

flume 与 kafka 的整合

需求：使用 flume 监控某一个文件夹下面的文件的产生，有了新文件，就将文件内容收集起来放到 kafka 消息队列当中

source: spoolDir Source

channel: memory channel

sink: 数据发送到 kafka 里面去

flume 与 kafka 的配置文件开发

第一步：flume 下载地址

<http://archive.cloudera.com/cdh5/cdh/5/flume-ng-1.6.0-cdh5.14.0.tar.gz>

第二步：上传解压 flume

第三步：配置 flume.conf

```
a1.sources = r1
a1.channels = c1
a1.sinks = k1
a1.sources.r1.channels = c1
a1.sources.r1.type = spooldir
a1.sources.r1.spoolDir = /export/servers/flumedata
a1.sources.r1.deletePolicy = never
a1.sources.r1.fileSuffix = .COMPLETED
a1.sources.r1.ignorePattern = ^(.)*\\.tmp$
a1.sources.r1.inputCharset = GBK
a1.channels.c1.type = memory
a1.sinks.k1.channel = c1
a1.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink
a1.sinks.k1.kafka.topic = test
a1.sinks.k1.kafka.bootstrap.servers = node01:9092,node02:9092,node03:9092
```

```
a1.sinks.k1.kafka.flumeBatchSize = 20  
a1.sinks.k1.kafka.producer.acks = 1
```

创建 flumedata 目录

```
mkdir -p /export/servers/flumedata
```

启动 flume

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

消费数据

```
bin/kafka-console-consumer.sh --bootstrap-server  
node01:9092,node02:9092,node03:9092 --from-beginning --topic test
```