

PARTE I - Streaming de tendencias con #hashtags en Twitter:

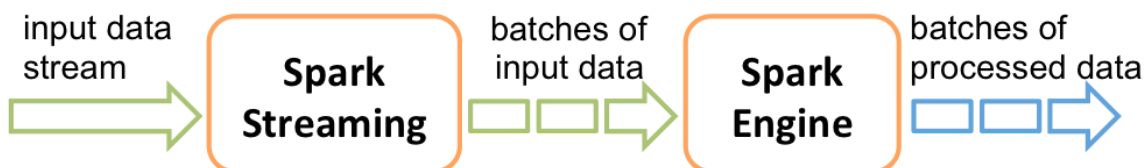
Arquitectura de Bloque twitter

Los datos pueden ser ingresados de muchas fuentes como Kafka, Flume, Kinesis o TCP, y pueden procesarse utilizando algoritmos complejos expresados con funciones de alto nivel como mapear, reducir, unir y ventana.

Spark Streaming



Internamente, Spark Streaming recibe flujos de datos de entrada en vivo y divide los datos en lotes, que luego son procesados por el motor Spark para generar el flujo final de resultados en lotes.

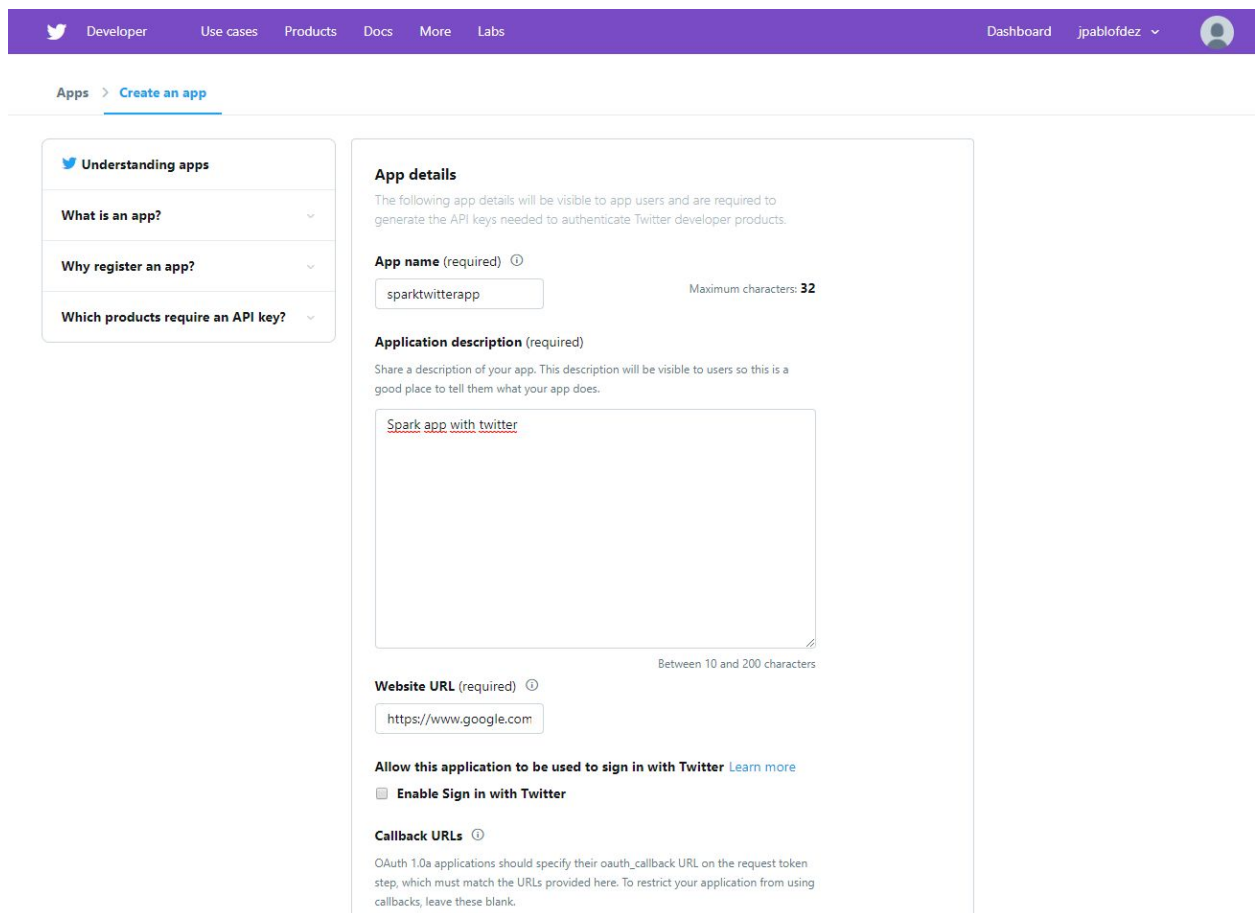


Modelo de Streaming con el Api de twitter

Pasos para utilizar el api streaming de twitter con pyspark

1. Primero de debe de crear una cuenta de desarrollador en twitter,
<http://developer.twitter.com/>

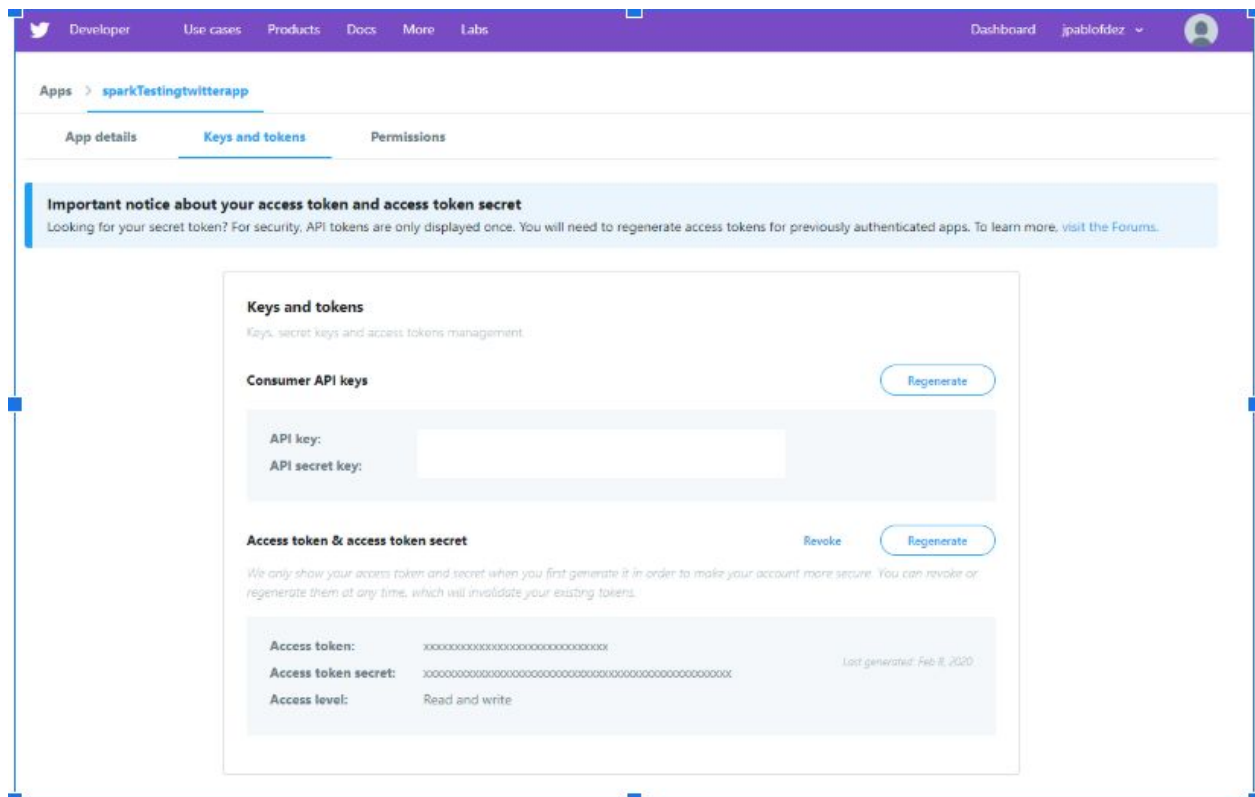
2. Se debe de crear una aplicación en el ambiente de desarrollador de twitter y así obtener las credenciales de acceso.



The screenshot shows the 'Create an app' page in the Twitter Developer Portal. The page has a purple header with navigation links: Developer, Use cases, Products, Docs, More, Labs, Dashboard, and a user profile for 'jpablofdez'. Below the header, the breadcrumb 'Apps > Create an app' is visible. On the left, there is a sidebar with 'Understanding apps' and three expandable sections: 'What is an app?', 'Why register an app?', and 'Which products require an API key?'. The main content area is titled 'App details' and contains the following fields:

- App name (required)**: A text input field containing 'sparktwitterapp'. A note indicates 'Maximum characters: 32'.
- Application description (required)**: A text area containing 'Spark app with twitter'. A note indicates 'Between 10 and 200 characters'.
- Website URL (required)**: A text input field containing 'https://www.google.com'.
- Allow this application to be used to sign in with Twitter**: A checkbox labeled 'Enable Sign in with Twitter' is currently unchecked. A 'Learn more' link is provided.
- Callback URLs**: A section with a note stating: 'OAuth 1.0a applications should specify their oauth_callback URL on the request token step, which must match the URLs provided here. To restrict your application from using callbacks, leave these blank.'

3. Se generan los acceso de seguridad (tokens para el api)



4. Una vez se crea la aplicación con las credenciales se instala el siguiente comando.

- Pip3 install tweepy

Estos son los pasos básicos para la interacción con el api stream y python con pyspark y jupyter notebook.