

## 目录

目录	1
云主机连接kafka集群	2
kafka manager使用指南	2
Kafka manager简介	2
基本功能	2
基本操作	3
集群管理	3
Topic管理	6
Broker管理	8
consumer管理	9

# 云主机连接kafka集群

创建单独的kafka集群后，往往需要通过其他云主机连接kafka集群，需要遵循以下步骤：

- 确保云主机可以访问KMR集群，参见[VPC网络设置](#)；
- 确保kafka集群的安全组中打开了kafka端口的入站规则，默认为6667；
- kafka broker hosts默认使用主机名而不是ip，因此跨集群默认情况下会出现消费者（consumer）无法联接broker的情况，为了保证消费者（consumer）正确联接broker，需要在消费者（consumer）所在机器的host文件中，加入kafka集群相关主机名与ip的映射：
  - 登入kafka集群的master-1主节点，`cat /etc/hosts`，打印出集群相关ip与host；
  - 登入消费者（consumer）所在机器，打开hosts文件 `vim /etc/hosts`，将kafka集群主节点中的ip及host填入hosts中。

测试云主机与kafka集群中各broker的联通性，可以使用如下的python脚本(需依赖python kafka包)：

```
import json
from kafka import KafkaProducer
port = '6667' # kafka监听的端口，默认为6667
for hn in ['x.x.x.x',
          'x.x.x.x',
          'x.x.x.x',
          'x.x.x.x']: # 填入broker的ip地址

    bootstrap_servers = '%s:%s' % (hn, port)
    try:
        producer = KafkaProducer(bootstrap_servers=bootstrap_servers)
        producer.send('test', b'some_message_bytes')
        print('\n=====done=====')
        print(bootstrap_servers)

    except Exception as e:
        print('\n*****')
        print(bootstrap_servers)
        print("Error:%s" % e)
```

## kafka manager使用指南

### Kafka manager简介

kafka Manager是由yahoo开源的一个Kafka管理工具，地址为：<https://github.com/yahoo/kafka-manager>

### 基本功能

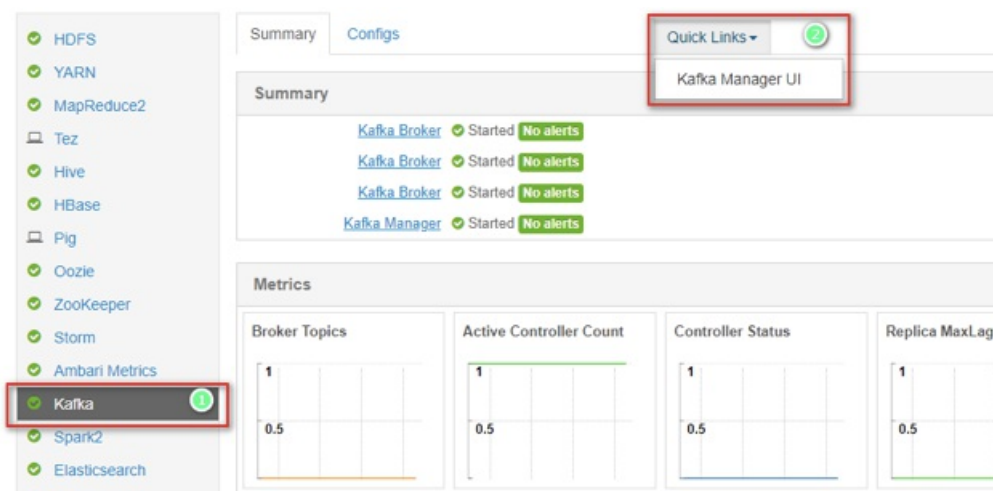
- 管理kafka集群
- 方便集群状态监控（包括topics, consumers, offsets, brokers, replica distribution, partition distribution）
- 方便选择分区副本
- 配置分区任务，包括选择使用哪些brokers
- 可以对分区任务重分配
- 提供不同的选项来创建及删除topic
- Topic list会指明哪些topic被删除
- 批量产生分区任务并且和多个topic和brokers关联
- 批量运行多个主题对应的多个分区

- 向已经存在的主题中添加分区
- 对已经存在的topic修改配置
- 可以在broker level和topic level的度量中启用JMX polling功能
- 可以过滤在ZK上没有ids/ owners/offsets/ directories的consumer

## 基本操作

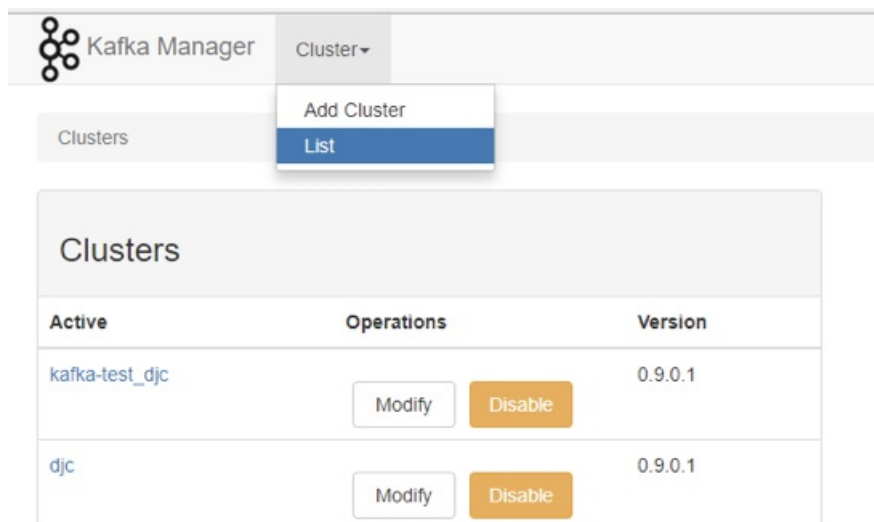
- 进入kafka manager web界面

登录Ambari控制台，在左侧选择Kafka后，在页面中部点击Quick Links，然后点击Kafka Manager UI进入其web页面。



- Kafka Manager 首页

在页面首页展示集群信息，可在顶端Cluster下拉菜单中进行集群创建及集群列表展示；在集群列表栏目可对集群进行修改及停止（Disable）/启动（Enable）。




## 集群管理

- 新建集群

依次点击顶部Cluster下拉菜单--> Add Cluster进入集群新建页面；依次填入名称、zookeeper hosts，选择kafka版本；最后点击最下边的save即可创建集群。

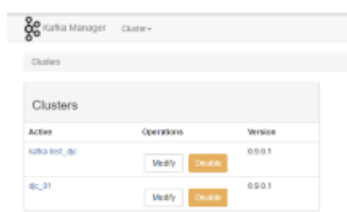
新建时各参数说明:

参数	说明
Cluster Name	用户自定义
Cluster Zookeeper Hosts	需与zookeeper.connect（在Amabri Service 页面中，选择Kafka服务，在zookeeper.connect中进行选择）中相对应，其可有一个或多个，注意后加:2181(冒号包含在内),若有多个host，各host间用逗号隔开 
Kafka Version	选择与Kafka版本最近的即可
Enable JMX Polling	JMX(Java Management Extensions, Java管理扩展)，通常使用JMX来监控系统的运行状态或管理系统的某些方面，比如清空缓存、重新加载配置文件等，建议选择
JMX Auth Username	JMX授权用户名
JMX Auth Password	JMX授权密码
JMX with SSL	为JMX使用SSL协议
Enable Logkafka	logkafka是一个日志收集代理，可以按行收集日志文件并发送到kafka
Poll consumer information	选举消费者信息（消费者多时不建议选择）
Filter out inactive consumers	过滤不活动消费者
Enable Active OffsetCache	激活Offset缓存
Display Broker and Topic Size	展示Broker和Topic大小

brokerViewUpdatePeriodSeconds	Broker视图周期更新时间/单位 (s)
clusterManagerThreadPoolSize	集群管理线程池大小
clusterManagerThreadPoolQueueSize	集群管理线程池队列大小
KafkaCommandThreadPoolSize	Kafka命令线程池大小
logkafkaCommandThreadPoolQueueSize	logkafka命令线程池队列大小
logkafkaUpdatePeriodSeconds	Logkafka周期更新时间/单位 (s)
partitionOffsetCacheTimeoutSeconds	Partition Offset缓存过期时间/单位 (s)
brokerViewThreadPoolSize	Broker视图线程池大小
brokerViewThreadPoolQueueSize	Broker视图线程池队列大小
offsetCacheThreadPoolSize	Offset缓存线程池大小
offsetCacheThreadPoolQueueSize	Offset缓存线程池队列大小
kafkaAdminClientThreadPoolSize	Kafka管理客户端线程池大小
kafkaAdminClientThreadPoolQueueSize	Kafka管理客户端线程池队列大小

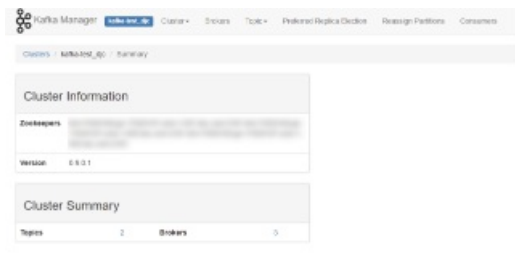
根据自己情况，填写对应参数，单击save，即可完成集群创建；

创建好的集群可在集群列表中展示；



- 查看集群信息

在集群列表中点击对应集群名称，可进入集群详情页面，在此页面上可查看集群版本，集群topic及broker信息；点击Cluster Summary中Topics和Brokers后的数字，可查看对应topic和broker信息；



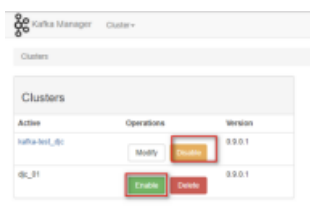
- 修改集群

在集群名称后点击Modify，即可对集群进行修改；除集群名称不可更改外，其他项（见新建集群）均可更改。

- 禁用/启用集群

在集群列表单击相应集群名称后的Disable按钮，可禁用集群

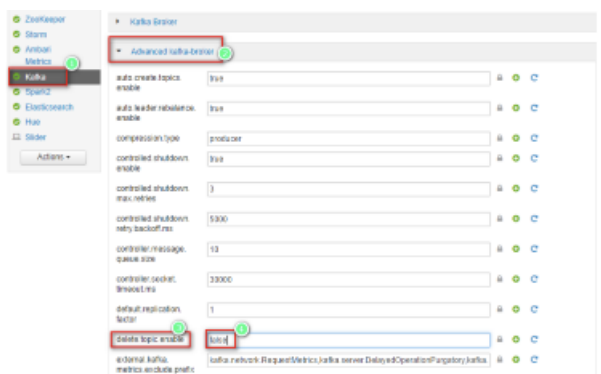
在集群列表单击相应集群名称后的Enable按钮，可启用集群



- 删除集群

在集群列表单击相应集群名称后的Delete按钮，可删除集群

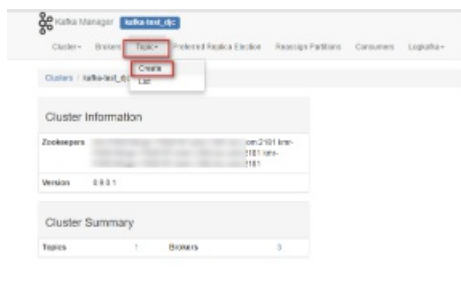
注：所删除集群需为禁用集群，需在Ambari Service界面中选择kafka，在config配置中将Advanced kafka-broker中的delete.topic.enable改为true



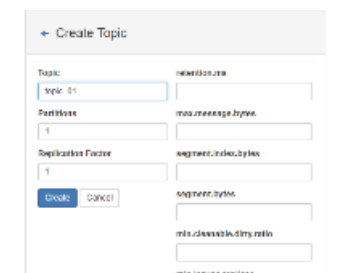
## Topic管理

- 新建topic

在集群列表单击相应集群名称后，单击上方Topic下拉菜单，单击Create



填写相应参数，单击Create，可对topic进行创建



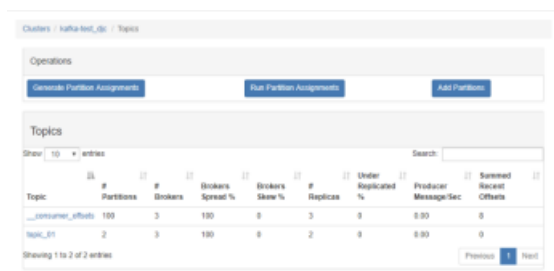
各参数说明如下：

参数	说明
Topic	Topic名称，用户自定义即可
Partitions	Topic的分区数， $\geq 1$ ，适当的分区数可以提高吞吐量
Replication Factor	副本数，用于保障kafka的高可用

其余参数可根据需要进行自定义

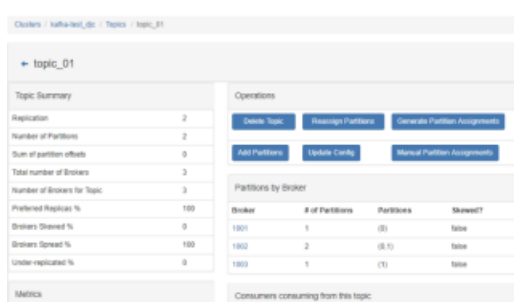
- 查看topic

在集群列表中点击相应集群名称，然后点击Cluster Summary下Topics后数字，可查看此集群的topic列表。



可查看topic名称、分区数、broker数、副本数等相关信息。

点击topic名称，可查看具体信息。

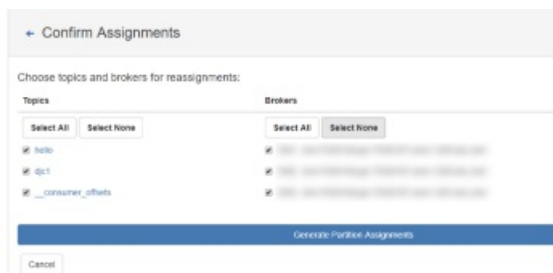


- 分区管理

## 重新分区

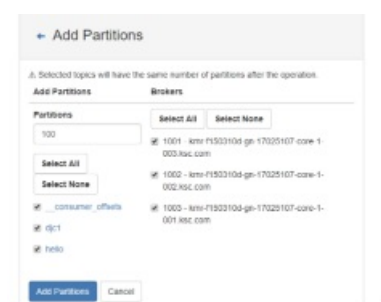
在点击Generate Partition Assignments可重新进行分区；

选择相应的Topics和Brokers，再点击下方的Generate Partition Assignment可完成重新分区；



## 增加分区

在topic列表上方单击Add Partitions，可对一个或多个topic进行增加分区操作；分区数必须大于原始topics中分区最大数。



单击topic名称进入topic详情页面，单击Add Partitions可对此topic进行增加分区操作；注意partitions中数目为分区后的总数，其应大于现有分区数。

## 分区详情页面操作

操作	解释
Delete Topic	删除topic
Reassign Partitions	平衡集群负载，可为assigned replica中选举一个新的leader，还可改变partition中的assigned replica
Add Partitions	增加分区
Update Config	Topic配置信息更新
Manual Partition Assignments	手动为每个分区下的副本分配broker，如下如，完成后点击Save Partition Assignment即可
Create Partition Assignments	系统自动为每个分区下的副本分配broker



## Broker管理

在每个集群信息页面单击Brokers后边的数字，可查看集群id、Host、端口等信息。



ID	Host	Part	Dir	Dir	Dir	Dir	Dir	Dir
		Path	Path	Path	Path	Path	Path	Path
1001	...	...	...	...	...	...	...	...
1002	...	...	...	...	...	...	...	...
1003	...	...	...	...	...	...	...	...

Metric	Mean	1 min	5 min	15 min
Messages in/sec	0.00	0.00	0.00	0.00
Bytes in/sec	0.00	0.00	0.00	0.00
Bytes out/sec	0.00	0.00	0.00	0.00
Bytes rejected/sec	0.00	0.00	0.00	0.00
Failed fetch request/sec	0.00	0.00	0.00	0.00
Failed produce request/sec	0.00	0.00	0.00	0.00

单击id栏相应id号，可查看其详细信息及所对应的topic等信息。

**Summary**

- # of Topics: 3
- # of Partitions: 106
- % of Messages: 0
- % of Incoming: 0
- % of Outgoing: 0

**Messages count**

**Per Topic Detail**

Topic	Replication	Total Partitions	Partitions on Broker	Partitions	Skewed?
__consumer_offsets	3	106	106	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99)	False
topic_01	2	5	2	(9, 2)	False
topic_02	3	4	3	(0, 1, 2, 3)	False

## consumer管理

- consumer组列表

在集群中勾选Poll consumer information (Not recommended for large # of consumers)，则在集群详情页面单击 Consumers可查看consumer相关信息。

**Consumers**

Show: 10 entries

Consumer	Type	Topics it consumes from
No data available in table		

Showing 0 to 0 of 0 entries

- topic 列表

点击consumer group列表的特定组，会显示这个consumer group中在消耗的topic列表，包括分区覆盖率，和总延迟量 (log)

← flume\_hdfs

Consumed Topic Information

Topic	Partitions Covered %	Total Lag
log	100	not available

- consumer列表

在consumer group > topic列表的点击具体topic，会进入到在消耗这个topic的具体consumer，显示的数据包括，总数据量，消费进度（offset），延迟量（log）等。

← flume\_hdfs / log

Topic Summary

Total Lag	0
% of Partitions assigned to a consumer instance	100

log

Partition	LogSize	Consumer Offset	Lag	Consumer Instance Owner
0	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
1	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
2	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
3	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
4	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
5	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
6	6144617	6144617	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0
7	6144616	6144616	0	flume_hdfs_kmr-1708d67e-gn-020aaadd-master-1-001.ksc.com-1535020474760-6a2579cd-0