Customer complaints about products

Use Case

edureka!

edureka!

Customer complaints about products

Table of Contents

Customer complaints about products

Problem Statement:

Consumer Complaint is a database that contains complaints received about financial products and services like Bank accounts or services, Consumer loans, Credit cards, Credit reporting, Debt collection, Money transfers, Mortgages, Private student loans, Payday loans, Prepaid cards, Other consumer loan (such as pawn & title loans), Other financial service (such as credit repair and debt settlement) to which company responds within a specific period of time. Here we are trying to analyze this data and below are the few problem statements that we have chosen:

a. Get the number of complaints filed under each products.

b. Get the total number of complaints filed from the state Florida (County Code: FL)

c. Get the list of complaints grouped by state which has no timely response.

Important Links:

Edureka VM Installation Guide:

Please refer to Installation guide section present in the LMS for accessing the Edureka VM Installation Guide.

Codes along with dataset:

https://edureka.wistia.com/medias/j61l5t4itz/download?media_file_id=67601485

Dataset Description:

http://www.consumerfinance.gov/complaintdatabase/technical-documentation/#field-reference

Tools and Technologies used:

- Map Reduce

- Pig

Implementation:

Use case 1:

Here we are writing the Map reduce program to get the number of complaints filed under each products.

Commands:

- **hadoop dfs -copyFromLocal /home/edureka/consumer_complaints.csv /**

- **hadoop dfs -ls /**

Copying the input file to the HDFS.

```
[edureka@localhost ~]$hadoop dfs -copyFromLocal /home/edureka/consumer_complaint
s.csv /    Copy input file to HDFS
DEPRECATED. use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:22:00 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[edureka@localhost ~]$hadoop dfs -ls /    Check if input file is in HDFS
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:23:45 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 31 items
-rw-r--r--    3 edurekalab supergroup         29 2014-12-18 11:41 /ashok.txt
drwxr-xr-x    - edurekalab supergroup          0 2014-12-22 16:21 /bank_project
drwxr-xr-x    - abhay      supergroup          0 2015-01-15 12:04 /binary
-rw-r--r--    3 edurekalab supergroup   53280007 2015-01-21 11:22 /consumer_compl
aints.csv
-rw-r--r--    3 abhay      supergroup        124 2015-01-18 15:55 /core-site.xml
```

We have list of people who have taken loan with particular grade. In mapper splitting the database based on comma delimiter after splitting: taking first part as complaint parts [0] taking second part as product parts [1].
In reducer for each key, we are running counter and taking the sum of all the values.

a). hadoop jar /home/edureka/consumerusecase1.jar /consumer_complaints.csv/consumerusecase1_output

Command to run map reduce program

```
[edureka@localhost ~]$hadoop jar /home/edureka/consumerusecase1.jar /consumer_complaints.csv
 /consumerusecase1_output
15/01/21 11:27:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
15/01/21 11:27:26 INFO client.RMProxy: Connecting to ResourceManager at opt-edu-057/192.168.
1.66:9003
15/01/21 11:27:26 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not perfor
med. Implement the Tool interface and execute your application with ToolRunner to remedy thi
s.
15/01/21 11:27:26 INFO input.FileInputFormat: Total input paths to process : 1
15/01/21 11:27:26 INFO mapreduce.JobSubmitter: number of splits:1
15/01/21 11:27:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1421491685121_
0051
15/01/21 11:27:27 INFO impl.YarnClientImpl: Submitted application application_1421491685121_
0051
15/01/21 11:27:27 INFO mapreduce.Job: The url to track the job: http://opt-edu-057:8088/prox
y/application_1421491685121_0051/
15/01/21 11:27:27 INFO mapreduce.Job: Running job: job_1421491685121_0051
15/01/21 11:27:31 INFO mapreduce.Job: Job job_1421491685121_0051 running in uber mode : fals
e
15/01/21 11:27:31 INFO mapreduce.Job:    map 0% reduce 0%
15/01/21 11:27:35 INFO mapreduce.Job:    map 100% reduce 0%
15/01/21 11:27:39 INFO mapreduce.Job:    map 100% reduce 100%
```

Output successful

- **hadoop dfs -ls /consumerusecase1_output**

- **hadoop dfs -cat /consumerusecase1_output/part-r-00000**

```
[edureka@localhost ~]$hadoop dfs -ls /consumerusecase1_output
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:29:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 edurekalab supergroup          0 2015-01-21 11:27 /consumerusecase1_output/_S
UCCESS
-rw-r--r--   3 edurekalab supergroup        185 2015-01-21 11:27 /consumerusecase1_output/pa
rt-r-00000
[edureka@localhost ~]$hadoop dfs -cat /consumerusecase1_output/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:30:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Bank account or service 39476
Consumer loan   10011
Credit card     42895
Credit reporting        43793
Debt collection 47920
Money transfers 1644
Mortgage        129548
Payday loan     1778
Student loan    9864
[edureka@localhost ~]$
```

Implementation:
Use case 2:

Here we are using the map reduce program to implement the total number of complaints filed from the state Florida (County Code: FL)

In the same data set. In mapper splitting database based on comma as delimiter and finding the complaints by state. In reducer for each key, we are running counter and taking the sum of all the values.

b). hadoop jar /home/edureka/consumerusecase2.jar /consumer_complaints.csv /consumerusecase2_output


```
[edureka@localhost ~]$hadoop jar /home/edureka/consumerusecase2.jar /consumer_complaints.csv
 /consumerusecase2_output
15/01/21 11:36:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
15/01/21 11:36:06 INFO client.RMProxy: Connecting to ResourceManager at opt-edu-057/192.168.
1.66:9003
15/01/21 11:36:06 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not perfor
med. Implement the Tool interface and execute your application with ToolRunner to remedy thi
s.
15/01/21 11:36:06 INFO input.FileInputFormat: Total input paths to process : 1
15/01/21 11:36:06 INFO mapreduce.JobSubmitter: number of splits:1
15/01/21 11:36:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1421491685121_
0052
15/01/21 11:36:06 INFO impl.YarnClientImpl: Submitted application application_1421491685121_
0052
15/01/21 11:36:06 INFO mapreduce.Job: The url to track the job: http://opt-edu-057:8088/prox
y/application_1421491685121_0052/
15/01/21 11:36:06 INFO mapreduce.Job: Running job: job_1421491685121_0052
15/01/21 11:36:11 INFO mapreduce.Job: Job job_1421491685121_0052 running in uber mode : fals
e
15/01/21 11:36:11 INFO mapreduce.Job:  map 0% reduce 0%
15/01/21 11:36:15 INFO mapreduce.Job:  map 100% reduce 0%
15/01/21 11:36:18 INFO mapreduce.Job:  map 100% reduce 100%
15/01/21 11:36:19 INFO mapreduce.Job: Job job_1421491685121_0052 completed successfully
```

Command to run map reduce program

- **hadoop dfs -ls /consumerusecase1_output**

- **hadoop dfs -cat /consumerusecase1_output/part-r-00000**

```
[edureka@localhost ~]$hadoop dfs -ls /consumerusecase2_output
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:39:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 edurekalab supergroup          0 2015-01-21 11:36 /consumerusecase2_output/_S
UCCESS
-rw-r--r--   3 edurekalab supergroup          9 2015-01-21 11:36 /consumerusecase2_output/pa
rt-r-00000
[edureka@localhost ~]$hadoop dfs -cat /consumerusecase2_output/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:39:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
FL      15616
```

Here we are loading the input file to Alias A, then input file is taken with no timely response
and that we are grouped by states.

c). A = load '/consumer_complaints.csv' using PigStorage(',') as (Complaintid:long,

**Product:chararray, Subproduct:chararray, Issue:chararray,Subissue:chararray,
State:chararray,
ZIPcode:int, Submittedvia:chararray, Datereceived:chararray,
Date_sent_to_company:chararray,Company:chararray, Companyresponse:chararray,
Timelyresponse:chararray, Consumerdisputed:chararray);
B = filter A by Timelyresponse=='No';**

C = group B by State;

D = foreach C generate group, B.Product; store D into '/consumerusecase3_output';

```
A = load '/consumer_complaints.csv' using PigStorage(',') as ( Complaintid:long, Product
:chararray, Subproduct:chararray, Issue:chararray,Subissue:chararray, State:chararray, Z
IPcode:int, Submittedvia:chararray, Datereceived:chararray, Date_sent_to_company:chararr
ay,Company:chararray, Companyresponse:chararray, Timelyresponse:chararray, Consumerdispu
ted:chararray);

B = filter A  by Timelyresponse=='No';         Filter the data by timely response

C = group B by State;     Group data by states

D = foreach C generate group, B.Product;

store D into '/consumerusecase3_output';       Store data
```

Command to load the data

pig /home/edureka/consumer_complaint_script.pig

> Command to run the program in Pig

```
[edureka@localhost ~]$pig /home/edureka/consumer_complaint_script.pig
15/01/21 11:49:54 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
15/01/21 11:49:54 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
15/01/21 11:49:54 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2015-01-21 11:49:54,805 [main] INFO  org.apache.pig.Main - Apache Pig version 0.13.0 (r1
606446) compiled Jun 29 2014, 02:27:58
2015-01-21 11:49:54,805 [main] INFO  org.apache.pig.Main - Logging error messages to: /h
ome/edurekalab/pig_1421821194802.log
2015-01-21 11:49:54,909 [main] WARN  org.apache.hadoop.util.NativeCodeLoader - Unable to
 load native-hadoop library for your platform... using builtin-java classes where applic
able
2015-01-21 11:49:54,996 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup fil
e /home/edurekalab/.pigbootup not found
2015-01-21 11:49:55,056 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

```
Input(s):
Successfully read 326929 records (53280373 bytes) from: "/consumer_complaints.csv"

Output(s):
Successfully stored 55 records (33941 bytes) in: "/consumerusecase3_output"

Counters:
Total records written : 55
Total bytes written : 33941
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1421491685121_0053


2015-01-21 11:51:41,412 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapRe
duceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED
 16821 time(s).
2015-01-21 11:51:41,412 [main] INFO  org.apa     d.hadoop.executionengine.mapRe
duceLayer.MapReduceLauncher - Success!
```

> Output Success

- **hadoop dfs -ls /consumerusecase3_output**
- **hadoop dfs cat /consumerusecase3_output/part-r-00000**

Command to check the output

```
[edureka@localhost ~]$hadoop dfs -ls /consumerusecase3_output
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:57:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y
our platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 edurekalab supergroup          0 2015-01-21 11:51 /consumerusecase3_outpu
t/_SUCCESS
-rw-r--r--   3 edurekalab supergroup      33941 2015-01-21 11:51 /consumerusecase3_outpu
t/part-r-00000
[edureka@localhost ~]$hadoop dfs -cat /consumerusecase3_output/part-r-00000
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/01/21 11:57:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y
our platform... using builtin-java classes where applicable
AK      {(Credit card),(Debt collection),(Credit card),(Student loan)}
AL      {(Student loan),(Bank account or service),(Debt collection),(Student loan),(Payd
ay loan),(Debt collection),(Debt collection),(Debt collection),(Bank account or service)
,(Debt collection),(Debt collection),(Debt collection),(Debt collection),(Debt collectio
n),(Debt collection),(Bank account or service),(Debt collection),(Debt collection),(Mort
gage),(Payday loan),(Credit card),(Debt collection),(Credit reporting),(Credit card),(Co
nsumer loan),(Credit card),(Credit card),(Mortgage),(Bank account or service),(Credit ca
rd),(Bank account or service),(Bank account or service)}
AR      {(Consumer loan),(Bank account or service),(Student loan),(Credit card),(Credit
```

We have successfully completed the project!!!!