild_64bit] [USA-Consu...] [acadgild@...] [acadgild@...] [flume.conf ...] [tmp] [spooldir] [Downloads]
To direct input to this VM, click inside or press Ctrl+G.
Windows Type here to search 14-Jul-2017 5:08 PM
```

Store Pig Latin output to HDFS

Store F into '/output/first' as PigStorage(',');

Table created in mysql:

```
Acadgild_64bit x
[Acadgild_64bit] org.apache.sqoop.tool.CodeGenTool.generateORM(CodeGenTool.java:96)
[Acadgild_64bit] at org.apache.sqoop.tool.ExportTool.exportTable(ExportTool.java:64)
[Acadgild_64bit] at org.apache.sqoop.tool.ExportTool.run(ExportTool.java:100)
[Acadgild_64bit] at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:143)
[Acadgild_64bit] at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
[Acadgild_64bit] at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:179)
[Acadgild_64bit] at org.apache.sqoop.Sqoop.runTool(Sqoop.java:218)
[Acadgild_64bit] at org.apache.sqoop.Sqoop.runTool(Sqoop.java:227)
[Acadgild_64bit] at org.apache.sqoop.Sqoop.main(Sqoop.java:236)
2017-07-14 16:33:21,517 ERROR [main] tool.ExportTool: Encountered IOException running export job: java.io.IOException: No columns to generate for ClassWriter
[acadgild64bit]@localhost ~ $ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 5
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> create table timely_response_table(timely String,complaints int);
ERROR 1064 (42000): You have an error in your SQL syntax, check the manual that corresponds to your MySQL server version for the right syntax to use near 'String,complaints int)' at line 1
mysql> create table timely_response_table(timely varchar(60),complaints int);
Query OK, 0 rows affected (0.02 sec)

mysql> flume.conf (~) - gedit
To direct input to this VM, click inside or press Ctrl+G.
```

Sqoop export of output in mysql table

```
Sqoop export --connect jdbc:mysql://localhost/db1 --username root -P --table timely_response_table -n 1 --export-dir '/output/first/'
```

```
at com.mysql.jdbc.Util.getInstance(Util.java:386)
at com.mysql.jdbc.SQLError.createSQLException(SQLError.java:1053)
at com.mysql.jdbc.MysqlIO.checkErrorPacket(MysqlIO.java:4074)
at com.mysql.jdbc.MysqlIO.checkErrorPacket(MysqlIO.java:4066)
at com.mysql.jdbc.MysqlIO.sendCommand(MysqlIO.java:2468)
at com.mysql.jdbc.MysqlIO.query(MysqlIO.java:2629)
at com.mysql.jdbc.ConnectionImpl.execSQL(ConnectionImpl.java:2719)
at com.mysql.jdbc.PreparedStatement.executeInternal(PreparedStatement.java:2155)
at com.mysql.jdbc.PreparedStatement.executeQuery(PreparedStatement.java:2318)
at org.apache.sqoop.manager.SqlManager.execute(SqlManager.java:750)
at org.apache.sqoop.manager.SqlManager.execute(SqlManager.java:759)
at org.apache.sqoop.manager.SqlManager.getColumnInfoForRawQuery(SqlManager.java:269)
at org.apache.sqoop.manager.SqlManager.getColumnsTypesForRawQuery(SqlManager.java:240)
at org.apache.sqoop.manager.SqlManager.getColumnTypes(SqlManager.java:226)
at org.apache.sqoop.manager.ConnManager.getColumnsTypes(ConnManager.java:295)
at org.apache.sqoop.orm.ClassWriter.generate(ClassWriter.java:1773)
at org.apache.sqoop.orm.ClassWriter.generate(ClassWriter.java:1578)
at org.apache.sqoop.tool.CodeGenTool.generateORM(CodeGenTool.java:96)
at org.apache.sqoop.tool.ExportTool.exportTable(ExportTool.java:64)
at org.apache.sqoop.tool.ExportTool.run(ExportTool.java:100)
at org.apache.sqoop.hadoop.util.ToolRunner.run(ToolRunner.java:70)
at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:179)
at org.apache.sqoop.Sqoop.runTool(Sqoop.java:218)
at org.apache.sqoop.Sqoop.runTool(Sqoop.java:227)
at org.apache.sqoop.Sqoop.main(Sqoop.java:236)
2017-07-14 09:26:14,375 [main] INFO [main] Sqoop: Running Sqoop version: 1.4.5
Enter password: [REDACTED]
To direct input to this VM, click inside or press Ctrl+G.
```

Select query shows no of complaints resolved timely is stored in table timely_response_table

```
at org.apache.sqoop.Sqoop.run(Sqoop.java:143)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:179)
at org.apache.sqoop.Sqoop.runTool(Sqoop.java:218)
at org.apache.sqoop.Sqoop.runTool(Sqoop.java:227)
at org.apache.sqoop.Sqoop.main(Sqoop.java:236)
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 9
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> select * from timely_response_table;
ERROR 1046 (3D000): No database selected
mysql> use db;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> select * from timely_response_table;
+-----+-----+
| timely | complaints |
+-----+-----+
| Yes   |    455383 |
+-----+-----+
1 row in set (0.00 sec)

mysql> [REDACTED]
```

Task 3

Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company

```
A = load '/user/acadgild/proj2/Consumer_Complaints.csv' using CSVLoader(',') as(f1:chararray, f2:chararray, f3:chararray , f4:chararray, f5:chararray, f6:chararray, f7:chararray, f8:chararray, f9:chararray, f10:chararray,f11:chararray,f12:chararray,f13:chararray,f14:chararray,f15:chararray,f16:chararray);
```

B = foreach A generate f1 as compdate,f12 as forwdate,f16 as id;

C =filter B by compdate==forwdate ;

D = group C all;

E= Foreach D generate COUNT(D.id)as same_day_forward;

Dump E;

No of complaints forwarded on same day of receiving

```

Input(s):
Successfully read 466874 records (892284792 bytes) from: "hdfs://localhost:9000/user/acadgild/proj2/Consumer_Complaints.csv"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp-1023326865/tmp-1942861270"

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

Job DAG:
job_local1517832922_0007

2017-07-14 10:28:23,925 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-07-14 10:28:23,926 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-07-14 10:28:23,926 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2017-07-14 10:28:23,951 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-07-14 10:28:23,951 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-14 10:28:23,951 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-07-14 10:28:23,951 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-07-14 10:28:23,951 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-07-14 10:28:24,159 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-07-14 10:28:24,159 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(161004)
grunt> flume.conf (~) -eedit
To direct input to this VM, click inside or press Ctrl+G.

```

store E into '/output/first' using PigStorage(',');

exit pig

hdfs dfs -ls /output/first

hdfs dfs -cat /output/first/part-m-00000

```

Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1499928453184_0021

2017-07-14 11:45:36,452 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 11:45:36,463 [main] INFO org.apache.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 11:45:36,661 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 11:45:36,754 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 11:45:36,998 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 11:45:37,004 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 11:45:37,260 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-07-14 11:45:37,260 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> quit;
2017-07-14 11:45:54,037 [main] INFO org.apache.pig.Main - Pig script completed in 5 minutes, 33 seconds and 701 milliseconds (333701 ms)
[acadgild@localhost ~]$ hdfs dfs -ls /output/second
17/07/14 11:46:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /output/first
17/07/14 11:46:20 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 acadgild supergroup 0 2017-07-14 11:45 /output/first/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 7 2017-07-14 11:45 /output/first/part-r-00000
[acadgild@localhost ~]$ hdfs dfs -ls /output/first/part-r-00000
17/07/14 11:46:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
-rw-r--r-- 1 acadgild supergroup 7 2017-07-14 11:45 /output/first/part-r-00000
[acadgild@localhost ~]$ hdfs dfs -cat /output/first/part-r-00000
17/07/14 11:46:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
161004
[acadgild@localhost ~]$ 

```

Table created in mysql same_day_forward with one column

Sqoop export of output in mysql table

```
Sqoop export --connect jdbc:mysql://localhost/db1 --username root -P --table same_day_forward -m 1 --export-dir '/output/first/'
```

```
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 10
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> create same_day_forward(same_day_fwd int);
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'same day forward(same_day_fwd int)' at line 1
mysql> create table same_day_forward(same_day_fwd int);
Query OK, 0 rows affected (0.09 sec)

mysql> exit;
Bye
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root -P --table same_day_forward -m 1 --export-dir '/output/first'
Warning: /usr/local/sqoop/_HCatalog does not exist! HCatalog jobs will fail.
Please set SHACCATL_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/_Accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/_zookeeper does not exist! Zookeeper imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
flume.conf(-)->edit
```

Select statement shows the output stored in HDFS from Pig is stored in table same_day_forward

```
Acadgild_64bit - VMware Workstation
File Edit View VMs Help ||| X
Acadgild_64bit X
[org.apache.sqoop.util.ToolRunner.run(Sqoop.java:143)
 at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
 at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:179)
 at org.apache.sqoop.Sqoop.runTool(Sqoop.java:218)
 at org.apache.sqoop.Sqoop.runTool(Sqoop.java:227)
 at org.apache.sqoop.Sqoop.main(Sqoop.java:236)
[acadgild@localhost ~]$ mysql -u root -p
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 13
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> select * from same_day_forward;
+-----+
| same_day_fwd |
+-----+
| 161004 |
+-----+
1 row in set (0.00 sec)

mysql> [spooldir]
Computer acadgild apache pig -... flume.conf (~...) tmp [spooldir] Downloads acadgild@loc...
To direct input to this VM, click inside or press Ctrl+G.
Windows Type here to search 7:48 PM 14-Jul-2017
```

Task 4

Write a pig script to find list of companies toping in complaint chart (companies with maximum number of complaints)

A = load'/Consumer_Complaints.csv' using CSVLoader(',');

B = foreach A generate \$7 as company,\$15 as id;

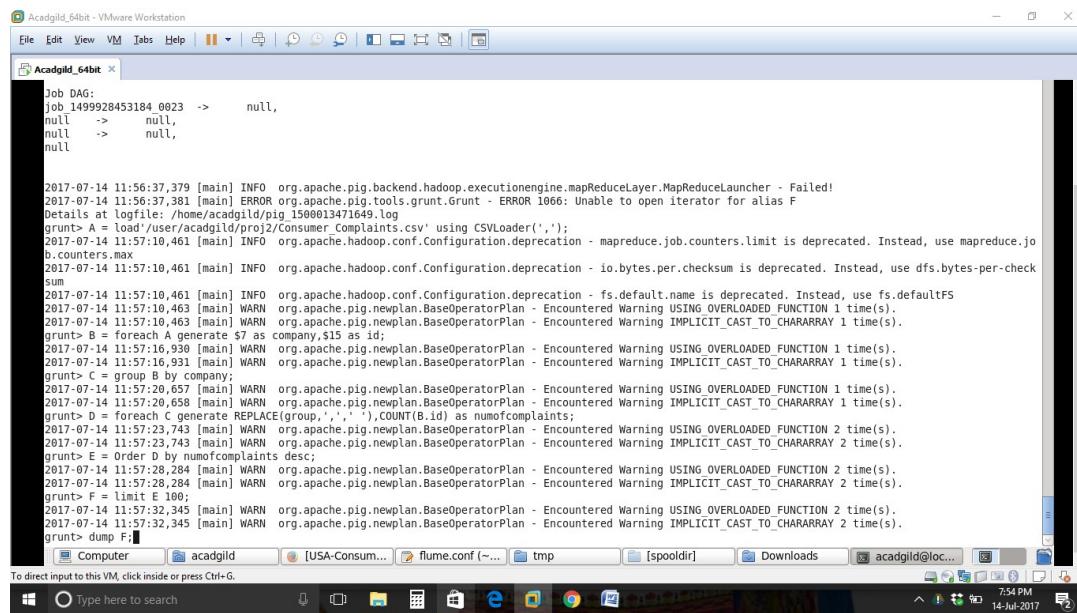
C = group B by company;

D = foreach C generate REPLACE(group,' ',''),COUNT(B.id) as numofcomplaints;

E = Order D by numofcomplaints desc;

F = limit E 100;

Dump F;

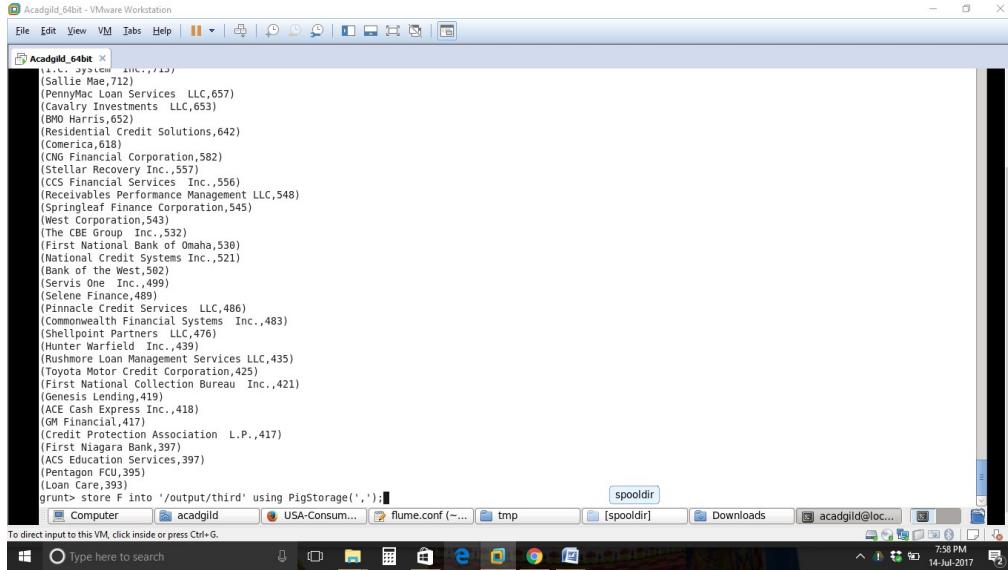


```
Job DAG:
    ab 1499928453184_0023 ->      null,
    null ->      null,
    null ->      null,
    null

2017-07-14 11:56:37,379 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Failed!
2017-07-14 11:56:37,381 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1066: Unable to open iterator for alias F
Details at logfile: /home/acadgild/proj2/Consumer Complaints.csv' using CSVLoader(',')
grunt> A = load'/user/acadgild/proj2/Consumer Complaints.csv' using CSVLoader(',');
2017-07-14 11:57:10,461 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.b.counters.max
2017-07-14 11:57:10,461 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-14 11:57:10,461 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-07-14 11:57:10,463 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-14 11:57:10,463 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> B = foreach A generate $7 as company,$15 as id;
2017-07-14 11:57:16,938 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-14 11:57:16,938 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> C = group B by company;
2017-07-14 11:57:20,659 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-14 11:57:20,659 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> D = foreach C generate REPLACE(group,' ',''),COUNT(B.id) as numofcomplaints;
2017-07-14 11:57:23,743 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2017-07-14 11:57:23,743 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> E = Order D by numofcomplaints desc;
2017-07-14 11:57:28,284 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2017-07-14 11:57:28,284 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> F = limit E 100;
2017-07-14 11:57:32,345 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 2 time(s).
2017-07-14 11:57:32,345 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
grunt> dump F;
```

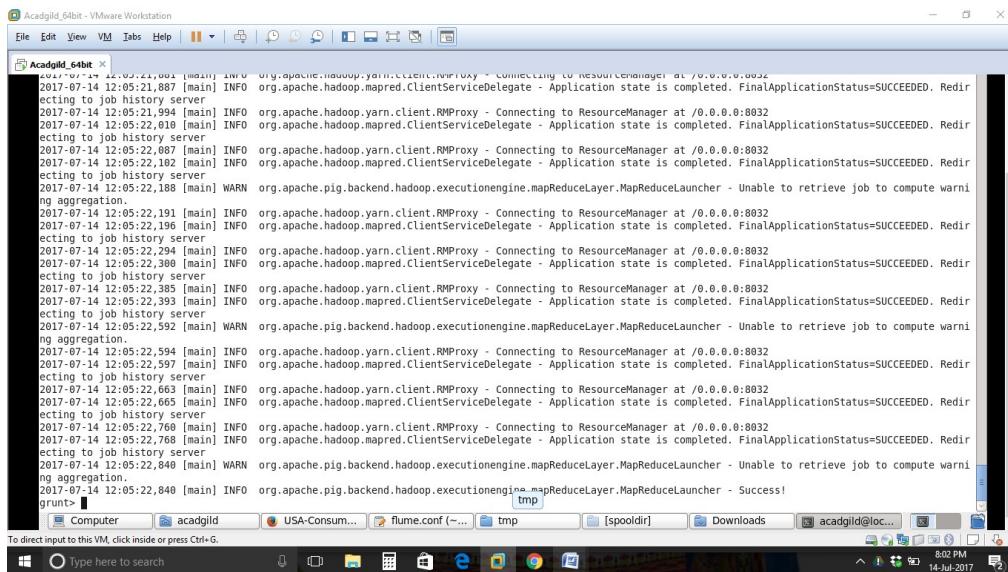
Top 100 companies with maximum number of complaints

Store F into '/output/Third' as PigStorage(',');



```
(Galilei,712)
(PennyMac Loan Services LLC,657)
(Cavalry Investments LLC,653)
(BMO Harris,652)
(Residential Credit Solutions,642)
(Comerica,618)
(CNG Financial Corporation,582)
(Stellar Recovery Inc.,557)
(TCS Financial Services Inc.,556)
(Rewables Performance Management LLC,548)
(Springleaf Finance Corporation,545)
(West Corporation,543)
(The CBE Group Inc.,532)
(First National Bank of Omaha,530)
(National Credit Systems Inc.,521)
(Bank of the West,502)
(Servis One Inc.,499)
(Sonesta Financial Services,486)
(Finagle Credit Services LLC,486)
(Commonwealth Financial Systems Inc.,483)
(Shellpoint Partners LLC,476)
(Hunter Warfield Inc.,439)
(Rushmore Loan Management Services LLC,435)
(Toyota Motor Credit Corporation,425)
(First National Collection Bureau Inc.,421)
(Genesis Lending,419)
(ACE Cash Express Inc.,418)
(OneMain Financial,417)
(Credit Protection Association L.P.,417)
(First Niagara Bank,397)
(ACS Education Services,397)
(Pentagon FCU,395)
(Loan Care,393)
grunt> store F into '/output/Third' using PigStorage(',');
```

Pig script stored in HDFS



```
2017-07-14 12:05:22,001 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,081 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,294 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,610 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,687 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,102 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,188 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation
2017-07-14 12:05:22,191 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,196 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,294 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,300 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,385 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,393 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,592 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation
2017-07-14 12:05:22,594 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,597 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,663 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,665 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,768 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,768 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:05:22,840 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation
2017-07-14 12:05:22,840 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

Table top_complaint_company with two columns company and complaints in mysql

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help ||| 
Acadgild_64bit X
Using MySQL 5.1.73 distribution.
2017-07-14 12:05:22,768 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:05:22,768 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redir
ecting to job history server
2017-07-14 12:05:22,849 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warni
ng aggregation
2017-07-14 12:05:22,849 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> quit;
2017-07-14 12:05:43,712 [main] INFO org.apache.pig.Main - Pig script completed in 11 minutes, 12 seconds and 360 milliseconds (672360 ms)
[acadgild@localhost ~]$ mysql -u root -p
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 14
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use dbl;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> create table top_complaint_company(name varchar(60),complaints int);
Query OK, 0 rows affected (0.05 sec)

mysql> commit;
Query OK, 0 rows affected (0.00 sec)

mysql> | Computer acadgild USA-Consum... flume.conf (~...) tmp [spooldir] Downloads acadgild@loc... To direct input to this VM, click inside or press Ctrl+G.
8:04 PM 14-Jul-2017
```

Sqoop export of output in mysql table

```
Sqoop export --connect jdbc:mysql://localhost/db1 --username root --password -P --table top_complaint_company -m 1 --export-dir '/output/Third/'
```

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help ||| 
Acadgild_64bit X
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 14
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use dbl;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> create table top_complaint_company(name varchar(60),complaints int);
Query OK, 0 rows affected (0.05 sec)

mysql> commit;
Query OK, 0 rows affected (0.00 sec)

mysql> exit;
Bye
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password -P --table top_complaint_company -m 1 --export-dir '/output/Third/'
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set SHCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
[acadgild@localhost ~]$
```

Select statement shows output of Pig script from HDFS is stored in top_complaint_company

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help || + | - | X | 
Acadgild_64bit
Copyright (c) 2000, 2015, Oracle and/or its affiliates. All rights reserved.
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> select * from top_complaint_company;
+-----+-----+
| name | complaints |
+-----+-----+
| Bank of America | 51127 |
| Wells Fargo | 37182 |
| JPMorgan Chase | 29583 |
| Experian | 24720 |
| Equifax | 24706 |
| Citibank | 22136 |
| TransUnion | 20054 |
| Ocwen | 18868 |
| Capital One | 13636 |
| Nationstar Mortgage | 11481 |
| U.S. Bancorp | 8267 |
| GE Capital Retail | 7898 |
| Ditech Financial LLC | 7455 |
| Navient | 6768 |
| PNC Bank | 6128 |
| HSBC | 5581 |
| Encore Capital Group | 5321 |
| Amex | 4756 |
+-----+-----+
spooldir
To direct input to this VM, click inside or press Ctrl+G.
Computer acadgild USA-Consum... flume.conf (~...) tmp [spooldir] Downloads acadgild@loc...
8:17 PM 14-Jul-2017
```

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help || + | - | X | 
Acadgild_64bit
Country Investments LLC | 650 |
BMO Harris | 652 |
Residential Credit Solutions | 642 |
Comerica | 618 |
CNG Financial Corporation | 582 |
Stellar Recovery Inc. | 557 |
CCS Financial Services Inc. | 556 |
Receivables Performance Management LLC | 548 |
Springleaf Finance Corporation | 545 |
West Corporation | 543 |
The CBE Group, Inc. | 532 |
First National Bank of Omaha | 530 |
National Credit Systems Inc. | 521 |
Bank of the West | 502 |
Servis One Inc. | 499 |
Selene Finance | 489 |
Pinnacle Credit Services LLC | 486 |
Commonwealth Financial Systems Inc. | 483 |
Shellpoint Partners LLC | 476 |
Hunter Warfield Inc. | 439 |
Rushmore Loan Management Services LLC | 435 |
Toyota Motor Credit Corporation | 425 |
First National Collection Bureau Inc. | 421 |
Gencis Lending | 419 |
ACE Cash Express Inc. | 418 |
GM Financial | 417 |
Credit Protection Association L.P. | 417 |
First Niagara Bank | 397 |
ACS Education Services | 397 |
Pentagon FCU | 395 |
Loan Care | 393 |
+-----+-----+
100 rows in set (0.00 sec)

mysql>
Computer acadgild USA-Consum... flume.conf (~...) tmp [spooldir] Downloads acadgild@loc...
8:18 PM 14-Jul-2017
```

Task 5

Write a pig script to find no of complaints filed with product type has “Debt collection” for the year 2015

```
A = load '/user/acadgild/proj2/Consumer_Complaints.csv' using CSVLoader(',') as  
(f1:chararray,f2:chararray);
```

```
B = filter A by f2=='Debt collection' and GetYear(ToDate(f1,'MM/dd/yyyy'))==2015;
```

```
C = foreach B generate f1 as complaintDate,f2 as product;
```

```
D = group C all;
```

```
E = Foreach D generate COUNT(C.product);
```

```
Dump E;
```

```
Acadgild_64bit - VMware Workstation  
File Edit View VM Tabs Help ||| 00:00:00.0000  
Acadgild_64bit x  
JVM-17992404-22104-0000  
2017-07-14 12:31:04,789 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-07-14 12:31:04,794 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redir  
ecting to job history server  
2017-07-14 12:31:04,933 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-07-14 12:31:04,943 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redir  
ecting to job history server  
2017-07-14 12:31:05,072 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032  
2017-07-14 12:31:05,079 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redir  
ecting to job history server  
2017-07-14 12:31:05,200 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warni  
ng aggregation  
2017-07-14 12:31:05,200 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2017-07-14 12:31:05,201 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
2017-07-14 12:31:05,201 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo  
b.counters.max  
2017-07-14 12:31:05,202 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2017-07-14 12:31:05,202 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2017-07-14 12:31:05,224 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2017-07-14 12:31:05,224 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1  
(3180)  
grunt> A = load '/user/acadgild/proj2/Consumer_Complaints.csv' using CSVLoader(',') as (f1:chararray,f2:chararray);  
2017-07-14 12:31:27,981 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo  
b.counters.max  
2017-07-14 12:31:27,981 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2017-07-14 12:31:27,981 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt> B = filter A by f2=='Debt collection' and GetYear(ToDate(f1,'MM/dd/yyyy'))==2015;  
grunt> C = foreach B generate f1 as complaintDate,f2 as product;  
grunt> D = group C all;  
grunt> E = foreach D generate COUNT(C.product);  
grunt> dump E;  
Downloads  
Computer acadgild USA-Consum... flume.conf (~...) tmp [spooldir] Downloads acadgild@loc...  
To direct input to this VM, click inside or press Ctrl+G.  
Windows Type here to search 829 PM 14-Jul-2017
```

Store E into '/output/Four' as PigStorage(',');

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help || + - x

Acadgild_64bit x
000 total time in seconds.
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A
alias Feature Outputs
job_1499928453184_0038 1 1 20 20 20 13 13 13 13 A,B,C,D,E GROUP_BY,COMBINER /output/Four,
Inputs:
Successfully read 0 records from: "/user/acadgild/proj2/Consumer_Complaints.csv"

Output(s):
Successfully stored 0 records in: "/output/Four"

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

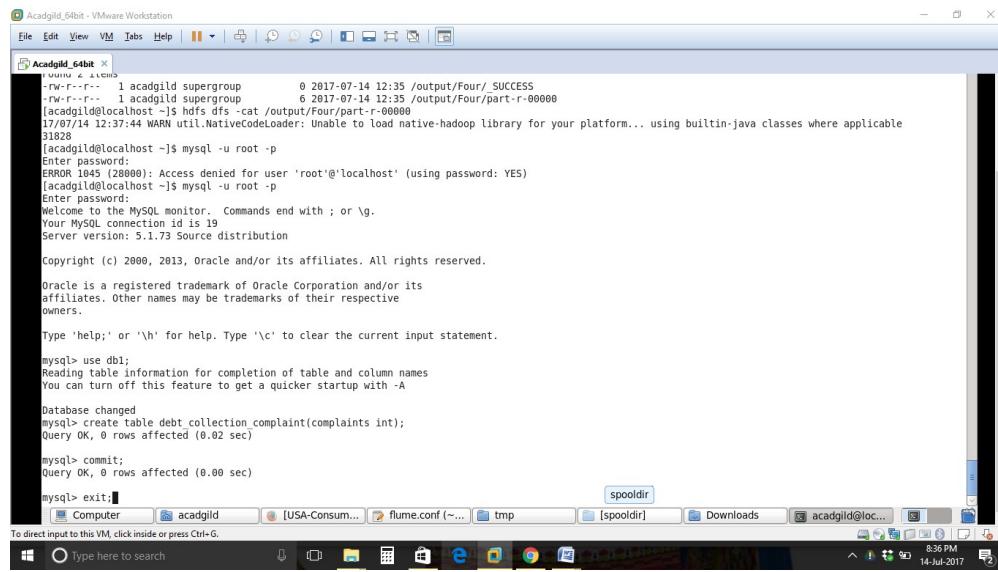
Job DAG:
job_1499928453184_0038

2017-07-14 12:35:56,291 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:35:56,299 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:35:56,392 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:35:56,403 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:35:56,524 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-07-14 12:35:56,555 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-07-14 12:35:56,599 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-07-14 12:35:56,599 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grant> Downloads
```

```
hdfs dfs -ls /output/Four
```

```
hdfs dfs -cat /output/Four/part-m-00000
```

Table created in mysql debt_collection_complaint



```
-rw-r--r-- 1 acadgild supergroup 0 2017-07-14 12:35 /output/Four/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 6 2017-07-14 12:35 /output/Four/part-r-00000
[acadgild@localhost ~]$ hdfs dfs -cat /output/Four/part-r-00000
17/07/14 12:37:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
31828
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
ERROR 1045 (28000): Access denied for user 'root'@'localhost' (using password: YES)
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 19
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

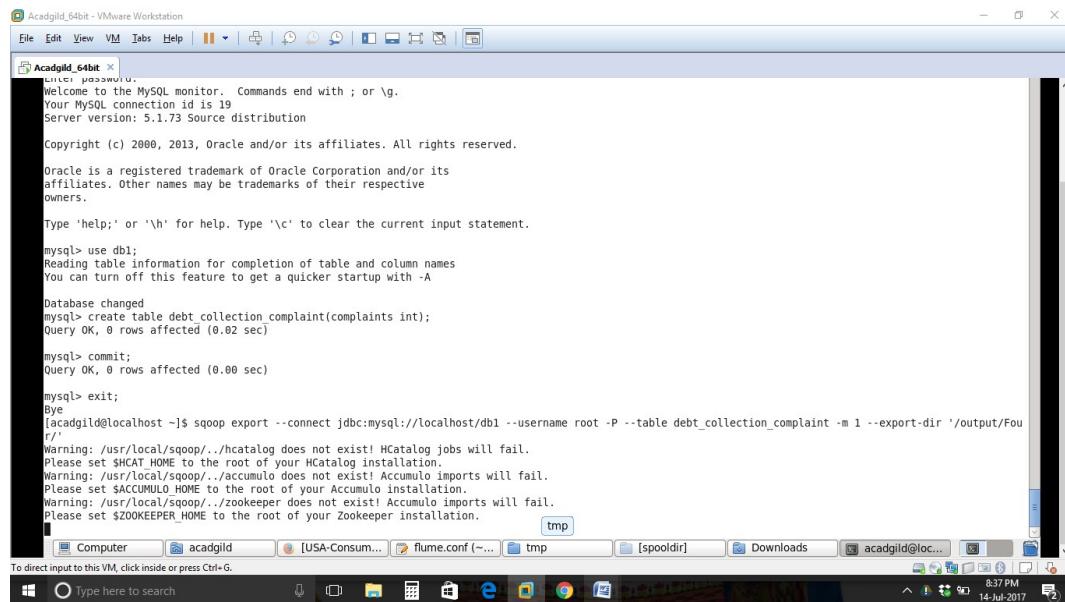
Database changed
mysql> create table debt_collection_complaint(complaints int);
Query OK, 0 rows affected (0.02 sec)

mysql> commit;
Query OK, 0 rows affected (0.00 sec)

mysql> exit;[spooldir]
[Computer] [acadgild] [USA-Consum...] [flume.conf (~...)] [tmp] [spooldir] [Downloads] [acadgild@loc...]
To direct input to this VM, click inside or press Ctrl+G.
```

Sqoop export of output in mysql table

```
Sqoop export --connect jdbc:mysql://localhost/db1 --username root -P --table timely_response_table -n 1 --export-dir '/output/first/'
```



```
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 19
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> create table debt_collection_complaint(complaints int);
Query OK, 0 rows affected (0.02 sec)

mysql> commit;
Query OK, 0 rows affected (0.00 sec)

mysql> exit;
Bye
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root -P --table debt_collection_complaint -m 1 --export-dir '/output/Four/'
Warning: /usr/local/sqoop..//hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop..//accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop..//zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
[Computer] [acadgild] [USA-Consum...] [flume.conf (~...)] [tmp] [spooldir] [Downloads] [acadgild@loc...]
To direct input to this VM, click inside or press Ctrl+G.
```

Select statement shows pig script output stored in HDFS is exported to mysql table debt_collection_complaints