

**Tipo** : Enunciado de laboratorio  
**Capítulo** : Componentes de Hadoop  
**Duración** : 30 minutos

---

## I. OBJETIVO

Conocer cómo implementar Flume, realizando la recolección de datos de Twitter.

## II. REQUISITOS

Los siguientes elementos de software son necesarios para la realización del laboratorio:

- Virtual Box
- Hortonworks Sandbox for HDP
- Putty
- Winscp

## III. EJECUCIÓN DEL LABORATORIO

1. Ingresar al servidor Hortonworks, para esto deberá ingresar a la siguiente dirección:  
<http://127.0.0.1:8080>
2. Ir al menú de Canvas
3. Seleccionar la opción File View
4. Luego de esto navegar hasta seleccionar la carpeta /user/maria\_dev, luego de esto dar click en el botón New Folder, crear el directorio **twitter\_data**, clic al botón +Add.
5. Para realizar el proceso de ingesta en twitter deberá crear una aplicación en twitter.
6. Luego de esto ingresar por putty al servidor
7. Sincronizar la hora para esto ejecutar
8. Creamos el archivo twitter.conf, con la configuración
9. Ir al directorio de configuración de Flume
10. Ejecutar el agente de flume, indicando el archivo de configuración **twitter.conf**, esperar que finalice la inicialización (Mensaje: Starting accept handler)
11. Ver el contenido de la carpeta HDFS /user/maria\_dev/twitter\_data
12. Para salir de flume, recuerde presionar CTRL+C, en la pantalla que se ejecuta e agente.