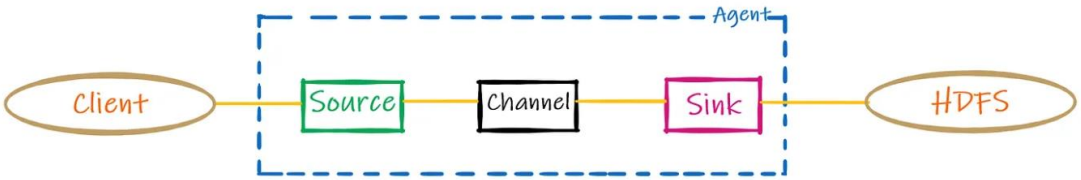
AYMANE HINANE EMSI CENTRE G5 4IIR

Flume est un système de collecte de donne , est de transfert de donne nom structure vers plusieurs source

Architecture



Il y'a 3 composant principaux dans flume :

Source :
Extrait des donnees de plusieur source on se basant sur le pattern envent driven
Est transfert les donnees vers à travers un Channel vers une destination
Les donnees reçu peuvent être soit des log ou de click générer d'un site web

Channel :
Permet le transfert des donnees depuis une source vers une destination
Il y'a 3 type de Channel : memory (développement) , File (production) , JDBC
Flume suis une approche transactionnel dans le stockage le transfert de donne
Un évènement ne serait pas supprimer du Channel jusqu' à ce que il soit reçu par
La destination est que le Sink renvoie une réponse
NB :
Parfois le Sink peut être plus lent que la Source ce qui peut entrainer une Corruption

Sink :
Est la destination qui reçois la donner, elle peut être soit : HDFS,Hbase,Kafka, Flume agent Log

Exemple 1 Netcat:

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/simple-example# ls
example.conf  node_modules  package-lock.json  package.json
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/simple-example# cat example.conf

a1.sources = r1
a1.sinks = k1
a1.channels = c1


a1.sources.r1.type = netcat
a1.sources.r1.bind = 0.0.0.0
a1.sources.r1.port = 44444

a1.sinks.k1.type = logger

a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
a1.channels.c1.transactionCapacity = 100

a1.sources.r1.channels = c1
a1.sinks.k1.channel = c1
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/simple-example#
```

Dans cette exemple j'ai utilisé une source de type netcat
Netcat permet d'écouter un port est return chaque ligne écrit vers un évènement Flume
Les paramètre que prend Netcat :
-type : netcat
-bind: host or ip address to bind to
-port : le port que je vais écouter

Le type du Channel est memory
Les évènement sont stockes dans une file d'attente en mémoire avec une taille maximale configurable.
Capacity: the maximum number of events stored in the channel
transactionCapacity: le maximum d'évènement que le Channel peut transférer

le type du Sink est logger
les evenement passer par le channel seront envoyer vers flume agent log

Execution

flume-ng agent --name=a1 --conf-file example.conf

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/simple-example# flume-ng agent --name=a1 --conf-file example.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/hdoop/hadoop-3.3.5/bin/hadoop) for HDFS access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/lib/jvm/java-8-openjdk-arm64/bin/java -Xmx20m -cp '/home/hdoop/apache-flume-1.11.0-bin/lib/*:/home/hdoop/hadoop-3.3.5/etc/hadoop:/home/hdoop/hadoop
-3.3.5/share/hadoop/common/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/common/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs:/home/hdoop/hadoop-3.3.5/share/hadoo
p/hdfs/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs/*:/home/hdoop/hadoop-3.3.5/share/hadoop/mapreduce/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn:/home/hd
oop/hadoop-3.3.5/share/hadoop/yarn/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn/*:/lib/*' -Djava.library.path=:/home/hdoop/hadoop-3.3.5/lib/native org.apac
he.flume.node.Application --name=a1 --conf-file example.conf
2023-04-03 22:58:31,466 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Added sinks: k1 Agent: a1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 22:58:31,467 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 22:58:31,467 WARN conf.FlumeConfiguration: Agent configuration for 'a1' has no configfilters.
2023-04-03 22:58:31,480 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [a1]
2023-04-03 22:58:31,480 INFO node.AbstractConfigurationProvider: Creating channels
2023-04-03 22:58:31,487 INFO channel.DefaultChannelFactory: Creating instance of channel c1 type memory
2023-04-03 22:58:31,489 INFO node.AbstractConfigurationProvider: Created channel c1
2023-04-03 22:58:31,490 INFO source.DefaultSourceFactory: Creating instance of source r1, type netcat
2023-04-03 22:58:31,496 INFO sink.DefaultSinkFactory: Creating instance of sink: k1, type: logger
2023-04-03 22:58:31,498 INFO node.AbstractConfigurationProvider: Channel c1 connected to [r1, k1]
2023-04-03 22:58:31,498 INFO node.Application: Initializing components
2023-04-03 22:58:31,501 INFO node.Application: Starting new configuration:{ sourceRunners:{r1=EventDrivenSourceRunner: { source:org.apache.flume.source.NetcatS
ource{name:r1,state:IDLE} }} sinkRunners:{k1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@14d3bc22 counterGroup:{ name:null counters:{} } }}
 channels:{c1=org.apache.flume.channel.MemoryChannel{name: c1}} }
2023-04-03 22:58:31,502 INFO node.Application: Starting Channel c1
2023-04-03 22:58:31,503 INFO node.Application: Waiting for channel: c1 to start. Sleeping for 500 ms
2023-04-03 22:58:31,546 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: c1: Successfully registered new MBean.
2023-04-03 22:58:31,546 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: c1 started
2023-04-03 22:58:32,005 INFO node.Application: Starting Sink k1
2023-04-03 22:58:32,007 INFO node.Application: Starting Source r1
2023-04-03 22:58:32,008 INFO source.NetcatSource: Source starting
2023-04-03 22:58:32,026 INFO source.NetcatSource: Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/0.0.0.0:44444]
```

Test

To get the host name
hostname -f

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/simple-example# telnet 82ab94650db5 44444
Trying 172.17.0.2...
Connected to 82ab94650db5.
Escape character is '^]'.
hello
OK
good by
OK
```

```
2023-04-03 22:58:32,026 INFO source.NetcatSource: Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/0.0.0.0:44444]
2023-04-03 22:59:02,051 INFO sink.LoggerSink: Event: { headers:{} body: 68 65 6C 6C 6F 0D                         hello. }
2023-04-03 23:08:28,405 INFO sink.LoggerSink: Event: { headers:{} body: 67 6F 6F 64 20 62 79 0D                   good by. }
```

Exemple 2 Sequence Generator Source:

Se type de source permet de generer des evenement d'une facon continue l'evement commence par 0 et s'incremente par 1 a chaque fois jusequ'a ce que le totalEvents soit atteint

```
a1.sources = r1
a1.sinks = k1
a1.channels = c1


a1.sources.r1.type = seq
a1.sources.r1.totalEvents = 4

a1.sinks.k1.type = logger

a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
a1.channels.c1.transactionCapacity = 100

a1.sources.r1.channels = c1
a1.sinks.k1.channel = c1
```

flume-ng agent --name=a1 --conf-file exampleSeq.conf

```
2023-04-03 23:25:07,577 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: c1: Successfully registered new MBean.
2023-04-03 23:25:07,578 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: c1 started
2023-04-03 23:25:08,033 INFO node.Application: Starting Sink k1
2023-04-03 23:25:08,035 INFO node.Application: Starting Source r1
2023-04-03 23:25:08,038 INFO source.SequenceGeneratorSource: Sequence generator source do starting
2023-04-03 23:25:08,039 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: r1: Successfully registered new MBean.
2023-04-03 23:25:08,045 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: r1 started
2023-04-03 23:25:08,054 INFO sink.LoggerSink: Event: { headers:{} body: 30                                             0 }
2023-04-03 23:25:08,054 INFO sink.LoggerSink: Event: { headers:{} body: 31                                             1 }
2023-04-03 23:25:08,055 INFO sink.LoggerSink: Event: { headers:{} body: 32                                             2 }
2023-04-03 23:25:08,055 INFO sink.LoggerSink: Event: { headers:{} body: 33                                             3 }
```

Exemple 2 Générateur de Log:

Dans cette exemple j'ai créé un script qui génère un log et l'écrit dans un fichier tous les 10s

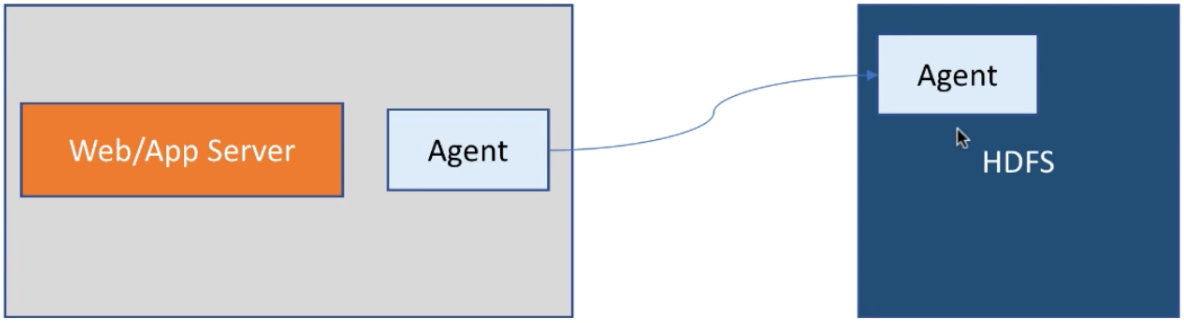
```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo# ls
logs  logs-generator  multi-simple  multi-sink  simple-example
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo# cd logs-generator
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/logs-generator# ls
data1.txt  generate_log.sh  start_log.sh  stop_log.sh
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/logs-generator#
```

generate_log.sh : génère les logs
start_log .sh : démarre le script log
stop_log.sh : arrête le script log

```
data1.txt  generate_log.sh  start_log.sh  stop_log.sh
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/logs-generator# tail -f data1.txt
data1 + 2023-04-03 23:29:01
data1 + 2023-04-03 23:29:11
data1 + 2023-04-03 23:29:21
data1 + 2023-04-03 23:29:31
data1 + 2023-04-03 23:29:41
data1 + 2023-04-03 23:29:51
data1 + 2023-04-03 23:30:01
data1 + 2023-04-03 23:30:11
data1 + 2023-04-03 23:30:21
data1 + 2023-04-03 23:30:31
data1 + 2023-04-03 23:30:41
```

Explication du test :

Une source peut avoir plusieur type :
-Hdfs
-syslog : pour lire les logs d'un conteneur externe
-avro : push data to another flume agent
-exec: this approche are used if the sender of log and the receiver of log are in the same Cotainer
Exec prend une commande, la sortie de cette de cette commande serrait pris on entrer
Pour l'entrée de la source : tail -f <file-name>



Channelize your data using Flume.

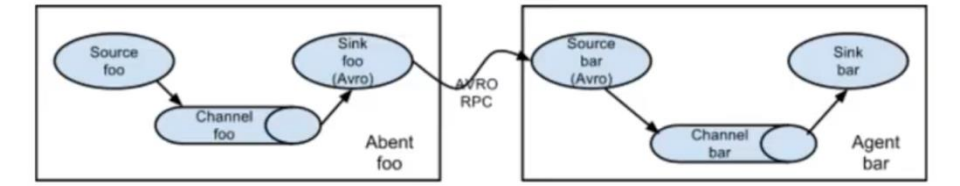
- As the web/app server are geographically separated – one of the strategy is setting up multiple agents
- First agent read log data from web or app server and push to avro sink
- Second agent read data from avro source (sink of first agent) and push to what ever sink you want to.

Cette approche est utiliser si j'ai 2 conteneur qui veulent communiquer entre eux

Dans cette approche j'aurais 2 agent , un pour chaque conteneur
Seul les agents flume peuvent communiquer entre eux
la source de l'agent 1 serait lier la destination de l'agent 2
une destination et une source qui communique entre eux sont de type avro

Flume Multi agent flows

-simple multi agent flow



avro: push data to another flume agent
NB:
L'agent qui reçois la donner doit être démarrer on premier
Sinon l'agent 2 ne connaitra pas la destination lors de son démarrage

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# cat second-agent.conf

sa.sources = r1
sa.sinks = k1
sa.channels = c1


sa.sources.r1.type = exec
sa.sources.r1.command = tail -F /home/hdoop/apache-flume-1.11.0-bin/flume-demo/logs-generator/data1.txt

sa.sinks.k1.type = avro
sa.sinks.k1.hostname = 82ab94650db5
sa.sinks.k1.port = 44444



sa.channels.c1.type = memory
sa.channels.c1.capacity = 1000
sa.channels.c1.transactionCapacity = 100

sa.sources.r1.channels = c1
sa.sinks.k1.channel = c1
```

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# cat first-agent.conf

fa.sources = r1
fa.sinks = k1
fa.channels = c1


fa.sources.r1.type = avro
fa.sources.r1.bind = 82ab94650db5
fa.sources.r1.port = 44444

fa.sinks.k1.type = logger

fa.channels.c1.type = memory
fa.channels.c1.capacity = 1000
fa.channels.c1.transactionCapacity = 100

fa.sources.r1.channels = c1
fa.sinks.k1.channel = c1
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple#
```

second agent

source : il prend la sortie de la commande tail
data1.txt est le ficher qui reçoit des logs tous les 10 s
type : exec
sink : puisque les 2 agents sont situer dans le même conteneur la destination
serrât la même machine vers le port 44444
type: avro

first agent

source : reçois le flux de donner depuis le port 44444
type : avro
sink : revoie les évènement vers flume agent
type : logger

exécution:

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# flume-ng agent --name=fa --conf-file first-agent.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/hdoop/hadoop-3.3.5/bin/hadoop) for HDFS access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/lib/jvm/java-8-openjdk-arm64/bin/java -Xmx20m -cp '/home/hdoop/apache-flume-1.11.0-bin/lib/*:/home/hdoop/hadoop-3.3.5/etc/hadoop:/home/hdoop/hadoop
-3.3.5/share/hadoop/common/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/common/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs:/home/hdoop/hadoop-3.3.5/share/hadoo
p/hdfs/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs/*:/home/hdoop/hadoop-3.3.5/share/hadoop/mapreduce/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn:/home/hd
oop/hadoop-3.3.5/share/hadoop/yarn/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn/*:/lib/*' -Djava.library.path=:/home/hdoop/hadoop-3.3.5/lib/native org.apac
he.flume.node.Application --name=fa --conf-file first-agent.conf
2023-04-03 23:57:47,148 INFO conf.FlumeConfiguration: Added sinks: k1 Agent: fa
2023-04-03 23:57:47,148 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:57:47,149 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 23:57:47,149 WARN conf.FlumeConfiguration: Agent configuration for 'fa' has no configfilters.
2023-04-03 23:57:47,164 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [fa]
2023-04-03 23:57:47,164 INFO node.AbstractConfigurationProvider: Creating channels
2023-04-03 23:57:47,169 INFO channel.DefaultChannelFactory: Creating instance of channel c1 type memory
2023-04-03 23:57:47,172 INFO node.AbstractConfigurationProvider: Created channel c1
2023-04-03 23:57:47,172 INFO source.DefaultSourceFactory: Creating instance of source r1, type avro
2023-04-03 23:57:47,183 INFO sink.DefaultSinkFactory: Creating instance of sink: k1, type: logger
2023-04-03 23:57:47,185 INFO node.AbstractConfigurationProvider: Channel c1 connected to [r1, k1]
2023-04-03 23:57:47,185 INFO node.Application: Initializing components
2023-04-03 23:57:47,189 INFO node.Application: Starting new configuration:{ sourceRunners:{r1=EventDrivenSourceRunner: { source:Avro source r1: { bindAddress:
82ab94650db5, port: 44444 } }} sinkRunners:{k1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@229d10bd counterGroup:{ name:null counters:{} }
}} channels:{c1=org.apache.flume.channel.MemoryChannel{name: c1}} }
2023-04-03 23:57:47,189 INFO node.Application: Starting Channel c1
2023-04-03 23:57:47,191 INFO node.Application: Waiting for channel: c1 to start. Sleeping for 500 ms
2023-04-03 23:57:47,233 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: c1: Successfully registered new MBean.
2023-04-03 23:57:47,233 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: c1 started
2023-04-03 23:57:47,693 INFO node.Application: Starting Sink k1
2023-04-03 23:57:47,694 INFO node.Application: Starting Source r1
2023-04-03 23:57:47,695 INFO source.AvroSource: Starting Avro source r1: { bindAddress: 82ab94650db5, port: 44444 }...
2023-04-03 23:57:48,023 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: r1: Successfully registered new MBean.
```

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple#  flume-ng agent --name=sa --conf-file second-agent.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/hdoop/hadoop-3.3.5/bin/hadoop) for HDFS access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/lib/jvm/java-8-openjdk-arm64/bin/java -Xmx20m -cp '/home/hdoop/apache-flume-1.11.0-bin/lib/*:/home/hdoop/hadoop-3.3.5/etc/hadoop:/home/hdoop/hadoop-3.3.5/share/hadoop/
common/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/common/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs/lib/*:/home/hdoop/hadoop-3.3.5/sh
are/hadoop/hdfs/*:/home/hdoop/hadoop-3.3.5/share/hadoop/mapreduce/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn/lib/*:/home/hdoop/hadoop-
3.3.5/share/hadoop/yarn/*:/lib/*' -Djava.library.path=:/home/hdoop/hadoop-3.3.5/lib/native org.apache.flume.node.Application --name=sa --conf-file second-agent.conf
2023-04-03 23:58:55,519 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Added sinks: k1 Agent: sa
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:c1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:k1
2023-04-03 23:58:55,520 INFO conf.FlumeConfiguration: Processing:r1
2023-04-03 23:58:55,520 WARN conf.FlumeConfiguration: Agent configuration for 'sa' has no configfilters.
2023-04-03 23:58:55,532 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [sa]
2023-04-03 23:58:55,532 INFO node.AbstractConfigurationProvider: Creating channels
2023-04-03 23:58:55,535 INFO channel.DefaultChannelFactory: Creating instance of channel c1 type memory
2023-04-03 23:58:55,537 INFO node.AbstractConfigurationProvider: Created channel c1
2023-04-03 23:58:55,538 INFO source.DefaultSourceFactory: Creating instance of source r1, type exec
2023-04-03 23:58:55,543 INFO sink.DefaultSinkFactory: Creating instance of sink: k1, type: avro
2023-04-03 23:58:55,550 INFO sink.AbstractRpcSink: Connection reset is set to 0. Will not reset connection to next hop
2023-04-03 23:58:55,551 INFO node.AbstractConfigurationProvider: Channel c1 connected to [r1, k1]
2023-04-03 23:58:55,551 INFO node.Application: Initializing components
2023-04-03 23:58:55,552 INFO node.Application: Starting new configuration:{ sourceRunners:{r1=EventDrivenSourceRunner: { source:org.apache.flume.source.ExecSource{name:r1,state:ID
LE} }} sinkRunners:{k1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@309e345f counterGroup:{ name:null counters:{} } }} channels:{c1=org.apache.flume.channel.Mem
oryChannel{name: c1}} }
2023-04-03 23:58:55,552 INFO node.Application: Starting Channel c1
2023-04-03 23:58:55,553 INFO node.Application: Waiting for channel: c1 to start. Sleeping for 500 ms
2023-04-03 23:58:55,591 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: c1: Successfully registered new MBean.
2023-04-03 23:58:55,591 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: c1 started
2023-04-03 23:58:56,054 INFO node.Application: Starting Sink k1
2023-04-03 23:58:56,056 INFO sink.AbstractRpcSink: Starting RpcSink k1 { host: 82ab94650db5, port: 44444 }...
2023-04-03 23:58:56,057 INFO node.Application: Starting Source r1
2023-04-03 23:58:56,057 INFO source.ExecSource: Exec source starting with command: tail -F /home/hdoop/apache-flume-1.11.0-bin/flume-demo/logs-generator/data1.txt
2023-04-03 23:58:56,058 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: k1: Successfully registered new MBean.
2023-04-03 23:58:56,058 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: k1 started
2023-04-03 23:58:56,058 INFO sink.AbstractRpcSink: Rpc sink k1: Building RpcClient with hostname: 82ab94650db5, port: 44444
```

```
2023-04-03 23:57:47,693 INFO node.Application: Starting Sink k1
2023-04-03 23:57:47,694 INFO node.Application: Starting Source r1
2023-04-03 23:57:47,695 INFO source.AvroSource: Starting Avro source r1: { bindAddress: 82ab94650db5, port: 44444 }...
2023-04-03 23:57:48,023 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: r1: Successfully registered new MBean.
2023-04-03 23:57:48,023 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: r1 started
2023-04-03 23:57:48,024 INFO source.AvroSource: Avro source r1 started.
2023-04-03 23:59:05,751 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,752 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,752 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,752 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,753 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,753 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,753 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,753 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,753 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:05,754 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:08,118 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:18,133 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:33,145 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
2023-04-03 23:59:38,169 INFO sink.LoggerSink: Event: { headers:{} body: 64 61 74 61 31 20 2B 20 32 30 32 33 2D 30 34 2D data1 + 2023-04- }
```

RAM 0.36 GB CPU 0.99% Disk 49.54 GB avail. of 65.70 GB Connected to Hub v4.17.0

Exemple 3 Sink type Hdfs:

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-sink# ls
logstomulti.conf
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-sink# cat logstomulti.conf

lm.sources = r1
lm.sinks = k1
lm.channels = c1


lm.sources.r1.type = exec
lm.sources.r1.command = tail -f /home/hdoop/apache-flume-1.11.0-bin/flume-demo/logs-generator/data1.txt

lm.sinks.k1.type = hdfs
lm.sinks.k1.hdfs.path = hdfs://localhost:9000/user/log1_%Y-%m-%d
lm.sinks.k1.hdfs.filePrefix = log
lm.sinks.k1.hdfs.fileSuffix = .txt
lm.sinks.k1.hdfs.rollInterval = 60
lm.sinks.k1.hdfs.rollSize = 0
lm.sinks.k1.hdfs.rollCount = 100
lm.sinks.k1.hdfs.fileType = DataStream
lm.sinks.k1.hdfs.useLocalTimeStamp = true


lm.channels.c1.type = memory
lm.channels.c1.capacity = 1000
lm.channels.c1.transactionCapacity = 100

lm.sources.r1.channels = c1
lm.sinks.k1.channel = c1
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-sink#
```

Source:
  Type:exec

Sinks:
  Type: hdfs
  Path: on définit le port de hdfs est le répertoire de stockage

Fileprefix : le nom du fichier
fileSuffix : le type du fichier
rolIntervall : un nouveau fichier de stockage serait créer toute les minute
rolCount : le nombre d'évènement maximal a écrire dans un fichier
useLocalTimeStamp : permet d'ajouter une date au nom du répertoire de stockage
fileType : mes donnees seront stocker sans compressage

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-sink# flume-ng agent --name=lm --conf-file logstomulti.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/hdoop/hadoop-3.3.5/bin/hadoop) for HDFS access
Info: Including Hive libraries found via () for Hive access
+ exec /usr/lib/jvm/java-8-openjdk-arm64/bin/java -Xmx20m -cp '/home/hdoop/apache-flume-1.11.0-bin/lib/*:/home/hdoop/hadoop-3.3.5/etc/hadoop:/home/hdoop/hadoop
-3.3.5/share/hadoop/common/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/common/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs:/home/hdoop/hadoop-3.3.5/share/hadoo
p/hdfs/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/hdfs/*:/home/hdoop/hadoop-3.3.5/share/hadoop/mapreduce/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn:/home/hd
oop/hadoop-3.3.5/share/hadoop/yarn/lib/*:/home/hdoop/hadoop-3.3.5/share/hadoop/yarn/*:/lib/*' -Djava.library.path=:/home/hdoop/hadoop-3.3.5/lib/native org.apac
he.flume.node.Application --name=lm --conf-file logstomulti.conf
2023-04-04 00:12:34,250 INFO conf.FlumeConfiguration: Processing:c1
2023-04-04 00:12:34,251 INFO conf.FlumeConfiguration: Processing:c1
2023-04-04 00:12:34,251 INFO conf.FlumeConfiguration: Processing:r1
```

```
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# hadoop fs -ls  /user
Found 4 items
drwxr-xr-x   - root supergroup          0 2023-04-03 19:46 /user/log-geneartor
drwxr-xr-x   - root supergroup          0 2023-04-03 21:52 /user/log1
drwxr-xr-x   - root supergroup          0 2023-04-03 22:19 /user/log1_2023-04-03
drwxr-xr-x   - root supergroup          0 2023-04-04 00:14 /user/log1_2023-04-04
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# hadoop fs -ls /user/log1_2023-04-03
Found 3 items
-rw-r--r--   1 root supergroup        448 2023-04-03 22:17 /user/log1_2023-04-03/log.1680560210453.txt
-rw-r--r--   1 root supergroup        168 2023-04-03 22:18 /user/log1_2023-04-03/log.1680560273536.txt
-rw-r--r--   1 root supergroup         56 2023-04-03 22:19 /user/log1_2023-04-03/log.1680560335591.txt
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# hadoop fs -ls /user/log1_2023-04-03 | wc -l
4
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple# hadoop fs -cat /user/log1_2023-04-03/log.1680560210453.txt
data1 + 2023-04-03 22:15:13
data1 + 2023-04-03 22:15:23
data1 + 2023-04-03 22:15:33
data1 + 2023-04-03 22:15:43
data1 + 2023-04-03 22:15:53
data1 + 2023-04-03 22:16:03
data1 + 2023-04-03 22:16:13
data1 + 2023-04-03 22:16:23
data1 + 2023-04-03 22:16:33
data1 + 2023-04-03 22:16:43
data1 + 2023-04-03 22:16:53
data1 + 2023-04-03 22:17:03
data1 + 2023-04-03 22:17:13
data1 + 2023-04-03 22:17:23
data1 + 2023-04-03 22:17:33
data1 + 2023-04-03 22:17:43
root@82ab94650db5:/home/hdoop/apache-flume-1.11.0-bin/flume-demo/multi-simple#
```