# flume实践手册

1. 我们需要创建一个目录来存放我们要采集的文件，此处设定的目录为/root/flume\_test

|  |
| --- |
| mkdir /root/flume\_test |

1. 我们在/root/flume\_test目录里创建一个名为11.txt的文件

|  |
| --- |
| vi /root/flume\_test/11.txt |

然后在文件里输入一些字符，保存退出，作为需要采集的文件。

1. 我们还需要在Hdfs上创建一个目录存来放我们采集后的文件，在此我们设定的目录为/flume\_test

|  |
| --- |
| hadoop fs -mkdir /flume\_test |

1. 接下来我们来配置flume

Flume配置文件在生产中一般习惯性放在/usr/lib/flume-ng/conf/目录（不过也可以根据自己的习惯来设定目录）。

此处我们定义flume agent的名字为a1（可以根据自己的习惯定义，简单明了即可）

通过vi打开/usr/lib/flume-ng/conf/flume.conf文件，在文件里做如下配置：

a1.sources = r1 #定义source的名字为r1

a1.sinks = k1 #定义sink的名字为k1

a1.channels = c1 #定义channel的名字为c1

# Describe/configure the source

a1.sources.r1.type = spooldir #定义source的类型为spooldir

a1.sources.r1.spoolDir = /root/flume\_test #定义source获取数据的目录

a1.sources.r1.deletePolicy=immediate #数据处理完后，删除文件

a1.sources.r1.fileHeader = true #保留文件头，如果false是删除文件头（文件的第一行）

a1.sources.r1.channels = c1 #source对应的channel

# Use a channel which buffers events in memory

a1.channels.c1.type = memory #定义channel类型

a1.channels.c1.capacity = 1000 #定义channel容量

a1.channels.c1.transactionCapacity = 100 #定义每次传输的event数量（记录数）

a1.sinks.k1.type = hdfs #定义sink的类型为hdfs

a1.sinks.k1.channel = c1 #sink对应的channel

a1.sinks.k1.hdfs.path = /flume\_test #定义目标路径（数据最终写入路径）

#a1.sinks.k1.hdfs.filePrefix = events- #定义写入文件的前缀

a1.sinks.k1.hdfs.fileType = DataStream #定义写入数据类型

1. 启动flume进程

flume-ng是flume agent的启动脚本，一般在flume安装目录下的bin文件夹下，flume-ng启动一般格式 ：

flume-ng agent --conf <flume配置自身配置路径> --conf-file <agent 配置文件> --name <agent名字> -Dflume.root.logger= <日志级别>

本次启动脚本如下：

|  |
| --- |
| flume-ng agent --conf /usr/lib/flume-ng/conf --conf-file /usr/lib/flume-ng/conf/flume.conf --name a1 -Dflume.root.logger=DEBUG,console |

6、检查结果

7、我们继续往/root/flume\_test目录里放置文件，进行测试（但文件不能直接在/root/flume\_test创建，需要在别的地方创建后再mv到/root/flume\_test）

8、更多agent参数大家可以查看官方文档，官网地址：

http://flume.apache.org/FlumeUserGuide.html