**网易：大数据工程师实验报告**

云课堂昵称：大石ykt1507814549683 实验日期： 2018-09-03

一、实验题目

数据获取和预处理：Flume

二、实验要求

1、学员通过教学视频安装flume客户端

2、安装完成后打印flume-ng version

3、通过教学视频中给定的文件地址，启动flume将文件自动同步到HDFS

4、通过hadoop fs命令验证文件同步成功

5、将操作过程及截图写入实验报告文档中，并将代码带包一起以附件形式上传

注：实验报告文档模板已上传至QQ群，同学们请前往群文件自行下载哦

作业提交要求：

1、格式：上传实验报告的Word文档/PDF或压缩包

2、文件命名：“网易大数据-数据获取和预处理：Flume-云课堂昵称

三、操作步骤

1，下载并解压、配置安装：

主要命令如下：

下载：

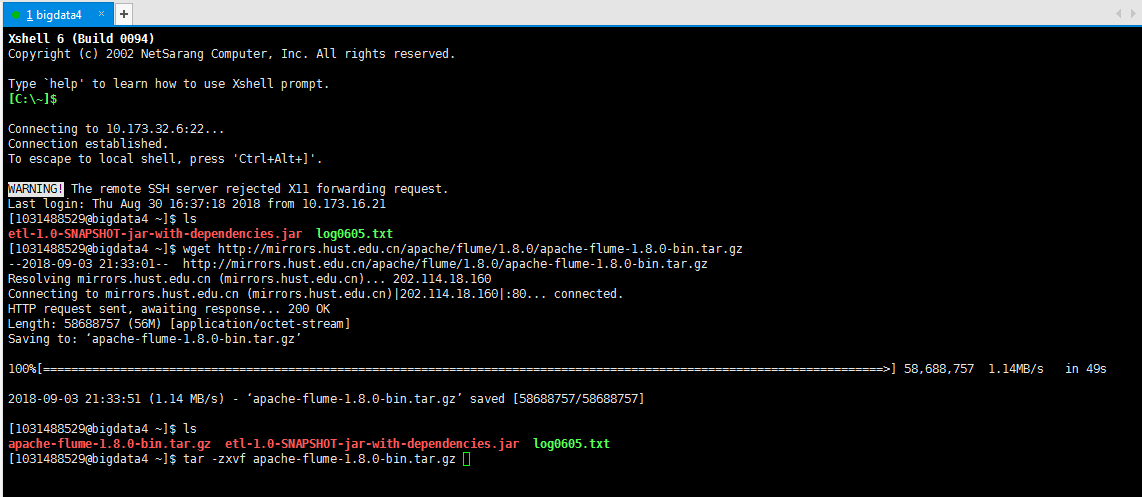
wget <http://mirrors.hust.edu.cn/apache/flume/1.8.0/apache-flume-1.8.0-bin.tar.gz>

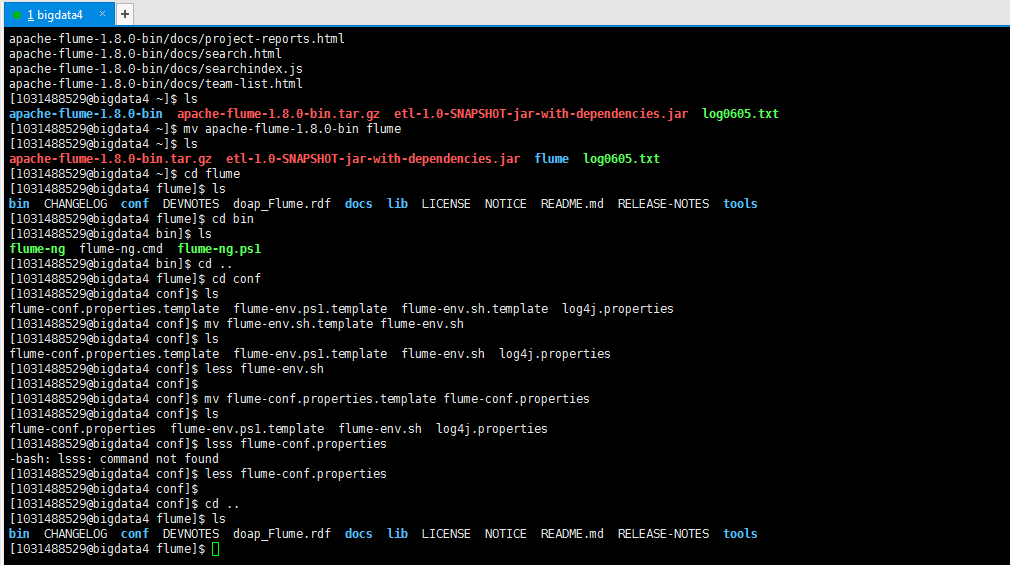
解压：

tar -zxvf apache-flume-1.8.0-bin.tar.gz

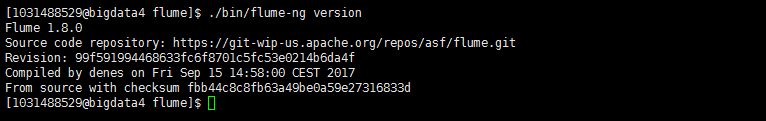
重命名：

mv apache-flume-1.8.0-bin flume



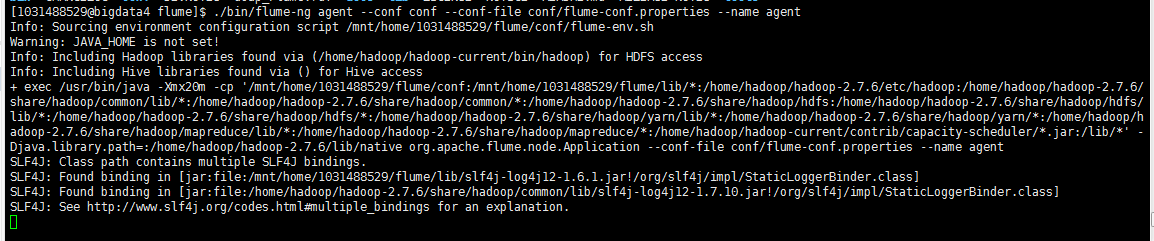


查看版本：



启动agent进程：

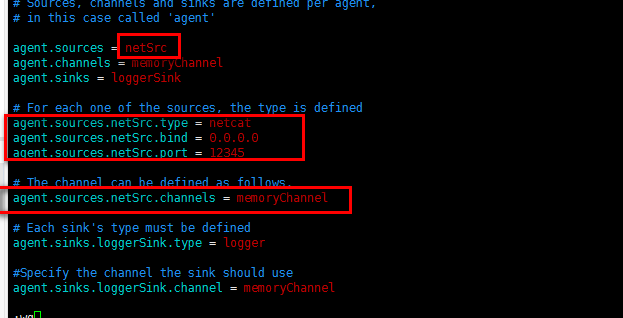
[1031488529@bigdata4 flume]$ ./bin/flume-ng agent --conf conf --conf-file conf/flume-conf.properties --name agent



示例一：

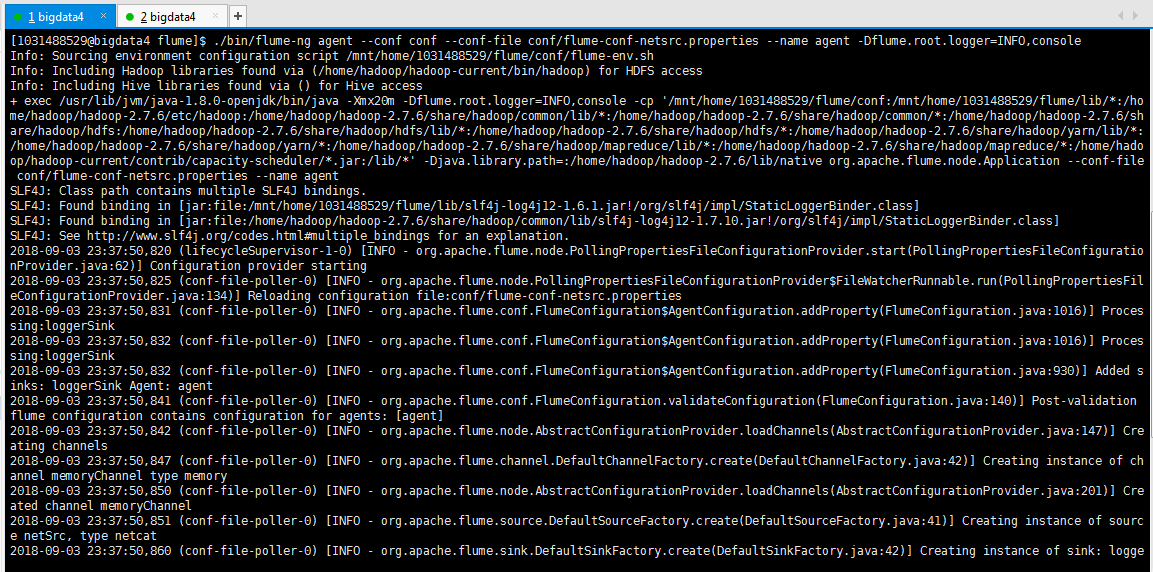
配置flume-conf-netsrc.properties

[1031488529@bigdata4 conf]$ vi flume-conf-netsrc.properties

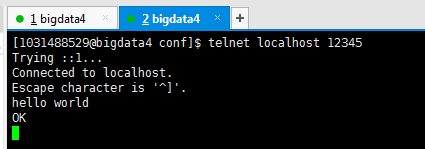


启动netsrc的agent进程：

[1031488529@bigdata4 flume]$ ./bin/flume-ng agent --conf conf --conf-file conf/flume-conf-netsrc.properties --name agent -Dflume.root.logger=INFO,console



telnet localhost 12345 测试

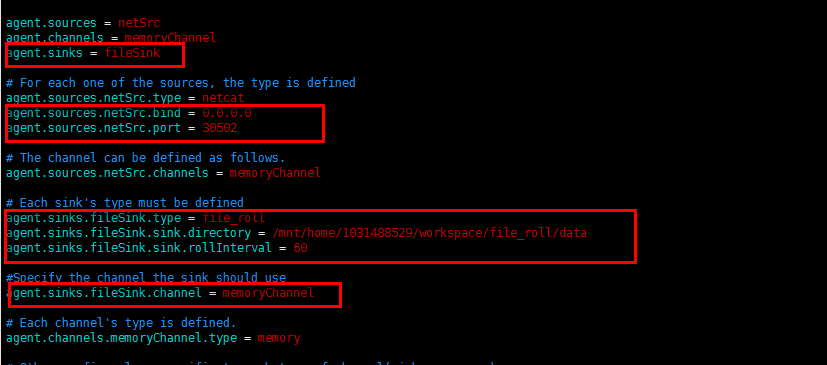


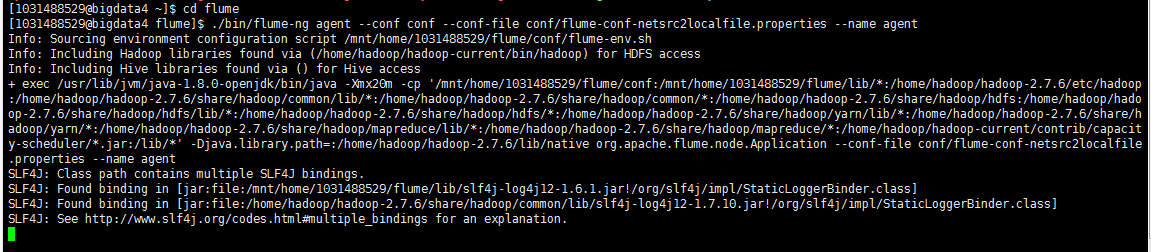


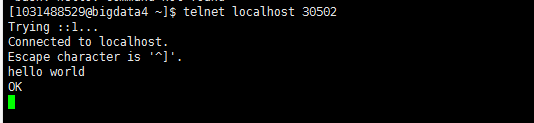
示例二：

配置：vi flume-conf-netsrc2localfile.properties

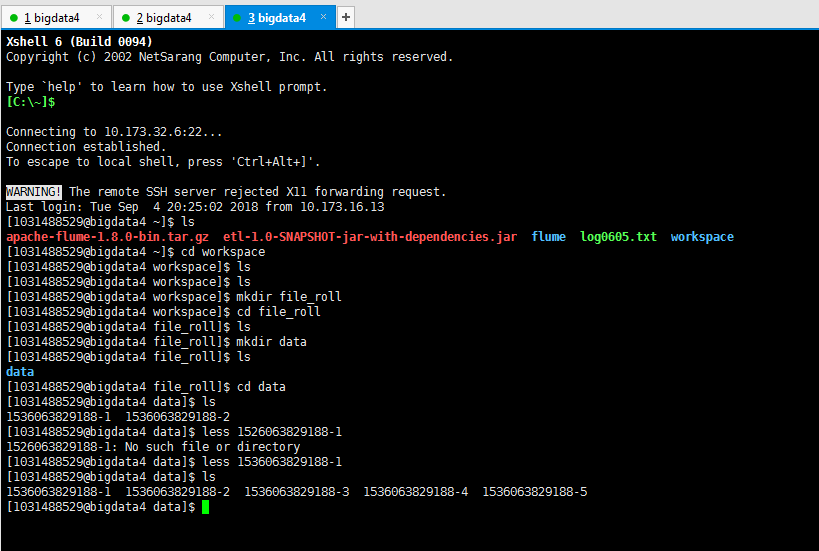
使用个人专用端口30502



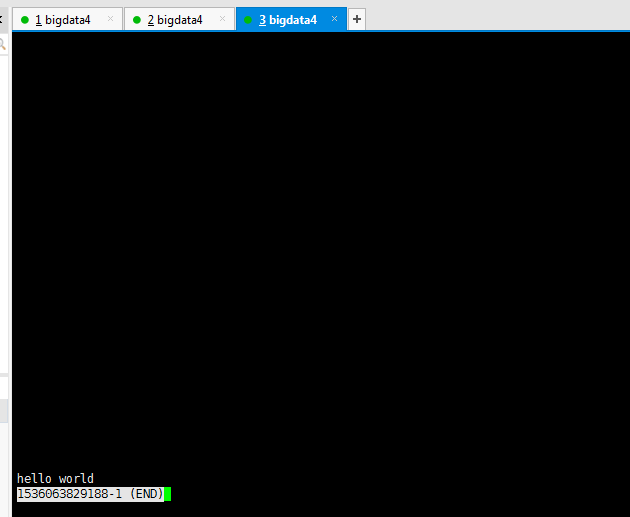


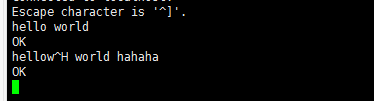


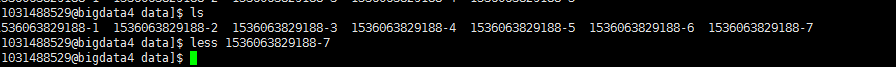
需要自行在个人根目录下创建 workspace/file\_roll/data目录

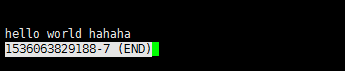


测试结果：



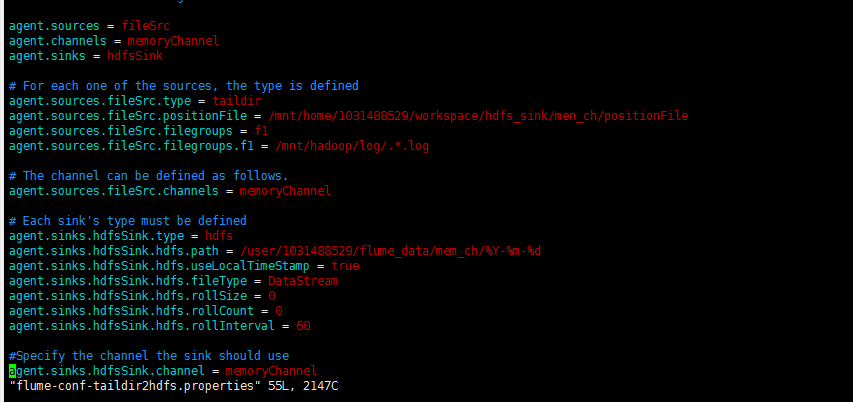






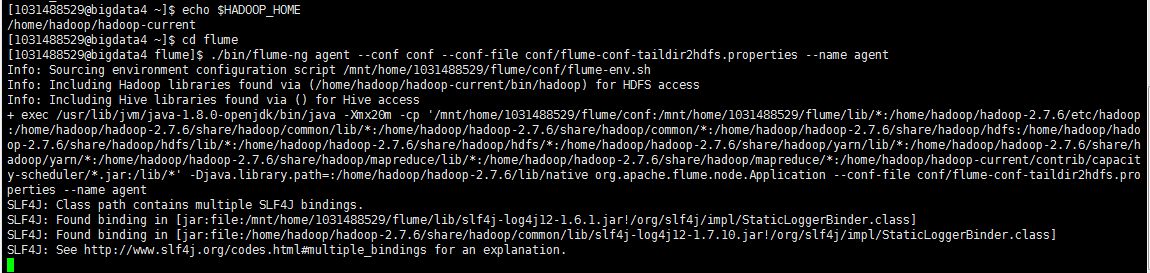
示例三：

配置文件：vi flume-conf-taildir2hdfs.properties



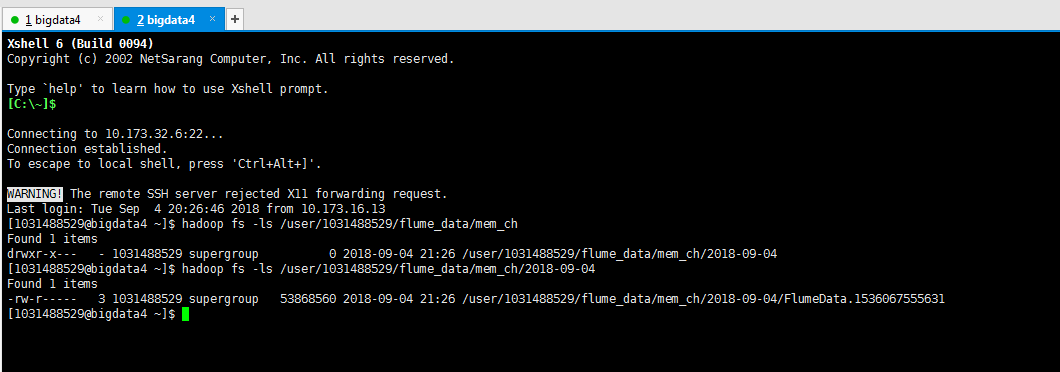
检查HADOOP\_HOME，并启动agent:

./bin/flume-ng agent --conf conf --conf-file conf/flume-conf-taildir2hdfs.properties --name agent



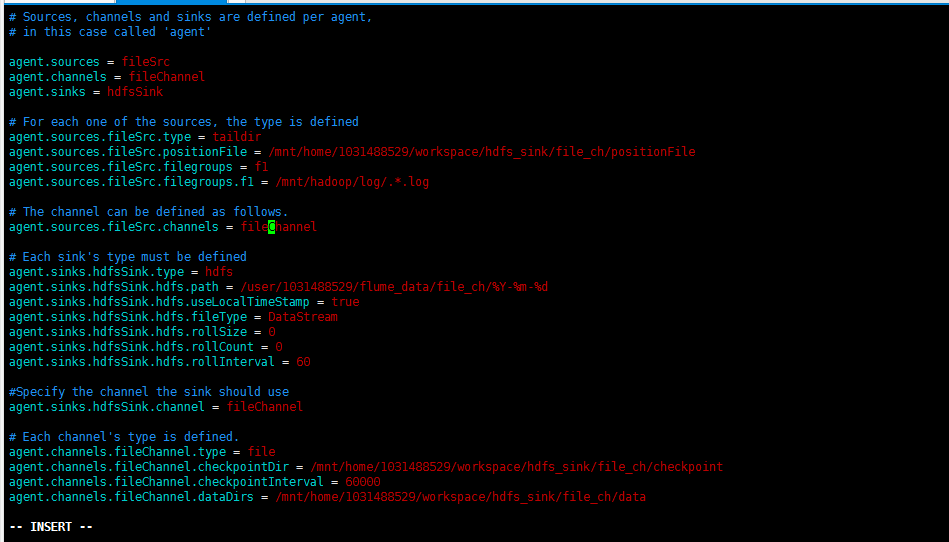
查看输出结果:

hadoop fs -ls /user/1031488529/flume\_data/mem\_ch



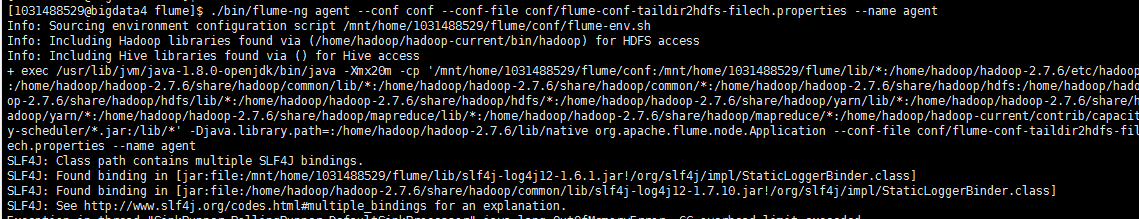
示例四：

配置文件：vi flume-conf-taildir2hdfs-filech.properties



启动进程：

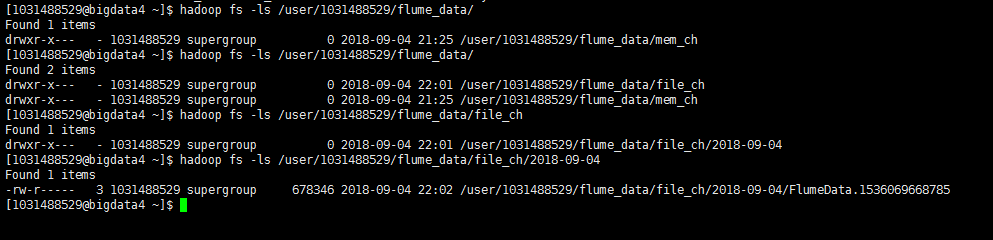
./bin/flume-ng agent --conf conf --conf-file conf/flume-conf-taildir2hdfs-filech.properties --name agent



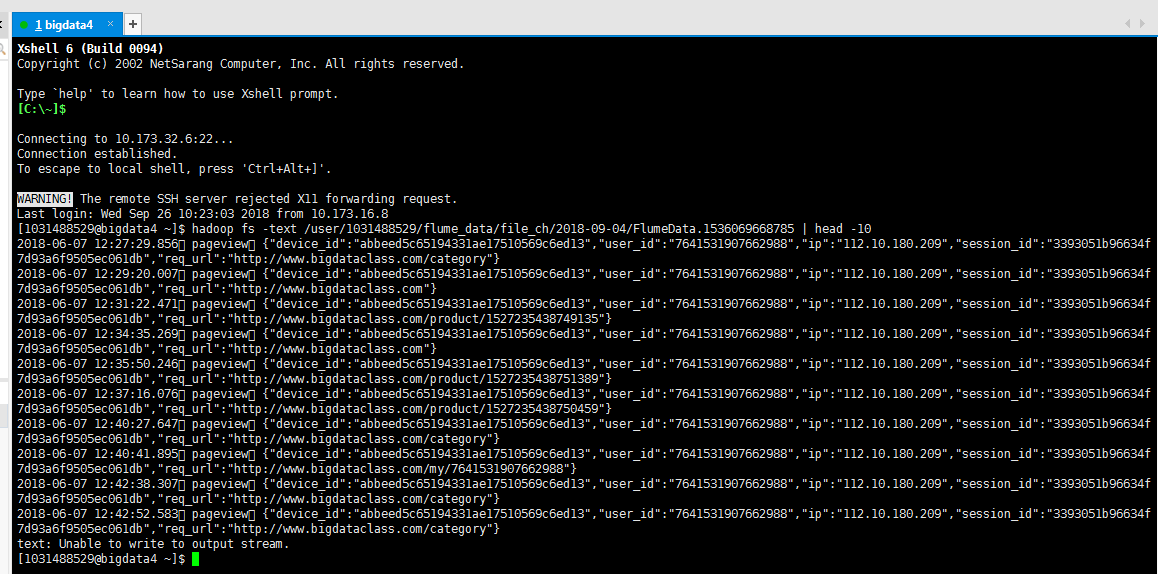
查看结果：

hadoop fs -ls /user/1031488529/flume\_data/file\_ch/2018-09-04

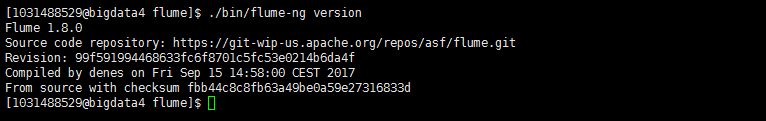
能看到已生成文件/user/1031488529/flume\_data/file\_ch/2018-09-04/FlumeData.1536069668785

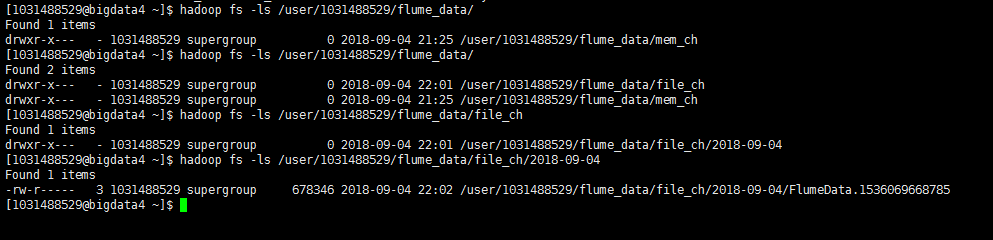


查看日志文件内容：



四、实验结果



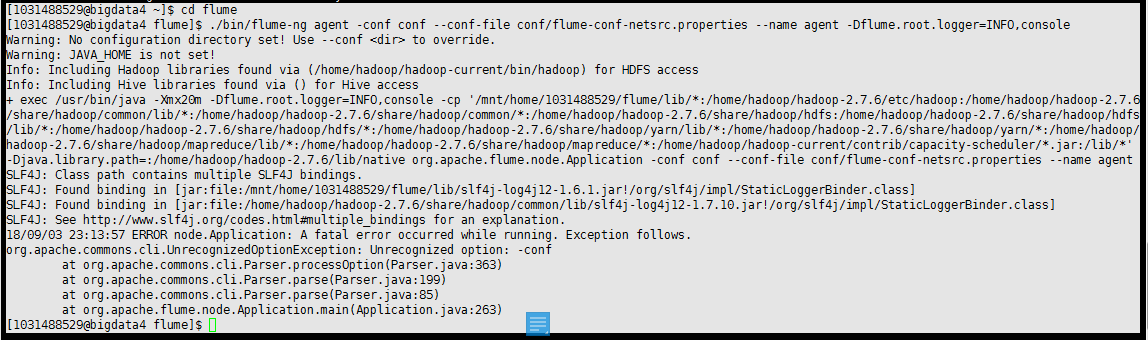


五、总结反思

1，这节课主要是配置properties文件，然后启动agent，启动agent的时候选择某个配置文件即可；

2，出现异常都是因为配置文件细节的错误，仔细检查即可；

3，一开始试验时犯了低级错误，因为启动agent的命令总--conf少敲了一个-，导致报错，需要细心、耐心和信心自己去查找原因；

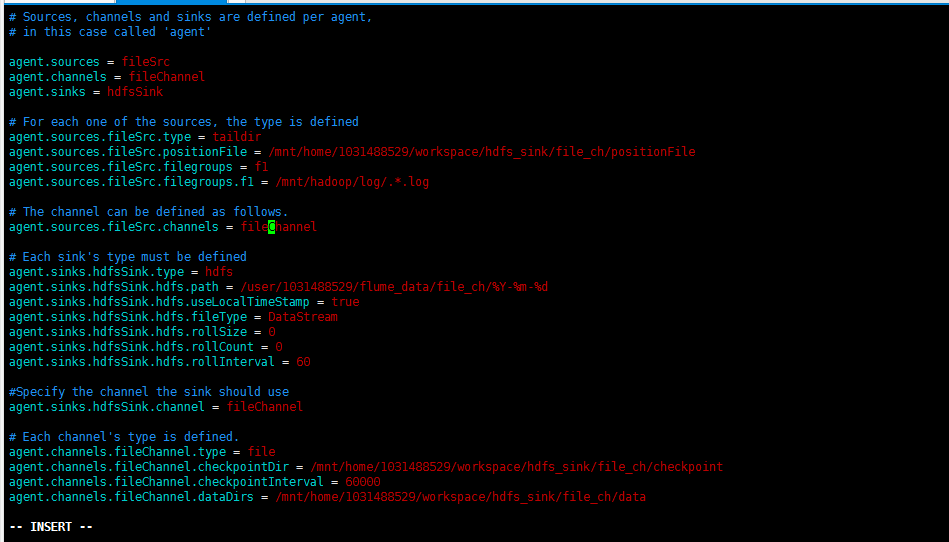


4，配置文件flume-conf-taildir2hdfs-filech.properties可以看到完整的Agent配置，包括source，channel，sink三个组件:

source：数据接收器，配置数据接收方式为taildir（即监听服务器上的日志文件），以及配置监听的目标文件夹等；

channel: 缓存容器，配置类型为fileChannel（即将文件缓存到本地文件，适合大容量、低吞吐、宕机可恢复），以及配置相关缓存的目标文件夹、时间间隔等；

sink：持久化组件，配置持久化目标为hdfs；



六、作业提交

1. 请将本文档按照《网易大数据\_X章X节\_云课堂昵称》的名称命名；

2. 请将本文档、数据文件打包以附件形式上传到课程作业部分