

# flume 的配置和使用

## 实验环境

- CentOS7.8 + oracle jdk8u341
- 软件包 [apache-flume-1.9.0-bin.tar.gz](http://archive.apache.org/dist/flume/1.9.0/apache-flume-1.9.0-bin.tar.gz)

## 实验步骤

1. 下载工具包并解压压缩包

```
mkdir -p /opt/module
wget http://archive.apache.org/dist/flume/1.9.0/apache-flume-1.9.0-bin.tar.gz
tar xzvf ./apache-flume-1.9.0-bin.tar.gz -C /opt/module
```

2. 添加环境变量

修改 `/etc/profile.d/flume.sh` 文件，在末尾添加变量

```
# flume ENV
export FLUME_HOME=/opt/module/apache-flume-1.9.0-bin
export PATH=${FLUME_HOME}/bin:$PATH
```

然后激活配置

```
source /etc/profile
```

这里为了和之后的一些环境变量设置相统一，将环境变量放在全局的配置文件中，事实上这里放在 `~/.bashrc` 中也是可以的

3. 配置 `source type` 为 `avro`，`sink type` 为 `logger`

编辑 `${FLUME_HOME}/conf/avro.conf`，内容如下：

```
# avro.conf 配置文件
agent1.sources = r1
agent1.sinks = k1
agent1.channels = c1
# 配置 Source 监听端口为 4141 的 avro 服务
agent1.sources.r1.type = avro
agent1.sources.r1.bind = 0.0.0.0
agent1.sources.r1.port = 4141
agent1.sources.r1.channels = c1
# 配置 Sink
agent1.sinks.k1.type = logger
agent1.sinks.k1.channel = c1
# 配置 Channel
agent1.channels.c1.type = memory
agent1.channels.c1.capacity = 1000
agent1.channels.c1.transactionCapacity = 100
```

4. 启动 flume

```
flume-ng agent --conf conf --conf-file conf/avro.conf --name agent1 -
Dflume.root.logger=INFO,console
```

5. 在**新的终端**中用户目录下新建文件 avro-input.txt 并写入信息，并使用 avro-client 向 agent1 监听的 avro 服务发送文件

```
flume-ng avro-client -c ${FLUME_HOME}/conf/ -H 0.0.0.0 -p 4141 -F ~/avro-
input.txt
```

文件内容应该是

```
Hello
Flume
[学号]
```

6. 在第 4 步中的输出中检查接收到的信息，截图并完成实验报告

## 思考题(额外加分项)

**背景介绍：**现在有如下脚本正在运行，该脚本的作用是每秒生成一个时间戳并输出到 /tmp/xxx.log 文件中。

```
#!/usr/bin/env bash
while :
do
    echo $(date '+%Y%m%d %T %s' [学号]) | tee -a /tmp/xxx.log
    sleep 1
done
```

**要求：**使用flume捕获到 /tmp/xxx.log 文件变化，输出到命令行中，并进行截图，需要贴出flume的相应配置文件。（查阅官方文档可以快速了解该题的解决方法）

沿用之前的配置，用 Linux 中的管道也能够实现这样的效果，但是我们这里希望大家使用 flume 完成上述要求

**注意：**这里的截图**必须**带有学号，并使用**红色方框**框出，否则做**无效**计。