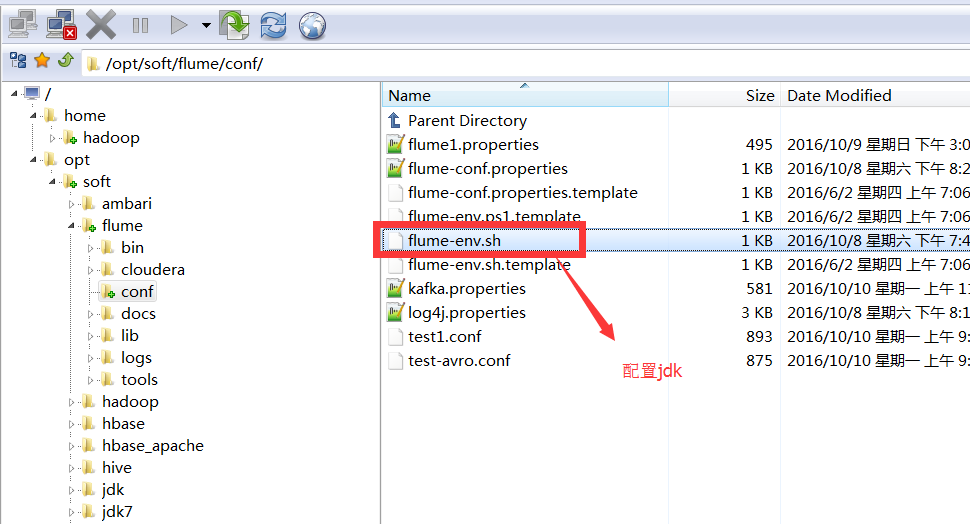
# Flume操作手册



**1. Flume采集某个日志的具体实例**

创建一个新的目录共平时日志采集放配置文件使用。

[hadoop@slavenode4 example]$ mkdir /opt/hadoop/flume-bin/example

1) 单节点flume直接写入hdfs，监控一个日志文件

[hadoop@slavenode4 example]$ cat flume\_directHDFS.conf

# Define a memory channel called ch1 on agent1

agent1.channels.ch1.type = memory

agent1.channels.ch1.capacity = 100000

agent1.channels.ch1.transactionCapacity = 100000

agent1.channels.ch1.keep-alive = 30

# Define an Avro source called avro-source1 on agent1 and tell it

#define source monitor a file

agent1.sources.avro-source1.type = exec

agent1.sources.avro-source1.shell = /bin/bash -c

agent1.sources.avro-source1.command =tail -n +0 -F /opt/hadoop/hadoop-2.7.2/logs/hadoop-hadoop-datanode-slavenode4.log

agent1.sources.avro-source1.channels = ch1

agent1.sources.avro-source1.threads = 5

# Define a logger sink that simply logs all events it receives

# and connect it to the other end of the same channel.

agent1.sinks.log-sink1.channel = ch1

agent1.sinks.log-sink1.type = hdfs

agent1.sinks.log-sink1.hdfs.path = hdfs://cluster-ha/flumeTest #需要手动创建

agent1.sinks.log-sink1.hdfs.writeFormat = Text

agent1.sinks.log-sink1.hdfs.fileType = DataStream

agent1.sinks.log-sink1.hdfs.rollInterval = 0

agent1.sinks.log-sink1.hdfs.rollSize = 1000000

agent1.sinks.log-sink1.hdfs.rollCount = 0

agent1.sinks.log-sink1.hdfs.batchSize = 1000

agent1.sinks.log-sink1.hdfs.txnEventMax = 1000

agent1.sinks.log-sink1.hdfs.callTimeout = 60000

agent1.sinks.log-sink1.hdfs.appendTimeout = 60000

# Finally, now that we've defined all of our components, tell

# agent1 which ones we want to activate.

agent1.channels = ch1

agent1.sources = avro-source1

agent1.sinks = log-sink1

[hadoop@masternode2 ~]$ hdfs dfs -mkdir hdfs://cluster-ha/flumeTest

启动如下命令，之后在hdfs上面看效果

[hadoop@slavenode4 example]$ ../bin/flume-ng agent --conf ../conf/ -f flume\_directHDFS.conf -n agent1 -Dflume.root.logger=INFO,console 生产环境不用用该选项

当出现如下日志说明配置已成功

2016-10-10 11:18:20,126 (agent-shutdown-hook) [INFO - org.apache.flume.sink.hdfs.HDFSEventSink.stop(HDFSEventSink.java:492)] Closing hdfs://cluster-ha/flumeTest/FlumeData

2016-10-10 11:18:20,126 (agent-shutdown-hook) [INFO - org.apache.flume.sink.hdfs.BucketWriter.close(BucketWriter.java:363)] Closing hdfs://cluster-ha/flumeTest/FlumeData.1476069384812.tmp

2016-10-10 11:18:20,129 (ResponseProcessor for block BP-116503124-192.168.237.230-1475137189670:blk\_1073742252\_1429) [WARN - org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer$ResponseProcessor.run(DFSOutputStream.java:738)] Slow ReadProcessor read fields took 79233ms (threshold=30000ms); ack: seqno: 19 reply: SUCCESS reply: SUCCESS reply: SUCCESS downstreamAckTimeNanos: 1112603 flag: 0 flag: 0 flag: 0, targets: [DatanodeInfoWithStorage[192.168.237.235:50010,DS-52b1f74d-56d9-4684-a763-447181035b6c,DISK], DatanodeInfoWithStorage[192.168.237.238:50010,DS-ba8c7923-a66f-4a09-91ed-bc1007194d08,DISK], DatanodeInfoWithStorage[192.168.237.234:50010,DS-d00c87be-54b4-4320-9805-ba7c334d4344,DISK]]

2016-10-10 11:18:20,170 (hdfs-log-sink1-call-runner-7) [INFO - org.apache.flume.sink.hdfs.BucketWriter$8.call(BucketWriter.java:629)] Renaming hdfs://cluster-ha/flumeTest/FlumeData.1476069384812.tmp to hdfs://cluster-ha/flumeTest/FlumeData.1476069384812

HDFS上查看效果如下。自动出来一些日志文件

[hadoop@masternode2 ~]$ hdfs dfs -ls hdfs://cluster-ha/flumeTest/

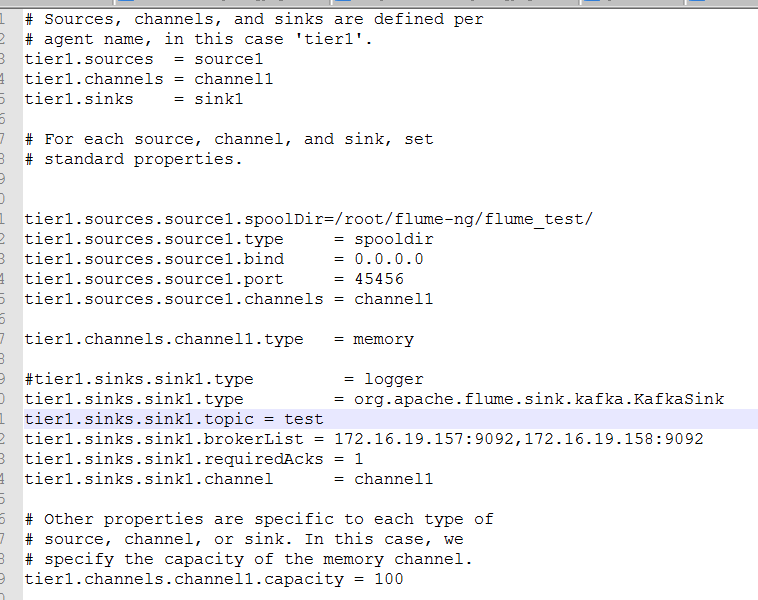
Found 2 items

-rw-r--r-- 3 hadoop supergroup 1004251 2016-10-10 11:16 hdfs://cluster-ha/flumeTest/FlumeData.1476069384811

-rw-r--r-- 3 hadoop supergroup 179996 2016-10-10 11:16 hdfs://cluster-ha/flumeTest/FlumeData.1476069384812.tmp

CDH平台flume测试：

Test-avro.config配置: 



启动agent服务:



