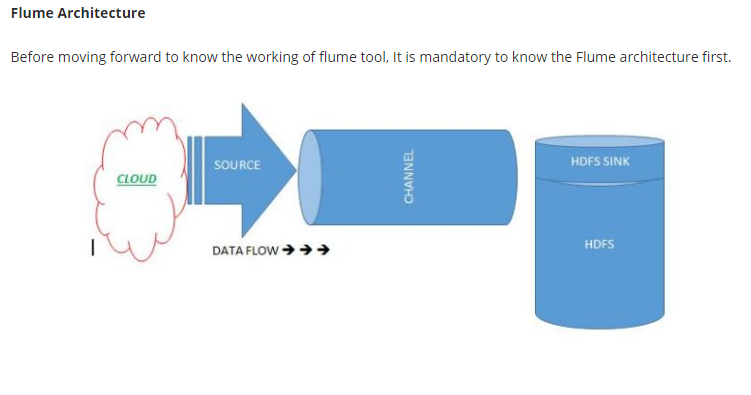
**Assignment 11.3**

Problem Statement:

Streaming twitter data into hdfs using flume



Flume is composed of the following components.

**Flume Event:**It is the main unit of the data that is transported inside the **Flume**(Typically a single log entry). It contains a payload of the byte array that is to be transported from the source path to the destination path which could be accompanied by optional headers.

**Flume Agent:** Is an independent Java virtual machine daemon process which receives the data (events) from clients and transports to the subsequent destination (sink or agent).

**Source:**Is the component of Flume agent which receives data from the data generators say, twitter, facebook, weblogs from different sites and transfers this data to one or more channels in the form of Flume event.

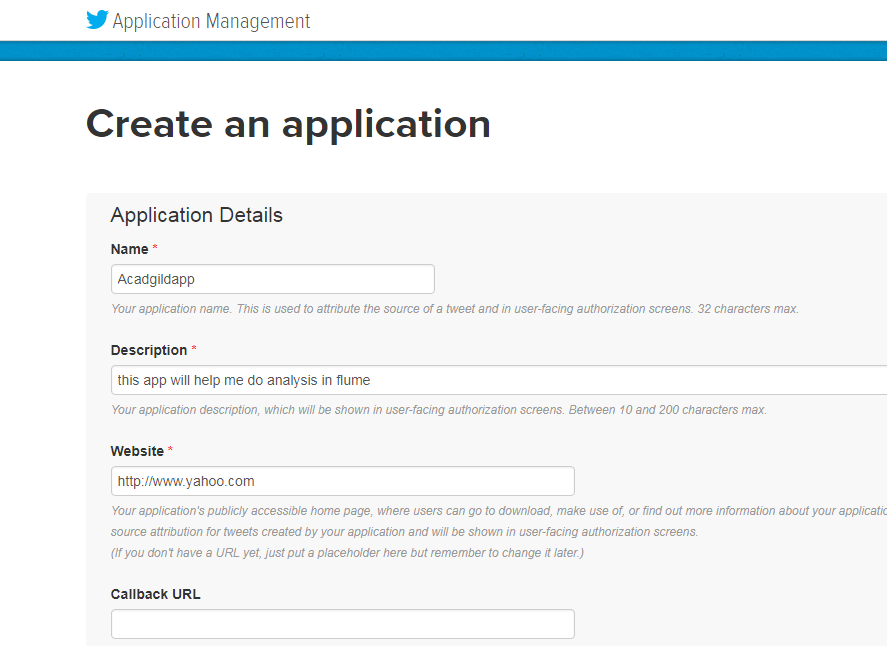
The external source sends data to Flume in a format that is recognized by the target Flume source. Example, an Avro Flume source can be used to receive Avro data from Avro clients or other Flume agents in the flow that send data from an Avro sink, or the Thrift Flume source will receive data from a Thrift sink, or a Flume Thrift RPC client or Thrift Clients are written in any language generated from the Flume thrift protocol.

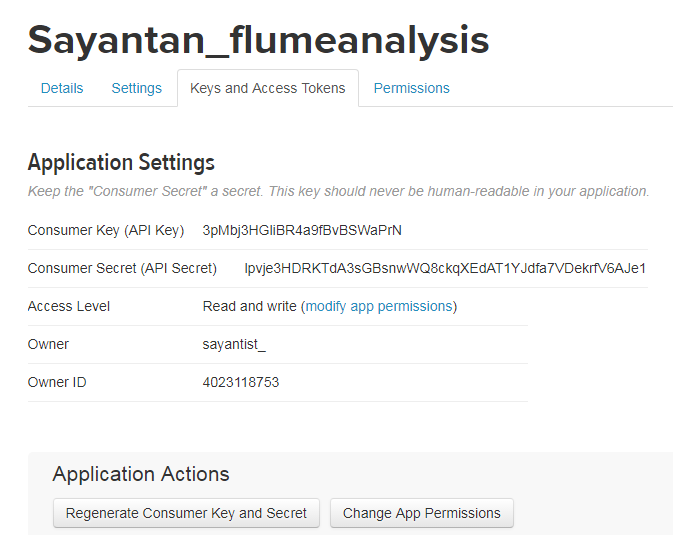
**Channel:**Once, the Flume source receives an Event, it stores this data into one or more channel and buffers them till they are consumed by sinks. It acts as a bridge between the source and sinks. These channels are implemented to handle any number of sources and sinks.  
**Sink:**It stores the data into the centralized stores like HDFS and HBase.

**Streaming Twitter Data**

To stream data to our database from twitter we should have the following pre-requisites.

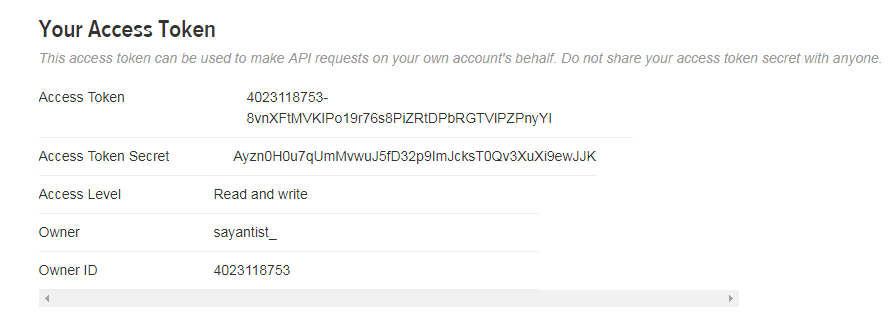
* Twitter account
* Hadoop cluster





Consumer Key (API Key) 3pMbj3HGliBR4a9fBvBSWaPrN

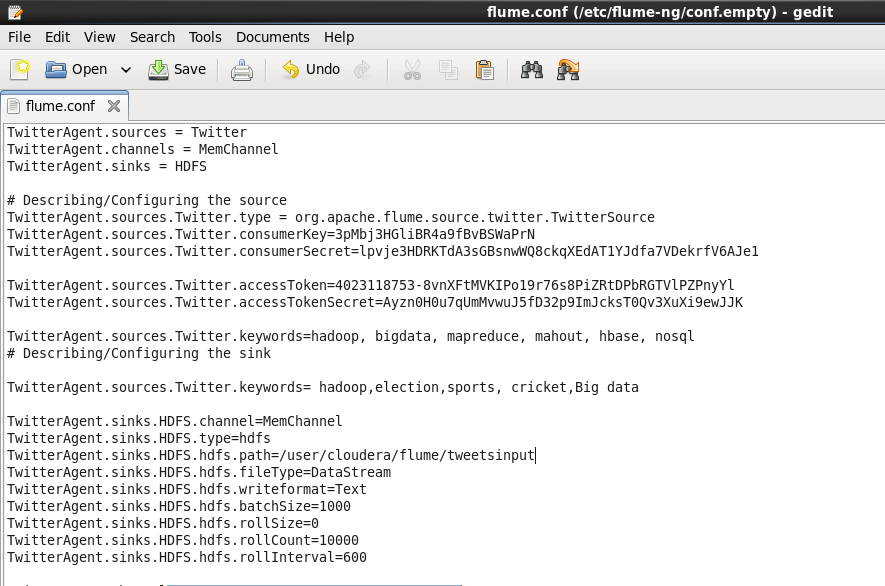
Consumer Secret (API Secret) lpvje3HDRKTdA3sGBsnwWQ8ckqXEdAT1YJdfa7VDekrfV6AJe1



Access Token 4023118753-8vnXFtMVKIPo19r76s8PiZRtDPbRGTVlPZPnyYl

Access Token Secret Ayzn0H0u7qUmMvwuJ5fD32p9ImJcksT0Qv3XuXi9ewJJK

Now create a new file in the flume conf directory



Now create a directory in hdfs to get the tweets

Hadoop fs –mkdir –p /user/cloudera/flume/tweets

Now open Mozilla browser to see the contents



