Module: 5 Assignment - 1

Find out the 'Tweet' aggregation data for your client regarding current trends in computer hardware.

Solution:

```
[training@quickstart ~]$ sudo mkdir -p /usr/lib/flume-ng/plugins.d/twitter-streaming/lib/
[training@quickstart ~]$ sudo mkdir -p /var/lib/flume-ng/plugins.d/twitter-streaming/lib/
[training@quickstart ~]$ sudo cp flume-sources-1.0-SNAPSHOT.jar /usr/lib/flume-ng/plugins.d/twitter-stre
aming/lib
[training@quickstart ~]$ sudo cp flume-sources-1.0-SNAPSHOT.jar /var/lib/flume-ng/plugins.d/twitter-stre
aming/lib
[training@quickstart ~]$ _
```

```
# Naming the components on the current agent.
TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=rUJVSkazumPiDlT4yVseElCvd
TwitterAgent.sources.Twitter.consumerSecret=Vma7u5vz8ACOX0R5rAAxJJBQs9kWllHIqJwmJE1gxrEZ5dLhq6
TwitterAgent.sources.Twitter.accessToken=36333727-6z1e6wdnZjZJgeEOmUqNzn1fzFzb5H9Br7o7AtaQY
TwitterAgent.sources.Twitter.accessTokenSecret=3ZCGaQGRBnFZvaDAVNkEbubTnSgHARUysFVgaJp5U8Nbm
TwitterAgent.sources.Twitter.keywords = corona,covid,vaccine,covishield
# Describing/Configuring the sink

TwitterAgent.sinks.HDFS.type = hdfs
TwitterAgent.sinks.HDFS.hdfs.path = /user/training/twitter_data
TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream
TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text
TwitterAgent.sinks.HDFS.hdfs.batchSize = 100
TwitterAgent.sinks.HDFS.hdfs.rollSize = 0
TwitterAgent.sinks.HDFS.hdfs.rollCount = 100

# Describing/Configuring the channel
TwitterAgent.channels.MemChannel.type = memory
TwitterAgent.channels.MemChannel.capacity = 10000
TwitterAgent.channels.MemChannel.transactionCapacity = 10000

# Binding the source and sink to the channel
TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sinks.HDFS.channel = MemChannel
```

```
[training@quickstart ~]$ mkdir tweets
[training@quickstart ~]$ cd tweets
[training@quickstart tweets]$ hdfs dfs -get twitter_data/* .
```
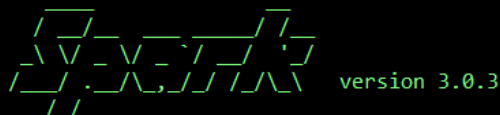
Flume Data:

```
[training@quickstart tweets]$ ls
FlumeData.1470903929980    FlumeData.1470905176680    FlumeData.1470906082940
FlumeData.1470903963401    FlumeData.1470905214802    FlumeData.1470906122881
FlumeData.1470903999254    FlumeData.1470905245296    FlumeData.1470906162262
FlumeData.1470904039070    FlumeData.1470905279976    FlumeData.1470906194012
FlumeData.1470904070593    FlumeData.1470905312972    FlumeData.1470906232091
FlumeData.1470904111578    FlumeData.1470905343135    FlumeData.1470906268425
FlumeData.1470904144695    FlumeData.1470905375654    FlumeData.1470906312221
FlumeData.1470904175333    FlumeData.1470905407650    FlumeData.1470906345218
FlumeData.1470904206419    FlumeData.1470905440691    FlumeData.1470906383031
FlumeData.1470904238309    FlumeData.1470905473696    FlumeData.1470906416588
FlumeData.1470904269932    FlumeData.1470905505378    FlumeData.1470906460955
FlumeData.1470904304203    FlumeData.1470905542889    FlumeData.1470906503524
FlumeData.1470904335330    FlumeData.1470905576197    FlumeData.1470906534756
FlumeData.1470904367783    FlumeData.1470905613160    FlumeData.1470906567502
FlumeData.1470904398555    FlumeData.1470905645142    FlumeData.1470906598555
FlumeData.1470904428795    FlumeData.1470905697375    FlumeData.1470906637365
FlumeData.1470904460731    FlumeData.1470905731992    FlumeData.1470906675512
FlumeData.1470904506442    FlumeData.1470905763075    FlumeData.1470906708410
FlumeData.1470904538115    FlumeData.1470905793423    FlumeData.1470906743501
FlumeData.1470904570645    FlumeData.1470905825504    FlumeData.1470906775758
FlumeData.1470904605718    FlumeData.1470905856201    FlumeData.1470906806249
FlumeData.1470904635814    FlumeData.1470905888006    FlumeData.1470906839189
FlumeData.1470904678371    FlumeData.1470905918112    FlumeData.1470906872439
FlumeData.1470904709229    FlumeData.1470905949739    FlumeData.1470906903050
FlumeData.1470904740793    FlumeData.1470905980383    FlumeData.1470906934882
FlumeData.1470904779108    FlumeData.1470906012608    FlumeData.1470906965846
FlumeData.1470905143957    FlumeData.1470906045826
```

Reading Flume Data:

[training@quickstart tweets]$ cat FlumeData.1470903929980
{"filter_level":"low","retweeted":false,"in_reply_to_screen_name":null,"truncated":false,"lang":"en","in_
reply_to_status_id_str":null,"id":763652545684406272,"in_reply_to_user_id_str":null,"timestamp_ms":"14709
03929526","in_reply_to_status_id":null,"created_at":"Thu Aug 11 08:25:29 +0000 2016","favorite_count":0,"
place":null,"coordinates":null,"text":"RT @JatinKiDuniya: My budget is 27k & I want to buy Mobile, Ha
rddisk, backpack, shoes & watch which ones should i buy from @flipkart? #Flip\u2026","contributors":n
ull,"retweeted_status":{"filter_level":"low","contributors":null,"text":"My budget is 27k & I want to
buy Mobile, Harddisk, backpack, shoes & watch which ones should i buy from @flipkart? #FlipkartFreed
omSale","geo":null,"retweeted":false,"in_reply_to_screen_name":null,"truncated":false,"lang":"en","entiti
es":{"symbols":[],"urls":[],"hashtags":[{"text":"FlipkartFreedomSale","indices":[123,143]}],"user_mention
s":[{"id":57947109,"name":"Flipkart","indices":[112,121],"screen_name":"Flipkart","id_str":"57947109"}]},
"in_reply_to_status_id_str":null,"is_quote_status":false,"id":763646872820920323,"source":"<a href=\"http
://twitter.com\" rel=\"nofollow\">Twitter Web Client<\/a>","in_reply_to_user_id_str":null,"favorited":fal
se,"in_reply_to_status_id":null,"retweet_count":36,"created_at":"Thu Aug 11 08:02:57 +0000 2016","in_repl
y_to_user_id":null,"favorite_count":18,"id_str":"763646872820920323","place":null,"user":{"location":"New
Delhi / Mumbai, India","default_profile":false,"profile_background_tile":false,"statuses_count":86951,"l
ang":"en","profile_link_color":"3B94D9","profile_banner_url":"https://pbs.twimg.com/profile_banners/38928
5374/1426368325","id":389285374,"following":null,"protected":false,"favourites_count":31302,"profile_text
_color":"000000","verified":false,"description":"Social Media Outreach Specialist #Traveler #Blogger #Foo
die #Techie! #Cricket #F1 #MSD #Yuvi #Kohli #Hrithik #PC Maniac!#FollowBack https://goo.gl/rOHNlJ","contr
ibutors_enabled":false,"profile_sidebar_border_color":"000000","name":"Jatin Israni","profile_background_
color":"000000","created_at":"Wed Oct 12 04:49:07 +0000 2011","default_profile_image":false,"followers_co
unt":27875,"profile_image_url_https":"https://pbs.twimg.com/profile_images/755468591521230848/FrcH00ph_no
rmal.jpg","geo_enabled":true,"profile_background_image_url":"http://abs.twimg.com/images/themes/theme1/bg
.png","profile_background_image_url_https":"https://abs.twimg.com/images/themes/theme1/bg.png","follow_re
quest_sent":null,"url":"http://jatinisrani.com/","utc_offset":-36000,"time_zone":"Hawaii","notifications"
:null,"profile_use_background_image":false,"friends_count":2062,"profile_sidebar_fill_color":"000000","sc
reen_name":"JatinKiDuniya","id_str":"389285374","profile_image_url":"http://pbs.twimg.com/profile_images/
755468591521230848/FrcH00ph_normal.jpg","listed_count":115,"is_translator":false},"coordinates":null},"ge
o":null,"entities":{"symbols":[],"urls":[],"hashtags":[{"text":"FlipkartFreedomSale","indices":[142,148]}
],"user_mentions":[{"id":389285374,"name":"Jatin Israni","indices":[3,17],"screen_name":"JatinKiDuniya","
id_str":"389285374"},{"id":57947109,"name":"Flipkart","indices":[131,140],"screen_name":"Flipkart","id_st
r":"57947109"}]},"is_quote_status":false,"source":"<a href=\"http://twitter.com/download/android\" rel=\"
nofollow\">Twitter for Android<\/a>","favorited":false,"in_reply_to_user_id":null,"retweet_count":0,"id_s
tr":"763652545684406272","user":{"location":"Kolkata, India","default_profile":false,"profile_background_
tile":false,"statuses_count":67276,"lang":"en","profile_link_color":"FF691F","profile_banner_url":"https:
//pbs.twimg.com/profile_banners/2963430699/1469080239","id":2963430699,"following":null,"protected":false
,"favourites_count":11891,"profile_text_color":"000000","verified":false,"description":"Ice-cream addict|

Viewing Data in Spark Dataframe:

```
21/11/06 17:52:54 WARN Utils: Your hostname, debian resolves to a loopback address: 127.0.0.1; using 192
.168.1.41 instead (on interface enp0s3)
21/11/06 17:52:54 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
21/11/06 17:52:56 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using
 builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://192.168.1.41:4040
Spark context available as 'sc' (master = local[*], app id = local-1636221191444).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.0.3
      /_/

Using Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_302)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val df = spark.read.json("tweets")
21/11/06 17:53:25 WARN SharedState: Not allowing to set spark.sql.warehouse.dir or hive.metastore.wareho
use.dir in SparkSession's options, it should be set statically for cross-session usages
21/11/06 17:53:37 WARN package: Truncated the string representation of a plan since it was too large. Th
is behavior can be adjusted by setting 'spark.sql.debug.maxToStringFields'.
df: org.apache.spark.sql.DataFrame = [contributors: string, coordinates: struct<coordinates: array<doubl
e>, type: string> ... 29 more fields]

scala> val df1 = df.select("text")
df1: org.apache.spark.sql.DataFrame = [text: string]
```

Output:

```
scala> df1.show(100,false)
+------------------------------------------------------------------------------
----------------------------------------+
|text
                                        |
+------------------------------------------------------------------------------
----------------------------------------+
|Don't miss the opportunity to win Flipkart vouchers.Be a part of #HitToKill contest nd win big with @HI
TKPKD .Visit https://t.co/IU6pZXzOKv    |
|RT @savemoneyindiaa: Flipkart – Buy Marvo Scorpion Unicorn Wired Headset at Rs 674 only https://t.co/rf
Tqw4giLZ https://t.co/xk6w3DWchs        |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @mrjintheuk: Buy some flip-flops and... done. https://t.co/wijHi79heg
                                        |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|Real-time updates for https://t.co/3tRK45n8lU online marketing strategy. https://t.co/wbgHxJnfMp
                                        |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|RT @HITKPKD: @JatinKiDuniya That's definitely a cool way! Share more ways of killing to win flipkart vo
uchers! #HitToKill https://t.co/CCH4…    |
|Reliance LYF Flame 8, Wind 3 on Flipkart; offers Jio 4G SIM with free data - The Indian Express https:/
/t.co/KQm9nrEDuf                         |
|Amazon, Flipkart, Snapdeal sale violates FDI norms: Traders body CAIT https://t.co/h5GWPkAezD https://t
```

Output:

```
                                              |
|Online #marketing voor #webwinkels, voor #tips en #advies over #interactie en #communicatie sta ik voor
 je klaar!   https://t.co/UWMtcnIRnO        |
|I ♥ this product on Flipkart! Hawkeyed 10 Card Holder https://t.co/P55A1gxWOw #buyonflipkart via @Flipk
art                                         |
|RT @SpeedySteamer: Affordable Online and Offline Marketing https://t.co/KJC5kdsmQg via @YouTube
                                              |
|Respite For #Flipkart: US Fund Valic Marks Up Its Valuation By 10%. Read more: https://t.co/b8OeNlVva3
                                              |
|RT @SpeedySteamer: Affordable Online and Offline Marketing https://t.co/KJC5kdsmQg via @YouTube
                                              |
|RT @SpeedySteamer: Affordable Online and Offline Marketing https://t.co/pB1O35gPZ3 via @YouTube
                                              |
|RT @krmayank13: @JatinKiDuniya @Flipkart Ye lo Rs. 27 ko bachat ke saath apke liye branded products fro
m #FlipkartFreedomSale. https://t.co…      |
|RT @SpeedySteamer: Affordable Online and Offline Marketing https://t.co/KJC5kdsmQg via @YouTube
                                              |
|Marketing with Strategic Empathy by Claire Brooks Paperback Book (English) https://t.co/bEpBjoa38f http
s://t.co/paVsb2b5Xz                         |
|Online Marketing Manager (m/w) für Karlsruhe gesucht -  https://t.co/GxkS3VRdVC
                                              |
|RT @ahmershahab: @kamlesh_maru @Flipkart welldone
                                              |
|RT @SpeedySteamer: Affordable Online and Offline Marketing https://t.co/KJC5kdsmQg via @YouTube
                                              |
+---------------------------------------------------------------------------------------------------
---------------------------------------+
only showing top 100 rows
```