How to Guide

Twitter Scrapper (DP)

Version: R-4.5

Contents

1. Component Overview

Twitter Scrapper helps in scrapping Twitter data. Twitter Scrapper provides two types of scrapping:

1. Real-Time
2. History Data

Real-Time – By using this option, the user can fetch the live Tweets by using twitter API along with specific search filters.

History Data – By using this option, the user can provide or fetch twitter data of past 7 days (provided limit is up to 1000 tweets). It also provides us a facility to fetch data according to specific tags.

2. Step by Step Process to Use Twitter Scrapper

The steps to configure the Twitter Scrapper component is described in this section.

i) Navigate to the Pipeline Editor.
ii) Drag the Twitter Scrapper from the Ingestion section provided under the Components Pallet and drop to the Pipeline Workspace.
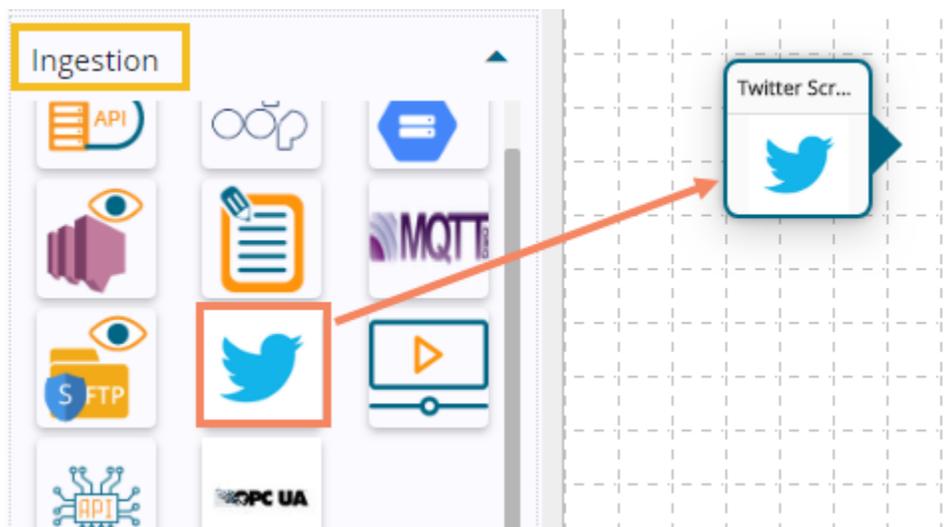
iii) Click the dragged Twitter Scrapper component to get the Configuration fields:
iv) The Basic Information tab opens by default.
   a. Select the invocation type (at present only '**Real-Time**' option is provided)
   b. Deployment Type: It comes preselected based on the component.
   c. Container Image Version: It comes preselected based on the component.
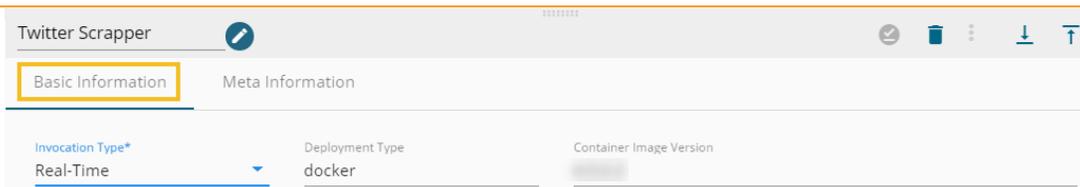
v) Click on the Meta Information tab to open the configuration fields for the selected columns.
   a. Consumer API Key: This is always is going to be unique depending upon user's account. It is essential in both the cases of scrapping Real-time and History Data.
   b. Consumer API Secret Key: This is always is going to be unique depending upon user's account. It is essential in both the cases of scrapping Real-time and History Data.
   c. Filter Text: This field is used to specify separate tags which user wants to fetch from a twitter. It provides the choice to the user for fetching multiple tags data at a time, to accomplish this user need to provide tags name with a comma separator. For instance, #India, #Cricket
   d. Twitter Data Type: This field provides dropdown menu from where the consumer can select the scrapping strategy(real-time/history).
   e. Access Token: This field is consumable only when user wants to fetch real-time data from a scrapper.
   f. Access Token Secret: This field is consumable only when user wants to fetch real-time data from a scrapper.
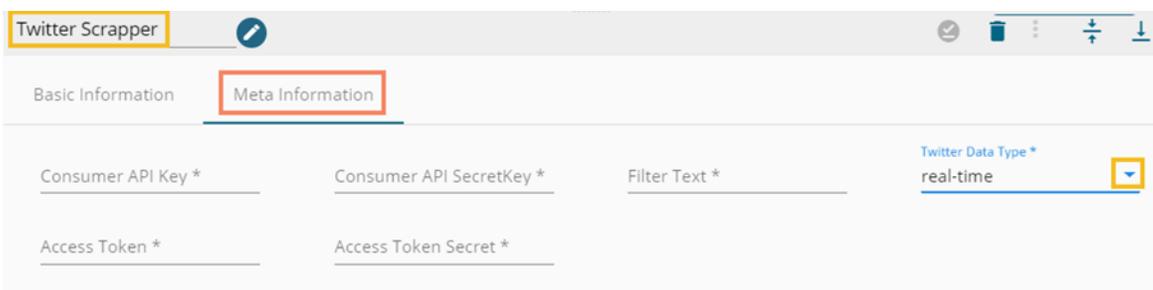
vi) Click on the Configuration tab to configure the required information.
   a. Limit: Provide Memory and CPU details to configure the limit
   b. Request: Provide Memory and CPU details to configure the request
   c. Click the '**Save**' icon.
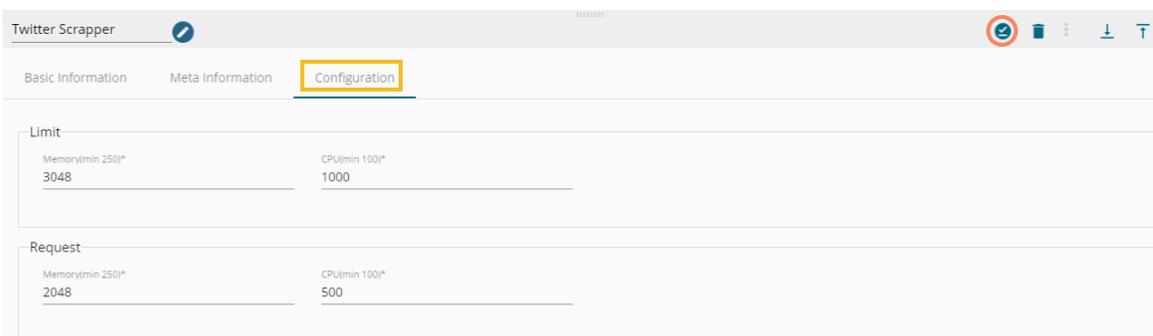
Note: The user should know Consumer API Key, Consumer Secret Key, Access Token, Access Token Secret from a developer account of twitter.