Flume搭建

前提条件线开启kafka

1. kafka创建主题

|  |
| --- |
| bin/kafka-topics.sh --create --bootstrap-server bigdata1:9092 --replication-factor 1 --partitions 4 --topic ods\_mall\_log  或者  kafka-topics.sh --zookeeper bigdata1:2181 --create --topic lxw --partitions 4 --replication-factor 1 |

1. 解压缩flume

|  |
| --- |
| tar -zxvf apache-flume-1.9.0-bin.tar.gz -C /opt/model/ |
| 重命名  mv apache-flume-1.9.0-bin.tar.gz flume-1.9.0 |

1. 在flume的conf下创建kafka的配置文件

|  |
| --- |
| vi /opt/model/flume-1.9.0/kafka.conf  #配置文件：  a1.sources = r1  a1.sinks = k1  a1.channels = c1  #这些是你定义的 Flume Agent 中的源（source）、汇（sink）和通道（ channel）的名称。  a1.sources.r1.type = netcat  a1.sources.r1.bind = localhost  a1.sources.r1.port = 25001 要采集的端口号  #这部分定义了一个名为 r1 的源，其类型为 netcat，它将监听位于本地主机上端口号为 25001 的数据。你可以在该端口上发送数据给 Flume 进行采集。  a1.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink  a1.sinks.k1.kafka.bootstrap.servers = bigdata1:9092  a1.sinks.k1.kafka.topic = ods\_mall\_log #命名为刚刚kafka的topic创建的日志文件  #这部分定义了一个名为 k1 的汇，其类型为 KafkaSink，它将数据发送到 Kafka 主题中。kafka.bootstrap.servers 指定了 Kafka 集群的引导服务器地址和端口，这里设置为 bigdata1:9092。kafka.topic 指定了数据要发送到的 Kafka 主题，这里设置为 ods\_mall\_log  a1.channels.c1.type = memory  a1.channels.c1.capacity = 1000  a1.channels.c1.transactionCapacity = 100  #这部分定义了一个名为 c1 的内存通道，它用于在源和汇之间缓冲数据。type 设置为 memory 表示使用内存作为通道的存储方式。capacity 指定了通道的最大容量，这里设置为 1000 条事件。transactionCapacity 指定了每次事务提交的最大事件数，这里设置为 100。  a1.sources.r1.channels = c1  a1.sinks.k1.channel = c1  #这部分指定了源 r1 所使用的通道为 c1，以及汇 k1 所使用的通道也是 c1。这样数据从源流向汇时会经过通道进行缓冲。  #通过以上配置，Flume Agent 将在 localhost:25001 监听数据，并将采集到的数据发送到 Kafka 集群的 ods\_mall\_log 主题中。 |

1. 启动flume

|  |
| --- |
| /opt/model/flume-1.9.0/bin/flume-ng agent -c conf/ -n a1 -f conf/kafka.conf -Dflume.root.logger=INFO,console  解释/usr/flume/bin/flume-ng：启动Flume-ng二进制文件。  agent：指定要启动的Flume组件类型为代理。  -n a1：给Flume代理指定一个名称为a1。  --conf conf：设置Flume配置文件所在的目录为当前目录下的conf目录。  --conf-file /usr/flume/conf/mykafka.conf：指定Flume代理使用的配置文件路径和名称。  -Dflume.root.logger=INFO,console：设置Flume代理的日志级别为INFO，并将日志输出到控制台。 |

1. 测试（启动生产者）在kafka中

|  |
| --- |
| kafka-console-producer.sh --broker-list slave1:9092 --topic ods\_mall\_log  解释  这是一个 Kafka 命令行生产者的示例命令，用于在 ods\_mall\_log 主题上发送消息到 Kafka 集群。  kafka-console-producer.sh 是 Kafka 提供的一个命令行工具，它可以用来发送消息到 Kafka 集群。  --broker-list hadoop102:9092 指定了 Kafka 集群的引导服务器地址和端口号。在这里，hadoop102:9092 表示 Kafka 集群中的一个 Broker，可以根据实际情况修改。  --topic ods\_mall\_log 指定了要将消息发送到的 Kafka 主题名称。在这里，消息将会被发送到名为 first 的主题。  使用这个命令后，将会进入一个交互式控制台，你可以在控制台中输入消息内容并按下回车键发送到 Kafka 集群。如果一切正常，你可以通过消费者工具来验证消息已经成功地被发送到了 Kafka 集群中的相应主题中。 |
|  |

6. 启动消费者

|  |
| --- |
| kafka-console-consumer.sh --bootstrap-server slave1:9092 --topic ods\_mall\_log --from-beginning  或者  nc localhost 25001  使用消费Kafka的topic，观察数据是否成功上传到Kafka中： |

7. 查看俩条数据

|  |
| --- |
| kafka-console-consumer.sh --bootstrap-server slave1:9092 --topic ods\_mall\_log --max-messages 2  解释8. 解释：#消费Kafka指定的topic 并取2条  --max-message 2 最多获取2条消息  --from-beginning 从最开始消费，如果不输则是实时消费 |