

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 `~/.bashrc` 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 ~/.bashrc
```

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

背景介绍：在生产环境中，日常 web 应用通常分布在上百个服务器，甚至上千个、上万个服务器。产生的日志，处理起来也非常麻烦。用 flume 的这种组合方式能很好的解决这一问题，每台服务器部署一个 flume 采集日志，传送到一个集中收集日志的 flume，再由此 flume 上传到 hdfs、hive、hbase 等进行日志分析。

问题：尝试聚合多个 flume agent 到一个 fluagent 中，并在实验报告中写明步骤和展示必要截图。

在三个 agent 上分别输入 [学号]-1, [学号]-2, [学号]-3，在聚合的 agent 上应当能看到这三个输出。

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

