flume 使用spool 读数据进kafka简单用例

1. 在flume安装目录下添加配置文件，执行操作：

cd conf/

vi cisco.conf

在新建的cisco.conf文件中添加如下内容：

# cisco

cisco.sources = rcisco

cisco.sinks = kcisco

cisco.channels = ccisco

# Describe/configure the source

cisco.sources.rcisco.type = spooldir

cisco.sources.rcisco.spoolDir = /var/log/apache/flumeSpool/cisco

cisco.sources.rcisco.fileHeader = true (注：fileHeader是否在event的Header中添加文件名，boolean类型)

# Describe the sink

cisco.sinks.kcisco.type = org.apache.flume.sink.kafka.KafkaSink

cisco.sinks.kcisco.topic =cisco

cisco.sinks.kcisco.brokerList =192.168.100.251:9092

cisco.sinks.kcisco.requiredAcks = 1

cisco.sinks.kcisco.batchSize = 20

# Use a channel which buffers events in memory

cisco.channels.ccisco.type = memory

cisco.channels.ccisco.capacity = 1000

cisco.channels.ccisco.transactionCapacity = 100

# Bind the source and sink to the channel

cisco.sources.rcisco.channels = ccisco

cisco.sinks.kcisco.channel = ccisco

然后保存并退出该文件。

1. 进入kafka目录，创建相应topic并查看，执行命令：

cd kafka\_2.11-0.10.0.0

bin/kafka-topics.sh --create --zookeeper 192.168.100.251:2181 --replication-factor 1 --partitions 1 --topic cisco

bin/kafka-topics.sh --list --zookeeper 192.168.100.251:2181

1. ~~创建kafka消费者，执行命令：~~

~~bin/kafka-console-consumer.sh --zookeeper 192.168.100.251:2181 --topic cisco~~

1. 创建spool指定目录，执行命令：

mkdir /var/log/apache/flumeSpool/cisco

1. 启动flume agent，执行命令：

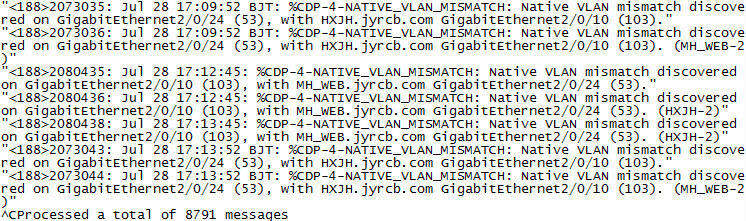
cd apache-flume-1.6.0-bin/

bin/flume-ng agent -n cisco -c conf -f conf/cisco.conf -Dflume.root.logger=ERROR,console

1. ~~手动向spool指定目录中添加cisco类别的日志文件，执行命令：~~

~~mv sdbexprtcsv0728-cisco /var/log/apache/flumeSpool/cisco~~

1. ~~如在consumer中收到文件内容表示flume spool已经部署完成。结果示例如下：~~



注： 上述步骤中的添加删除线的3,6,7均为测试flume部署是否成功所需，在实际部署中并不需要执行；此外上述所有步骤中所提到的ip地址均为本次测试所用，在实际部署中需改成实际ip。