# Flume安装及部署

Apache Flume可用于从不同的数据源可靠有效地加载数据流到HDFS中。Flume具有分布式、高可靠、高容错、易于定制和扩展的特点，是适用于各种方式数据收集的轻量级工具。

本次安装的Flume版本为1.7.0

## 1.1 安装前准备

1.上传安装包并解压

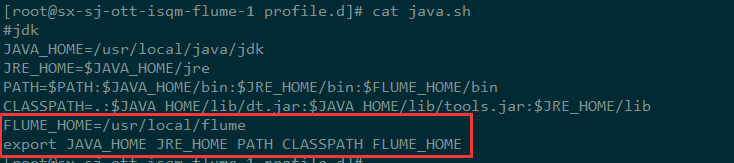
tar -zxvf apache-flume-1.7.0-bin.tar.gz

删除安装包 apache-flume-1.7.0-bin.tar.gz

移动至home目录 mv flume/ /usr/local

2.配置环境变量

vi /etc/profile.d/java.sh



source /etc/profile使配置生效

## 1.2 配置flume-env.sh文件

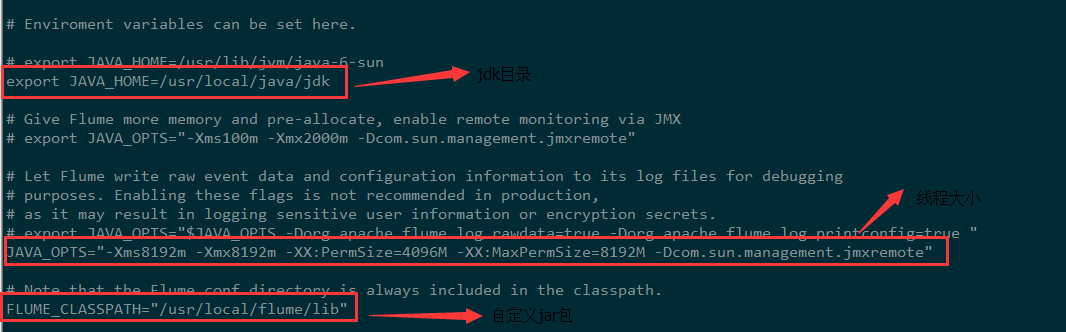
1.进入flume配置目录下

cd /usr/local/flume/conf/

2.拷贝flume-env.sh.template，将其修改为flume-env.sh

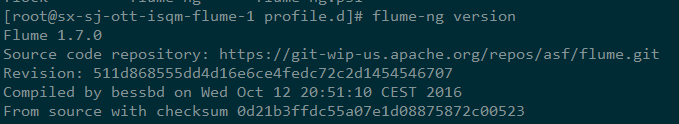
cp flume-env.sh.template flume-env.sh

3.配置JDK路径



## 1.3 验证

flume-ng version



## 1.4 Flume测试实例

Flume可以通过Avro监听某个端口并捕获传输的数据，具体示例如下：

// 创建一个Flume配置文件

cd /home/usr/local/apache-flume-1.7.0-bin/conf

mkdir example

cd example/

touch netcat.conf

// 配置netcat.conf用于实时获取另一终端输入的数据

vim example/netcat.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 that buffers events in memory

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

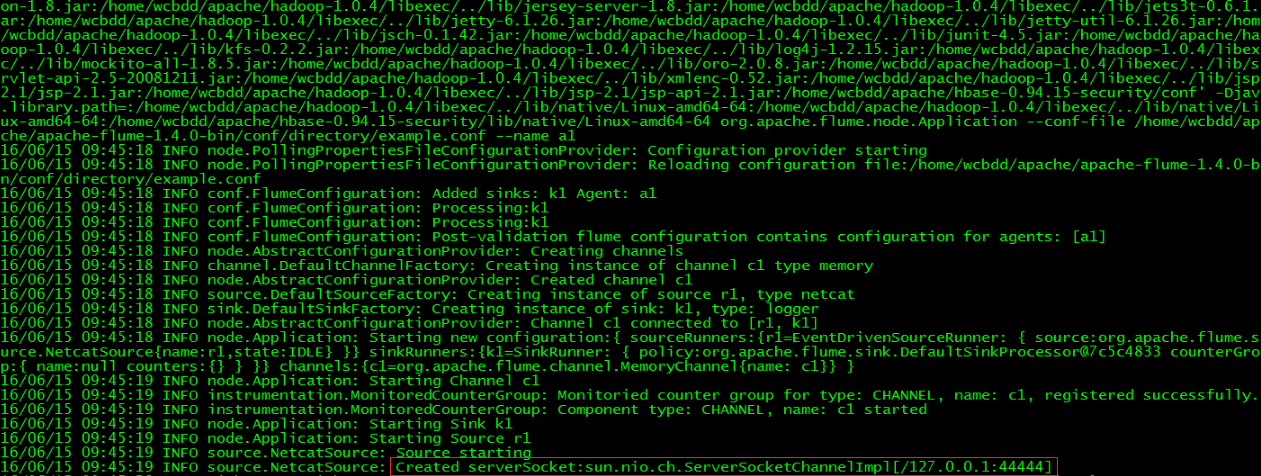
# Bind the source and sink to the channel

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

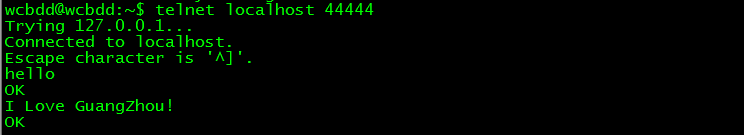
// 运行FlumeAgent，监听本机的44444端口

flume-ng agent -c conf -f example/netcat.conf -n a1 -Dflume.root.logger=INFO,console

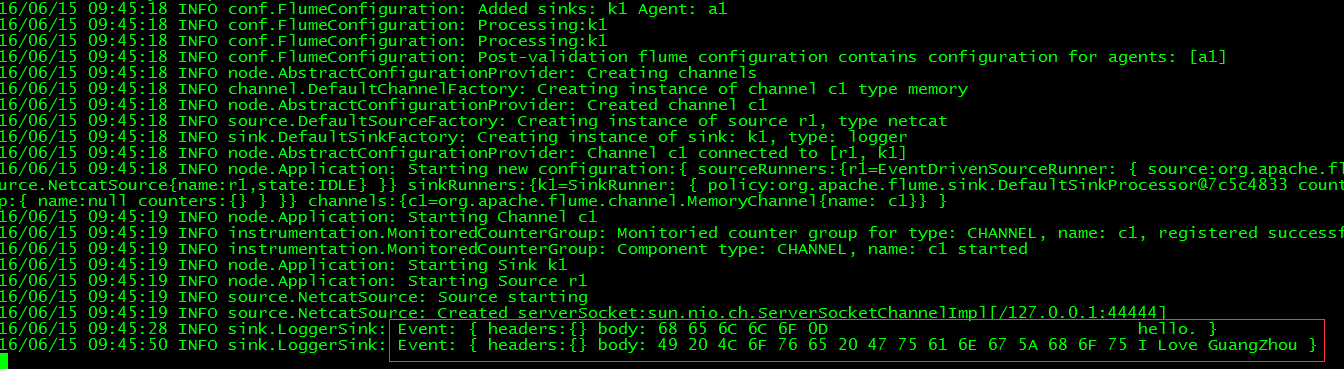


// 打开另一终端，通过telnet登录localhost的44444，输入测试数据

telnet localhost 44444



// 查看flume收集数据情况



参考资料：<http://blog.csdn.net/carl810224/article/details/52472831>