# **利用Flume将本地文件数据中收集到HDFS**

## 1 准备Flume安装包

准备Flume安装包apache-flume-1.7.0-bin.tar.gz，并解压。

## 2 启动Flume agent

**步骤1：在本地创建/tmp/logs文件夹**

|  |
| --- |
| mkdir /tmp/logs |

**步骤2：在HDFS上创建/data/logs目录，并确认创建成功**

|  |
| --- |
| bin/hdfs dfs -mkdir -p /data/logs |

如果没有权限创建该目录，则使用：

|  |
| --- |
| HADOOP\_USER\_NAME=hdfs bin/hdfs dfs -mkdir -p /data/logs |

**步骤3：在conf/目录下编写logagent.properties (注意，请将hdfs://master:8020换成自己namenode的地址，具体查看core-site.xml中的fs.defaultFS配置，有的同学配置的可能是hdfs://master:9000)**

|  |
| --- |
| **LogAgent.sources = mysource**  **LogAgent.channels = mychannel**  **LogAgent.sinks = mysink**  **LogAgent.sources.mysource.type = spooldir**  **LogAgent.sources.mysource.channels = mychannel**  **LogAgent.sources.mysource.spoolDir =/tmp/logs**  **LogAgent.sinks.mysink.channel = mychannel**  **LogAgent.sinks.mysink.type = hdfs**  **LogAgent.sinks.mysink.hdfs.path = hdfs://master:8020/data/newlogs/%Y/%m/%d/%H/**  **LogAgent.sinks.mysink.hdfs.batchSize = 1000**  **LogAgent.sinks.mysink.hdfs.rollSize = 0**  **LogAgent.sinks.mysink.hdfs.rollCount = 10000**  **LogAgent.sinks.mysink.hdfs.useLocalTimeStamp = true**  **LogAgent.channels.mychannel.type = memory**  **LogAgent.channels.mychannel.capacity = 10000**  **LogAgent.channels.mychannel.transactionCapacity = 100** |

**步骤4：在flume安装目录下运行以下命令启动agent：**

|  |
| --- |
| **bin/flume-ng agent --conf conf --conf-file conf/logagent.properties --name LogAgent -Dflume.root.logger=DEBUG,console** |

**步骤5： 向/tmp/logs拷入几个文本文件（不重名，只能是文本文件）查看HDFS上的/data/logs是否产生数据，如果存在，且配置成功，可继续往/tmp/logs拷贝文件**

|  |
| --- |
| **bin/hdfs dfs -ls /data/logs** |