

Flume配置参数解读：<https://blog.csdn.net/qq_39160721/article/details/80255194>

安装教程参考博客：https://blog.csdn.net/lzxlfly/article/details/80672267

三、下载安装

去http://flume.apache.org/download.html下载，这里下载最新稳定版apache-flume-1.8.0-bin.tar.gz

在s1下，/usr/local/soft/ 路径下，解压

tar -zxvf apache-flume-1.8.0-bin.tar.gz

为了方便，重命名下

mv apache-flume-1.8.0-bin flume

flume安装目录：/usr/local/soft

配置文件目录： /usr/local/soft/flume/conf/

四、配置flume

1、配置环境变量

在/usr/local/soft/flume/conf/下，默认flume-env.sh.template，复制一份命名flume-env.sh

cp flume-1.8.0/conf/flume-env.sh.template flume-1.8.0/conf/flume-env.sh

编辑 vim flume-1.8.0/conf/flume-env.sh,指定jdk路径

export JAVA\_HOME=$JAVA\_HOME

或直接指定jdk路径 export JAVA\_HOME=/usr/local/jdk1.8

2、配置启动参数

在s1，/usr/local/soft/flume/conf/ 下，默认flume-conf.properties.template，复制一份命名flume-conf.properties

cp flume/conf/flume-conf.properties.template flume/conf/flume-conf.properties

由于node01、node02、node03上的flume是Agent角色，所以这三台flume-conf.properties配置一样。编辑 vim flume/conf/flume-conf.properties,设置相关参数。

**s3、s4、s5配置**

*# Name the components on this agent*

a1.channels = c1

a1.sources = r1

a1.sinks = k1 k2

# Use a channel which buffers events **in** memory

a1.channels.c1.type =file

#信道最大容量

a1.channels.c1.capacity = 1000000

#通道支持的事务的最大大小

a1.channels.c1.transactionCapacity =1000

a1.channels.c1.byteCapacity = 800000

#等待放置操作时间（秒）

a1.channels.c1.keep-alive = 60

a1.channels.c1.maxFileSize = 900000000

#存储文件检查点目录

a1.channels.hdfs\_c2.checkpointDir = /data/flume/checkpoint

#文件检查点的备份目录，不能和checkpointDir相同

a1.channels.hdfs\_c2.backupCheckpointDir = /data/flume/backupcheckpoint

#用于存储日志文件目录，多目录用逗号分隔

a1.channels.hdfs\_c2.dataDirs = /data/flume/data

#数据源为avro类型

a1.sources.r1.type=avro

*#any address to listen*监听其它flume端口

a1.sources.r1.bind=0.0.0.0

a1.sources.r1.port=52020

a1.sources.r1.channels=c1

*#define the sink 1*

a1.sinks.k1.type=avro

a1.sinks.k1.hostname=192.168.1.77

a1.sinks.k1.port=52020

a1.sinks.k1.channel=c1

*#define the sink 2*

a1.sinks.k2.type=avro

a1.sinks.k2.hostname=192.168.1.136

a1.sinks.k2.port=52020

a1.sinks.k2.channel=c1

#define sinkgroups

a1.sinkgroups=g1

a1.sinkgroups.g1.sinks=k1 k2

*#a1.sinkgroups.g1.processor.type=failover*

*#a1.sinkgroups.g1.processor.priority.k1=10*

*#a1.sinkgroups.g1.processor.priority.k2=5*

*#a1.sinkgroups.g1.processor.maxpenalty=10000*

*#负载均衡配置*

a1.sinkgroups.g1.processor.type=load\_balance

#如果开启，则将失败的 sink 放入黑名单

a1.sinkgroups.g1.processor.backoff=true

# 另外还支持 random 随机round\_robin 是轮训的意思

a1.sinkgroups.g1.processor.selector=round\_robin

### s1、s2配置

a1.sources = r1

a1.channels = hdfs\_c2

a1.sinks = hdfs\_k2

*#properties of avro-AppSrv-source*

a1.sources.r1.type = avro

a1.sources.r1.bind = s1

a1.sources.r1.port = 52020

*#设置sources的channels*

a1.sources.r1.channels=hdfs\_c2

*#增加拦截器 所有events,增加头,类似json格式里的"headers":{" key":" value"}*

*#拦截器名字*

*#a1.sources.r1.interceptors = i1*

*#拦截器类型*

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

*#自定义*

*#a1.sources.r1.interceptors.i1.key = Collector*

*#自定义*

*#a1.sources.r1.interceptors.i1.value =s1*

*#set hdfs channel*

*#a1.channels.hdfs\_c2.type = memory*

a1.channels.hdfs\_c2.type = file

a1.channels.hdfs\_c2.capacity = 1000000

a1.channels.hdfs\_c2.transactionCapacity = 100

a1.channels.hdfs\_c2.keep-alive = 60

a1.channels.hdfs\_c2.byteCapacity = 1000000

a1.sinks.hdfs\_k2.hdfs.idleTimeout=60

a1.channels.hdfs\_c2.checkpointDir = /data/flume/checkpoint

a1.channels.hdfs\_c2.backupCheckpointDir = /data/flume/backupcheckpoint

a1.channels.hdfs\_c2.dataDirs = /data/flume/data

*#set sink to hdfs*

*#传输到hdfs*

a1.sinks.hdfs\_k2.type=hdfs

*#传输的channel名*

a1.sinks.hdfs\_k2.channel=hdfs\_c2

*#所以这里写的是集群的Namespace*

*#a1.sinks.hdfs\_k2.hdfs.path=hdfs://114.115.170.6:9000/data/hadoop/test/%Y%m%d*

a1.sinks.hdfs\_k2.hdfs.path=hdfs://s2:8020/data/hadoop/testflume/%{hostname}/%Y%m%d

a1.sinks.hdfs\_k2.hdfs.fileType=DataStream

*#文本格式*

*#a1.sinks.hdfs\_k2.hdfs.writeFormat=TEXT*

*#a1.sinks.hdfs\_k2.hdfs.writeFormat=Text*

*#这个flume保存在hdfs上的文件的大小的*

a1.sinks.hdfs\_k2.hdfs.rollInterval=30

*#a1.sinks.hdfs\_k2.hdfs.rollSize=2097152*

*#超过六十秒关闭无效的文件*

a1.sinks.hdfs\_k2.hdfs.idleTimeout=60

*#a1.sinks.hdfs\_k2.hdfs.rollCount=10*

a1.sinks.hdfs\_k2.hdfs.rollCount=0

a1.sinks.hdfs\_k2.hdfs.rollSize=0

*#失败1s回滚*

a1.sinks.hdfs\_k2.hdfs.roundValue =1

*#文件名前缀*

a1.sinks.hdfs\_k2.hdfs.filePrefix=%Y-%m-%d

*#文件名后缀*

a1.sinks.hdfs\_k2.hdfs.fileSuffix=.txt

a1.sinks.hdfs\_k2.hdfs.useLocalTimeStamp = true

a1.sinks.hdfs\_k2.hdfs.callTimeout = 40000

a1.sinks.hdfs\_k2.hdfs.threadsPoolSize = 50

a1.sinks.hdfs\_k2.hdfs.batchSize = 100

### 启动命令

#s1、s2启动flume命令：

nohup flume/bin/flume-ng agent --**conf** /usr/**local**/soft/flume/**conf** --**conf**-**file** /usr/**local**/soft/flume/**conf**/flume-**conf**.properties --name a1 -Dflume.root.logger=INFO,console > flume/logs/flume-server.**log** 2>&1 &

#s3、s4、s5启动flume命令：

nohup flume/bin/flume-ng agent --**conf** **conf** --**conf**-**file** flume/**conf**/flume-**conf**.properties --name a1 -Dflume.root.logger=DEBUG,console >flume/logs/flume-server.**log** 2>&1 &