# 实验1.2：大数据采集技术−Kafka

## **一、实验目的**

1. 了解kafka概念及特性，队列模式以及订阅模式；
2. 下载部署Kafka以及zookeeper；
3. 使用Kafka、zookeeper完成队列模式消息传输实例。

## **二、实验内容**

1. 简单介绍kafka概念及特性，队列模式以及订阅模式；
2. 操作演示Kafka以及zookeeper的下载和配置；
3. 操作演示Kafka、zookeeper完队列消息实例。

## **三、实验过程**

**任务一：认识什么是Kafka？**

在大数据系统中，经常会碰到一个问题，整个大数据是由各个子系统组成，数据需要在各个子系统中高性能、低延迟的不停流转。传统的企业消息系统并不适合大规模的数据处理。为了能够同时处理在线应用（消息）和离线应用（数据文件、日志等），Kafka由此诞生。 Kafka是一个分布式、支持分区的（partition）、多副本的（replica）、基于Zookeeper协调的分布式消息系统。

一般消息队列有两种模式的消费方式，分别是 队列模式和订阅模式。

队列模式：一对一，就是一个消息只能被一个消费者消费，不能重复消费。一般情况队列支持存在多个消费者，但是对于一个消息，只会有一个消费者可以消费它。

订阅模式：一对多，一个消息可能被多次消费，消息生产者将消息发布到Topic中，只要是订阅改Topic的消费者都可以消费。

**任务二：下载Kfaka及zookeeper**

打开浏览器，在浏览器地址栏中输入Kfaka服务器后台地址“kafka.apache.org”,进入下载页面，按住【Ctrl】+【F】键使用搜索功能,输入版本号“2.11-0.10.2”，找到相应版本号，点击下载。同理下载zookeeper。输入zookeeper服务器后台地址“mirror.bit.edu.cn”,进入下载页面,点击【stable】,点击“zookeeper-3.4.12.tar.gz”下载。如图所示

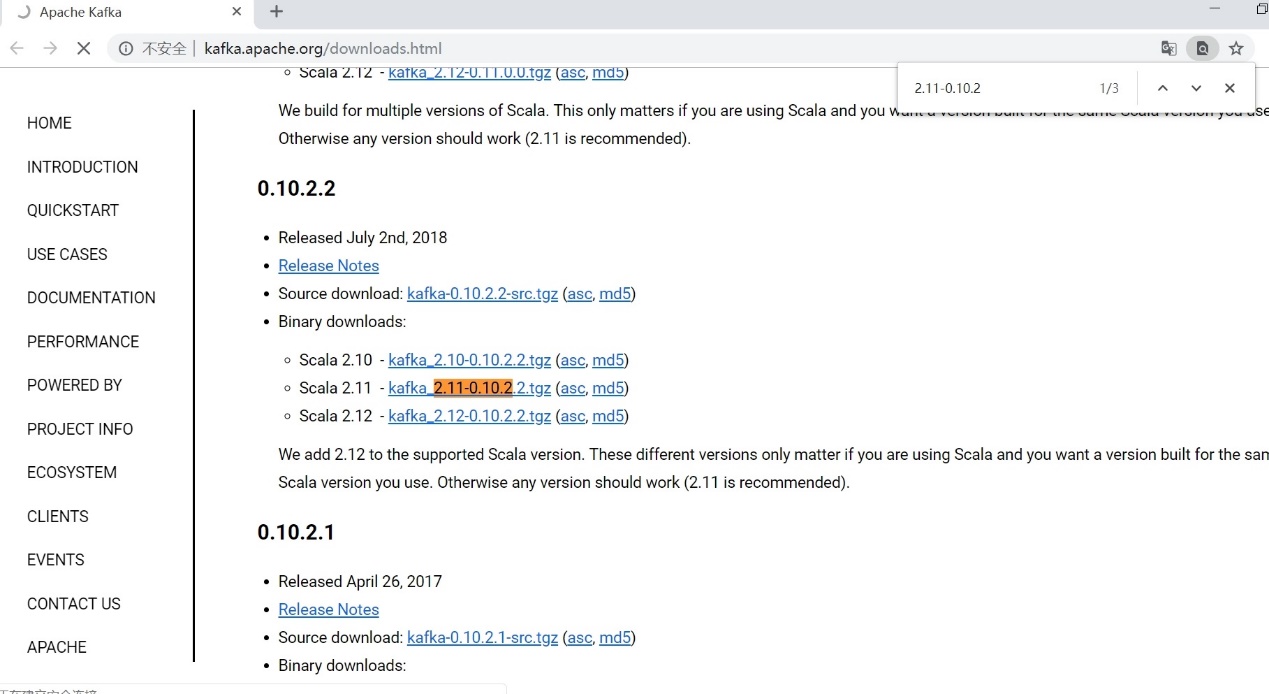


图4-1 下载Kfka

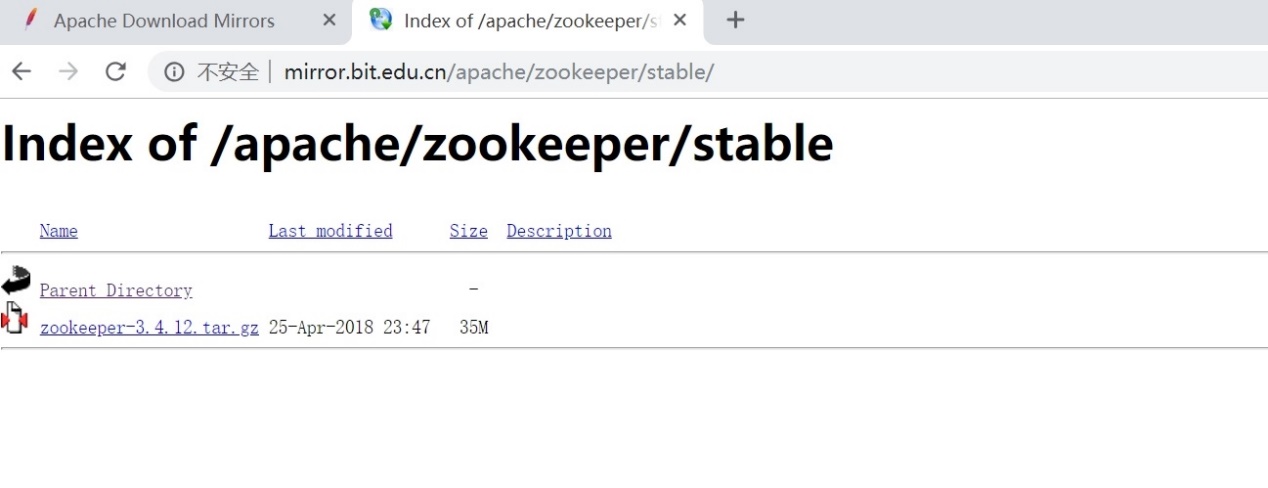


图4-2 下载zookeeper

**任务二：配置Kfaka及zookeeper**

1.打开下载好的压缩包，分别解压。右键点击【此电脑】，依次选择【属性】、【高级系统设置】、【高 级】、【环境变量】。如图所示

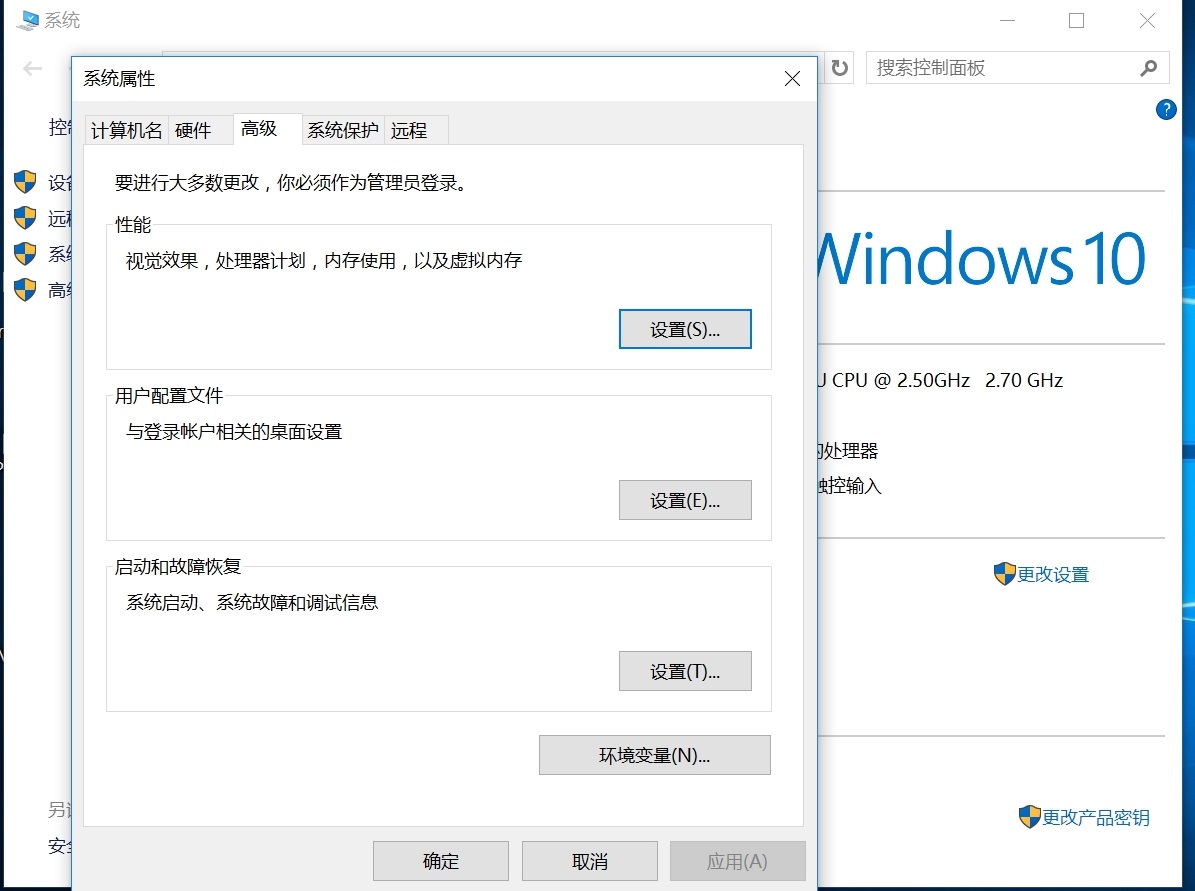


图4-3 配置zookeeper

2.点击【新建】，新建一个系统变量，输入变量名【ZOOKEEPER\_HOME】,将zookeeper的根目录作为变量值，点击【确定】。如图所示

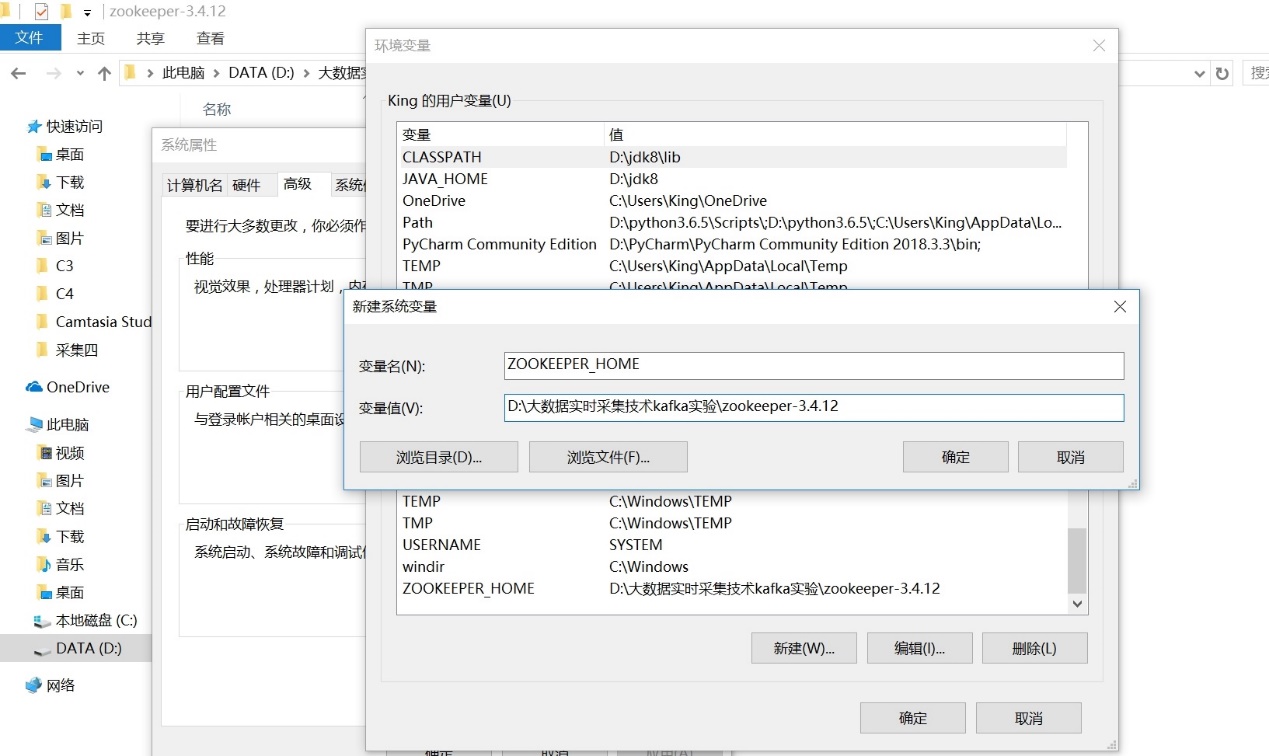


图4-4 新建系统变量

双击【path】变量，进入【编辑环境变量】，点击【新建】，添加zookeeper下bin目录作为文件路径，点击【确定】。如图所示

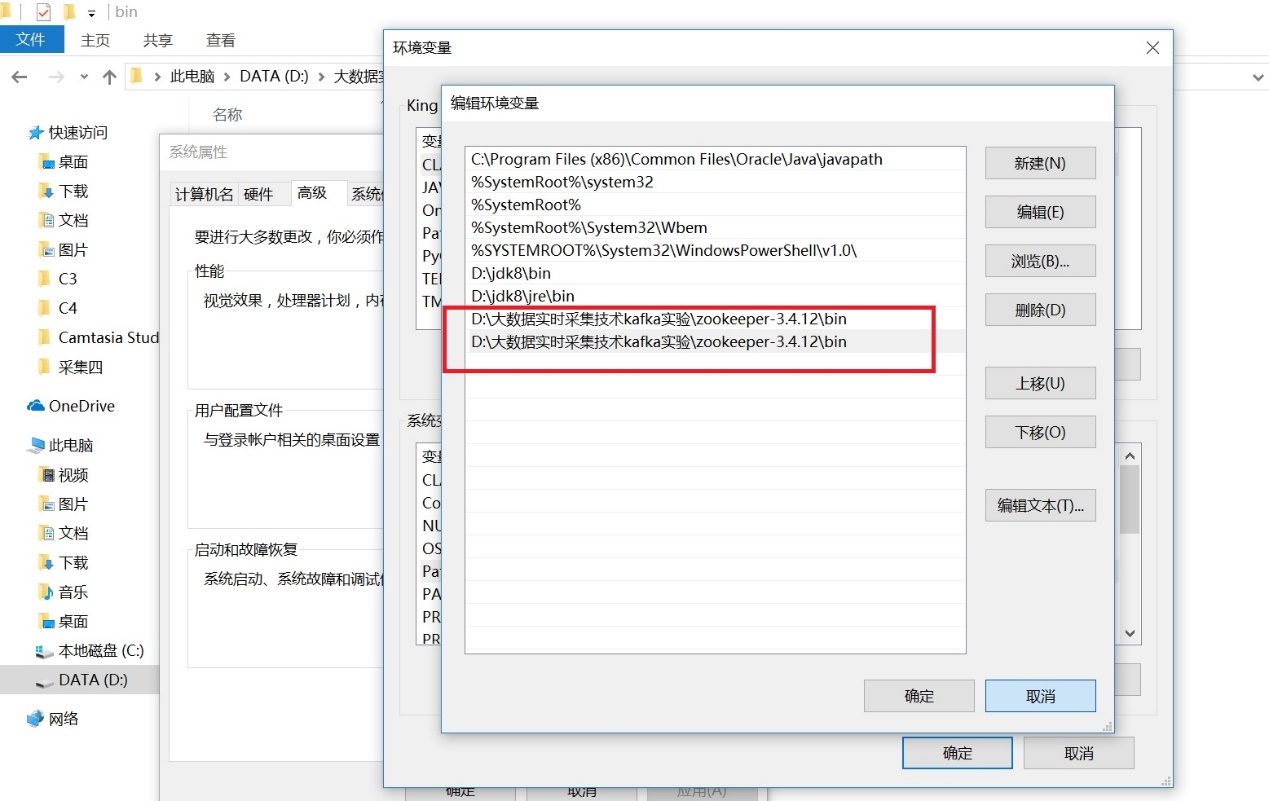


图4-5 编辑环境变量

点击【conf】文件夹，复制该文件夹下【zoo\_sample.cfg】文件,粘贴到该文件夹下,并改名为“zoo.cfg”。如图所示

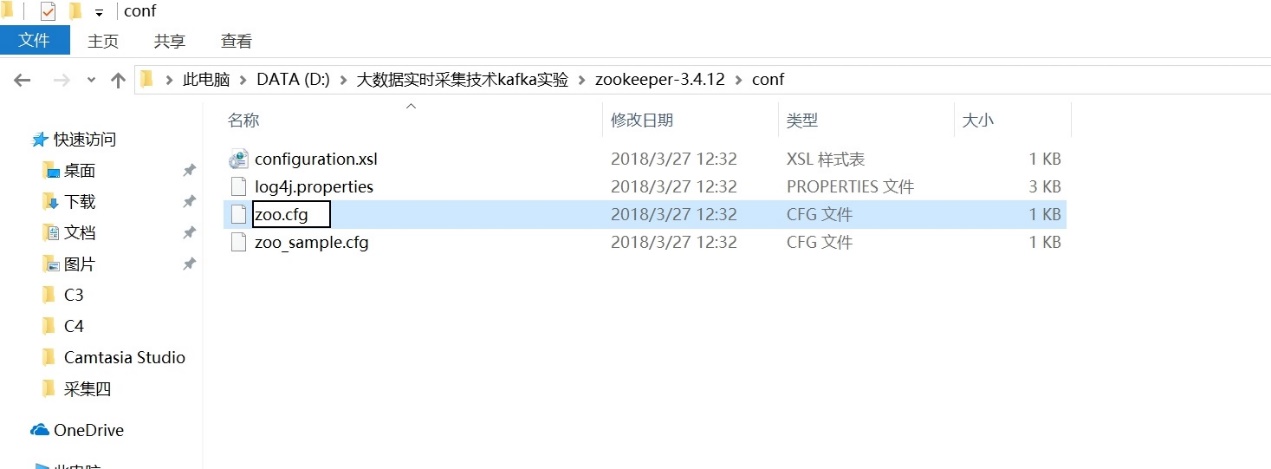


图4-6 配置zookeeper

点击打开【zoo.cfg】,将文件中“dataDir=/tmp/zookeeper”修改为zookeeper的根目录.如图所示

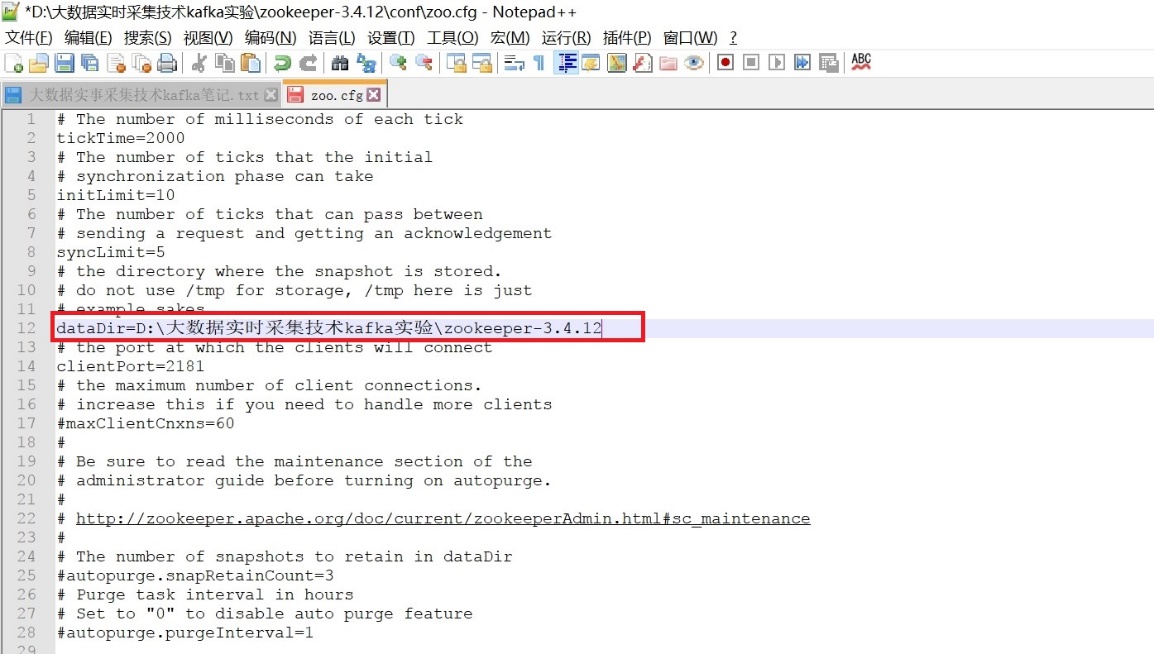


图4-7 修改存储位置

3.进入zookeeper的bin文件,按住【shift】不放,右键点击文件空白处,再点击“在此处打开powershell”，如图所示

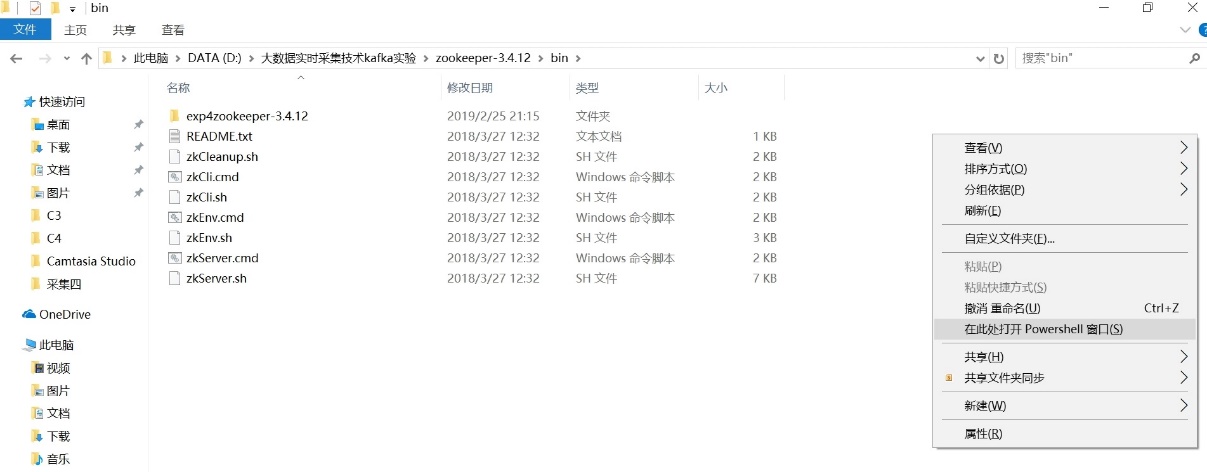


图4-8 打开zookeeper

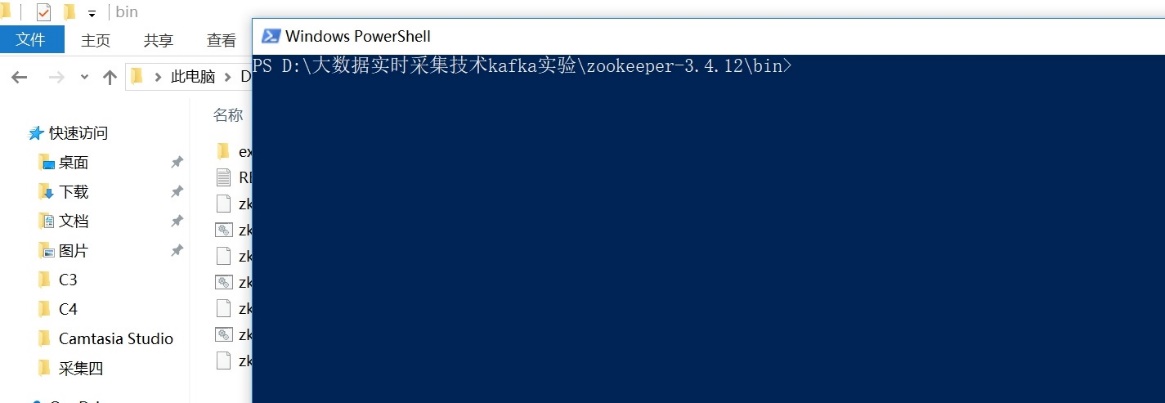


图4-9 zookeeper主页面

输入【zkserver】，点击【Enter】，成功启动zookeeper。如图所示

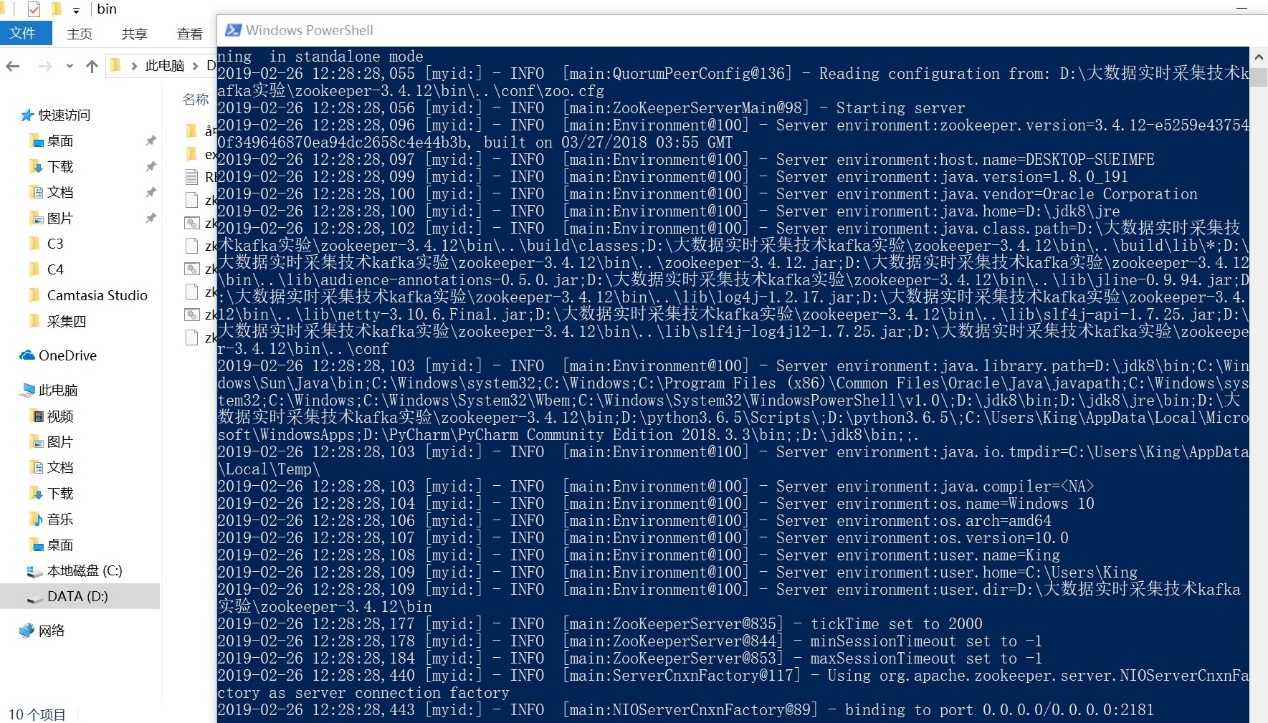


图4-10 zookeeper启动成功

进入Kfka，点击打开【config】文件夹，打开【sever.properties】。如图所示

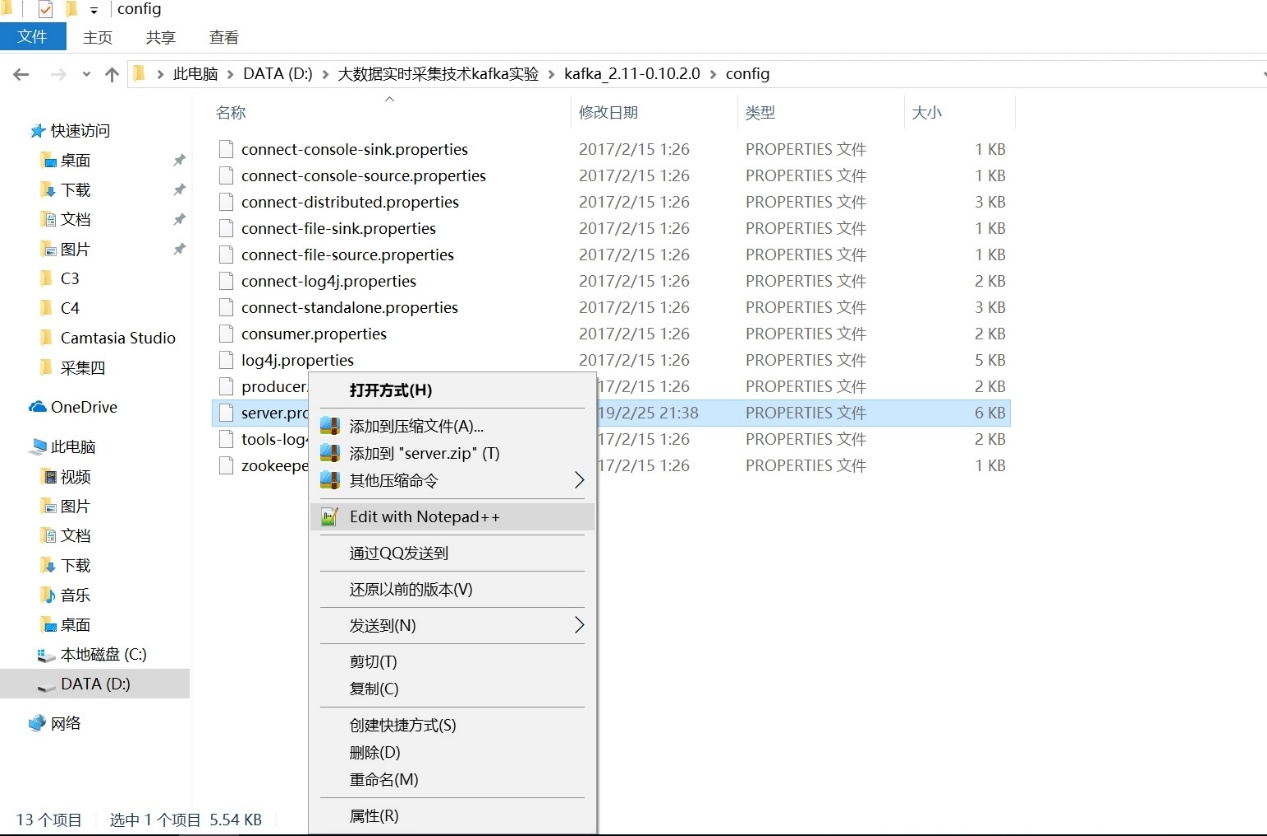


图4-11 配置Kfka

打开后修改【log.dirs】路径为kafka的根目录路径,点击【保存】。如图所示

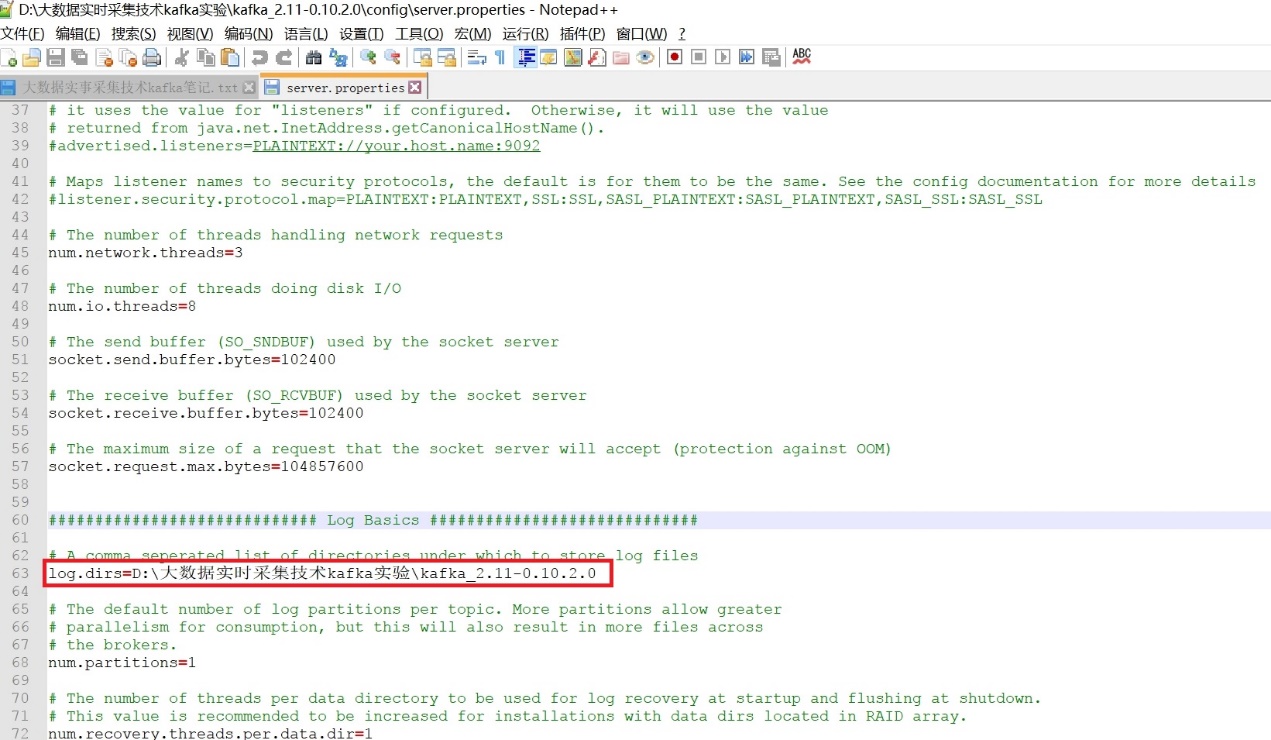


图4-12 修改kfka存储路径

打开【Windows】文件夹，按住【shift】不放,右键点击文件空白处,再点击“在次数打开powershell”。如图所示

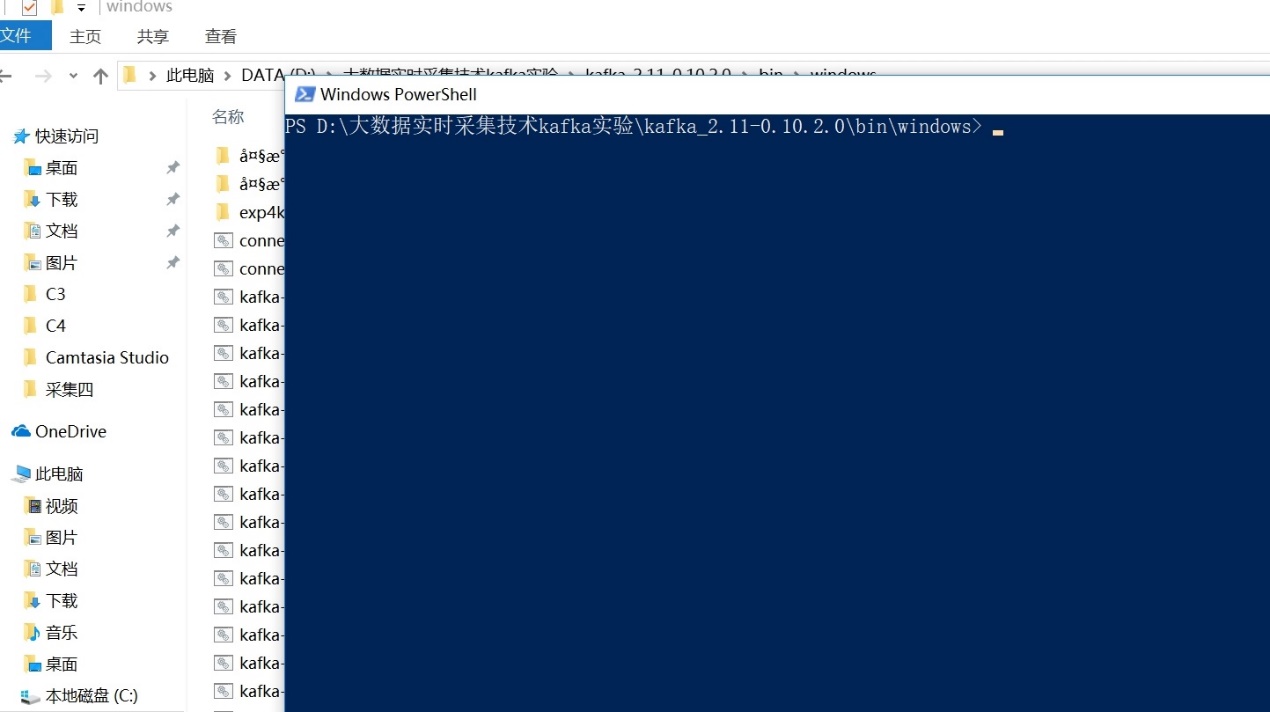


图4-13 kafka主界面

输入【.\kafka-server-atart.bat ..\..\config\server.properties】指令，点击【Enter】,启动成功。如图所示

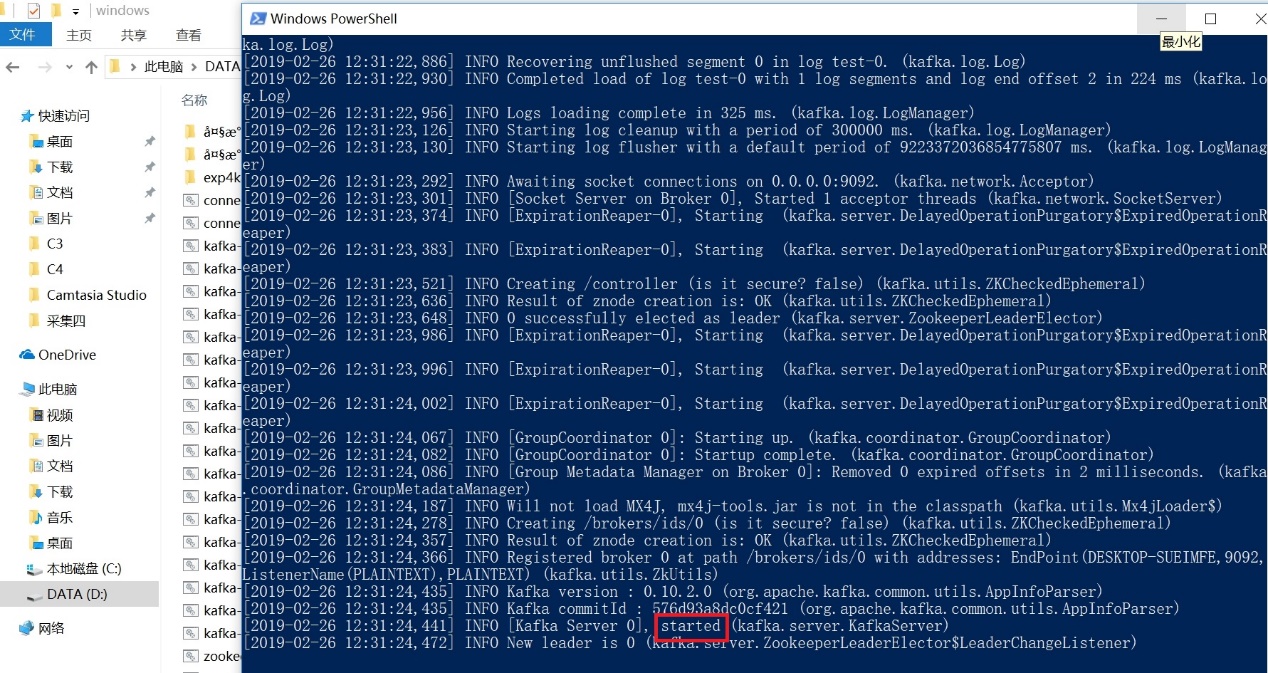


图4-14 kafka启动成功

**任务三：实例演示**

1.打开【Windows】文件夹，按住【shift】不放,右键点击文件空白处,再点击“在次数打开powershell”，输入准备好的指令创建一个名为“test”的topic，点击【Enter】，创建成功。如图所示

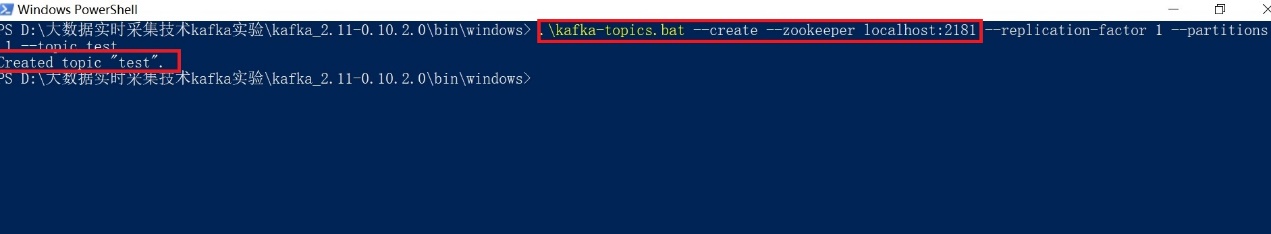


图4-15 topic创建成功

2.打开【Windows】文件夹，按住【shift】不放,右键点击文件空白处,再点击“在次数打开powershell”，输入准备好的指令创建一个生产者，点击【Enter】，创建成功。

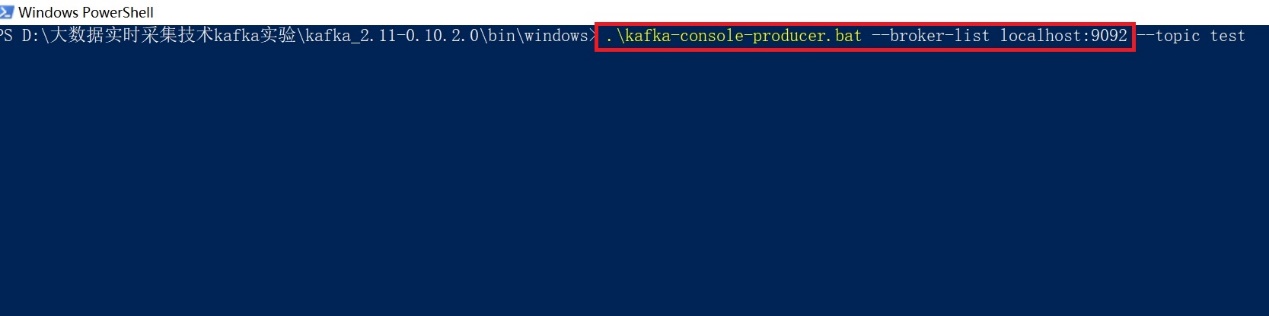


图4-16 创建“生产者”命令窗口

3.打开【Windows】文件夹，按住【shift】不放,右键点击文件空白处,再点击“在次数打开powershell”，输入准备好的指令创建一个消费者，点击【Enter】，创建成功。

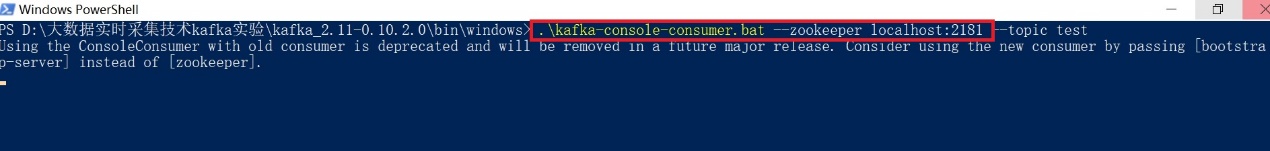


图4-17 创建“消费者”命令窗口

4.打开生产者命令窗口，输入【hello】，消费者窗口也收到了此消息。再次输入几行消息，可以看到测试成功。如图所示

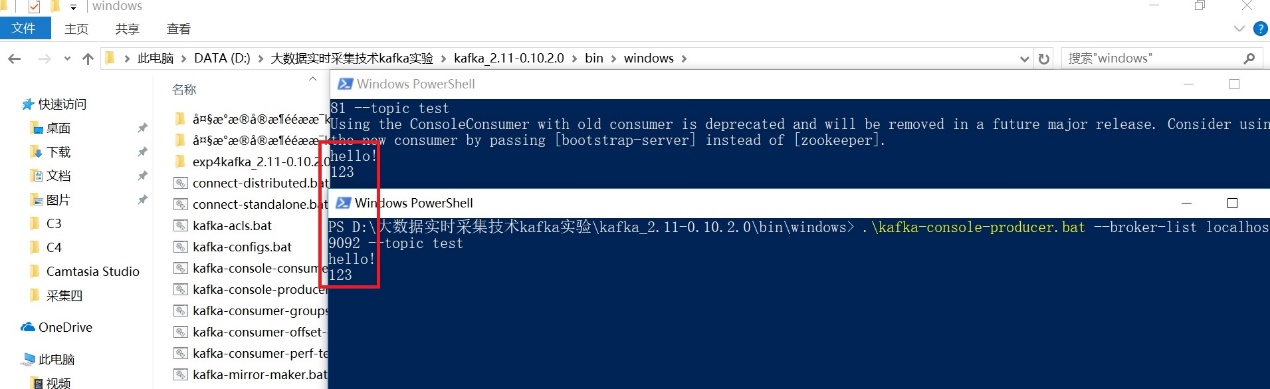


图4-18 队列消息测试成功

## **四、实验小结**

通过本次实验学习了大数据实时采集技术−Kafka的使用。下载并部署了Kafka以及Zookeeper。了解了两种消息消费模式。使用kafka代理及Zookeeper完成了队列模式的消息传输实例。