Kafka+Flume配置文档

搭建人：卢居辉

文档时间：2018.7.20

目 录

一　Kafka

1 版本选择

CDH-3.7.5 的组件版本：KAFKA-3.1.0-1.3.1.0.p0.35

2 集群搭建规划

	Manager	Namenode	Datanode
ZooKeeper	是	是	是
Kafka(broker)	是	是	是
Flume(agent)	是	---	---

【注意：必须配置好主机（免密登录、修改 host 主机名）】

3 修改 Kafka 相关配置文件

第一步：

```
cd /opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/etc/kafka/conf.dist
```

第二步：

```
vi  server.properties
```

第三步：(主要修改其中的 6 个参数)

```
broker.id=0  //标示符（三台主机的 id 分别为 0,1,2）
host.name=manager //绑定的主机
log.dirs=/usr/local/soft/kafka/kafka-logs  //数据保存的位置
log.retention.hours=168  //数据的保留时间(168 hours=7 天)
zookeeper.connect=manager:2181,namenode:2181,datanode:2181
delete.topic.enable=true //可以删除已创建主题
```

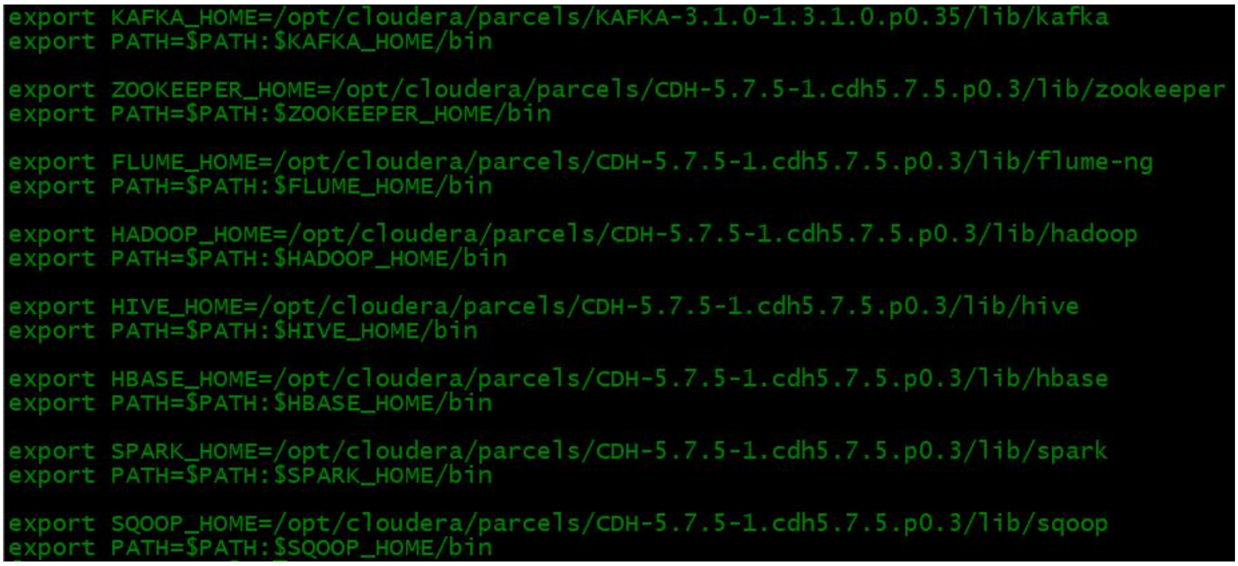
4 配置全局变量

（1）vi .bashrc

（2）添加配置

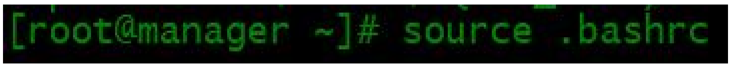
```
export KAFKA_HOME=/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35
```

export PATH=$PATH:$KAFKA_HOME/bin



```
export KAFKA_HOME=/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka
export PATH=$PATH:$KAFKA_HOME/bin

export ZOOKEEPER_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin

export FLUME_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/flume-ng
export PATH=$PATH:$FLUME_HOME/bin

export HADOOP_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop
export PATH=$PATH:$HADOOP_HOME/bin

export HIVE_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hive
export PATH=$PATH:$HIVE_HOME/bin

export HBASE_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hbase
export PATH=$PATH:$HBASE_HOME/bin

export SPARK_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/spark
export PATH=$PATH:$SPARK_HOME/bin

export SQOOP_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
```

（3）：退出保存，执行以下命令



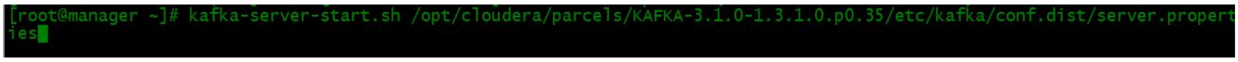
```
[root@manager ~]# source .bashrc
```

（4）：将配置好的.bashrc 文件分发给 Namenode、DataNode 主机

```
scp -r .bashrc namenode:$PWD
scp -r .bashrc datanode:$PWD
```

5 启动 Kafka 集群

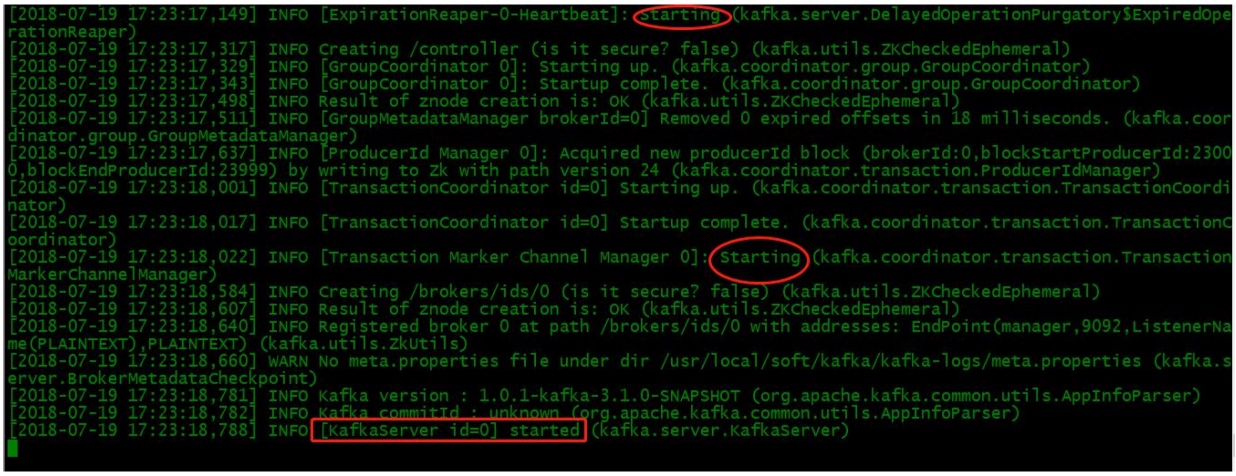
（1）启动命令：



```
[root@manager ~]# kafka-server-start.sh /opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/etc/kafka/conf.dist/server.propert
ies
```

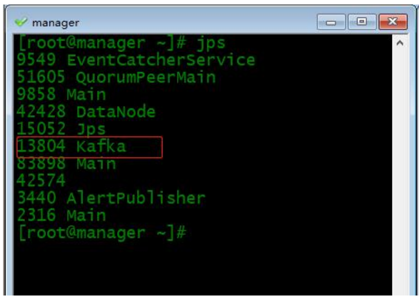
kafka-server-start.sh /opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/etc/kafka/conf.dist/server.properties

（2）启动成功：

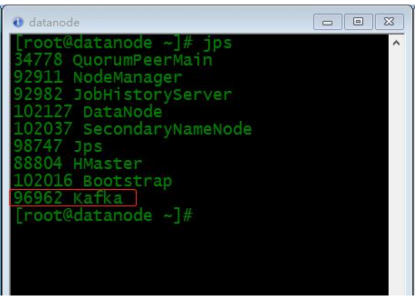


```
[2018-07-19 17:23:17,149] INFO [ExpirationReaper-0-Heartbeat]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOpe
rationReaper)
[2018-07-19 17:23:17,317] INFO Creating /controller (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2018-07-19 17:23:17,329] INFO [GroupCoordinator 0]: Starting up. (kafka.coordinator.group.GroupCoordinator)
[2018-07-19 17:23:17,343] INFO [GroupCoordinator 0]: Startup complete. (kafka.coordinator.group.GroupCoordinator)
[2018-07-19 17:23:17,498] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2018-07-19 17:23:17,511] INFO [GroupMetadataManager brokerId=0] Removed 0 expired offsets in 18 milliseconds. (kafka.coor
dinator.group.GroupMetadataManager)
[2018-07-19 17:23:17,637] INFO [ProducerId Manager 0]: Acquired new producerId block (brokerId:0,blockStartProducerId:2300
0,blockEndProducerId:23999) by writing to Zk with path version 24 (kafka.coordinator.transaction.ProducerIdManager)
[2018-07-19 17:23:18,001] INFO [TransactionCoordinator id=0] Starting up. (kafka.coordinator.transaction.TransactionCoordi
nator)
[2018-07-19 17:23:18,017] INFO [TransactionCoordinator id=0] Startup complete. (kafka.coordinator.transaction.TransactionC
oordinator)
[2018-07-19 17:23:18,022] INFO [Transaction Marker Channel Manager 0]: Starting (kafka.coordinator.transaction.Transaction
MarkerChannelManager)
[2018-07-19 17:23:18,584] INFO Creating /brokers/ids/0 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)
[2018-07-19 17:23:18,607] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)
[2018-07-19 17:23:18,640] INFO Registered broker 0 at path /brokers/ids/0 with addresses: EndPoint(manager,9092,ListenerNa
me(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)
[2018-07-19 17:23:18,660] WARN No meta.properties file under dir /usr/local/soft/kafka/kafka-logs/meta.properties (kafka.s
erver.BrokerMetadataCheckpoint)
[2018-07-19 17:23:18,781] INFO Kafka version : 1.0.1-kafka-3.1.0-SNAPSHOT (org.apache.kafka.common.utils.AppInfoParser)
[2018-07-19 17:23:18,782] INFO Kafka commitId : unknown (org.apache.kafka.common.utils.AppInfoParser)
[2018-07-19 17:23:18,788] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)
```

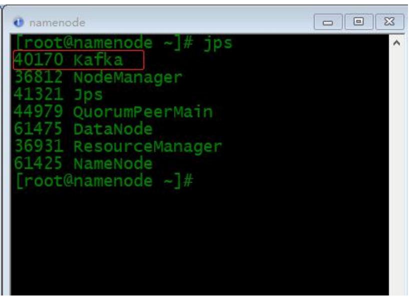
（3）使用 jps 命令查看



```
[root@manager ~]# jps
9549 EventCatcherService
51605 QuorumPeerMain
9858 Main
42428 DataNode
15052 Jps
13804 Kafka
83898 Main
42574
3440 AlertPublisher
2316 Main
[root@manager ~]#
```



```
[root@datanode ~]# jps
34778 QuorumPeerMain
92911 NodeManager
92982 JobHistoryServer
102127 DataNode
102037 SecondaryNameNode
98747 Jps
88804 HMaster
102016 Bootstrap
96962 Kafka
[root@datanode ~]#
```

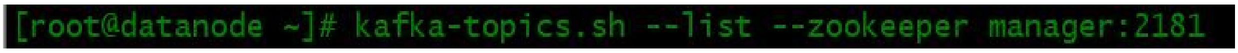


```
[root@namenode ~]# jps
40170 Kafka
36812 NodeManager
41321 Jps
44979 QuorumPeerMain
61475 DataNode
36931 ResourceManager
61425 NameNode
[root@namenode ~]#
```

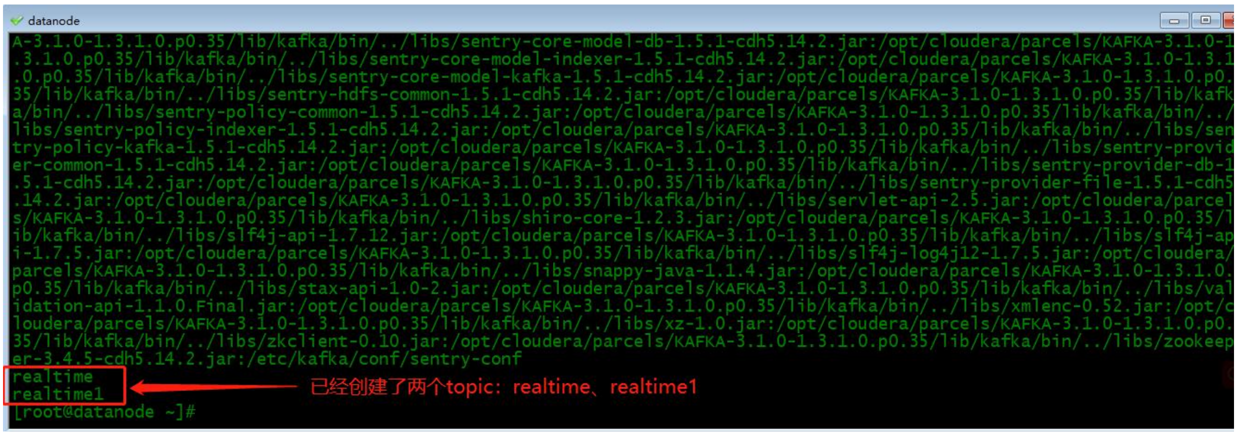
6 测试数据

（1）查看当前有哪些主题命令(此处选择 manager 主机的 zk)：

```
kafka-topics.sh --list --zookeeper manager:2181
```



```
[root@datanode ~]# kafka-topics.sh --list --zookeeper manager:2181
```

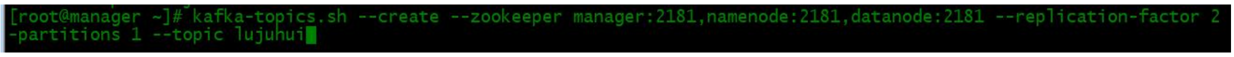


```
A-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-core-model-db-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1
.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-core-model-indexer-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1
.0.p0.35/lib/kafka/bin/../libs/sentry-core-model-kafka-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.
35/lib/kafka/bin/../libs/sentry-hdfs-common-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafk
a/bin/../libs/sentry-policy-common-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../
libs/sentry-policy-indexer-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sen
try-policy-kafka-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-provid
er-common-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-provider-db-1
.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-provider-file-1.5.1-cdh5
.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/servlet-api-2.5.jar:/opt/cloudera/parcel
s/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/shiro-core-1.2.3.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/l
ib/kafka/bin/../libs/slf4j-api-1.7.12.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-ap
i-1.7.5.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-log4j12-1.7.5.jar:/opt/cloudera/
parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/snappy-java-1.1.4.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.
p0.35/lib/kafka/bin/../libs/stax-api-1.0-2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/val
idation-api-1.1.0.Final.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xmlenc-0.52.jar:/opt/c
loudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xz-1.0.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.
35/lib/kafka/bin/../libs/zkclient-0.10.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/zookeep
er-3.4.5-cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
realtime
realtime1
[root@datanode ~]#
```

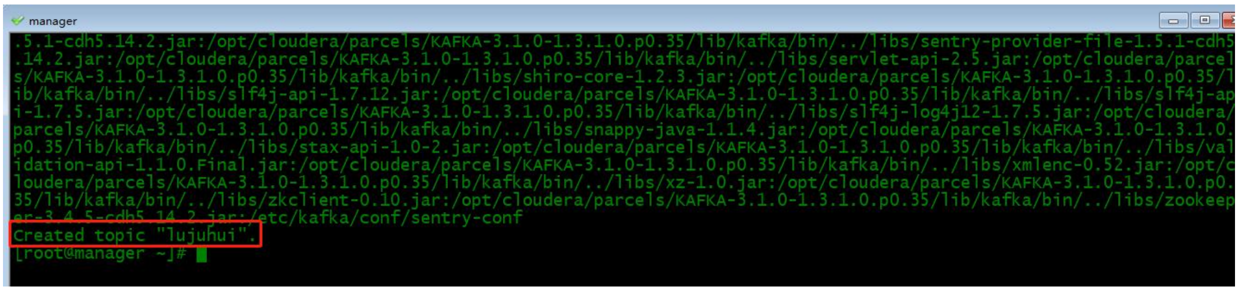
已经创建了两个topic：realtime、realtime1

（2）创建新的主题命令（设置为 3 个副本,1 个分区）：

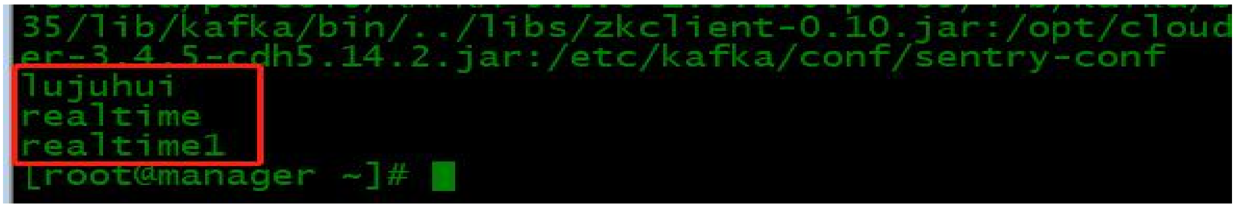
```
kafka-topics.sh --create --zookeeper manager:2181,namenode:2181,datanode:2181
--replication-factor 2 --partitions 1 --topic lujuhui
```



```
[root@manager ~]# kafka-topics.sh --create --zookeeper manager:2181,namenode:2181,datanode:2181 --replication-factor 2
-partitions 1 --topic lujuhui
```



```
.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-provider-file-1.5.1-cdh5
.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/servlet-api-2.5.jar:/opt/cloudera/parcel
s/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/shiro-core-1.2.3.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/l
ib/kafka/bin/../libs/slf4j-api-1.7.12.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-ap
i-1.7.5.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-log4j12-1.7.5.jar:/opt/cloudera/
parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/snappy-java-1.1.4.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.
p0.35/lib/kafka/bin/../libs/stax-api-1.0-2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/val
idation-api-1.1.0.Final.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xmlenc-0.52.jar:/opt/c
loudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xz-1.0.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.
35/lib/kafka/bin/../libs/zkclient-0.10.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/zookeep
er-3.4.5-cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
Created topic "lujuhui".
[root@manager ~]#
```



```
35/lib/kafka/bin/../libs/zkclient-0.10.jar:/opt/cloud
er-3.4.5-cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
lujuhui
realtime
realtime1
[root@manager ~]#
```

（3）在 manager 上模拟，往 kafka 的 lujuhui 主题里面发送数据，然后 datanode 上去消

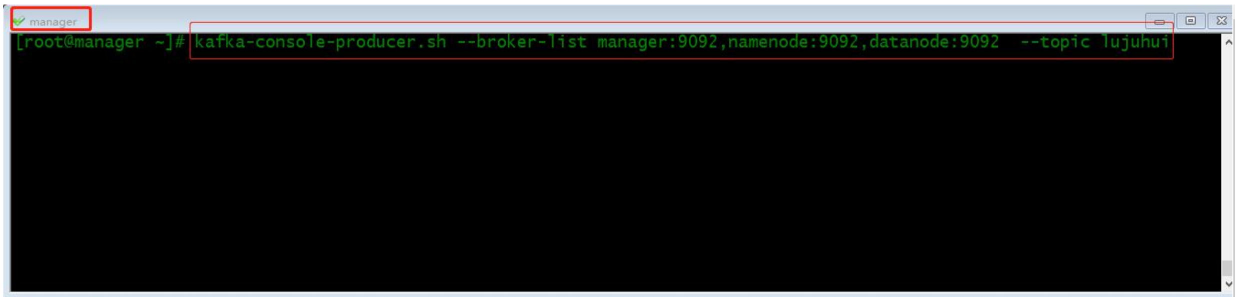
费这个数据:

Manager 发送数据命令:

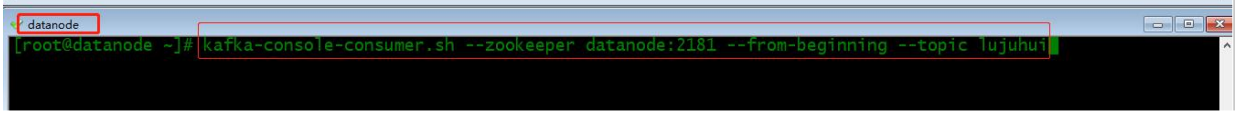
kafka-console-producer.sh --broker-list manager:9092,namenode:9092,datanode:9092 --topic lujuhui

Datanode 消费数据命令:

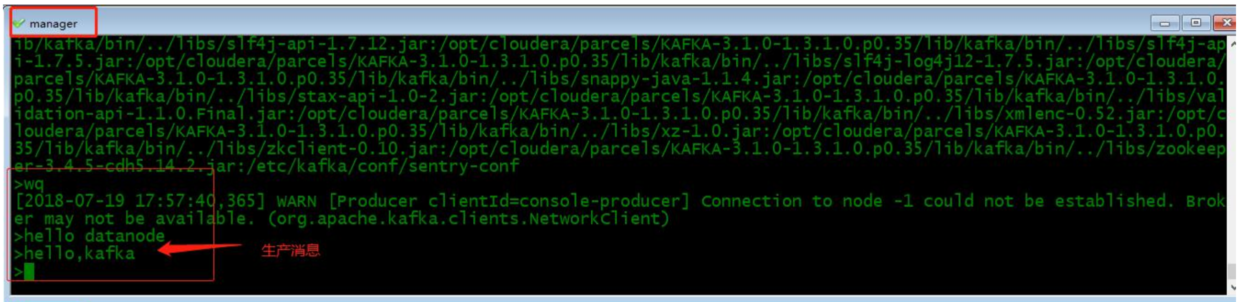
kafka-console-consumer.sh --zookeeper datanode:2181 --from-beginning --topic lujuhui



```
[root@manager ~]# kafka-console-producer.sh --broker-list manager:9092,namenode:9092,datanode:9092  --topic lujuhui
```

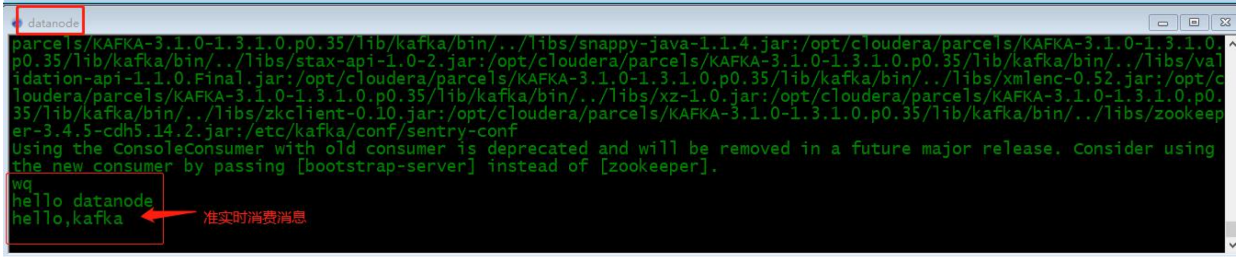


```
[root@datanode ~]# kafka-console-consumer.sh --zookeeper datanode:2181 --from-beginning --topic lujuhui
```



```
ib/kafka/bin/../libs/slf4j-api-1.7.12.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-ap
i-1.7.5.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-log4j12-1.7.5.jar:/opt/cloudera/
parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/snappy-java-1.1.4.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.
p0.35/lib/kafka/bin/../libs/stax-api-1.0-2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/val
idation-api-1.1.0.Final.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xmlenc-0.52.jar:/opt/c
loudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xz-1.0.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.
35/lib/kafka/bin/../libs/zkclient-0.10.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/zookeep
er-3.4.5-cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
>wq
[2018-07-19 17:57:40,365] WARN [Producer clientId=console-producer] Connection to node -1 could not be established. Brok
er may not be available. (org.apache.kafka.clients.NetworkClient)
>hello datanode
>hello,kafka
>
```

生产消息

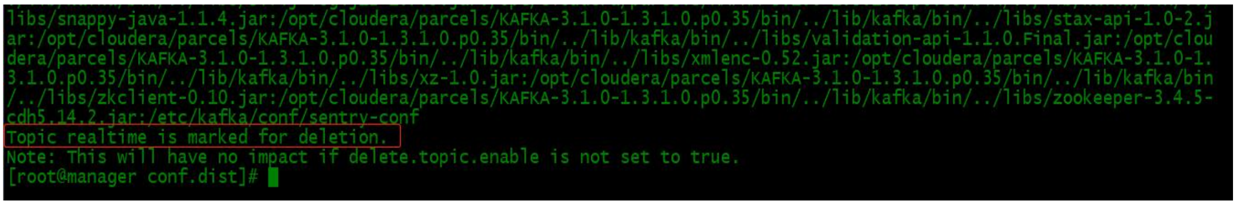


```
parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/snappy-java-1.1.4.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.
p0.35/lib/kafka/bin/../libs/stax-api-1.0-2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/val
idation-api-1.1.0.Final.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xmlenc-0.52.jar:/opt/c
loudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xz-1.0.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.
35/lib/kafka/bin/../libs/zkclient-0.10.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/zookeep
er-3.4.5-cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a future major release. Consider using
the new consumer by passing [bootstrap-server] instead of [zookeeper].
wq
hello datanode
hello,kafka
```

准实时消费消息

（4）删除主题命令:

kafka-topics --delete --topic realtime --zookeeper manager:2181



```
libs/snappy-java-1.1.4.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/bin/../lib/kafka/bin/../libs/stax-api-1.0-2.j
ar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/bin/../lib/kafka/bin/../libs/validation-api-1.1.0.Final.jar:/opt/clou
dera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/bin/../lib/kafka/bin/../libs/xmlenc-0.52.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.
3.1.0.p0.35/bin/../lib/kafka/bin/../libs/xz-1.0.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/bin/../lib/kafka/bin
/../libs/zkclient-0.10.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/bin/../lib/kafka/bin/../libs/zookeeper-3.4.5-
cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
Topic realtime is marked for deletion.
Note: This will have no impact if delete.topic.enable is not set to true.
[root@manager conf.dist]#
```

二 Flume

1 版本选择

CDH-3.7.5的组件版本

2 修改Flume相关配置文件

（1）首先，进入flume的配置目录：

/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/etc/flume-ng/conf.empty

（2）其次，新建一个文件：

flume-conf.properties

[注意：如果文件中有flume.conf,可以直接删除，新建的文件名字可以随意取，但是后缀必须是.properties]

（3）在flume-conf.properties文件中添加链接kafka的配置：

```
a1.sources = r1

a1.sinks = s1

a1.channels = c1


#sources 消息生产

a1.sources.r1.type = spooldir

a1.sources.r1.channels = c1

a1.sources.r1.spoolDir = /usr/local/soft/flume/flume_dir //用于存放收集的日志

a1.sources.r1.fileHeader = false

a1.sources.r1.interceptors = i1

a1.sources.r1.interceptors.i1.type = timestamp


#channels 消息传递

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000
```

```
a1.channels.c1.transactionCapacity = 100


#sinks 消息消费
a1.sinks.s1.type = org.apache.flume.sink.kafka.KafkaSink
a1.sinks.s1.brokerList = manager:9092,namenode:9092,datanode:9092 //链接 kafka
a1.sinks.s1.topics = lujuhui//flume 收集的日志分发给 kafka 的对应主题名称
a1.sinks.s1.requiredAcks = 1
a1.sinks.s1.batchSize = 20
a1.sinks.s1.channel = c1  //注意这里是 channel 不是 channels
```

3 配置全局变量

(1) 回到根目录输入命令：

```
vi .bashrc
```

(2) 添加配置：

```
export FLUME_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/flume-ng
export PATH=$PATH:$FLUME_HOME/bin
```

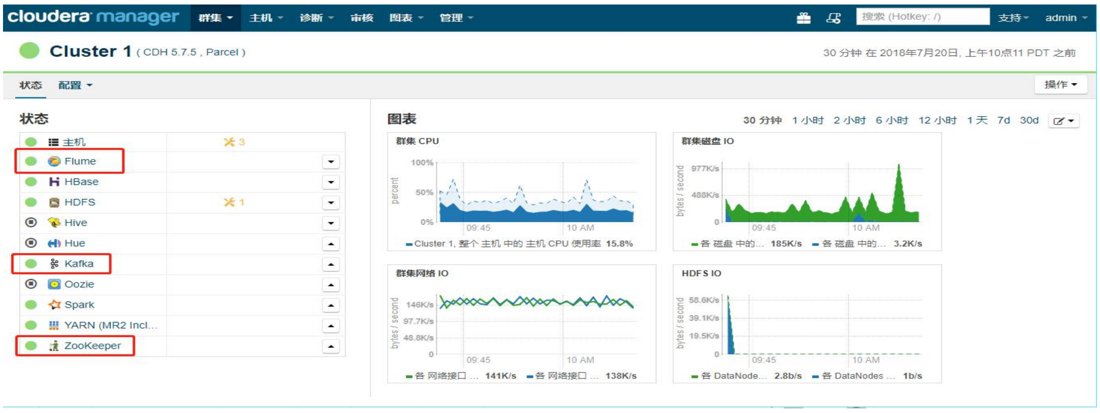
(3) 保存 --> 退出 --> 更新

```
source .bashrc
```

4 启动 Flume

（1）分别启动 zookeeper 集群、kafka 集群：

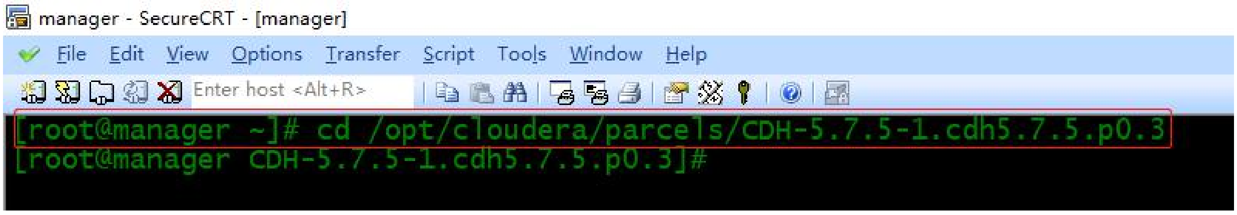
因为是在 Cloudera Manager 上直接添加的服务组件，所以可以直接在 CM 上启动。



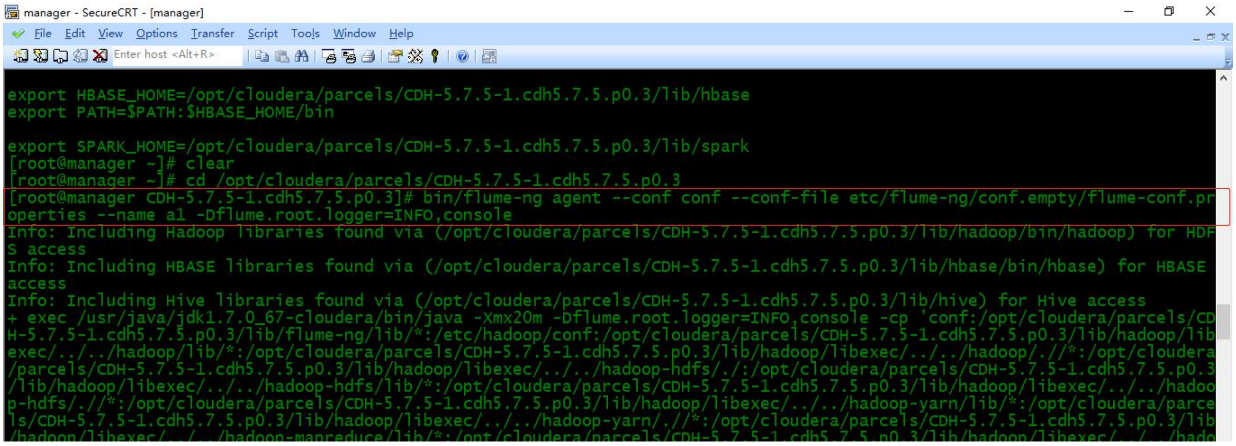
（2）因为是在主机 manager 上配置的 flume，因此启动 manager 主机上的 flume，启动命令如下：

1）cd /opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3

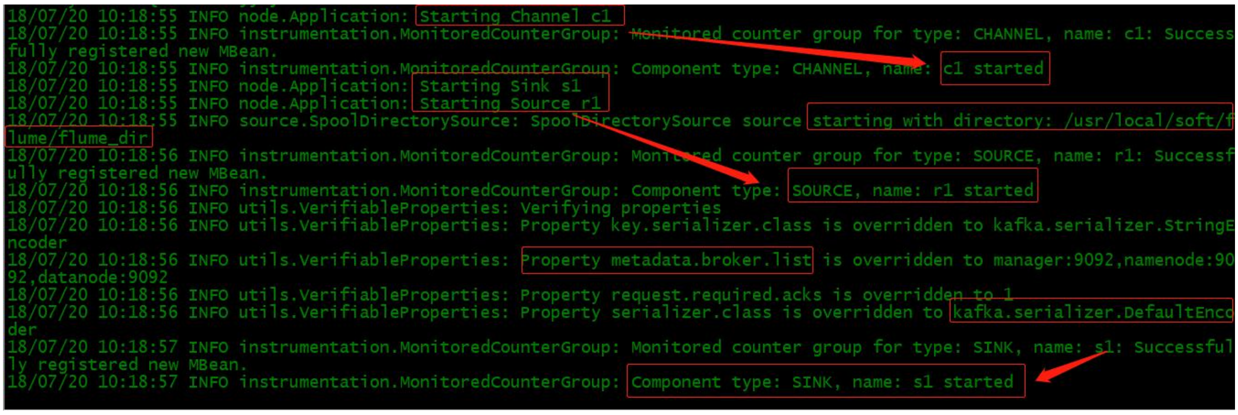
2)bin/flume-ng agent --conf conf --conf-file etc/flume-ng/conf.empty/flume-conf.properties --name a1 -Dflume.root.logger=INFO,console


manager - SecureCRT - [manager]
File Edit View Options Transfer Script Tools Window Help
Enter host <Alt+R>

```
[root@manager ~]# cd /opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3
[root@manager CDH-5.7.5-1.cdh5.7.5.p0.3]#
```


manager - SecureCRT - [manager]
File Edit View Options Transfer Script Tools Window Help
Enter host <Alt+R>

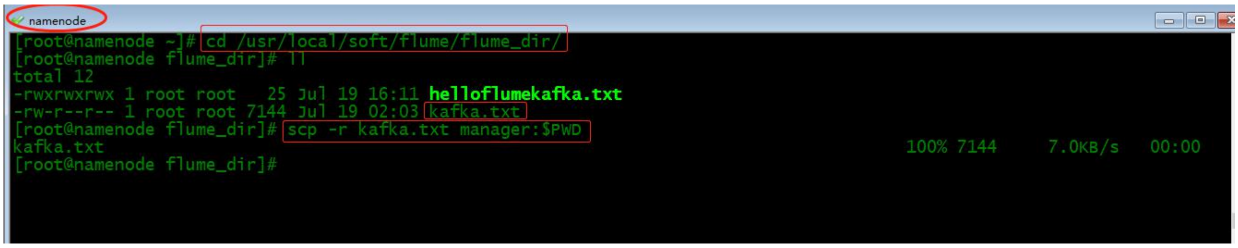
```
export HBASE_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hbase
export PATH=$PATH:$HBASE_HOME/bin

export SPARK_HOME=/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/spark
[root@manager ~]# clear
[root@manager ~]# cd /opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3
[root@manager CDH-5.7.5-1.cdh5.7.5.p0.3]# bin/flume-ng agent --conf conf --conf-file etc/flume-ng/conf.empty/flume-conf.pr
operties --name a1 -Dflume.root.logger=INFO,console
Info: Including Hadoop libraries found via (/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/bin/hadoop) for HDF
S access
Info: Including HBASE libraries found via (/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hbase/bin/hbase) for HBASE
access
Info: Including Hive libraries found via (/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hive) for Hive access
+ exec /usr/java/jdk1.7.0_67-cloudera/bin/java -Xmx20m -Dflume.root.logger=INFO,console -cp 'conf:/opt/cloudera/parcels/CD
H-5.7.5-1.cdh5.7.5.p0.3/lib/flume-ng/lib/*:/etc/hadoop/conf:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/lib
exec/../../hadoop/lib/*:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/libexec/../../hadoop/.//*:/opt/cloudera
/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/libexec/../../hadoop-hdfs/./:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3
/lib/hadoop/libexec/../../hadoop-hdfs/lib/*:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/libexec/../../hadoo
p-hdfs/.//*:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/libexec/../../hadoop-yarn/lib/*:/opt/cloudera/parce
ls/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/libexec/../../hadoop-yarn/.//*:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib
/hadoop/libexec/../../hadoop-mapreduce/lib/*:/opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3/lib/hadoop/libexec/../../hado
```



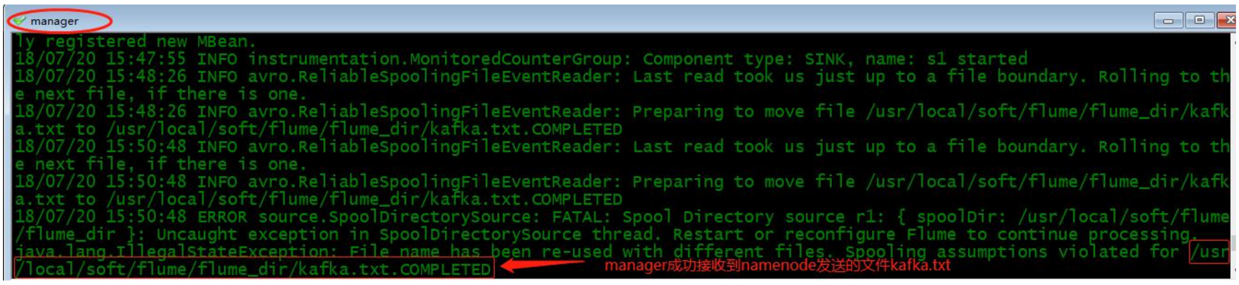
```
18/07/20 10:18:55 INFO node.Application: Starting Channel c1
18/07/20 10:18:55 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: c1: Success
fully registered new MBean.
18/07/20 10:18:55 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: c1 started
18/07/20 10:18:55 INFO node.Application: Starting Sink s1
18/07/20 10:18:55 INFO node.Application: Starting Source r1
18/07/20 10:18:55 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /usr/local/soft/f
lume/flume_dir
18/07/20 10:18:56 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: r1: Successf
ully registered new MBean.
18/07/20 10:18:56 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: r1 started
18/07/20 10:18:56 INFO utils.VerifiableProperties: Verifying properties
18/07/20 10:18:56 INFO utils.VerifiableProperties: Property key.serializer.class is overridden to kafka.serializer.StringE
ncoder
18/07/20 10:18:56 INFO utils.VerifiableProperties: Property metadata.broker.list is overridden to manager:9092,namenode:90
92,datanode:9092
18/07/20 10:18:56 INFO utils.VerifiableProperties: Property request.required.acks is overridden to 1
18/07/20 10:18:56 INFO utils.VerifiableProperties: Property serializer.class is overridden to kafka.serializer.DefaultEnco
der
18/07/20 10:18:57 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: s1: Successful
ly registered new MBean.
18/07/20 10:18:57 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: s1 started
```

5 测试数据

(1) 在 namenode 主机上添加一个 kafka.txt 文件，然后发送到 manager 主机上的/usr/local/soft/flume/flume_dir 目录中，这时可以发现 manager 主机上发生如下变化：

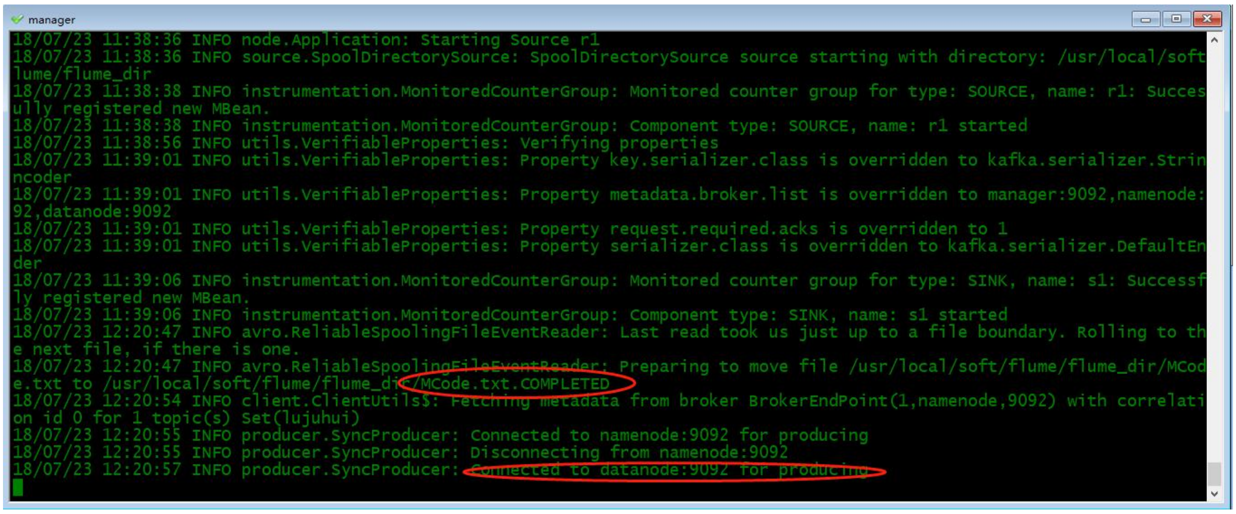


```
[root@namenode ~]# cd /usr/local/soft/flume/flume_dir/
[root@namenode flume_dir]# ll
total 12
-rwxrwxrwx 1 root root   25 Jul 19 16:11 helloflumekafka.txt
-rw-r--r-- 1 root root 7144 Jul 19 02:03 kafka.txt
[root@namenode flume_dir]# scp -r kafka.txt manager:$PWD
kafka.txt                                                            100% 7144     7.0KB/s   00:00
[root@namenode flume_dir]#
```



```
ly registered new MBean.
18/07/20 15:47:55 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: s1 started
18/07/20 15:48:26 INFO avro.ReliableSpoolingFileEventReader: Last read took us just up to a file boundary. Rolling to th
e next file, if there is one.
18/07/20 15:48:26 INFO avro.ReliableSpoolingFileEventReader: Preparing to move file /usr/local/soft/flume/flume_dir/kafk
a.txt to /usr/local/soft/flume/flume_dir/kafka.txt.COMPLETED
18/07/20 15:50:48 INFO avro.ReliableSpoolingFileEventReader: Last read took us just up to a file boundary. Rolling to th
e next file, if there is one.
18/07/20 15:50:48 INFO avro.ReliableSpoolingFileEventReader: Preparing to move file /usr/local/soft/flume/flume_dir/kafk
a.txt to /usr/local/soft/flume/flume_dir/kafka.txt.COMPLETED
18/07/20 15:50:48 ERROR source.SpoolDirectorySource: FATAL: Spool Directory source r1: { spoolDir: /usr/local/soft/flume
/flume_dir }: Uncaught exception in SpoolDirectorySource thread. Restart or reconfigure Flume to continue processing.
java.lang.IllegalStateException: File name has been re-used with different files. Spooling assumptions violated for /usr
/local/soft/flume/flume_dir/kafka.txt.COMPLETED
```

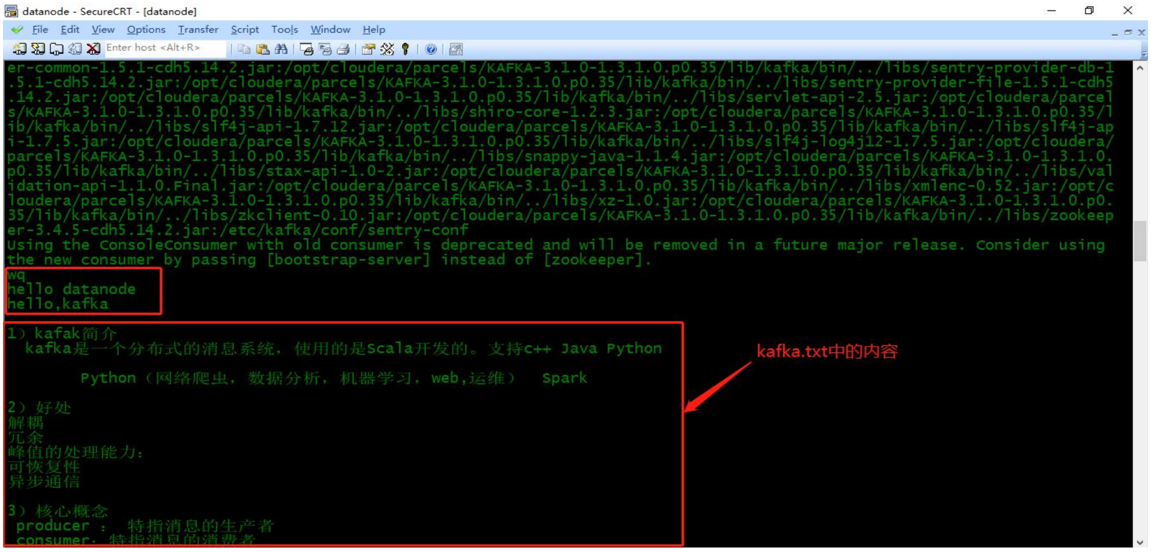
manager成功接收到namenode发送的文件kafka.txt



```
18/07/23 11:38:36 INFO node.Application: Starting Source r1
18/07/23 11:38:36 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /usr/local/soft
lume/flume_dir
18/07/23 11:38:38 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: r1: Succes
ully registered new MBean.
18/07/23 11:38:38 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: r1 started
18/07/23 11:38:56 INFO utils.VerifiableProperties: Verifying properties
18/07/23 11:39:01 INFO utils.VerifiableProperties: Property key.serializer.class is overridden to kafka.serializer.Strin
ncoder
18/07/23 11:39:01 INFO utils.VerifiableProperties: Property metadata.broker.list is overridden to manager:9092,namenode:
92,datanode:9092
18/07/23 11:39:01 INFO utils.VerifiableProperties: Property request.required.acks is overridden to 1
18/07/23 11:39:01 INFO utils.VerifiableProperties: Property serializer.class is overridden to kafka.serializer.DefaultEn
der
18/07/23 11:39:06 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: s1: Successf
ly registered new MBean.
18/07/23 11:39:06 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: s1 started
18/07/23 12:20:47 INFO avro.ReliableSpoolingFileEventReader: Last read took us just up to a file boundary. Rolling to th
e next file, if there is one.
18/07/23 12:20:47 INFO avro.ReliableSpoolingFileEventReader: Preparing to move file /usr/local/soft/flume/flume_dir/MCod
e.txt to /usr/local/soft/flume_dir/MCode.txt.COMPLETED
18/07/23 12:20:54 INFO client.ClientUtils$: Fetching metadata from broker BrokerEndPoint(1,namenode,9092) with correlati
on id 0 for 1 topic(s) Set(lujuhui)
18/07/23 12:20:55 INFO producer.SyncProducer: Connected to namenode:9092 for producing
18/07/23 12:20:55 INFO producer.SyncProducer: Disconnecting from namenode:9092
18/07/23 12:20:57 INFO producer.SyncProducer: Connected to datanode:9092 for producing
```

（2）随机选择 dataNode 主机，查看从 manager 主机上的 flume 传过来的数据：

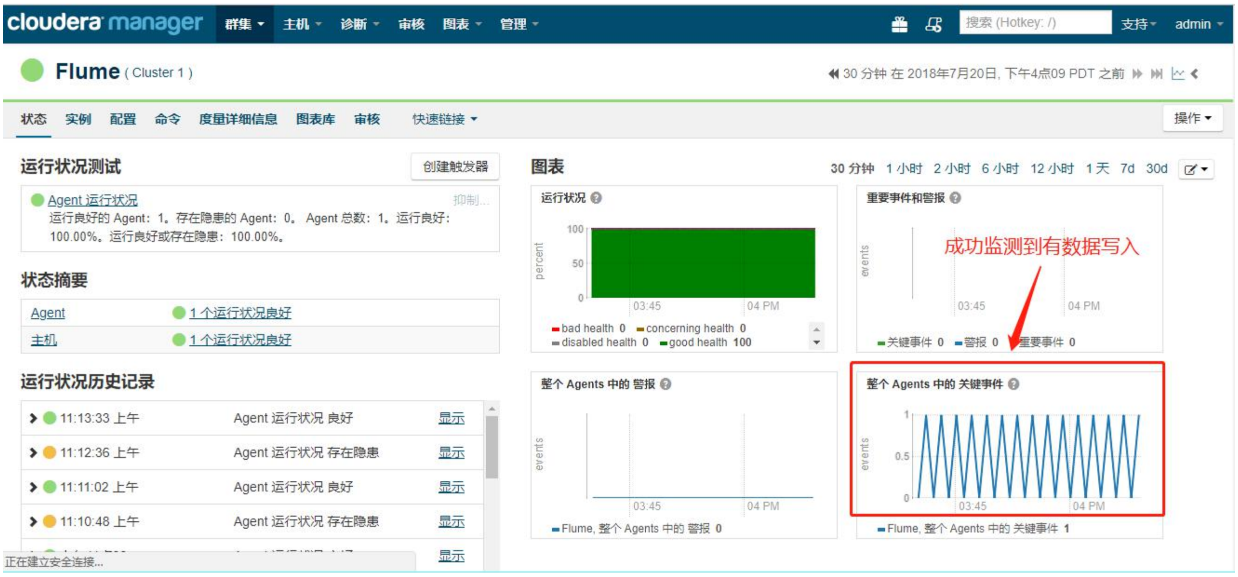
1）kafka-console-consumer.sh --zookeeper manager:2181,namenode:2181,datanode:2181 --from-beginning --topic lujuhui



```
er-common-1.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-provider-db-1
.5.1-cdh5.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/sentry-provider-file-1.5.1-cdh5
.14.2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/servlet-api-2.5.jar:/opt/cloudera/parcel
s/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/shiro-core-1.2.3.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/l
ib/kafka/bin/../libs/slf4j-api-1.7.12.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-ap
i-1.7.5.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/slf4j-log4j12-1.7.5.jar:/opt/cloudera/
parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/snappy-java-1.1.4.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.
p0.35/lib/kafka/bin/../libs/stax-api-1.0-2.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/val
idation-api-1.1.0.Final.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xmlenc-0.52.jar:/opt/c
loudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/xz-1.0.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.
35/lib/kafka/bin/../libs/zkclient-0.10.jar:/opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin/../libs/zookeep
er-3.4.5-cdh5.14.2.jar:/etc/kafka/conf/sentry-conf
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a future major release. Consider using
the new consumer by passing [bootstrap-server] instead of [zookeeper].
wq
hello datanode
hello,kafka

1）kafak简介
  kafka是一个分布式的消息系统，使用的是Scala开发的。支持c++ Java Python

        Python（网络爬虫，数据分析，机器学习，web,运维）   Spark

2）好处
解耦
冗余
峰值的处理能力：
可恢复性
异步通信

3）核心概念
 producer ： 特指消息的生产者
 consumer： 特指消息的消费者
```

kafka.txt中的内容

（3）Cloudera Manager 上的监控状态



运行状况测试

创建触发器

● Agent 运行状况　抑制...

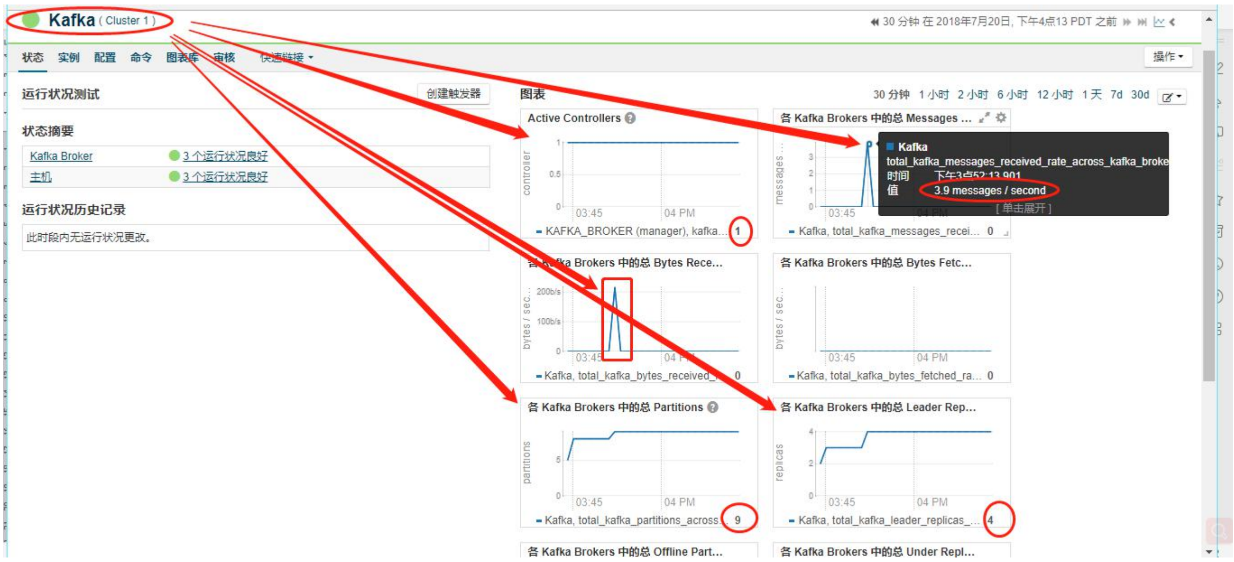
运行良好的 Agent：1。存在隐患的 Agent：0。 Agent 总数：1。运行良好：100.00%。运行良好或存在隐患：100.00%。

状态摘要

Agent	● 1 个运行状况良好
主机	● 1 个运行状况良好

运行状况历史记录

● 11:13:33 上午	Agent 运行状况 良好	显示
● 11:12:36 上午	Agent 运行状况 存在隐患	显示
● 11:11:02 上午	Agent 运行状况 良好	显示
● 11:10:48 上午	Agent 运行状况 存在隐患	显示


Kafka (Cluster 1)
30 分钟 在 2018年7月20日, 下午4点13 PDT 之前
状态
实例
配置
命令
图表库
审核
快速链接
操作
图表
30 分钟
1 小时
2 小时
6 小时
12 小时
1 天
7d
30d
Active Controllers
controller
1
0.5
0
03:45
04 PM
KAFKA_BROKER (manager), kafka... 1
各 Kafka Brokers 中的总 Messages ...
messages ...
Kafka
total_kafka_messages_received_rate_across_kafka_broke
时间 下午3点52:13.901
值 3.9 messages / second
[单击展开]
Kafka, total_kafka_messages_recei... 0
各 Kafka Brokers 中的总 Bytes Rece...
bytes / sec...
200b/s
100b/s
Kafka, total_kafka_bytes_received_... 0
各 Kafka Brokers 中的总 Bytes Fetc...
Kafka, total_kafka_bytes_fetched_ra... 0
各 Kafka Brokers 中的总 Partitions
partitions
5
Kafka, total_kafka_partitions_across... 9
各 Kafka Brokers 中的总 Leader Rep...
replicas
4
2
Kafka, total_kafka_leader_replicas_... 4
各 Kafka Brokers 中的总 Offline Part...
各 Kafka Brokers 中的总 Under Repl...

运行状况测试

创建触发器

状态摘要

Kafka Broker	● 3 个运行状况良好
主机	● 3 个运行状况良好

运行状况历史记录

此时段内无运行状况更改。

三 Flume + Kafka 测试启动命令

1 Flume 启动测试命令

```
1）cd /opt/cloudera/parcels/CDH-5.7.5-1.cdh5.7.5.p0.3
2) bin/flume-ng agent --conf conf
--conf-file etc/flume-ng/conf.empty/flume-conf.properties --name
a1 -Dflume.root.logger=INFO,console
```

2 Kafaka 启动测试命令

```
1）cd /opt/cloudera/parcels/KAFKA-3.1.0-1.3.1.0.p0.35/lib/kafka/bin
2)kafka-console-consumer.sh --zookeeper
manager:2181,namenode:2181,datanode:2181 --from-beginning --topic lujuhui
```