



BIG DATA
DEVELOPMENT

Project 2

ACADGILD

Project 2 - USA Consumer Forum Data Analysis

Table of Contents

1. Introduction3
2. Objective3
3. Prerequisites3
4. Associated Data Files3
5. Problem Statement4
6. Approximate Time to Complete Task4

1. Introduction

The dataset is in csv format and contains the attributes pertaining to resolution of the consumer complaints.

2. Objective

3. Prerequisites

You should have Hadoop cluster installed in your system.

4. Associated Data Files

<https://drive.google.com/file/d/0B1QaXx7tpw3SQTlnQ0MzVW5HajA/view?usp=sharing>

Dataset Description

Below is the description of the data set

Column heading	index	Description
Date received	0	date on which consumer filed the complaint
Product	1	Type of the product
Sub-product	2	Sub product type
Issue	3	Issue faced by the consumer
Sub-issue	4	Any sub issues if exists
Consumer complaint narrative	5	Detailed description of complaint
Company public response	6	Company's public response to the complaint
Company	7	Name of the company
State	8	State from which consumer filed the complaint
ZIP code	9	Zip code
Submitted via	10	Channel from which complaint was submitted
Date sent to company	11	Date on which consumer forum forwarded the complaint to company
Company response to consumer	12	Company's response to the consumer
Timely response?	13	
Consumer disputed?	14	
Complaint ID	15	Unique complaint id

This data is comma delimited.

5. Problem Statement

Note: You need to copy the dataset into HDFS using Flume and the results of the problem statements should be exported into RDBMS(Mysql) using sqoop.

The aim of this project is to analyze performance of various companies on aspects like:

1. Write a pig script to find no of complaints which got timely response
2. Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company
3. Write a pig script to find list of companies toping in complaint chart (companies with maximum number of complaints)
4. Write a pig script to find no of complaints filed with product type has “ Debt collection” for the year 2015

Submit the screenshots of all the solutions with the source code.

6. Approximate Time to Complete Task

240 Minutes