一、配置采集方案

1.在hadoop02,hadoop03的/export/servers/flume/conf目录下编写同样的日志采集方案

vim exec-avro\_logCollection.conf

a1.sources=r1 r2 r3

a1.sinks=k1

a1.channels=c1

a1.sources.r1.type=exec

a1.sources.r1.command=tail -F /root/logs/access.log

a1.sources.r1.interceptors=i1

a1.sources.r1.interceptors.i1.type=static

a1.sources.r1.interceptors.i1.key=type

a1.sources.r1.interceptors.i1.value=access

a1.sources.r2.type=exec

a1.sources.r2.command=tail -F /root/logs/nginx.log

a1.sources.r2.interceptors=i2

a1.sources.r2.interceptors.i2.type=static

a1.sources.r2.interceptors.i2.key=type

a1.sources.r2.interceptors.i2.value=nginx

a1.sources.r3.type=exec

a1.sources.r3.command=tail -F /root/logs/web.log

a1.sources.r3.interceptors=i3

a1.sources.r3.interceptors.i3.type=static

a1.sources.r3.interceptors.i3.key=type

a1.sources.r3.interceptors.i3.value=web

a1.channels.c1.type=memory

a1.channels.c1.capacity=2000000

a1.channels.c1.transactionCapacity=100000

a1.sinks.k1.type=avro

a1.sinks.k1.hostname=**hadoop01**

a1.sinks.k1.port=41414

a1.sources.r1.channels=c1

a1.sources.r2.channels=c1

a1.sources.r3.channels=c1

a1.sinks.k1.channel=c1

2.在hadoop-1的/export/servers/flume/conf目录下编写第二级采集方案

vim avro-hdfs\_logCollection.conf

a1.sources=r1

a1.sinks=k1

a1.channels=c1

a1.sources.r1.type=avro

a1.sources.r1.bind=**hadoop01**

a1.sources.r1.port=41414

a1.sources.r1.interceptors=i1

a1.sources.r1.interceptors.i1.type=timestamp

a1.channels.c1.type=memory

a1.channels.c1.capacity=20000

a1.channels.c1.transactionCapacity=10000

a1.sinks.k1.type=hdfs

a1.sinks.k1.hdfs.path=hdfs://**hadoop01**:9000/source/logs/%{type}/%Y%m%d

a1.sinks.k1.hdfs.filePrefix=events

a1.sinks.k1.hdfs.fileType=DataStream

a1.sinks.k1.hdfs.writeFormat=Text

a1.sinks.k1.hdfs.rollCount=0

a1.sinks.k1.hdfs.rollInterval=0

a1.sinks.k1.hdfs.rollSize=10485760

a1.sinks.k1.hdfs.batchSize=20

a1.sinks.k1.hdfs.threadsPoolSize=10

a1.sources.r1.channels=c1

a1.sinks.k1.channel=c1

二、启动hadoop集群

1.start-dfs.sh

start-yarn.sh

2.进入hadoop01:

cd /export/servers/flume/

flume-ng agent –c conf/ -f conf/avro-hdfs\_logCollection.conf \--name a1 -Dflume.root.logger=INFO,console

3.进入hadoop02和03

cd /export/servers/flume/

flume-ng agent –c conf/ -f conf/exec-avro\_logCollection.conf \--name a1 -Dflume.root.logger=INFO,console

4.在hadoop0203分别开三个克隆会话，分别执行下列命令：

Hadoop02执行:

cd /root/

mkdir logs

while true; do echo "access access ..." >> /root/logs/access.log ; \sleep 1;done

while true; do echo "nginx nginx ..." >> /root/logs/nginx.log ; \sleep 1;done

while true; do echo "web web ..." >> /root/logs/web.log ; \sleep 1;done

在hadoop03执行：

cd /root/

mkdir logs

while true; do echo "access access ..." >> /root/logs/access.log ; \sleep 1;done

while true; do echo "nginx nginx ..." >> /root/logs/nginx.log ; \sleep 1;done

while true; do echo "web web ..." >> /root/logs/web.log ; \sleep 1;done

三、打开UI界面查看

Source-logs-三个文件，打开一个代表成功