JPE_Twitter

Splunk App for Twitter Mining

May 2019

Juan Alejandro Pérez Chandía
Computer Engineering
jalejandro.ingeniero@gmail.com

Jocelyn Lisette Flores Scheffer
English Teacher
Translation Support
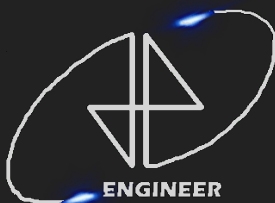
ENGINEER

1 Introduction

JPE_Twitter is a Splunk Application developed in Python, whose purpose is the Twitter Mining and complete integration with the log analysis tool previously mentioned. With JPE_Twitter you can extract of Twitter important information based on words filter, phrases, hashtag, among others and get data, that indicate what's happening with our business or service on Twitter, building a development base to sentiment analysis on Twitter.

Contents

2   Basic concepts of development

The application is developed in Python 2.7, with the purpose of sharing Python base installation of the Splunk platform (6.x or higher) and has five essential modules:

- **config :** The application has a configuration file called *twitter.conf*, where config module can read and write the necessary parameter to the authentication with Twitter API, log files configuration, encrypt, among others.
- **security :** For the security of confidential data (authentication token), this module has what it needs to encrypt and decrypt, based on the key that the user defines in the configuration of the tokens in Splunk.
- **splunk_read :** The purpose of this modules is take all the entered parameters from the Splunk platform and take them to the corresponding module (config, security, service, among others).
- **service :** Responsible to start, stop and restart the service, besides to obtain its status.
- **twitter :** Corresponding to the main module, who will return everything filtered on Twitter.

Besides, JPE_Twitter has its own log files, to make process debug through Splunk or directly looking at log files:

- **config.log :** Show the ran process by *config* module, including its problem (if there is any).
- **service.log :** Show the service state and its PID, besides include problems that this service might have.
- **twitter.log :** Show the authentication state with Twitter API and filtered messages from this social network.
- **splunk_read.log :** Show the ran process by *splunk_read* module, including its problem (if there is any).
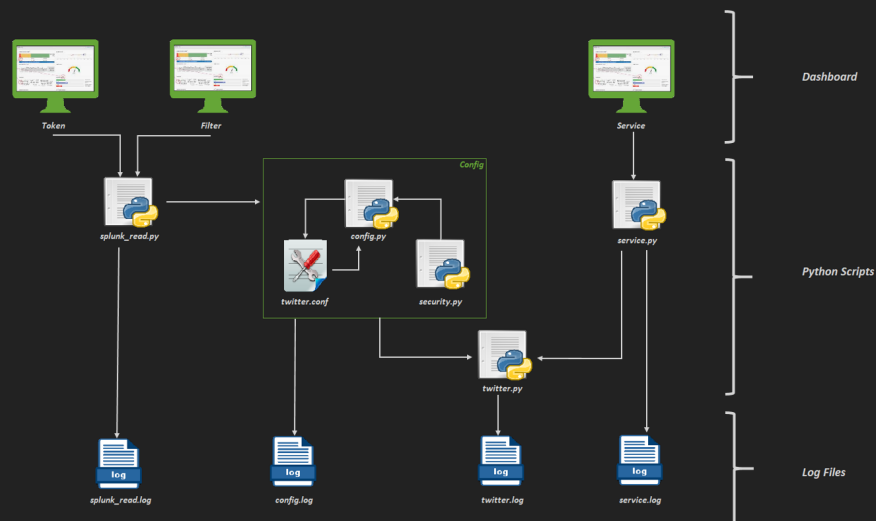

Fig. 1: JPE_Twitter General Diagram

Important: All the necessary libraries to the proper running of the application, are included on the installation package.

2.1 Dashboards

To the administration of Twitter listening process exists three important dashboards, that allows to configure the authentication tokens and filters, also allows to start, stop and restart the service or check its status:

- Token and Security

- Filter

- Service

Besides, have three basic dashboards, that allows to show Twitter general information base on the filters configured, such as user with most followers, hashtag the most used and users mention:

- Twitter Users

- Twitter Hashtag

- Twitter Comments

2.1.1 Token and Security

This dashboard allows to configure the authentication of Twitter API tokens, which are: *consumer key, consumer secret, access token and access secret.* This tokens are stored in the configuration file *twitter.conf* and related to security, the authentication tokens can be encrypted to prevent that other people access to them, only is necessary to write an encryption key in the same panel, after to indicate that you want to encrypt the tokens.
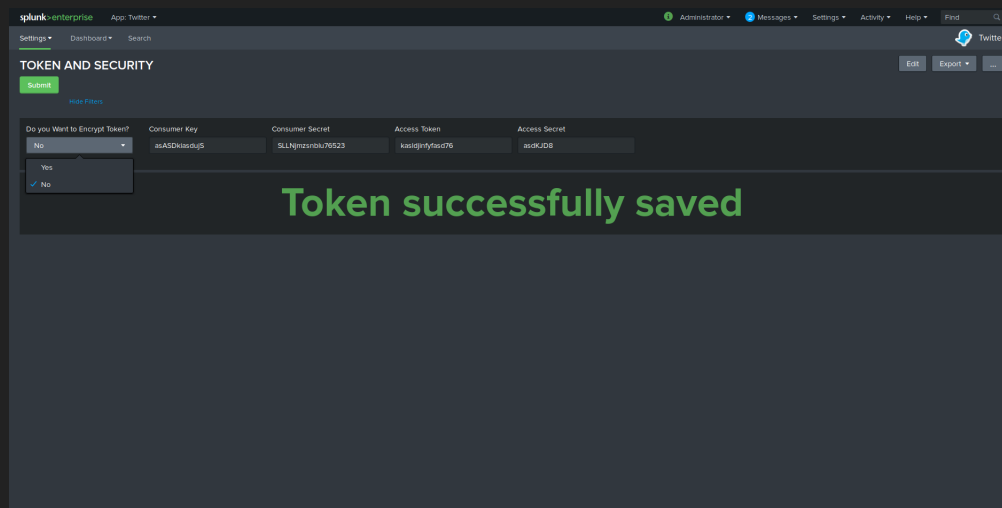

Fig. 2: Setting: Token and Security Dashboard

It is important to consider that: to obtain tokens, you must create a developer account on Twitter (`https://developer.twitter.com`)

2.1.2   Filter

This dashboard allows to configure filters that will be apply in the listening process. Configuration option included adding a new filter to what is already configured or create a new filter, replacing the previous ones, by choosing *add o new* respectively.

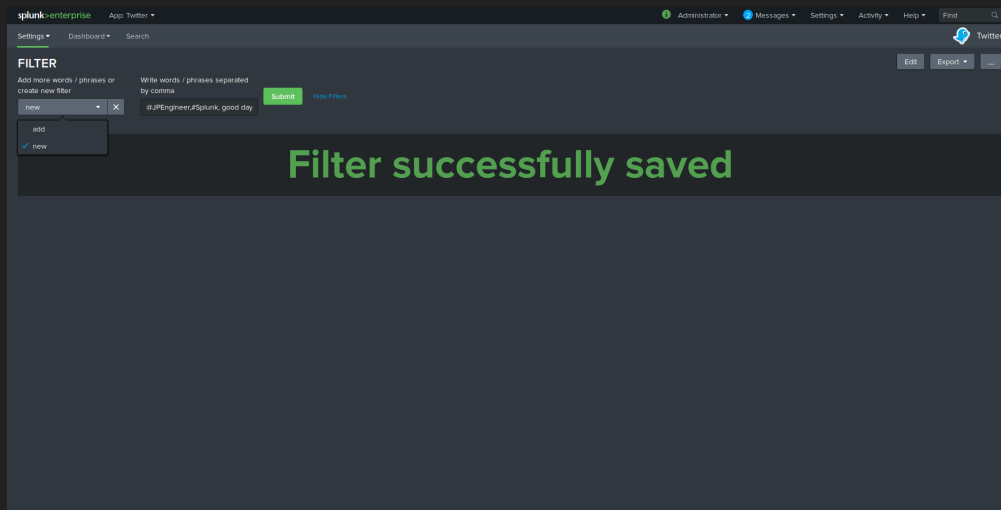Entered filters can be words, phrases, hashtag, among others and should be separated by comma.

Fig. 3: Setting: filter Dashboard

2.1.3   Service

This dashboard allows the action of run over Twitter service. Start option allows to initiate the service and begin to index the data obtained from Twitter and as previous requirement, you will need to have an authentication tokens and filters configured. Stop option is responsible to stop the service. Restart option allows automatically to stop and start the service if it is require for own developments or other reasons. Status option shows service information (active or stopped).
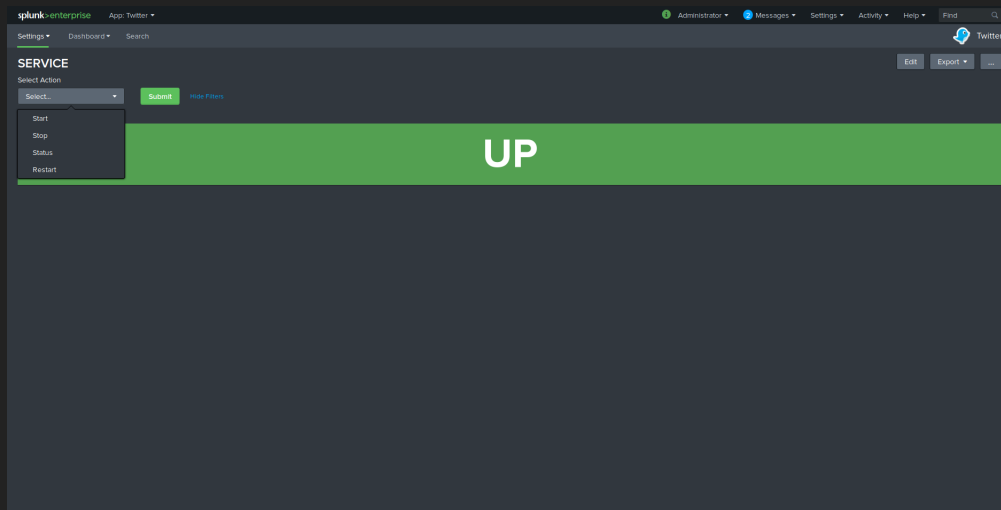

Fig. 4: Setting: Service Dashboard

2.1.4　Twitter Users

This dashboard shows Twitter users information, as followers number, name, alias and location. On the same dashboard new developments can be applied to obtain additional information.
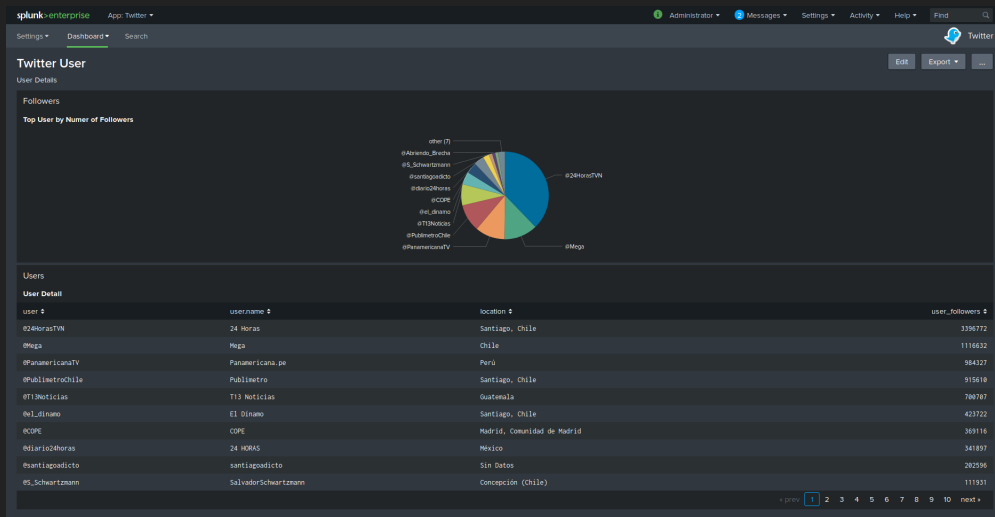

Fig. 5: Twitter User Dashboard

2.1.5　Twitter Hashtag

This dashboard shows hashtag information (trending topic). On the same dashboard new developments can be applied to obtain additional information.
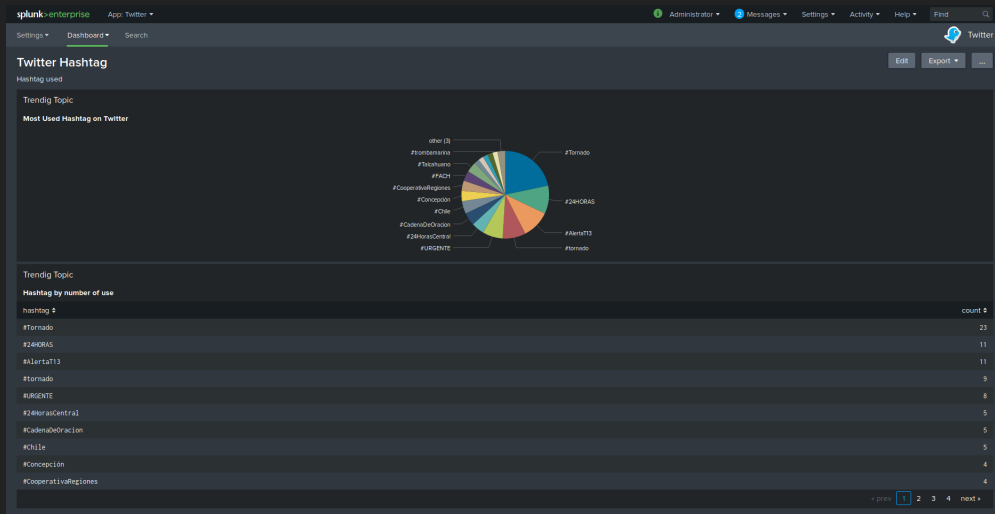

Fig. 6: Twitter Hashtag Dashboard

2.1.6   Twitter Comments

This dashboard shows the details of the most commented users, comments made and user information. On the same dashboard new developments can be applied to obtain additional information.

Fig. 7: Twitter Comments Dashboard

3 Installation

JPE_Twitter was tested in Splunk 6.x versions or higher and was development on Splunk 7, which will be used in the following installation guide.
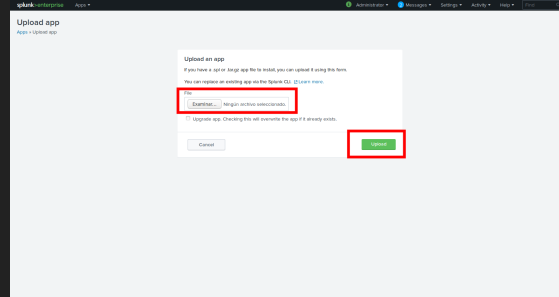
Considering that the application is downloaded, you must follow the next steps:

(a) Make click on top left of the Splunk platform, says **App: Search & Reporting** > **Manage Apps**

(b) Make click on top right of the Splunk platform, on the buttom **Install Apps from File**

(c) Make click on **Examine** and select JPE_Twitter.zip, later make click on **Upload**
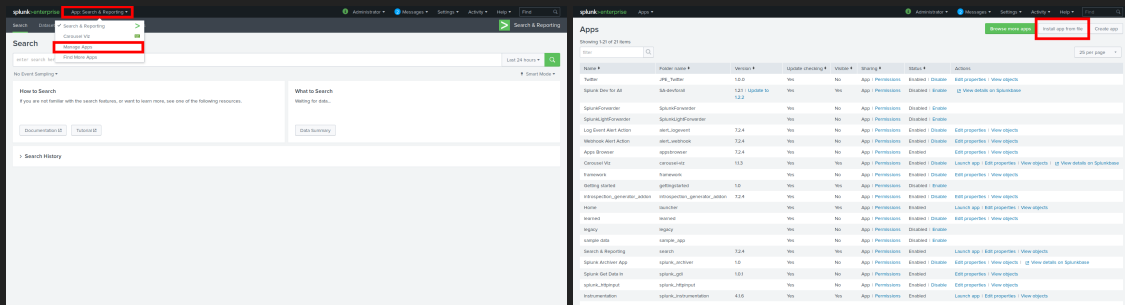
**(a)** Manage Apps   **(b)** Install Apps from File

(c) Upload

Fig. 8: JPE_Twitter Installation

With these simple steps, JPE_Twitter will be installed. The recomendation is to restart Splunk.
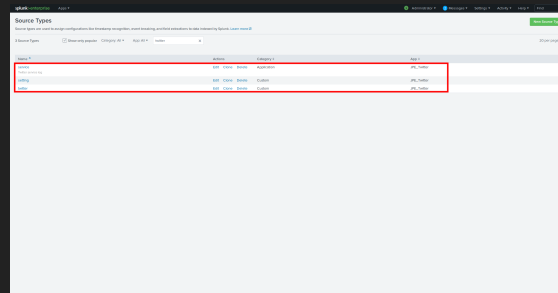
4 Validations

In order to validate the installation, you must follow the next steps:

(a) Make click on top right of Splunk platform, in **Setting** > **Indexes** section.

(b) Validate that **twitter** index exist.

(c) Make click on **Setting** > **Source Types** and validate that **twitter, setting** y **service** source types exist.
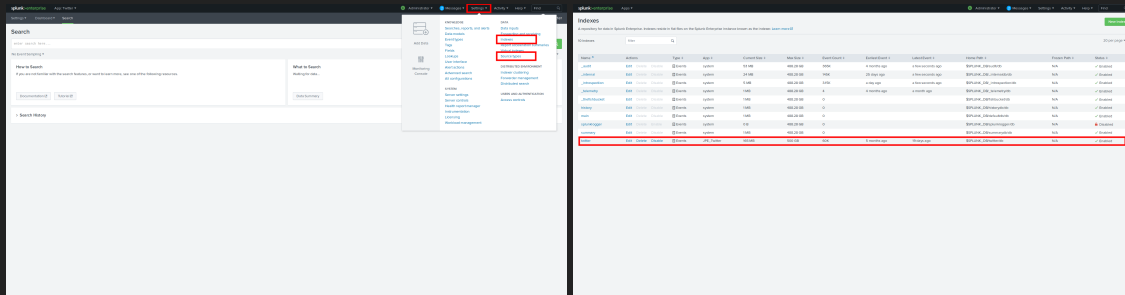
(a) Setting: Index - Sourcetype

(b) Index

(c) Sourcetype

Fig. 9: Validations

Index twitter: Data repository of installed application.
Source Type twitter: Identify the structure of the filtered data on twitter.
Source Type service: Identify the structure of the generated data by *service.py* module.
Source Type setting: Identify the structure of the generated data by *twitter.py* y *security.py* modules.

5 Support, questions and comments

Full Name: Juan Alejandro Pérez Chandía
E-Mail: jalejandro.ingeniero@gmail.com
Languages: Spanish - English