

flume 测试环境使用文档

Hang Yan

March 15, 2014

1 简介

flume是结构化日志的传输通道，集成结构化日志客户端的组件，将日志发送给本机部署的flume后，本机并发送给solr建立索引(云知道)。搭建测试环境时，本机的flume收到结构化日志数据后，不再发往远端，由于flume传输的数据均为序列化后的二进制数据，所以写到本地的日志文件时不可读的，需要用avro工具

2 依赖

2.1 jdk

flume是java程序(windows任务管理器的程序名即为java。测试机上需要有完整的jdk环境。

2.2 chmod

测试flume模拟了线上写入hadoop集群的操作。中间对文件的操作需要调用linux上的chmod命令。linux目录下也会有此程序，将其加入环境变量中的PATH即可。

3 启动

3.1 windows

如果jdk版本为1.6.31的话，将整个jdk1.6.0_31/目录拷贝至apache-flume-1.5.0-SNAPSHOT-bin目录下.则直接点击run-windows.bat即可。如果不是以上情况，并确定java命令在系统path.bat中的"jdk1.6.0_31/bin/java.exe" 改为java.exe即可.然互点击run-windows.bat启动.

3.2 linux

确保java在PATH中即可，运行start.sh 即可。

4 运行

4.1 文件

flume启动后，会在本地创建一个data目录用来缓存日志文件，logs/flume.log是正在写的flume日志.如果data和flume占用空间过大，可删除这两个目录。logs下的日志文件过多的话也可删除.每次启动flume

4.2 页面

flume启动后，会启动两个端口,一个是用来接收客户端数据的4545端口。另一个是网页端口,用来展示运行.此页面展示了一个json文档,推荐浏览器中安装jsonview插件查看.客户端成功连接flume后，此页面会有lo

5 日志查看

flume收到的日志文件最终都写在flume文件夹下,正常的目录结构是 flume/s.n/s.v/date/hour/hostname
s.n s.v是结构化日志的结构名称和版本号。当前是msp/1.date/hour是loglib flush接口所传入的时间戳解
最深层的目录包含current和stable.current是flume正在写的日志文件,所有文件名有tmp后缀, 不可查看
并去除'tmp'后缀.

avro-tool jar包可以用来查看stable里面的的日志内容.具体命令为

```
java -jar avro-tools.jar tojson [logavro-file]
```

可将输出重定向到文本文件中查看。

6 停止

6.1 windows

flume 的java 启动窗口,Ctrl-C y 确认即可.

6.2 linux

ps aux|grep flume 找到相关的进程号.kill -15 或kill 即可.