Néstor Batista Díaz

Flume - listar directorio a logger

Creó el archivo agente.conf

```
# Nombramos los componentes del agente

ExecLoggerAgent.sources = Exec
ExecLoggerAgent.channels = MemChannel
ExecLoggerAgent.sinks = LoggerSink

# Describimos el tipo de origen
ExecLoggerAgent.sources.Exec.type = exec
ExecLoggerAgent.sources.Exec.command = ls /home/cloudera

# Describimos el destino
ExecLoggerAgent.sinks.LoggerSink.type = logger

# Describimos la configuración del canal
ExecLoggerAgent.channels.MemChannel.type = memory
ExecLoggerAgent.channels.MemChannel.capacity = 1000
ExecLoggerAgent.channels.MemChannel.transactionCapacity = 100

# Unimos el origen y el destino a través del canal
ExecLoggerAgent.sources.Exec.channels = MemChannel
ExecLoggerAgent.sinks.LoggerSink.channel = MemChannel
```

Antes de lanzar el agente Flume, hay que arrancar tanto Hadoop como YARN mediante el comando *start-all.sh*.

```
[cloudera@quickstart sbin]$ sudo ./start-all.sh
starting org.apache.spark.deploy.master.Master, logging to /var/log/spark/spark-root-org.apache.spark.deploy.master.Master-1-quickstart.cloudera.out
quickstart.cloudera: Warning: Permanently added 'quickstart.cloudera' (RSA) to the list of known hosts.
root@quickstart.cloudera's password:
quickstart.cloudera: starting org.apache.spark.deploy.worker.Worker, logging to /var/log/spark/spark-root-org.apache.spark.deploy.worker.Worker-1-quickstart
.cloudera.out
```

Lanzar agente Flume (la opción -n sirve para indicar el nombre del agente, y con -f indicamos el nombre del archivo de configuración):

flume-ng agent -n ExecLoggerAgent -f agente.conf

```
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 61 67 65 6E 74 2E 63 6F 6E 66                  agent.conf }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 61 67 65 6E 74 65 2E 63 6F 6E 66               agente.conf }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 61 72 63 68 69 76 6F 2E 74 78 74               archivo.txt }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 63 6C 6F 75 64 65 72 61 2D 6D 61 6E 61 67 65 72 cloudera-manager }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 63 6D 5F 61 70 69 2E 70 79                     cm_api.py }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 44 65 73 6B 74 6F 70                           Desktop }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 44 6F 63 75 6D 65 6E 74 73                     Documents }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 44 6F 77 6E 6C 6F 61 64 73                     Downloads }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 65 63 6C 69 70 73 65                           eclipse }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 65 6E 74 65 72 70 72 69 73 65 2D 64 65 70 6C 6F enterprise-deplo }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 65 78 70 72 65 73 73 2D 64 65 70 6C 6F 79 6D 65 express-deployme }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 6B 65 72 62 65 72 6F 73                        kerberos }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 6C 69 62                                      lib }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 6C 6F 63 61 6C 5F 66 69 6C 65 5F 63 6F 70 79 2E local_file_copy. }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 6C 6F 63 61 6C 5F 66 69 6C 65 2E 74 78 74       local_file.txt }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 4D 75 73 69 63                                Music }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 70 61 72 63 65 6C 73                         parcels }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 50 69 63 74 75 72 65 73                      Pictures }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 50 75 62 6C 69 63                            Public }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 73 74 61 6E 64 61 6C 6F 6E 65                standalone }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 73 74 61 6E 64 61 6C 6F 6E 65 2D 66 69 6C 65 2E standalone-file. }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 54 65 6D 70 6C 61 74 65 73                   Templates }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 56 69 64 65 6F 73                            Videos }
24/01/09 10:08:00 INFO sink.LoggerSink: Event: { headers:{} body: 77 6F 72 6B 73 70 61 63 65                   workspace }
```

Modificamos el Sink:

```
# Nombramos los componentes del agente

ExecLoggerAgent.sources = Exec
ExecLoggerAgent.channels = MemChannel
ExecLoggerAgent.sinks = LoggerSink

# Describimos el tipo de origen
ExecLoggerAgent.sources.Exec.type = exec
ExecLoggerAgent.sources.Exec.command = ls /home/cloudera

# Describimos el destino
ExecLoggerAgent.sinks.LoggerSink.type = hdfs
ExecLoggerAgent.sinks.LoggerSink.hdfs.path = hdfs:///flume/events
ExecLoggerAgent.sinks.LoggerSink.hdfs.filePrefix = flume-caso3-seqgen
ExecLoggerAgent.sinks.LoggerSink.hdfs.rollInterval = 0
ExecLoggerAgent.sinks.LoggerSink.hdfs.rollCount = 1000
ExecLoggerAgent.sinks.LoggerSink.hdfs.fileType = DataStream

# Describimos la configuración del canal
ExecLoggerAgent.channels.MemChannel.type = memory
ExecLoggerAgent.channels.MemChannel.capacity = 1000
ExecLoggerAgent.channels.MemChannel.transactionCapacity = 100

# Unimos el origen y el destino a través del canal
ExecLoggerAgent.sources.Exec.channels = MemChannel
ExecLoggerAgent.sinks.LoggerSink.channel = MemChannel
```

flume-ng agent -n ExecLoggerAgent -f agente.conf

```
/01/09 10:54:18 INFO node.Application: Starting Channel MemChannel
/01/09 10:54:19 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: MemChannel: Successfully registered new MBean.
/01/09 10:54:19 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: MemChannel started
/01/09 10:54:19 INFO node.Application: Starting Sink LoggerSink
/01/09 10:54:19 INFO node.Application: Starting Source Exec
/01/09 10:54:19 INFO source.ExecSource: Exec source starting with command: ls /home/cloudera
/01/09 10:54:19 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: Exec: Successfully registered new MBean.
/01/09 10:54:19 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: Exec started
/01/09 10:54:19 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: LoggerSink: Successfully registered new MBean.
/01/09 10:54:19 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: LoggerSink started
/01/09 10:54:19 INFO source.ExecSource: Command [ls /home/cloudera] exited with 0
/01/09 10:54:19 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
/01/09 10:54:19 INFO hdfs.BucketWriter: Creating hdfs:///flume/events/flume-caso3-seqgen.1704826459102.tmp
```

Comprobamos:

```
loudera@quickstart ~]$ hdfs dfs -ls /flume/events
und 2 items
w-r--r--   1 cloudera supergroup        285 2024-01-09 10:53 /flume/events/flume-caso3-seqgen.1704826396976.tmp
w-r--r--   1 cloudera supergroup        285 2024-01-09 10:54 /flume/events/flume-caso3-seqgen.1704826459102.tmp
loudera@quickstart ~]$
```

cloudera@quickstart:~