8.2FLUME基本使用

1.flume安装

tar zxvf apache-flume-1.8.0-bin.tar.gz -C /export/servers/

mv apache-flume-1.8.0-bin flume

2.flume配置

cd /export/servers/flume/conf/

cp flume-env.sh.template flume-env.sh

vim flume-env.sh

export JAVA\_HOME=/export/servers/jdk

3.vim /etc/profile

添加：

export FLUME\_HOME=/export/servers/flume

export PATH=$PATH:$FLUME\_HOME/bin:

刷新一下：

source /etc/profile

4.flume入门使用

cd /export/servers/flume/conf

编写测试文件netcat-logger.conf

# Name the components on this agent

a1.sources = r1

a1.sinks = k1

a1.channels = c1

# Describe/configure the source

a1.sources.r1.type = netcat

a1.sources.r1.bind = localhost

a1.sources.r1.port = 44444

# Describe the sink

a1.sinks.k1.type = logger

# Use a channel which buffers events in memory

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

# Bind the source and sink to the channel

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

使用指定采集方案启动flume

cd /export/servers/flume/

flume-ng agent –conf conf/ --conf-file conf/netcat-logger.conf \--name a1 -Dflume.root.logger=INFO,console

克隆对话

yum -y install telnet

telnet localhost 44444

8.4flume的可靠性保证

1.将集群主节点的配置文件分发到其他子节点

scp -r /export/servers/flume hadoop02:/export/servers/

scp -r /export/servers/flume hadoop03:/export/servers/

scp /etc/profile hadoop02:/etc/profile

scp /etc/profile hadoop03:/etc/profile

在hadoop02、03上执行source /etc/profile

2.

cd /export/servers/flume/conf/

vim exec-avro.conf

#配置一级采集方案

a1.sources=r1

#用空格分隔两个sink

a1.sinks=k1 k2

a1.channels=c1

#描述并配置sourcecs组件

a1.sources.r1.channels=c1

a1.sources.r1.type=exec

a1.sources.r1.command=tail -F /root/logs/123.log

#描述并配置channels

a1.channels.c1.type=memory

a1.channels.c1.capacity=1000

a1.channels.c1.transactionCapacity=100

#设置sink1，采集

a1.sinks.k1.channel=c1

a1.sinks.k1.type=avro

a1.sinks.k1.hostname=hadoop02

a1.sinks.k1.port=52020

#设置

a1.sinks.k2.channel=c1

a1.sinks.k2.type=avro

a1.sinks.k2.hostname=hadoop03

a1.sinks.k2.port=52020

#配置sink及处理器策略

a1.sinkgroups=g1

a1.sinkgroups.g1.sinks=k1 k2

s1.sinkgroups.g1.processor.type=load\_balance

s1.sinkgroups.g1.processor.backoff=true

s1.sinkgroups.g1.processor.selector=random

s1.sinkgroups.g1.processor.maxTimeOut=10000

3.在hadoop02的/export/servers/flume/conf目录下编写

cd /export/servers/flume/conf

vim avro-logger.conf

# Name the components on this agent

a1.sources = r1

a1.sinks = k1

a1.channels = c1

# Describe/configure the source

a1.sources.r1.type = avro

a1.sources.r1.channels = c1

a1.sources.r1.bind = hadoop02

a1.sources.r1.port = 52020

# Describe the sink

a1.sinks.k1.type = logger

# Use a channel which buffers events in memory

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

# Bind the source and sink to the channel

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

4.在hadoop03的/export/servers/flume/conf目录下编写

cd /export/servers/flume/conf

vim avro-logger.conf

# Name the components on this agent

a1.sources = r1

a1.sinks = k1

a1.channels = c1

# Describe/configure the source

a1.sources.r1.type = avro

a1.sources.r1.channels = c1

a1.sources.r1.bind = hadoop03

a1.sources.r1.port = 52020

# Describe the sink

a1.sinks.k1.type = logger

# Use a channel which buffers events in memory

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

# Bind the source and sink to the channel

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

5.在hadoop02、03上分别执行

cd /export/servers/flume/

flume-ng agent --conf conf/ --conf-file conf/avro-logger.conf \--name a1 -Dflume.root.logger=INFO,console

6.回到hadoop01上，执行如下命令

cd /export/servers/flume/

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

7.克隆hadoop01

cd /root/

mkdir logs

while true; do echo "access access ..." >> /root/logs/123.log ; \sleep 1;done

