Flume+Kafka

1.安装zookeeper

到https://www-eu.apache.org/dist/zookeeper/下载zookeeper-3.4.6.tar.gz

```
gzip -d zookeeper-3.4.6.tar.gz
tar -xvf zookeeper-3.4.6.tar
```

进入conf目录，将示例的配置文件zoo_sample.cfg改名为zoo.cfg，作为我们的配置文件使用，命令如下：

```
cd zookeeper-3.4.6/conf
mv zoo_sample.cfg zoo.cfg
vi zoo.cfg
```

修改zoo.cfg的端口为容器开放出来的端口例如2181

```
clientPort=2181 #默认使用2181
```

改完后进入bin目录，执行zkServer.sh的start命令启动Zookeeper服务

```
cd ..
cd bin
./zkServer.sh start
```

提示以下内容，启动成功

JMX enabled by default

Using config: /opt/zookeeper-3.4.6/bin/../conf/zoo.cfg

Starting zookeeper ... STARTED

2.安装kafka

到https://www.apache.org/dyn/closer.cgi?path=/kafka/2.1.1/kafka_2.11-2.1.1.tgz下载kafka_2.11-2.1.1.gz

2.1 执行以下命令，下载并解压Kafka

```
tar xvzf kafka_2.11-2.1.1.tgz
```

2.2 修改kafka_2.11-2.1.1/config目录下的server.properties文件

找到zookeeper.connect一项，修改为tc-host的地址，如下：

```
zookeeper.connect=10.170.66.52:2181 #zookeeper装在另一个容器中 所以通过宿主机的地址和端口访问
```

2.3 执行bin目录下的kafka-server-start.sh命令启动Kafka

以server.properties文件作为参数，启动Kafka

```
$ cd kafka_2.11-2.1.1
$ ./bin/kafka-server-start.sh ./config/server.properties &
```

命令后面的&符号是将启动的Kafka服务设置为后台进程，方便我们进一步的操作。

启动命令是以配置文件为参数，按照相关的配置来启动的。server.properties是默认的配置文件，几个比较常用的配置项包括：

(1)broker.id broker的id号

(2)port 端口

(3)zookeeper.connect zookeeper的连接地址

(4)log.dirs 日志的目录

2.4 简单功能验证

Kafka成功启动后，可以通过一些简单的命令来验证一下功能。

(1)创建一个名为test的topic

```
./bin/kafka-topics.sh --create --zookeeper 10.170.66.52:2181 --replication-factor 1 --partitions 1 --topic test
```

create命令的replication-factor是设置该topic在多少个broker上存储。

(2)查询topic的属性

```
./bin/kafka-topics.sh --describe --zookeeper 10.170.66.52:2181 --topic test
```

describe命令的返回信息中，罗列了所有partition的信息，其中：

(1)Partition是编号

(2)Leader是一个broker的编号，该broker存储了当前partition，并且被选举为broker列表中的Leader。在Kafka中，只有Leader节点会负责消息的读和写，其他broker只是做备份

(3)Replicas是存储了该partition的broker列表

(4)Isr是当前可用的broker列表

(3)生产者连接broker发送消息

```
./bin/kafka-console-producer.sh --broker-list localhost:9092 --topic test
```

(4)消费者连接broker接收消息

```
bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic test --from-beginning
```

3. flume+kafka

3.1 配置flume

在flume的conf目录中新建nginx.conf实现转发nginx日志文件给kafka：

nginx.conf如下所示：

```
agent1.sources = seqGenSrc
agent1.channels = memoryChannel
```

```
agent1.sinks = loggerSink

# For each one of the sources, the type is defined
agent1.sources.seqGenSrc.type = exec
agent1.sources.seqGenSrc.command = tail -f /var/log/nginx/access.log

# The channel can be defined as follows.
agent1.sources.seqGenSrc.channels = memoryChannel
agent1.sources.seqGenSrc.batchSize = 10000
agent1.sources.seqGenSrc.batchTimeput = 1000

# Each sink's type must be defined
agent1.sinks.loggerSink.type = org.apache.flume.sink.kafka.KafkaSink
agent1.sinks.loggerSink.topic = test #依据你创建的topic
agent1.sinks.loggerSink.brokerList = 10.170.66.52:9092 #容器不同通过宿主机ip端口
agent1.sinks.loggerSink.requiredAcks = 1
agent1.sinks.loggerSink.batchSize = 1000


#Specify the channel the sink should use
agent1.sinks.loggerSink.channel = memoryChannel

# Each channel's type is defined.
agent1.channels.memoryChannel.type = memory

# Other config values specific to each type of channel(sink or source)
# can be defined as well
# In this case, it specifies the capacity of the memory channel
agent1.channels.memoryChannel.capacity = 10000
agent1.channels.memoryChannel.transactionCapacity = 10000
```

3.2 zookeeper容器中启动zookeeper后，转到kafka容器中启动kafka

```
./bin/kafka-server-start.sh ./config/server.properties &
```

并创建一个名为test的topic

```
./bin/kafka-topics.sh --create --zookeeper 10.170.66.52:2181 --replication-factor 1 --partitions 1 --topic test
```

3.3 flume容器中启动flume发送日志

```
flume-ng agent --conf /usr/local/flume/apache-flume-1.9.0-bin/conf --conf-file /usr/local/flume/apache-flume-1.9.0-bin/conf/nginx.conf --name agent1 -Dflume.root.logger=INFO,console
```

3.4 kafka容器中消费者连接broker接收消息

```
bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic test ---topic test --from-beginning
```