

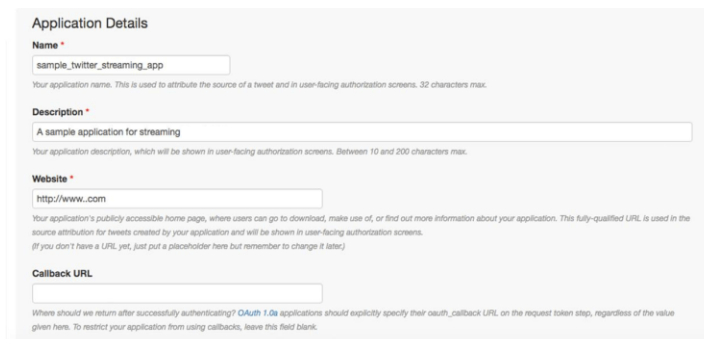
MD FAKRUL ISLAM (613839)

I will show how to build a simple application that reads online streams from Twitter using Python, then processes the tweets using Apache Spark Streaming to identify hashtags.

1. Creating Credentials for Twitter APIs

To get tweets from Twitter, I registered on **TwitterApps** by clicking on “Create new app” and then fill the below form click on “Create your Twitter app.”

Create an application



The screenshot shows the 'Create an application' form on the Twitter developer portal. It includes fields for Name, Description, Website, and Callback URL, each with a small text box and a larger description box. The Name field contains 'sample_twitter_streaming_app'. The Description field contains 'A sample application for streaming'. The Website field contains 'http://www.com'. The Callback URL field is empty. Below the form, there is a note about OAuth 1.0a applications.

Application Details

Name *
sample_twitter_streaming_app
Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *
A sample application for streaming
Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *
http://www.com
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later)

Callback URL
Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

2. Second, I went to my newly created app and opened the “Keys and Access Tokens” tab. Then click on “Generate my access token.”

The new access tokens will appear below.

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token

Access Token Secret

Access Level Read and write

Owner

Owner ID

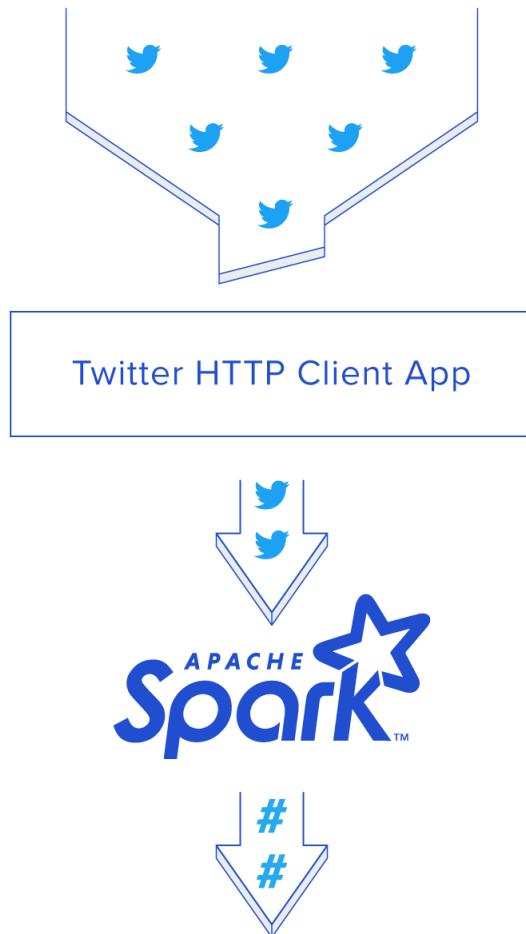
And now I am ready for the next step.

3. Building the Twitter HTTP Client

In this step, I build a simple client that will get the tweets from Twitter API using Python and pass them to the Spark Streaming instance.

4. Setting Up Our Apache Spark Streaming Application

I made Spark streaming app that will do real-time processing for the incoming tweets, extract the hashtags from them, and calculate how many hashtags have been mentioned. I configured all those in my azure account.



Finally, here is a sample output of the Spark Streaming while running and printing the hashtag_counts_df, the output is printed exactly every two seconds as per the batch intervals.

hashtag	hashtag_count
#Hiring	703
#job	603
#CareerArc	591
#Job	261
#Jobs	233
#hiring!	180
#Hospitality	142
#Veterans	115
#hiring	100
#Retail	99

hashtag	hashtag_count
#Hiring	710
#job	608
#CareerArc	597
#Job	268
#Jobs	239
#hiring!	182
#Hospitality	147
#Veterans	118
#hiring	104
#Retail	99