

Flume. Práctica 3: Flujo multi-agentes

Si no se tiene instalado, previamente se instala el agente de Flume en Cloudera en el host node1.

En esta práctica aprendemos a conectar varios agentes entre sí. Llevamos a cabo un caso de uso donde se lee información de tres fuentes distintas:

- una fuente de **Netcat** con un canal basado en ficheros,
- otra que realice **spooling** de una carpeta, vigila una carpeta y cuando haya algún archivo, lo ingesta y lo elimina utilizando un canal en memoria
- y un tercero que **ejecuta un comando** utilizando también un canal en memoria.

Como agente de consolidación que una la información de las tres fuentes de datos, se reutiliza el agente **AvroHdfsAgent** que hemos creado en la práctica 2 de Flume.

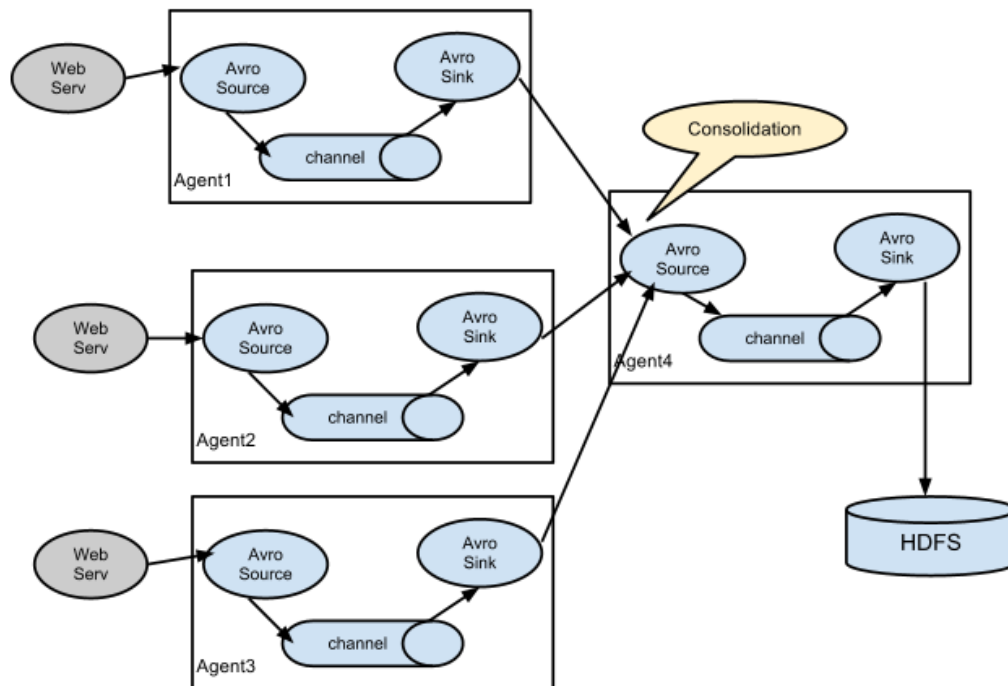


Figure 1: Esquema de funcionamiento de la práctica

Los cuatro agentes se configuran y arrancan en el mismo host, en *pasarela*.

Configuración de los agentes conectados

Los tres agentes que escriben en el mismo sumidero (Avro) se configuran en el mismo archivo **multiagent-avro.conf**.

Como novedad en la configuración, al describir los componentes, se definen los nombres de las tres fuentes, canales y sumideros a la vez, en las mismas variables de configuración.

```
# Nombramos las tres fuentes con sus tres sumideros
```

```
MultiAgent.sources = Netcat SpoolDir Exec
```

```
MultiAgent.channels = FileChannel MemChannel1 MemChannel2
```

```
MultiAgent.sinks = AvroSink1 AvroSink2 AvroSink3
```

```
# Describimos el primer agente Netcat
```

```
MultiAgent.sources.Netcat.type = netcat
```

```
MultiAgent.sources.Netcat.bind = localhost
```

```
MultiAgent.sources.Netcat.port = 10004
```

```
# Describimos el segundo agente SpoolDir
```

```
MultiAgent.sources.SpoolDir.type = spoolDir
```

```
MultiAgent.sources.SpoolDir.spoolDir = /home/alumno/practicass_curso/flume/spoolDir
```

```
MultiAgent.sources.SpoolDir.deletePolicy = immediate
```

```

# Describimos el tercer agente Exec
MultiAgent.sources.Exec.type = exec
MultiAgent.sources.Exec.command = cat /home/alumno/practicas_curso/flume/datos/constitucion.txt

# Describimos los tres destinos como Avro en localhost:10003
MultiAgent.sinks.AvroSink1.type = avro
MultiAgent.sinks.AvroSink1.hostname = localhost
MultiAgent.sinks.AvroSink1.port = 10003

MultiAgent.sinks.AvroSink2.type = avro
MultiAgent.sinks.AvroSink2.hostname = localhost
MultiAgent.sinks.AvroSink2.port = 10003

MultiAgent.sinks.AvroSink3.type = avro
MultiAgent.sinks.AvroSink3.hostname = localhost
MultiAgent.sinks.AvroSink3.port = 10003

# Describimos los canales
MultiAgent.channels.FileChannel.type = file
MultiAgent.channels.FileChannel.dataDir = /home/alumno/practicas_curso/data
MultiAgent.channels.FileChannel.checkpointDir = /home/alumno/practicas_curso/flume/checkpoint
MultiAgent.channels.MemChannel1.type = memory
MultiAgent.channels.MemChannel2.type = memory

# Unimos los orígenes y destinos
MultiAgent.sources.Netcat.channels = FileChannel
MultiAgent.sources.Spooledir.channels = MemChannel1
MultiAgent.sources.Exec.channels = MemChannel2
MultiAgent.sinks.AvroSink1.channel = FileChannel
MultiAgent.sinks.AvroSink2.channel = MemChannel1
MultiAgent.sinks.AvroSink3.channel = MemChannel2

```

Hay que revisar todas las rutas, tanto en el sistema de archivos de linux como el de HDFS, para que coincidan con las que se tienen en la plataforma de trabajo

Lanzamiento del agente consolidador

En primer lugar lanzamos el agente consolidador para que Flume quede a la espera de mensajes Avro en localhost:10003

```

flume-ng agent --conf /etc/flume-ng/conf --conf-file avro-hdfs.conf \
  --name AvroHdfsAgent -Dflume.root.logger=INFO,console

```

Lanzamiento del multi agente

Para poder probar la correcta ejecución del multiagente es necesario para cada agente

- **Netstat:** ejecutar **nc localhost 10004** en un terminal una vez lanzado e introducir líneas
- **SpoolDir:** colocar un archivo de texto (por ejemplo, un documento CSV) en */home/alumno/practicas_curso/flume/spoolDir*
- **Exec:** el archivo */home/alumno/practicas_curso/flume/datos/constitucion.txt* está disponible

```

flume-ng agent --conf /etc/flume-ng/conf --conf-file multiagent-avro.conf \
  --name MultiAgent -Dflume.root.logger=INFO,console

```

Comprobación del funcionamiento

Una vez lanzados los dos agentes, en el consolidador (receptor) aparece el mensaje de que el source Avro está esperando

```
AvroSource: Successfully registered new MBean.  
.285 (lifecycleSupervisor-1-3) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.start(MonitoredCounterGroup.java:95)] Component type: SOURCE, name: Av  
.295 (lifecycleSupervisor-1-3) [INFO - org.apache.flume.source.AvroSource.start(AvroSource.java:223)] Avro source AvroSource started.
```

Figure 2: Avro Source arrancado

En el terminal donde se ha arrancado el multiagente aparecen multitud de mensajes ya que se están procesando los agentes Exec y SpoolDir inmediatamente.

Se pueden producir fallos en la ejecución debido a que los mensajes que se están generando saturan los canales utilizados de memoria.

```
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:169)] Starting Channel MemChannel2  
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:169)] Starting Channel FileChannel  
(lifecycleSupervisor-1-1) [INFO - org.apache.flume.channel.file.FileChannel.start(FileChannel.java:278)] Starting FileChannel FileChannel { dataDirs:  
/data] }...  
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:169)] Starting Channel MemChannel1  
(lifecycleSupervisor-1-2) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.register(MonitoredCounterGroup.java:119)] Monitored counter g
```

Figure 3: Canales de memoria y ficheros arrancados

```
(lifecycleSupervisor-1-2) [INFO - org.apache.flume.sink.AbstractRpcSink.createConnection(AbstractRpcSink.java:215)] Rpc sink AvroSink1: Building RpcClient with hostn  
003  
(lifecycleSupervisor-1-2) [INFO - org.apache.flume.sink.AvroSink.initializeRpcClient(AvroSink.java:126)] Attempting to create Avro Rpc client.  
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:196)] Starting Sink AvroSink2  
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:207)] Starting Source Exec  
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:207)] Starting Source Netcat  
(conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:207)] Starting Source SpoolDir  
(lifecycleSupervisor-1-6) [INFO - org.apache.flume.sink.AbstractRpcSink.start(AbstractRpcSink.java:297)] Starting RpcSink AvroSink2 { host: localhost, port: 10003 }.
```

Figure 4: Los tres Sources aparecen arrancados

Recoge en varios pantallazos los terminales en funcionamiento y los archivos creados en HDFS. Recuerda que se debe ver tu nombre en la imagen.

```

(lifecycleSupervisor-1-6) [INFO - org.apache.flume.sink.AbstractRpcSink.start(AbstractRpcSink.java:311)] Rpc sink AvroSink2 started.
(lifecycleSupervisor-1-2) [INFO - org.apache.flume.sink.AbstractRpcSink.start(AbstractRpcSink.java:311)] Rpc sink AvroSink1 started.
(lifecycleSupervisor-1-1) [INFO - org.apache.flume.sink.AbstractRpcSink.start(AbstractRpcSink.java:311)] Rpc sink AvroSink3 started.
(pool-7-thread-1) [WARN - org.apache.flume.source.SpooledDirectorySource$SpooledDirectoryRunnable.run(SpooledDirectorySource.java:275)] The channel is full,
will try again after 250 milliseconds
(pool-8-thread-1) [ERROR - org.apache.flume.source.ExecSource$ExecRunnable.run(ExecSource.java:356)] Failed while running command: cat /home/alumno/pr

```

Figure 5: Sinks disponibles

Por último queda verificar que los ficheros se ha escrito correctamente en el directorio de HDFS `/user/alumno/flume/avro_data/`. Deben aparecer multitud de ficheros de texto, cuyos contenidos alternan entre las líneas de *filmoteca.csv* y *constitucion.txt*

Home

/ user / alumno / flume / avro_data

Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		alumno	supergroup	drwxr-xr-x	March 27, 2022 03:23 PM
<input type="checkbox"/>	.		alumno	supergroup	drwxr-xr-x	March 27, 2022 03:24 PM
<input type="checkbox"/>	FlumeData.1648419809788	408 bytes	alumno	supergroup	-rw-r--r--	March 27, 2022 03:23 PM
<input type="checkbox"/>	FlumeData.1648419809789	833 bytes	alumno	supergroup	-rw-r--r--	March 27, 2022 03:23 PM
<input type="checkbox"/>	FlumeData.1648419809790	384 bytes	alumno	supergroup	-rw-r--r--	March 27, 2022 03:23 PM
<input type="checkbox"/>	FlumeData.1648419809791	695 bytes	alumno	supergroup	-rw-r--r--	March 27, 2022 03:23 PM

Figure 6: Ficheros generados