Module 3: Assignment - 1

Flume JSON feed dataset

Solution:

Creating Table and Schema for loading the data

```
hive> CREATE TABLE if not exists tweets3 (
    > id BIGINT,
    > created_at STRING,
    > source STRING,
    > favorited BOOLEAN,
    > retweeted_status STRUCT<
    > text:STRING,
    > current_user_info:STRUCT<screen_name:STRING,name:STRING>,
    > retweet_count:INT>,
    > entities STRUCT<
    > urls:ARRAY<STRUCT<expanded_url:STRING>>,
    > user_mentions:ARRAY<STRUCT<screen_name:STRING,name:STRING>>,
    > hashtags:ARRAY<STRUCT<text:STRING>>>,
    > text STRING,
    > current_user STRUCT<
    > screen_name:STRING,
    > name:STRING,
    > friends_count:INT,
    > followers_count:INT,
    > statuses_count:INT,
    > verified:BOOLEAN,
    > utc_offset:INT,
    > time_zone:STRING>,
    > in_reply_to_screen_name STRING)
    > ROW FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe'
    > STORED AS TEXTFILE;
OK
Time taken: 0.081 seconds
```

Loading the Twitter Data from JSON file

```
hive> load data local inpath '/home/bitnami/Data/FlumeData.1470903929980' into table tweets3;
Loading data to table aug.tweets3
OK
Time taken: 0.573 seconds
```

Output:

```
hive>  select * from tweets3 limit 5;
OK
763652545684406272      Thu Aug 11 08:25:29 +0000 2016  <a href="http://twitter.com/download/android" rel="nofollow">Twitter f
or Android</a>  false   {"text":"My budget is 27k &amp; I want to buy Mobile, Harddisk, backpack, shoes &amp; watch which ones
 should i buy from @flipkart? #FlipkartFreedomSale","current_user_info":null,"retweet_count":36}         {"urls":[],"user_menti
ons":[{"screen_name":"JatinKiDuniya","name":"Jatin Israni"},{"screen_name":"Flipkart","name":"Flipkart"}],"hashtags":[{"text":
"FlipkartFreedomSale"}]}          RT @JatinKiDuniya: My budget is 27k &amp; I want to buy Mobile, Harddisk, backpack, shoes &amp
; watch which ones should i buy from @flipkart? #Flip…  NULL    NULL
763652595550392320      Thu Aug 11 08:25:41 +0000 2016  <a href="https://roundteam.co" rel="nofollow">RoundTeam</a>     false{
"text":"Trump vs. Clinton. Welke presidentskandidaat heeft de beste online marketing aanpak? https://t.co/9IF2nKET72 by @bloov
i","current_user_info":null,"retweet_count":1}  {"urls":[{"expanded_url":"http://bloo.vi/sqNDil"}],"user_mentions":[{"screen_n
ame":"PDM9","name":"Pieter De Moor"},{"screen_name":"bloovi","name":"Bloovi"}],"hashtags":[]}   RT @PDM9: Trump vs. Clinton. W
elke presidentskandidaat heeft de beste online marketing aanpak? https://t.co/9IF2nKET72 by @bloovi      NULL    NULL
763652619520925696      Thu Aug 11 08:25:47 +0000 2016  <a href="http://twitter.com/download/android" rel="nofollow">Twitter f
or Android</a>  false   {"text":"So, the #Flipkartfreedomsale is on &amp; I need some help. If you help me shop, I'll give 5 o
f you flipkart vouchers worth Rs.1000 each. Ready?","current_user_info":null,"retweet_count":2} {"urls":[],"user_mentions":[{"
screen_name":"lady_gabbar","name":"Puddin'"}],"hashtags":[{"text":"Flipkartfreedomsale"}]}       RT @lady_gabbar: So, the #Flip
kartfreedomsale is on &amp; I need some help. If you help me shop, I'll give 5 of you flipkart vouchers worth Rs.…     NULL N
ULL
763652639556894720      Thu Aug 11 08:25:51 +0000 2016  <a href="http://publicize.wp.com/" rel="nofollow">WordPress.com</a>  f
alse    NULL    {"urls":[{"expanded_url":"http://shopmava.in/computersoffice-gaming/sandisk-cruzer-blade-16-gb-utility-pendriv
e-offer-flipkart/"}],"user_mentions":[],"hashtags":[{"text":"Sandisk"},{"text":"Cruzer"},{"text":"Utility"},{"text":"Pendrive"
}]}     #Sandisk #Cruzer Blade 16 GB #Utility #Pendrive Flat 35% Off https://t.co/xNE5LsHqww https://t.co/nZZeWlmezL     NULL N
ULL
763652640999743488      Thu Aug 11 08:25:52 +0000 2016  <a href="http://twitter.com/download/android" rel="nofollow">Twitter f
or Android</a>  false   NULL    {"urls":[],"user_mentions":[{"screen_name":"JatinKiDuniya","name":"Jatin Israni"},{"screen_nam
e":"Flipkart","name":"Flipkart"}],"hashtags":[{"text":"FlipkartFreedomSale"}]}  @JatinKiDuniya you should buy all these things
 in your budget RS. 27000
@Flipkart #FlipkartFreedomSale https://t.co/YQ3EtKk2DS  NULL    JatinKiDuniya
Time taken: 0.201 seconds, Fetched: 5 row(s)
```