PRÁCTICA Nº4.5
Actividad 4.5. Agente Flume con sumidero en HDFS

- Nombre y apellidos: Alvaro Lucio-Villegas De Cea

Índice

Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

Enunciado

Vamos a crear un agente Flume, de tal forma que registre eventos que vamos a generar modificando un el fichero de texto que estamos utilizando como si fuera un fichero de Log.

La idea es que se cree el fichero, y se vaya incrementando su contenido de forma dinámica, de tal forma que los cambios registrados en el fichero, no se muestren en la consola, sino que se vuelquen en ficheros de texto de HDFS. Como sabemos, el fichero de log es /tmp/log.txt.

Vamos a preparar el Fichero de Configuración.

- Utiliza el directorio flumeconf en la carpeta /home/cloudera, que ya utilizamos en la actividad 4.4

- Dentro de ese directorio, crea el fichero agenteficherologhdfs.conf.

- Utiliza como nombre del agente agenteficherologhdfs.

- Ese fichero debe contener todos los elementos de configuración que hemos visto en teoría. Vamos a concretarlos:

Como puede verse en el contenido del fichero que se muestra a continuación, aparece una configuración de propiedades del SINK, del sumidero mucho más elaborada, con cierto grado de complejidad. Este SINK o SUMIDERO es el destino en el que volcamos el contenido detectado por cada evento.

A continuación, una descripción de lo que representa cada elemento:

Propiedad	Valor	Significado
.sinks.sink1.type	**hdfs**	Tipo de destino, en este caso hdfs.
.sinks.sink1.hdfs.path	**/flume/events/%y-%m-%d/%H%M**	Ruta del cluster hdfs en la que depositará los eventos leidos de la fuente (source)
sink1.hdfs.filePrefix	**events**	Esto es un texto que utiliza flume para nombrar al fichero que genera en cada evento.
sinks.sink1.hdfs.round	**true**	Indica si debe redondear la marca de tiempo utilizada para construir los ficheros y carpetas que genera
sinks.sink1.hdfs.roundValue	**1**	Utiliza esta cantidad para partir las capturas de evento en ficheros y carpetas. La unida que se utiliza es la que se indique en **roundUnit**, que en este caso son minutos.
sinks.sink1.hdfs.roundUnit	**minute**	Unidad para partir las carpetas
sinks.sink1.hdfs.useLocalTimeStamp	**true**	Usa el tiempo del servidor local, en lugar del tiempo que venga en la cabecera del evento.
sinks.sink1.hdfs.fileType	**DataStream**	Tipo de fichero generado. Se admiten los siguientes: SequenceFile, DataStream or CompressedStream

Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

Fichero agenteficherologhdfs.conf:

```
# Definición de componentes del agente
agenteficherolog.sources = source1
agenteficherolog.sinks = sink1
agenteficherolog.channels = ch1
# Configuración de propiedades del source
agenteficherolog.sources.source1.type = exec
agenteficherolog.sources.source1.command = tail -F /tmp/log.txt
agenteficherolog.sources.source1.shell = /bin/bash -c
# Configuración de propiedades del sink
agenteficherologhdfs.sinks.sink1.type = hdfs
agenteficherologhdfs.sinks.sink1.hdfs.path = /flume/events/%y-%m-%d/%H%M
agenteficherologhdfs.sinks.sink1.hdfs.filePrefix = events
agenteficherologhdfs.sinks.sink1.hdfs.round = true
agenteficherologhdfs.sinks.sink1.hdfs.roundValue = 1
agenteficherologhdfs.sinks.sink1.hdfs.roundUnit = minute
agenteficherologhdfs.sinks.sink1.hdfs.useLocalTimeStamp = true
agenteficherologhdfs.sinks.sink1.hdfs.fileType = DataStream
# Configuración de un canal de tipo memoria
agenteficherolog.channels.ch1.type = memory
agenteficherolog.channels.ch1.capacity = 1000
agenteficherolog.channels.ch1.transactionCapacity = 100
# Vincular source y sink al canal creado
agenteficherolog.sources.source1.channels = ch1
agenteficherolog.sinks.sink1.channel = ch1
```


Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

agenteficherologhdfs.conf | log.txt

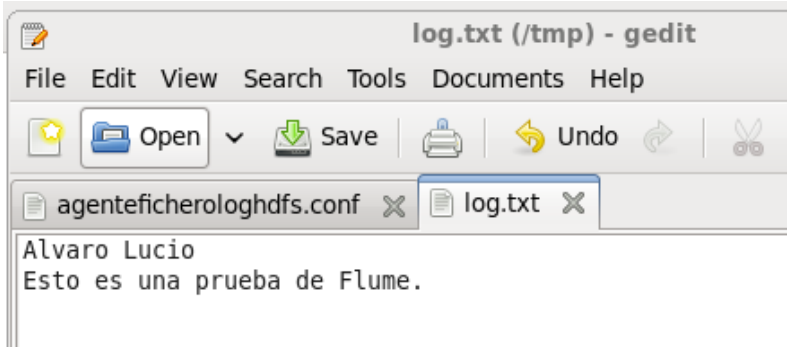
```
# Definición de componentes del agente
agenteficherologhdfs.sources = source1
agenteficherologhdfs.sinks = sink1
agenteficherologhdfs.channels = ch1

# Configuración de propiedades del source
agenteficherologhdfs.sources.source1.type = exec
agenteficherologhdfs.sources.source1.command = tail -F /tmp/log.txt
agenteficherologhdfs.sources.source1.shell = /bin/bash -c

# Configuración de propiedades del sink
agenteficherologhdfs.sinks.sink1.type = hdfs
agenteficherologhdfs.sinks.sink1.hdfs.path = /flume/events/%y-%m-%d/%H%M
agenteficherologhdfs.sinks.sink1.hdfs.filePrefix = events
agenteficherologhdfs.sinks.sink1.hdfs.round = true
agenteficherologhdfs.sinks.sink1.hdfs.roundValue = 1
agenteficherologhdfs.sinks.sink1.hdfs.roundUnit = minute
agenteficherologhdfs.sinks.sink1.hdfs.useLocalTimeStamp = true
agenteficherologhdfs.sinks.sink1.hdfs.fileType = DataStream

# Configuración de un canal de tipo memoria
agenteficherologhdfs.channels.ch1.type = memory
agenteficherologhdfs.channels.ch1.capacity = 1000
agenteficherologhdfs.channels.ch1.transactionCapacity = 100

# Vincular source y sink al canal creado
agenteficherologhdfs.sources.source1.channels = ch1
agenteficherologhdfs.sinks.sink1.channel = ch1
```

Prepara el fichero log.txt:

- Crea el fichero log.txt en la ruta /tmp con el editor de texto de la máquina Cloudera.

- Añade un texto cualquiera

- Cierra el Editor



Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

Crea el directorio de Volcado en HDFS:

- Crea el directorio /flume/events en HFDS, utilizando un terminal y las líneas de comando, mediante el comando "hdfs dfs.....".

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /flume
[cloudera@quickstart ~]$ hdfs dfs -mkdir /flume/events
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 2 items
drwxr-xr-x   - cloudera cloudera          0 2023-03-20 13:00 _sqoop
drwxr-xr-x   - cloudera cloudera          0 2023-03-20 12:36 dataset-Telefonia
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 10 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 12:09 /flume
drwxr-xr-x   - hbase    supergroup          0 2023-03-13 12:16 /hbase
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxr-xr-x   - cloudera supergroup          0 2023-03-13 12:24 /tablascolhdfs
drwxr-xr-x   - cloudera supergroup          0 2023-03-29 12:02 /tablashdfs
drwxr-xr-x   - root     supergroup          0 2023-03-29 11:39 /tablashdfsalv
drwxrwxrwt   - hdfs     supergroup          0 2023-02-22 12:34 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$
```

Crea el Agente Flume:

Siguiendo las indicaciones que aparecen en teoría para crear el agente Flume, construye el

comando correspondiente, que tenga en cuenta lo siguiente:

- Nombre del Agente: agenteficherologhdfs
- Ruta del Directorio de Configuración: /home/cloudera/flumeconf
- Nombre del Fichero de configuración: agenteficherologhdfs.conf

```
flume-ng agent - agenteficherologhdfs -f
/home/cloudera/flumeconf/agenteficherologhdfs.conf
-Dflume.root.logger=INFO,console–Xmx512m
```


Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

```
[cloudera@quickstart ~]$ flume-ng agent -n agenteficherologhdfs -f /home/cloudera/flumeconf/agenteficherologhdfs.conf -Dflume.root.logger=INFO,console--Xmx512m
Warning: No configuration directory set! Use --conf <dir> to override.
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 -Dflume.root.logger=INFO,console--Xmx512m -cp '/usr/lib/flume-ng/lib/*:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/.//*:/usr/lib/hado
/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/ja
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
ib/hbase/bin/../lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/hbase/bin/../lib/api-util-1.0.0-M20.jar:/usr/lib/hbase/bin/../lib/asm-3.2.jar:/usr/lib/hbase/bin/../lib/avro.jar:/usr/lib/hbase/bin/../lib/aws-
-1.11.134.jar:/usr/lib/hbase/bin/../lib/commons-beanutils-1.9.2.jar:/usr/lib/hbase/bin/../lib/commons-beanutils-core-1.8.0.jar:/usr/lib/hbase/bin/../lib/commons-cli-1.2.jar:/usr/lib/hbase/bin/../lib/
.9.jar:/usr/lib/hbase/bin/../lib/commons-collections-3.2.2.jar:/usr/lib/hbase/bin/../lib/commons-compress-1.4.1.jar:/usr/lib/hbase/bin/../lib/commons-configuration-1.6.jar:/usr/lib/hbase/bin/../lib/c
.0.13.jar:/usr/lib/hbase/bin/../lib/commons-digester-1.8.jar:/usr/lib/hbase/bin/../lib/commons-el-1.0.jar:/usr/lib/hbase/bin/../lib/commons-httpclient-3.1.jar:/usr/lib/hbase/bin/../lib/commons-io-2.4
base/bin/../lib/commons-lang-2.6.jar:/usr/lib/hbase/bin/../lib/commons-logging-1.2.jar:/usr/lib/hbase/bin/../lib/commons-math-2.1.jar:/usr/lib/hbase/bin/../lib/commons-math3-3.1.1.jar:/usr/lib/hbase/
ons-net-3.1.jar:/usr/lib/hbase/bin/../lib/core-3.1.1.jar:/usr/lib/hbase/bin/../lib/curator-client-2.7.1.jar:/usr/lib/hbase/bin/../lib/curator-framework-2.7.1.jar:/usr/lib/hbase/bin/../lib/curator-rec
/usr/lib/hbase/bin/../lib/disruptor-3.3.0.jar:/usr/lib/hbase/bin/../lib/findbugs-annotations-1.3.9-1.jar:/usr/lib/hbase/bin/../lib/gson-2.2.4.jar:/usr/lib/hbase/bin/../lib/guava-12.0.1.jar:/usr/lib/h
/hamcrest-core-1.3.jar:/usr/lib/hbase/bin/../lib/hbase-annotations-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-annotations-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/../lib/hbase-client-1.2
r:/usr/lib/hbase/bin/../lib/hbase-common-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-common-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/../lib/hbase-examples-1.2.0-cdh5.13.0.jar:/usr/lib/hba
base-external-blockcache-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-hadoop2-compat-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-hadoop2-compat-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase
se-hadoop-compat-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-hadoop-compat-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/../lib/hbase-it-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-it-
-tests.jar:/usr/lib/hbase/bin/../lib/hbase-prefix-tree-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-procedure-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-protocol-1.2.0-cdh5.13.0.jar:/
in/../lib/hbase-resource-bundle-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-rest-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-rsgroup-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbas
-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/../lib/hbase-server-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-server-1.2.0-cdh5.13.0-tests.jar:/usr/lib/hbase/bin/../lib/hbase-shell-1.2.0-cdh5.13.0.
ase/bin/../lib/hbase-spark-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/hbase-thrift-1.2.0-cdh5.13.0.jar:/usr/lib/hbase/bin/../lib/high-scale-lib-1.1.1.jar:/usr/lib/hbase/bin/../lib/hsqldb-1.8.0.10.
ase/bin/../lib/htrace-core-3.2.0-incubating.jar:/usr/lib/hbase/bin/../lib/htrace-core4-4.0.1-incubating.jar:/usr/lib/hbase/bin/../lib/htrace-core.jar:/usr/lib/hbase/bin/../lib/httpclient-4.2.5.jar:/u
n/../lib/httpcore-4.2.5.jar:/usr/lib/hbase/bin/../lib/jackson-annotations-2.2.3.jar:/usr/lib/hbase/bin/../lib/jackson-core-2.2.3.jar:/usr/lib/hbase/bin/../lib/jackson-core-asl-1.8.8.jar:/usr/lib/hbas
ckson-databind-2.2.3.jar:/usr/lib/hbase/bin/../lib/jackson-jaxrs-1.8.8.jar:/usr/lib/hbase/bin/../lib/jackson-mapper-asl-1.8.8.jar:/usr/lib/hbase/bin/../lib/jackson-xc-1.8.8.jar:/usr/lib/hbase/bin/../
me-2.4.1.jar:/usr/lib/hbase/bin/../lib/jasper-compiler-5.5.23.jar:/usr/lib/hbase/bin/../lib/jasper-runtime-5.5.23.jar:/usr/lib/hbase/bin/../lib/java-xmlbuilder-0.4.jar:/usr/lib/hbase/bin/../lib/jaxb-
r/lib/hbase/bin/../lib/jaxb-impl-2.2.3-1.jar:/usr/lib/hbase/bin/../lib/jcodings-1.0.8.jar:/usr/lib/hbase/bin/../lib/jersey-client-1.9.jar:/usr/lib/hbase/bin/../lib/jersey-core-1.9.jar:/usr/lib/hbase/
ey-json-1.9.jar:/usr/lib/hbase/bin/../lib/jersey-server-1.9.jar:/usr/lib/hbase/bin/../lib/jets3t-0.9.0.jar:/usr/lib/hbase/bin/../lib/jettison-1.3.3.jar:/usr/lib/hbase/bin/../lib/jetty-6.1.26.cloudera
/hbase/bin/../lib/jetty-sslengine-6.1.26.cloudera.4.jar:/usr/lib/hbase/bin/../lib/jetty-util-6.1.26.cloudera.4.jar:/usr/lib/hbase/bin/../lib/joni-2.1.2.jar:/usr/lib/hbase/bin/../lib/jruby-cloudera-1.
b/hbase/bin/../lib/jsch-0.1.42.jar:/usr/lib/hbase/bin/../lib/jsp-2.1-6.1.14.jar:/usr/lib/hbase/bin/../lib/jsp-api-2.1-6.1.14.jar:/usr/lib/hbase/bin/../lib/jsp-api-2.1.jar:/usr/lib/hbase/bin/../lib/ju
sr/lib/hbase/bin/../lib/leveldbjni-all-1.8.jar:/usr/lib/hbase/bin/../lib/libthrift-0.9.3.jar:/usr/lib/hbase/bin/../lib/log4j-1.2.17.jar:/usr/lib/hbase/bin/../lib/metrics-core-2.2.0.jar:/usr/lib/hbase
ty-all-4.0.23.Final.jar:/usr/lib/hbase/bin/../lib/paranamer-2.3.jar:/usr/lib/hbase/bin/../lib/protobuf-java-2.5.0.jar:/usr/lib/hbase/bin/../lib/servlet-api-2.5-6.1.14.jar:/usr/lib/hbase/bin/../lib/se
ar:/usr/lib/hbase/bin/../lib/slf4j-api-1.7.5.jar:/usr/lib/hbase/bin/../lib/slf4j-log4j12.jar:/usr/lib/hbase/bin/../lib/snappy-java-1.0.4.1.jar:/usr/lib/hbase/bin/../lib/spymemcached-2.11.6.jar:/usr/l
/lib/xmlenc-0.52.jar:/usr/lib/hbase/bin/../lib/xz-1.0.jar:/usr/lib/hbase/bin/../lib/zookeeper.jar:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/.//*:/usr/lib/hadoop-hdfs/./:/usr/lib/hadoop-h
lib/hadoop-hdfs/.//*:/usr/lib/hadoop-yarn/lib/*:/usr/lib/hadoop-yarn/.//*:/usr/lib/hadoop-mapreduce/lib/*:/usr/lib/hadoop-mapreduce/.//*:/etc/hadoop/conf:/usr/lib/hadoop/*:/usr/lib/hadoop/lib/*:/usr/
:/usr/lib/zookeeper/lib/*::/conf:/lib/*:/usr/lib/flume-ng/../search/lib/antlr-2.7.7.jar:/usr/lib/flume-ng/../search/lib/antlr-runtime-3.4.jar:/usr/lib/flume-ng/../search/lib/apacheds-kerberos-codec-2
sr/lib/flume-ng/../search/lib/apache-mime4j-core-0.7.2.jar:/usr/lib/flume-ng/../search/lib/apache-mime4j-dom-0.7.2.jar:/usr/lib/flume-ng/../search/lib/api-asn1-api-1.0.0-M20.jar:/usr/lib/flume-ng/../
util-1.0.0-M20.jar:/usr/lib/flume-ng/../search/lib/argparse4j-0.4.3.jar:/usr/lib/flume-ng/../search/lib/asm-3.2.jar:/usr/lib/flume-ng/../search/lib/asm-4.1.jar:/usr/lib/flume-ng/../search/lib/asm-com
sr/lib/flume-ng/../search/lib/asm-debug-all-4.1.jar:/usr/lib/flume-ng/../search/lib/aspectjrt-1.6.5.jar:/usr/lib/flume-ng/../search/lib/avro.jar:/usr/lib/flume-ng/../search/lib/bcmail-jdk15-1.45.jar:
```

```
23/04/10 12:20:02 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
23/04/10 12:20:02 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:/home/cloudera/flumeconf/agenteficherologhdfs.conf
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Processing:sink1
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Added sinks: sink1 Agent: agenteficherologhdfs
23/04/10 12:20:02 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agenteficherologhdfs]
23/04/10 12:20:02 INFO node.AbstractConfigurationProvider: Creating channels
23/04/10 12:20:02 INFO channel.DefaultChannelFactory: Creating instance of channel ch1 type memory
23/04/10 12:20:02 INFO node.AbstractConfigurationProvider: Created channel ch1
23/04/10 12:20:02 INFO source.DefaultSourceFactory: Creating instance of source source1, type exec
23/04/10 12:20:02 INFO sink.DefaultSinkFactory: Creating instance of sink: sink1, type: hdfs
23/04/10 12:20:02 INFO node.AbstractConfigurationProvider: Channel ch1 connected to [source1, sink1]
23/04/10 12:20:02 INFO node.Application: Starting new configuration:{ sourceRunners:{source1=EventDrivenSourceRunner: { source:org.apache.flume.source.ExecSource{name:source1,state:IDLE} }} sinkRunners:{sink1=SinkR
unner: { policy:org.apache.flume.sink.DefaultSinkProcessor@7da707a0 counterGroup:{ name:null counters:{} } }} channels:{ch1=org.apache.flume.channel.MemoryChannel{name: ch1}} }
23/04/10 12:20:02 INFO node.Application: Starting Channel ch1
23/04/10 12:20:02 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: ch1: Successfully registered new MBean.
23/04/10 12:20:02 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: ch1 started
23/04/10 12:20:02 INFO node.Application: Starting Sink sink1
23/04/10 12:20:02 INFO node.Application: Starting Source source1
23/04/10 12:20:02 INFO source.ExecSource: Exec source starting with command: tail -F /tmp/log.txt
23/04/10 12:20:02 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: sink1: Successfully registered new MBean.
23/04/10 12:20:02 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: sink1 started
23/04/10 12:20:02 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: source1: Successfully registered new MBean.
23/04/10 12:20:02 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: source1 started
23/04/10 12:20:06 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
23/04/10 12:20:06 INFO hdfs.BucketWriter: Creating /flume/events/23-04-10/1220/events.1681154406607.tmp
23/04/10 12:20:38 INFO hdfs.BucketWriter: Closing /flume/events/23-04-10/1220/events.1681154406607.tmp
23/04/10 12:20:38 INFO hdfs.BucketWriter: Renaming /flume/events/23-04-10/1220/events.1681154406607.tmp to /flume/events/23-04-10/1220/events.1681154406607
23/04/10 12:20:38 INFO hdfs.HDFSEventSink: Writer callback called.
```

Al lanzar el agente, debería verse en la terminal correspondiente lo siguiente:

A partir de este momento, el agente está escuchando en las operaciones que se hagan sobre el fichero /tmp/log.txt, y reaccionará a cada cambio del fichero de forma periódica, creando, si no existe una carpeta con el día actual, una carpeta con el minuto actual y dentro generará ficheros que tendrán nombres como “events.12331242142”.

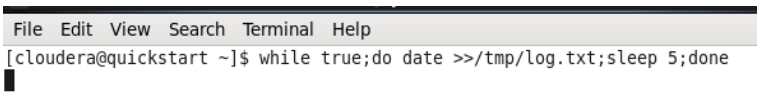
```
[cloudera@quickstart ~]$ hdfs dfs -ls /flume/events
Found 1 items
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 12:20 /flume/events/23-
04-10
[cloudera@quickstart ~]$ hdfs dfs -ls /flume/events
Found 1 items
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 12:20 /flume/events/23-04-10
[cloudera@quickstart ~]$ hdfs dfs -ls /flume/events/23-04-10
Found 1 items
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 12:20 /flume/events/23-04-10/1220
[cloudera@quickstart ~]$ hdfs dfs -ls /flume/events/23-04-10/1220
Found 1 items
-rw-r--r--   1 cloudera supergroup         43 2023-04-10 12:20 /flume/events/23-04-10/1220/events.1681154406607
[cloudera@quickstart ~]$ 
```

Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

Modifica el Fichero /temp/log.txt

- Para modificar el fichero, vamos a ejecutar un código que añada líneas de forma periódica al fichero. Para eso, utilizaremos la siguiente instrucción, que añade una nueva línea con fecha y hora cada 5 segundos:

```
while true; do date >> /tmp/log.txt; sleep 5; done
```



Comprobamos en HDFS que se están registrando todos los cambios con fecha y hora

```
[cloudera@quickstart ~]$ hdfs dfs -ls /flume/events/23-04-10/
Found 2 items
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 12:20 /flume/events/23-04-10/1220
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 12:25 /flume/events/23-04-10/1225
[cloudera@quickstart ~]$ hdfs dfs -ls /flume/events/23-04-10/1225
Found 2 items
-rw-r--r--   1 cloudera supergroup        203 2023-04-10 12:25 /flume/events/23-04-10/1225/events.1681154705892
-rw-r--r--   1 cloudera supergroup         29 2023-04-10 12:25 /flume/events/23-04-10/1225/events.1681154736018.tmp
[cloudera@quickstart ~]$
```

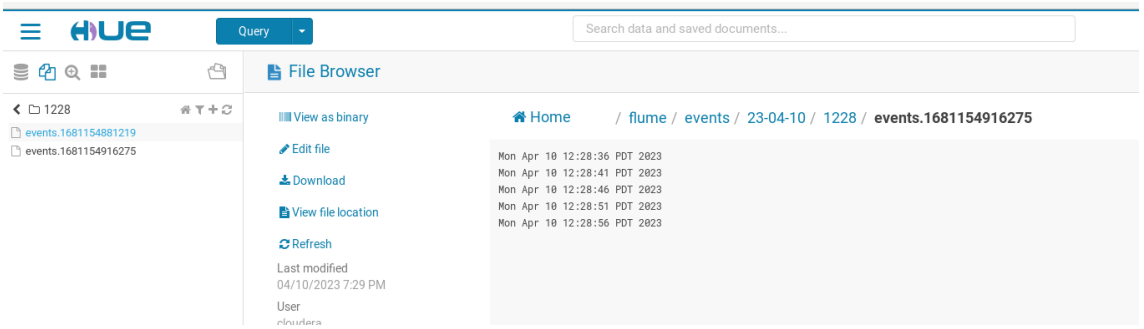

Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

Responde a las siguientes cuestiones:

- Observa la terminal donde has lanzado el agente flume. ¿Qué ocurre?

-Se está mostrando toda la información sobre los sucesos del fichero.

- Observa el directorio HDFS desde el navegador, mediante la herramienta HUE. Consulta el contenido de los directorios y de los ficheros.



- ¿Cada cuanto se genera un nuevo directorio en el directorio /flume/events del HDFS?

–Cada dia

¿Y en el directorio /flume/events/%y-%m-%d?

-Cada Minuto

- ¿Cada cuanto se genera un nuevo fichero de evento? ¿A qué crees que se debe? Revisa la propiedad rollInterval en la documentación de Flume

-Por defecto está puesto cada 30 segundos.Y esto sucede por el parámetro “rollintervall” (https://flume.apache.org/FlumeUserGuide.html#flume-sinks)

.

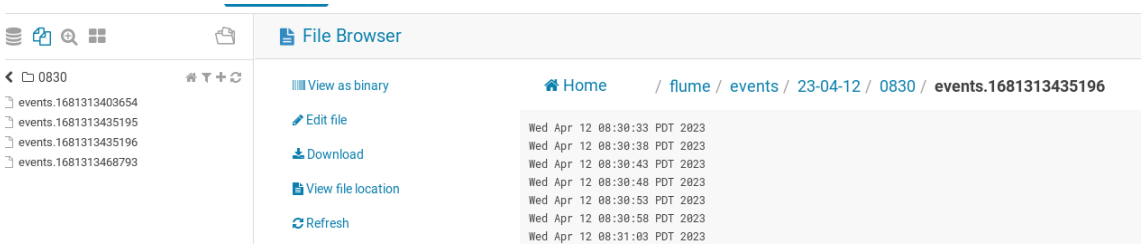
Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

Cambia para que se generen 4 ficheros de eventos cada minuto.

Se agrega en el fichero de configuración el parámetro “rollintervall”

agenteficherologhdfs.conf | log.txt

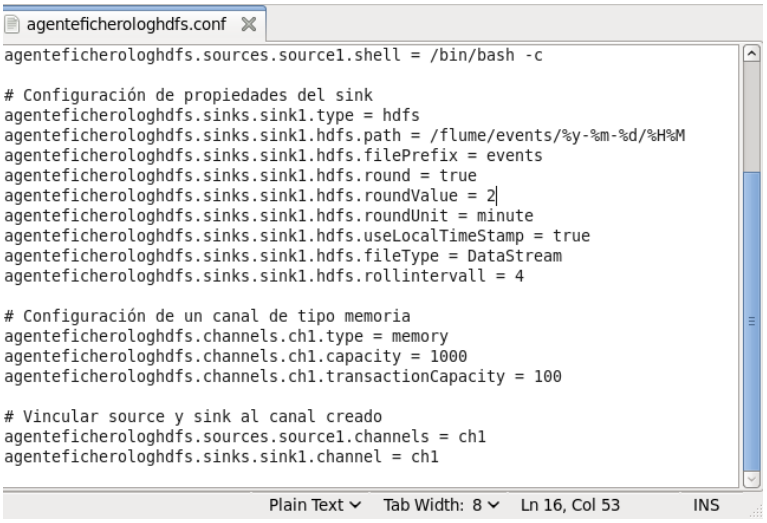
```
# Definición de componentes del agente
agenteficherologhdfs.sources = source1
agenteficherologhdfs.sinks = sink1
agenteficherologhdfs.channels = ch1

# Configuración de propiedades del source
agenteficherologhdfs.sources.source1.type = exec
agenteficherologhdfs.sources.source1.command = tail -F /tmp/log.txt
agenteficherologhdfs.sources.source1.shell = /bin/bash -c

# Configuración de propiedades del sink
agenteficherologhdfs.sinks.sink1.type = hdfs
agenteficherologhdfs.sinks.sink1.hdfs.path = /flume/events/%y-%m-%d/%H%M
agenteficherologhdfs.sinks.sink1.hdfs.filePrefix = events
agenteficherologhdfs.sinks.sink1.hdfs.round = true
agenteficherologhdfs.sinks.sink1.hdfs.roundValue = 1
agenteficherologhdfs.sinks.sink1.hdfs.roundUnit = minute
agenteficherologhdfs.sinks.sink1.hdfs.useLocalTimeStamp = true
agenteficherologhdfs.sinks.sink1.hdfs.fileType = DataStream
agenteficherologhdfs.sinks.sink1.hdfs.rollintervall = 14

# Configuración de un canal de tipo memoria
agenteficherologhdfs.channels.ch1.type = memory
agenteficherologhdfs.channels.ch1.capacity = 1000
agenteficherologhdfs.channels.ch1.transactionCapacity = 100

# Vincular source y sink al canal creado
agenteficherologhdfs.sources.source1.channels = ch1
agenteficherologhdfs.sinks.sink1.channel = ch1
```



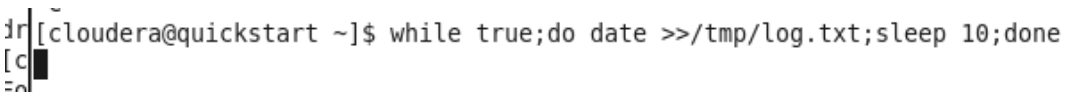
Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

- Cambia a 2 minutos la agrupación de eventos en carpetas. Prueba a hacerlo sin parar el agente flume ¿Qué ocurre?

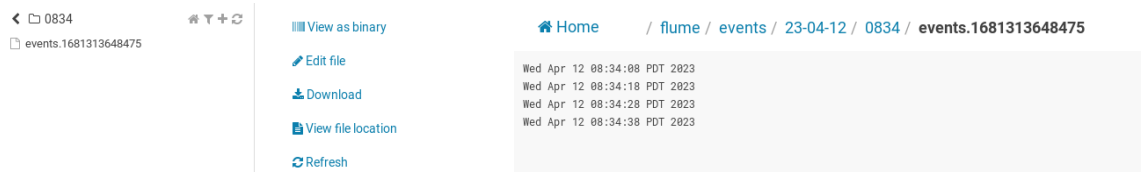
Se almacenan más eventos en una sola carpeta.



- Modifica el proceso que añade líneas al fichero de log, para que escriba cada 10 segundos, en lugar de cada 5 segundos ¿Qué ocurre con los ficheros generados?



Se generan menos entradas en el log.

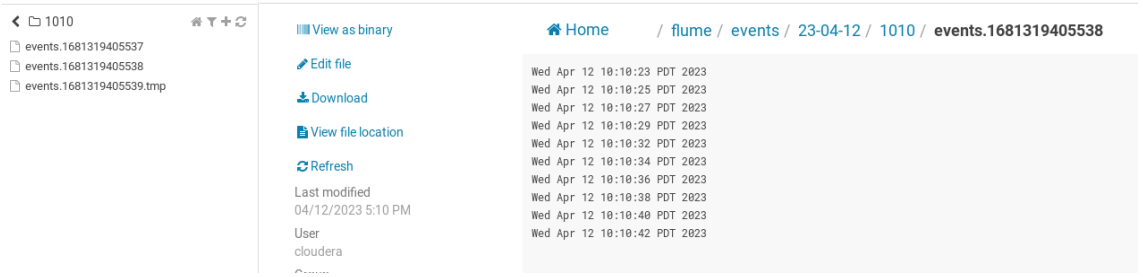


Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS

-Cámbialo a 2 segundos y explica lo que ocurre con los ficheros generados en esta ocasión

```
[cloudera@quickstart ~]$ while true;do date >>/tmp/log.txt;sleep 2;done
```

Se generan más entradas en el log.



Nombre y Apellidos: Alvaro-Lucio-Villegas de Cea
Título de la práctica: Actividad 4.5. Agente Flume con sumidero en HDFS