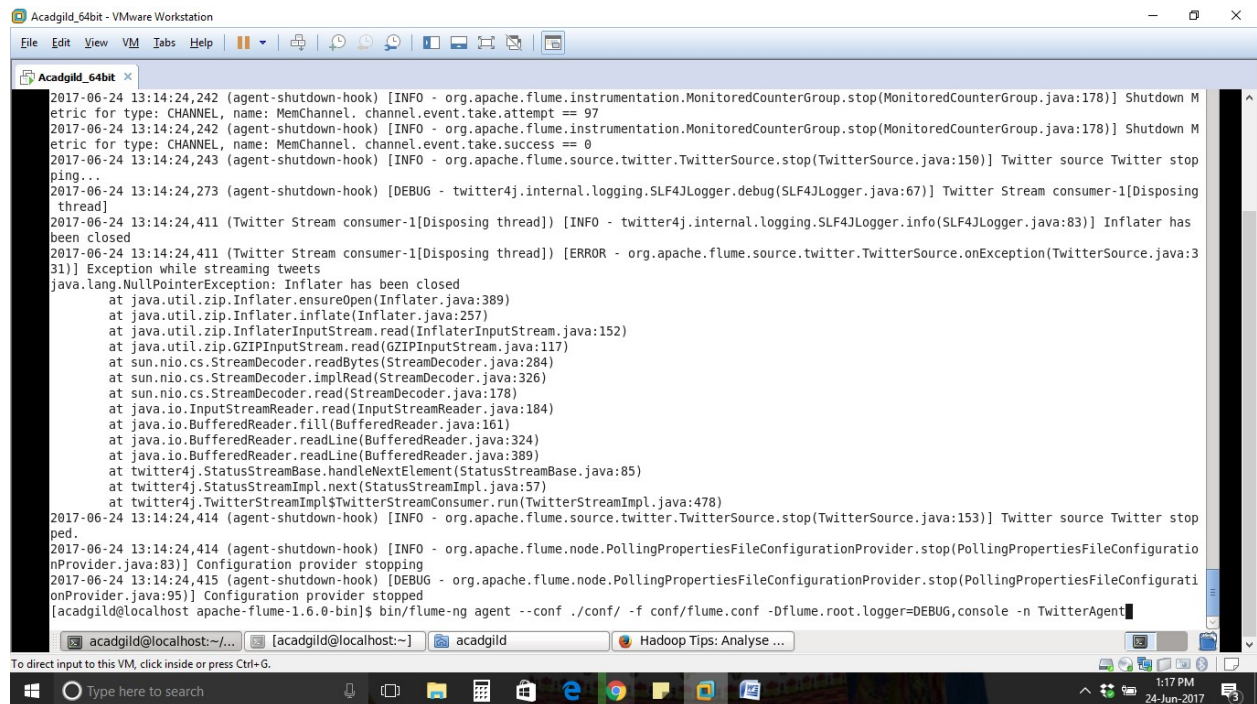


Streaming Twitter Data in HDFS using Flume

Create a flume agent that streams data from Twitter and stores in the HDFS.

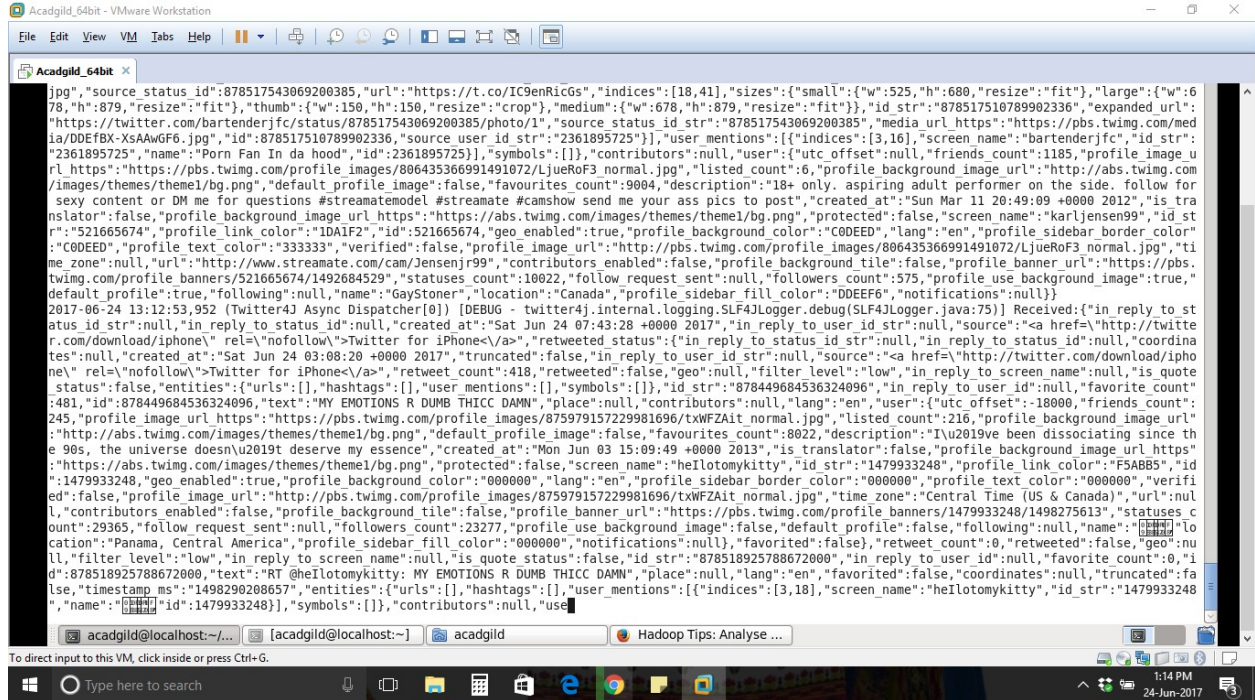
```
bin/flume-ng agent --conf ./conf/ -f conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent
```

Starting the Twitter Agent

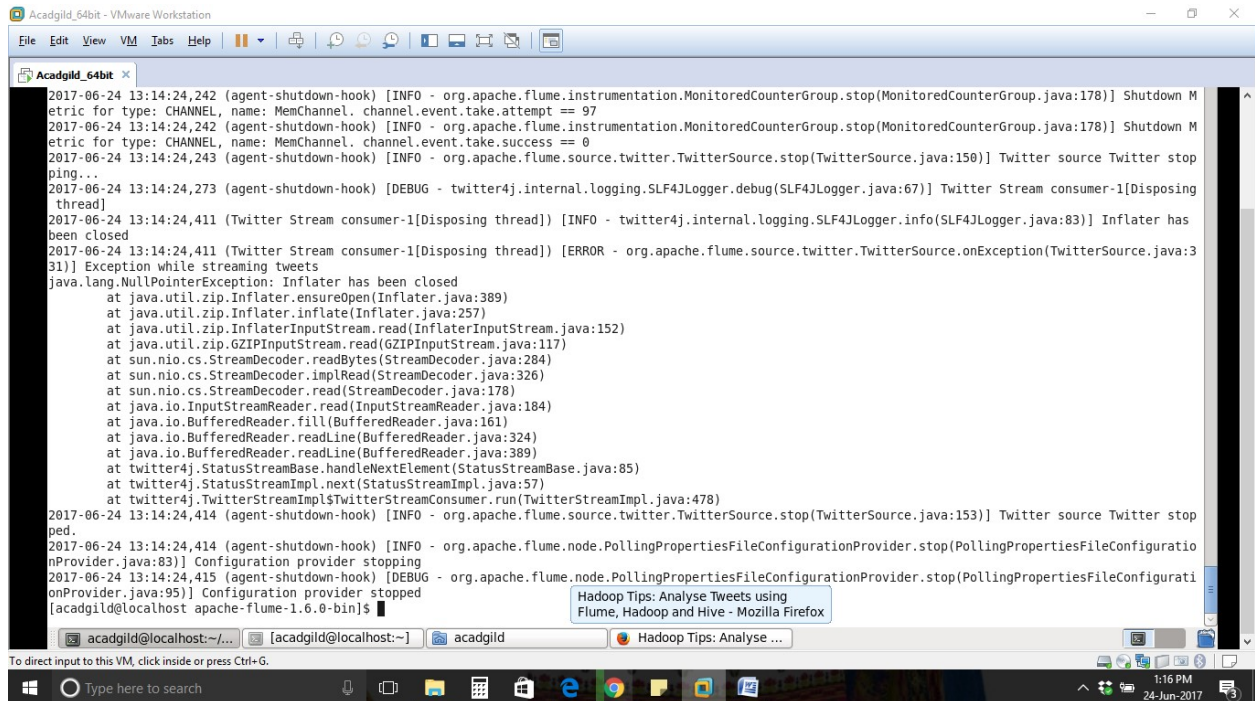


```
2017-06-24 13:14:24,242 (agent-shutdown-hook) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.stop(MonitoredCounterGroup.java:178)] Shutdown Metric for type: CHANNEL, name: MemChannel. channel.event.take.attempt == 97
2017-06-24 13:14:24,242 (agent-shutdown-hook) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.stop(MonitoredCounterGroup.java:178)] Shutdown Metric for type: CHANNEL, name: MemChannel. channel.event.take.success == 0
2017-06-24 13:14:24,243 (agent-shutdown-hook) [INFO - org.apache.flume.source.twitter.TwitterSource.stop(TwitterSource.java:150)] Twitter source Twitter stop ping...
2017-06-24 13:14:24,273 (agent-shutdown-hook) [DEBUG - twitter4j.internal.logging.SLF4JLogger.debug(SLF4JLogger.java:67)] Twitter Stream consumer-1[Disposing thread]
2017-06-24 13:14:24,411 (Twitter Stream consumer-1[Disposing thread]) [INFO - twitter4j.internal.logging.SLF4JLogger.info(SLF4JLogger.java:83)] Inflater has been closed
2017-06-24 13:14:24,411 (Twitter Stream consumer-1[Disposing thread]) [ERROR - org.apache.flume.source.twitter.TwitterSource.onException(TwitterSource.java:331)] Exception while streaming tweets
java.lang.NullPointerException: Inflater has been closed
    at java.util.zip.Inflater.ensureOpen(Inflater.java:389)
    at java.util.zip.Inflater.inflate(Inflater.java:257)
    at java.util.zip.InflaterInputStream.read(InflaterInputStream.java:152)
    at java.util.zip.GZIPInputStream.read(GZIPInputStream.java:117)
    at sun.nio.cs.StreamDecoder.readBytes(StreamDecoder.java:284)
    at sun.nio.cs.StreamDecoder.implRead(StreamDecoder.java:326)
    at sun.nio.cs.StreamDecoder.read(StreamDecoder.java:178)
    at java.io.InputStreamReader.read(InputStreamReader.java:184)
    at java.io.BufferedReader.fill(BufferedReader.java:161)
    at java.io.BufferedReader.readLine(BufferedReader.java:324)
    at java.io.BufferedReader.readLine(BufferedReader.java:389)
    at twitter4j.StatusStreamBase.handleNextElement(StatusStreamBase.java:85)
    at twitter4j.StatusStreamImpl.next(StatusStreamImpl.java:57)
    at twitter4j.TwitterStreamImpl$TwitterStreamConsumer.run(TwitterStreamImpl.java:478)
2017-06-24 13:14:24,414 (agent-shutdown-hook) [INFO - org.apache.flume.source.twitter.TwitterSource.stop(TwitterSource.java:153)] Twitter source Twitter stop ped.
2017-06-24 13:14:24,414 (agent-shutdown-hook) [INFO - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.stop(PollingPropertiesFileConfigurationProvider.java:83)] Configuration provider stopping
2017-06-24 13:14:24,415 (agent-shutdown-hook) [DEBUG - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.stop(PollingPropertiesFileConfigurationProvider.java:95)] Configuration provider stopped
[acadgild@localhost apache-flume-1.6.0-bin]$ bin/flume-ng agent --conf ./conf/ -f conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent
```

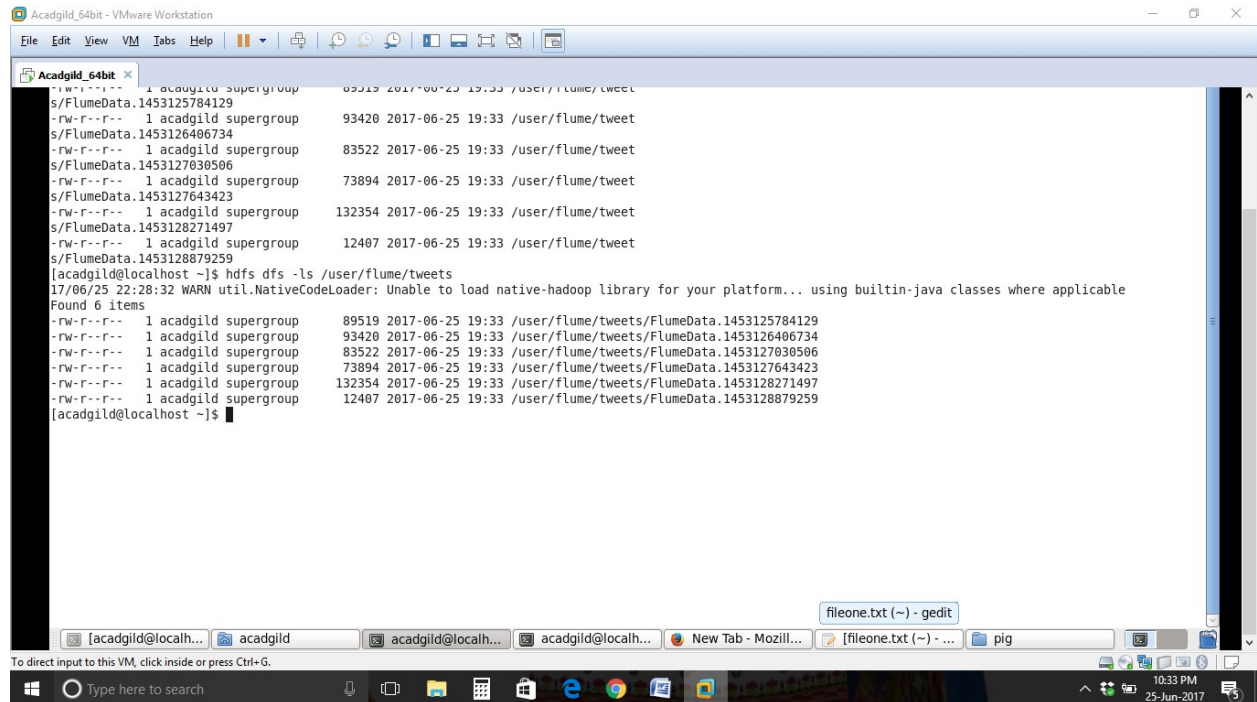
Data being streamed into HDFS



Stop streaming by pressing Ctrl +C

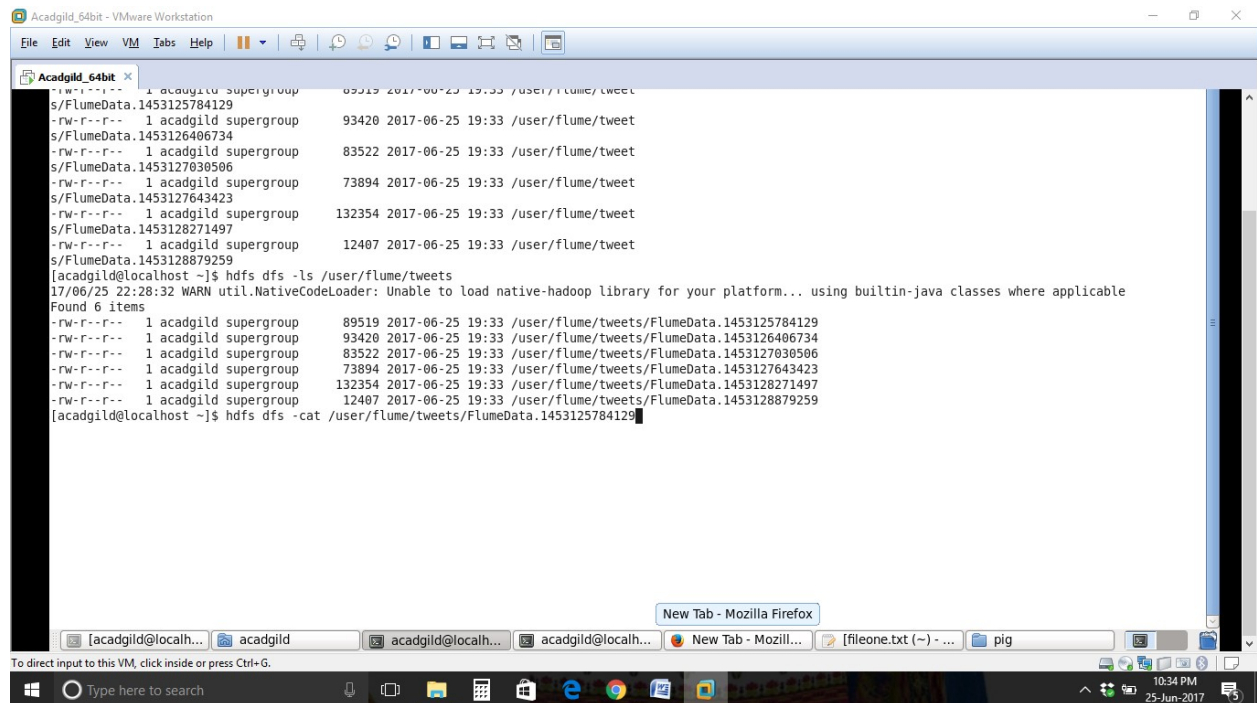


Listing contents of /user/flume/tweets



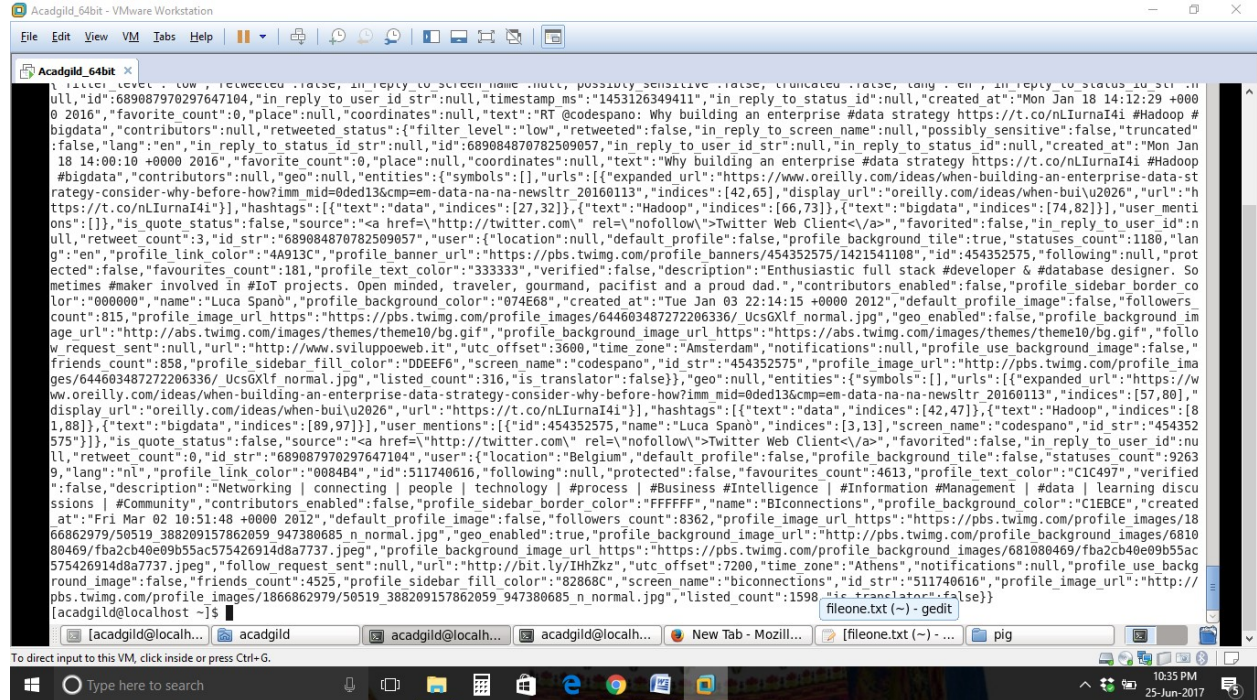
```
Acadgild_64bit x
-rw-r--r-- 1 acadgild supergroup 89519 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453125784129
-rw-r--r-- 1 acadgild supergroup 93420 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453126406734
-rw-r--r-- 1 acadgild supergroup 83522 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453127030506
-rw-r--r-- 1 acadgild supergroup 73894 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453127643423
-rw-r--r-- 1 acadgild supergroup 132354 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453128271497
-rw-r--r-- 1 acadgild supergroup 12407 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453128879259
[acadgild@localhost ~]$ hdfs dfs -ls /user/flume/tweets
17/06/25 22:28:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 6 items
-rw-r--r-- 1 acadgild supergroup 89519 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453125784129
-rw-r--r-- 1 acadgild supergroup 93420 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453126406734
-rw-r--r-- 1 acadgild supergroup 83522 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453127030506
-rw-r--r-- 1 acadgild supergroup 73894 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453127643423
-rw-r--r-- 1 acadgild supergroup 132354 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453128271497
-rw-r--r-- 1 acadgild supergroup 12407 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453128879259
[acadgild@localhost ~]$
```

Command to display Contents of file in which the Twitter Data is collected



```
Acadgild_64bit x
-rw-r--r-- 1 acadgild supergroup 89519 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453125784129
-rw-r--r-- 1 acadgild supergroup 93420 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453126406734
-rw-r--r-- 1 acadgild supergroup 83522 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453127030506
-rw-r--r-- 1 acadgild supergroup 73894 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453127643423
-rw-r--r-- 1 acadgild supergroup 132354 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453128271497
-rw-r--r-- 1 acadgild supergroup 12407 2017-06-25 19:33 /user/flume/tweets/FlumeData.1453128879259
[acadgild@localhost ~]$ hdfs dfs -cat /user/flume/tweets/FlumeData.1453125784129
[acadgild@localhost ~]$
```


Contents of the File in which Twitter Data is Streamed



```
filter_level": "low", "retweeted": false, "in_reply_to_screen_name": null, "possibly_sensitive": false, "truncated": false, "lang": "en", "in_reply_to_status_id_str": null, "id": "689087970297647104", "in_reply_to_user_id_str": null, "timestamp_ms": "1453126349411", "in_reply_to_status_id": null, "created_at": "Mon Jan 18 14:12:29 +0000 2016", "favorite_count": 0, "place": null, "coordinates": null, "text": "RT @codespano: Why building an enterprise #data strategy https://t.co/nLIurnaI4i #Hadoop #bigdata", "contributors": null, "retweeted_status": {"filter_level": "low", "retweeted": false, "in_reply_to_screen_name": null, "possibly_sensitive": false, "truncated": false, "lang": "en", "in_reply_to_status_id_str": null, "id": "689084870782509057", "in_reply_to_user_id_str": null, "in_reply_to_status_id": null, "created_at": "Mon Jan 18 14:00:10 +0000 2016", "favorite_count": 0, "place": null, "coordinates": null, "text": "Why building an enterprise #data strategy https://t.co/nLIurnaI4i #Hadoop #bigdata", "contributors": null, "geo": null, "entities": {"symbols": [], "urls": [{"expanded_url": "https://www.oreilly.com/ideas/when-building-an-enterprise-data-strategy-consider-why-before-how?imm_mid=0ded13&cmp=em-data-na-na-newsltr_20160113", "indices": [42, 65], "display_url": "oreilly.com/ideas/when-building-an-enterprise-data-strategy-consider-why-before-how?imm_mid=0ded13&cmp=em-data-na-na-newsltr_20160113", "indices": [27, 32]}, {"text": "Hadoop", "indices": [66, 73]}, {"text": "bigdata", "indices": [74, 82]}], "user_mentions": [], "is_quote_status": false, "source": "<a href='\"http://twitter.com/\"' rel='\"nofollow/\">Twitter Web Client</a>", "favorited": false, "in_reply_to_user_id": null, "retweet_count": 3, "id_str": "689084870782509057", "user": {"location": null, "default_profile": false, "profile_background_tile": true, "statuses_count": 1180, "language": "en", "profile_link_color": "4A913C", "profile_banner_url": "https://pbs.twimg.com/profile_banners/454352575/1421541108", "id": 454352575, "following": null, "protected": false, "favourites_count": 181, "profile_text_color": "333333", "verified": false, "description": "Enthusiastic full stack #developer & #database designer. Sometimes #maker involved in #IoT projects. Open minded, traveler, gourmand, pacifist and a proud dad.", "contributors_enabled": false, "profile_sidebar_border_color": "000000", "name": "Luca Spanò", "profile_background_color": "074E68", "created_at": "Tue Jan 03 22:14:15 +0000 2012", "default_profile_image": false, "followers_count": 815, "profile_image_url_https": "https://pbs.twimg.com/profile_images/644603487272206336/UcsGXlf_normal.jpg", "geo_enabled": false, "profile_background_image_url": "http://abs.twimg.com/images/themes/theme10/bg.gif", "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme10/bg.gif", "follow_request_sent": null, "url": "http://www.sviluppoeweb.it", "utc_offset": -3600, "time_zone": "Amsterdam", "notifications": null, "profile_use_background_image": false, "friends_count": 858, "profile_sidebar_fill_color": "DDEEF6", "screen_name": "codespano", "id_str": "454352575", "profile_image_url": "http://pbs.twimg.com/profile_images/644603487272206336/UcsGXlf_normal.jpg", "listed_count": 316, "is_translator": false}, "geo": null, "entities": {"symbols": [], "urls": [{"expanded_url": "https://www.oreilly.com/ideas/when-building-an-enterprise-data-strategy-consider-why-before-how?imm_mid=0ded13&cmp=em-data-na-na-newsltr_20160113", "indices": [57, 80], "display_url": "oreilly.com/ideas/when-building-an-enterprise-data-strategy-consider-why-before-how?imm_mid=0ded13&cmp=em-data-na-na-newsltr_20160113", "indices": [42, 47]}, {"text": "bigdata", "indices": [89, 97]}], "hashtags": [{"text": "data", "indices": [42, 47]}, {"text": "Hadoop", "indices": [66, 73]}, {"text": "bigdata", "indices": [89, 97]}], "user_mentions": [{"id": "454352575", "name": "Luca Spanò", "indices": [3, 13], "screen_name": "codespano", "id_str": "454352575"}], "is_quote_status": false, "source": "<a href='\"http://twitter.com/\"' rel='\"nofollow/\">Twitter Web Client</a>", "favorited": false, "in_reply_to_user_id": null, "retweet_count": 0, "id_str": "689087970297647104", "user": {"location": "Belgium", "default_profile": false, "profile_background_tile": false, "statuses_count": 92639, "language": "nl", "profile_link_color": "0084B4", "id": "511740616", "following": null, "protected": false, "favourites_count": 4613, "profile_text_color": "1C1C49", "verified": false, "description": "Networking | connecting | people | technology | #process | #Business Intelligence | #Information Management | #data | learning discussions | #Community", "contributors_enabled": false, "profile_sidebar_border_color": "FFFFFF", "name": "Biconnections", "profile_background_color": "C1EBCE", "created_at": "Fri Mar 02 10:51:48 +0000 2012", "default_profile_image": false, "followers_count": 8362, "profile_image_url_https": "https://pbs.twimg.com/profile_images/1866862979/50519_388209157862059_947380685_n_normal.jpg", "geo_enabled": true, "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/681080469/fba2cb40e09b55ac575426914d8a7737.jpeg", "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/681080469/fba2cb40e09b55ac575426914d8a7737.jpeg", "follow_request_sent": null, "url": "http://bit.ly/IhH2kz", "utc_offset": -7200, "time_zone": "Athens", "notifications": null, "profile_use_background_image": false, "friends_count": 4525, "profile_sidebar_fill_color": "82868C", "screen_name": "biconnections", "id_str": "511740616", "profile_image_url": "http://pbs.twimg.com/profile_images/1866862979/50519_388209157862059_947380685_n_normal.jpg", "listed_count": 1598, "is_translator": false}} [acagild@localhost ~]$
```

fileone.txt (-) - gedit

[acagild@localhost ~] acagild acagild@localhost... acagild@localhost... New Tab - Mozill... fileone.txt (-) - ... pig

Type here to search 10:35 PM 25-Jun-2017