

# 快学Spark 2.0

## 第31讲、Kafka集群搭建和测试

讲师：Cloudy（北风网版权所有）

答疑QQ: 2510337544

# CM里添加Kafka

低版本的cdh里不包含Kafka，新版本里才有。

<input type="radio"/>	 Hue	Hue 是与包括 Apache Hadoop 的 Cloudera Distribution 一起配合使用的图形用户界面（需要 HDFS、MapReduce）。
<input type="radio"/>	 Impala	Impala 为存储在 HDFS 和 HBase 中的数据提供了一个实时 SQL 查询接口。Impala 需要 Hive 服务，并与 Hue 共享元数据。
<input type="radio"/>	 Isilon	EMC Isilon is a distributed filesystem.
<input type="radio"/>	 Java KeyStore KMS	The Hadoop Key Management Service with file-based Java KeyStore. Maintains a single copy of keys, using similar to HBase. Requires CDH 5.3+. <b>Not recommended for production use.</b>
<input type="radio"/>	 Kafka	Apache Kafka is publish-subscribe messaging rethought as a distributed commit log. <b>Before adding this service, ensure that the Kafka package is installed.</b>

# 集群搭建步骤

## 1. Kafka下载:

如何选择兼容的Kafka版本? 基本都兼容, 不宜用高版本的。

基于scala 2.11的版本均可, 如kafka\_2.11-0.8.2.1, 我们继续采用这个版本即可。

```
wget http://apache.fayea.com/kafka/0.8.2.1/kafka_2.11-0.8.2.1.tgz
```

```
解压 tar zxvf kafka_2.11-0.8.2.1.tgz
```

kafka使用scala编写, 需要下载scala相关的库。

2.11是Scala版本, 0.8.2.1是Kafka版本

## 2. 下载安装sbt:

下载地址: <http://dl.bintray.com/sbt/rpm/>

选择sbt-0.13.1.rpm 版本及以上

安装yum install -y sbt-0.13.1.rpm 或 rpm -ivh sbt-0.13.1.rpm

# 集群搭建步骤

3. 更新scala环境:

```
cd kafka_2.11-0.8.2.1
```

```
sbt update
```

```
sbt package
```

更新比较慢，需要等几分钟，可以先配置第4步。

4. 配置config/server.properties

**broker.id**为依次增长的：0、1、2、3、4，集群中唯一id

**log.dirs**设置到大硬盘路径下

**num.network.threads**

**num.partitions** ，默认分区数

**num.io.threads** 建议值为机器的核数；

**zookeeper.connect** 设置为zookeeper Servers 列表，各节点以逗号分开；

# Kafka 启动、创建topic

在kafka的部署目录下，**在各个节点上**通过如下命令来启动：

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

```
[root@master ~]# jps  
3098 ZooKeeperMain  
3037 Kafka
```

## 创建topic:

```
$ bin/kafka-topics.sh --zookeeper 192.168.1.107:2181,192.168.1.108:2181  
--topic topicName --replication-factor 1 --partitions 1 --create
```

其中， --topic 定义topic名  
--replication-factor 定义副本数  
--partitions 定义分区数

## 查看全部Topic:

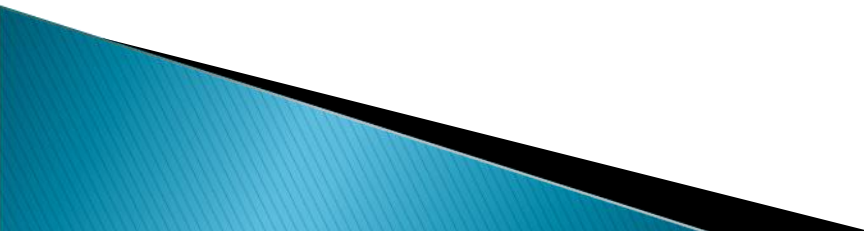
```
bin/kafka-topics.sh --zookeeper 192.168.1.107:2181,192.168.1.107:2181 --list
```

## 删除Topic:

```
./kafka-topics.sh --topic topicName --delete --zookeeper  
192.168.1.107:2181,192.168.1.107:2181
```

## 查看Topic明细:

```
./kafka-topics.sh --topic test --describe --zookeeper  
192.168.1.107:2181,192.168.1.107:2181
```



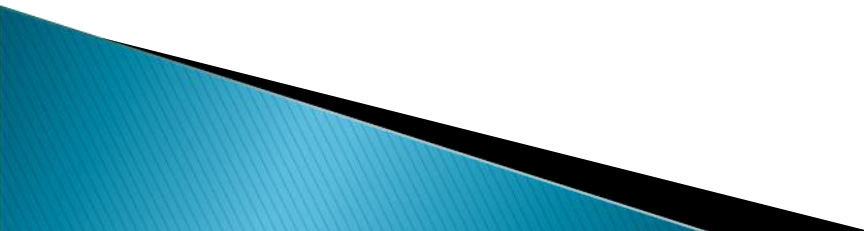
# 生产和消费

## Producer:

```
bin/kafka-console-producer.sh --broker-list  
192.168.1.107:9092,192.168.1.108:9092 --topic topicName
```

## Consumer:

```
bin/kafka-console-consumer.sh --zookeeper  
192.168.1.107:2181,192.168.1.107:2181 --topic topicName --from-beginning
```



# 异常处理

如果报异常：

Failed to load class "org.slf4j.impl.StaticLoggerBinder"

解决方法：

下载slf4j-1.7.6.zip

```
wget http://www.slf4j.org/dist/slf4j-1.7.6.zip
```

解压

```
unzip slf4j-1.7.6.zip
```

把slf4j-nop-1.7.6.jar 包复制到kafka libs目录下面



## 配置文件**config/server.properties** 中配置项:

<b>broker.id</b>	整数, 建议根据ip区分	
log.dirs	kafka存放消息文件的路径,	默认/tmp/kafka-logs
port	broker用于接收producer消息的端口	
<b>zookeeper.connect</b>	zookeeper连接	格式为 ip1:port,ip2:port,ip3:port
message.max.bytes	单条消息的最大长度	
num.network.threads	broker用于处理网络请求的线程数	如不配置默认为3, server.properties默认是2
num.io.threads	broker用于执行网络请求的IO线程数	如不配置默认为8, server.properties默认是2可适当增大,
queued.max.requests	排队等候IO线程执行的requests	默认为500
host.name	broker的hostname	默认null,建议写主机的ip,不然消费端不配置hosts会有麻烦
<b>num.partitions</b>	topic的默认分区数	默认2
<b>log.retention.hours</b>	消息被删除前保存多少小时	默认1周168小时
auto.create.topics.enable	是否可以程序自动创建Topic	默认true,建议false
default.replication.factor	消息备份数目	默认1不做复制, 建议修改
num.replica.fetchers	用于复制leader消息到follower的IO线程数	默认1

本课程版权归北风网所有

欢迎访问我们的官方网站  
[www.ibeifeng.com](http://www.ibeifeng.com)