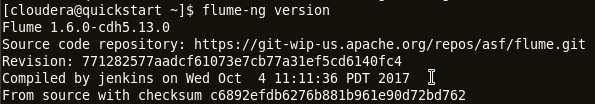
**1. Configuración del Entorno de Flume**

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.



**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*

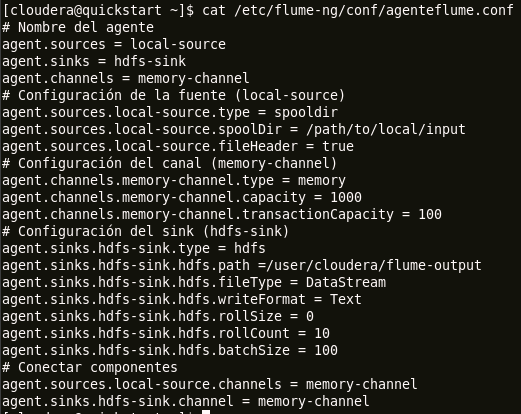
*# Conectar componentes*

*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:





**3. Preparar el Directorio Local de Entrada**

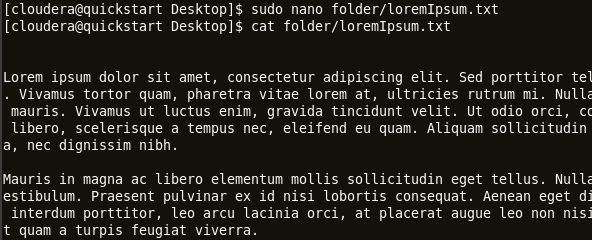
**Crea el Directorio de Entrada Local:**

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/



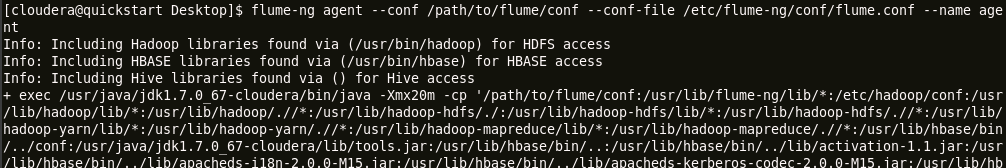
**4. Ejecutar Apache Flume**

**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



**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



**Muestra el Contenido Ingerido:**

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

