

1. Configuración del Entorno de Flume

1. **Verifica que Apache Flume esté instalado y configurado.** En Cloudera, puedes instalar Flume desde Cloudera Manager si aún no está disponible.

Podemos ver que está instalado.

```
[cloudera@quickstart ~]$ flume-ng version
Flume 1.6.0-cdh5.13.0
Source code repository: https://git-wip-us.apache.org/repos/asf/flume.git
Revision: 771282577aadc61073e7cb77a31ef5cd6140fc4
Compiled by jenkins on Wed Oct 4 11:11:36 PDT 2017
From source with checksum c6892efdb6276b881b961e90d72bd762
```

2. Crear una Configuración para Flume

Archivo de Configuración (*agenteflume.conf*):

```
# Nombre del agente
agent.sources = local-source
agent.sinks = hdfs-sink
agent.channels = memory-channel
# Configuración de la fuente (local-source)
agent.sources.local-source.type = spooldir
agent.sources.local-source.spoolDir = /path/to/local/input
agent.sources.local-source.fileHeader = true
# Configuración del canal (memory-channel)
agent.channels.memory-channel.type = memory
agent.channels.memory-channel.capacity = 1000
agent.channels.memory-channel.transactionCapacity = 100
# Configuración del sink (hdfs-sink)
agent.sinks.hdfs-sink.type = hdfs
agent.sinks.hdfs-sink.hdfs.path = /user/cloudera/flume-output
agent.sinks.hdfs-sink.hdfs.fileType = DataStream
agent.sinks.hdfs-sink.hdfs.writeFormat = Text
agent.sinks.hdfs-sink.hdfs.rollSize = 0
agent.sinks.hdfs-sink.hdfs.rollCount = 10
agent.sinks.hdfs-sink.hdfs.batchSize = 100
# Conectar componentes
agent.sources.local-source.channels = memory-channel
agent.sinks.hdfs-sink.channel = memory-channel
```

Tras investigar en internet, veo que el archivo suele tener otro nombre, así que ante la duda, creo dos archivos con el mismo contenido:

```
[cloudera@quickstart ~]$ sudo nano /etc/flume-ng/conf/flume.conf
[cloudera@quickstart ~]$ sudo nano /etc/flume-ng/conf/agenteflume.conf
```

```
[cloudera@quickstart ~]$ cat /etc/flume-ng/conf/agentflume.conf
# Nombre del agente
agent.sources = local-source
agent.sinks = hdfs-sink
agent.channels = memory-channel
# Configuración de la fuente (local-source)
agent.sources.local-source.type = spooldir
agent.sources.local-source.spoolDir = /path/to/local/input
agent.sources.local-source.fileHeader = true
# Configuración del canal (memory-channel)
agent.channels.memory-channel.type = memory
agent.channels.memory-channel.capacity = 1000
agent.channels.memory-channel.transactionCapacity = 100
# Configuración del sink (hdfs-sink)
agent.sinks.hdfs-sink.type = hdfs
agent.sinks.hdfs-sink.hdfs.path = /user/cloudera/flume-output
agent.sinks.hdfs-sink.hdfs.fileType = DataStream
agent.sinks.hdfs-sink.hdfs.writeFormat = Text
agent.sinks.hdfs-sink.hdfs.rollSize = 0
agent.sinks.hdfs-sink.hdfs.rollCount = 10
agent.sinks.hdfs-sink.hdfs.batchSize = 100
# Conectar componentes
agent.sources.local-source.channels = memory-channel
agent.sinks.hdfs-sink.channel = memory-channel
```

3. Preparar el Directorio Local de Entrada

Crea el Directorio de Entrada Local:

```
mkdir -p /path/to/local/input
```

Copia el Archivo de Texto al Directorio: Coloca **elquijote.txt** (u otros archivos de texto) en este directorio para que Flume lo procese:

```
cp /path/to/elquijote.txt /path/to/local/input/
```

```
[cloudera@quickstart Desktop]$ sudo nano folder/loremIpsum.txt
[cloudera@quickstart Desktop]$ cat folder/loremIpsum.txt
```

```

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed porttitor tel
. Vivamus tortor quam, pharetra vitae lorem at, ultricies rutrum mi. Nulla
mauris. Vivamus ut luctus enim, gravida tincidunt velit. Ut odio orci, co
libero, scelerisque a tempus nec, eleifend eu quam. Aliquam sollicitudin
a, nec dignissim nibh.
```

```

Mauris in magna ac libero elementum mollis sollicitudin eget tellus. Nulla
estibulum. Praesent pulvinar ex id nisi lobortis consequat. Aenean eget di
interdum porttitor, leo arcu lacinia orci, at placerat augue leo non nisi
t quam a turpis feugiat viverra.
```

4. Ejecutar Apache Flume

Inicia el Agente Flume: Ejecuta Flume con el archivo de configuración:

```
flume-ng agent --conf /path/to/flume/conf --conf-file /etc/flume-ng/conf/flume.conf --name agent
-Dflume.root.logger=INFO,console
```

Me devuelve muchísimo texto

```
[cloudera@quickstart Desktop]$ flume-ng agent --conf /path/to/flume/conf --conf-file /etc/flume-ng/conf/flume.conf --name agent
Info: Including Hadoop libraries found via (/usr/bin/hadoop) for HDFS access
Info: Including HBASE libraries found via (/usr/bin/hbase) for HBASE access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/java/jdk1.7.0_67-cloudera/bin/java -Xmx20m -cp '/path/to/flume/conf:/usr/lib/flume-ng/lib/*:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/*:/usr/lib/hadoop-hdfs/*:/usr/lib/hadoop-hdfs/lib/*:/usr/lib/hadoop-hdfs/*:/usr/lib/hadoop-yarn/lib/*:/usr/lib/hadoop-yarn/*:/usr/lib/hadoop-mapreduce/lib/*:/usr/lib/hadoop-mapreduce/*:/usr/lib/hbase/bin/./conf:/usr/java/jdk1.7.0_67-cloudera/lib/tools.jar:/usr/lib/hbase/bin/*:/usr/lib/hbase/bin/*:/lib/activation-1.1.jar:/usr/lib/hbase/bin/*:/lib/apacheds-i18n-2.0.0-M15.jar:/usr/lib/hbase/bin/*:/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/lib/hbase/bin/*'
[cloudera@quickstart Desktop]$
```

Verifica que los Datos Llegaron a HDFS: Lista los archivos en la ruta configurada en el sink ([/user/cloudera/flume-output](#)):

hdfs dfs -ls /user/cloudera/flume-output

```
[cloudera@quickstart Desktop]$ sudo hdfs dfs -ls /user/cloudera/flume-output
ls: '/user/cloudera/flume-output': No such file or directory
[cloudera@quickstart Desktop]$
```

Muestra el Contenido Ingerido:

hdfs dfs -cat /user/cloudera/flume-output/*

```
[cloudera@quickstart Desktop]$ sudo hdfs dfs -cat /user/cloudera/flume-output/*
cat: '/user/cloudera/flume-output/*': No such file or directory
[cloudera@quickstart Desktop]$
```