**《专高2第十一单元》练习手册**

编写人员：

编写时间：

专高二\_高性能架构\_第11单元
练习手册

一、判断题

1. (判断)大数据专业高级2-16.0课程11单元:kafka中的同一个partition可以被同一个消费组中多个消费者消费
A. 正确
B. 错误
2. (判断)大数据专业高级2-16.0课程11单元:kafka每个partition只能被同一个group中的同一个consumer消费
A. 正确
B. 错误
3. (判断)大数据专业高级2-16.0课程11单元:Zookeerper中保存这每个topic下的每个partition在每个group中消费的offset
A. 正确
B. 错误
4. (判断)大数据专业高级2-16.0课程11单元:kafka在传输数据的时候，采用的是“零拷贝”技术
A. 正确
B. 错误
5. (判断)大数据专业高级2-16.0课程11单元:消息生产者producer和消费者consumer可以在多个Broker上生产/消费topic
A. 正确
B. 错误
6. (判断)大数据专业高级2-16.0课程11单元:Kafka的生产者发送消息时，key的值不能为null
A. 正确
B. 错误
7. (判断)大数据专业高级2-16.0课程11单元:kafka中每个partition都有一个leader负责读写操作
A. 正确
B. 错误
8. (判断)大数据专业高级2-16.0课程11单元:kafka中的topic是逻辑存在的，它被分成多个partition
A. 正确
B. 错误
9. (判断)大数据专业高级2-16.0课程11单元:Kafka中的message是以topic为基本单位进行组织的
A. 正确
B. 错误
10. (判断)大数据专业高级2-16.0课程11单元:kafka中不同的topic之间是互相独立的
A. 正确
B. 错误

二、单选题

1. (单选)大数据专业高级2-16.0课程11单元:以下哪个组件可以实现作为消息系统使用？
A. mysql
B. kafka
C. oracle
D. tomcat
2. (单选)大数据专业高级2-16.0课程11单元:配置kafka集群，除了需要kafka本身，还需要用到以下哪个组件？
A. zookeeper
B. hdfs
C. jvm
D. namenode
3. (单选)大数据专业高级2-16.0课程11单元:Kafka中的offset指的是哪个概念？
A. 消费偏移量
B. 生产者
C. 消费者
D. 注册中心
4. (单选)大数据专业高级2-16.0课程11单元:kafka传输数据比较快的一个原因是，它使用（）技术
A. 零拷贝
B. 单机化
C. 本地化
D. 本地存储
5. (单选)大数据专业高级2-16.0课程11单元:kafka中的一个partition，可以被同一个group消费组中的（）个消费者进行消费
A. 2
B. 3
C. 1
D. 多个
6. (单选)大数据专业高级2-16.0课程11单元:kafka的可以将数据缓存到硬盘上，在默认情况下，数据能够存储（）天
A. 1天
B. 2天
C. 3天
D. 7天
7. (单选)大数据专业高级2-16.0课程11单元:kafka中的broker会存储生产者生产的数据，broker会将数据具体存放到哪里？
A. zookeeper
B. channel
C. partition
D. map
8. (单选)大数据专业高级2-16.0课程11单元:kafka中存储的数据，会对以下哪种格式的数据采用hash策略？
A. String格式
B. （key:value）格式
C. int格式
D. long格式
9. (单选)大数据专业高级2-16.0课程11单元:kafka中采用轮训策略进行存储的是哪种格式的数据？
A. String格式
B. （key:value）格式
C. int格式
D. long格式
10. (单选)大数据专业高级2-16.0课程11单元:Kafka中的partition可以分为多个（）？
A. segment file
B. broker
C. zookeeper
D. event
11. (单选)大数据专业高级2-16.0课程11单元:Kafka中负责partition所有的读写操作的是（）
A. zookeeper
B. leader
C. producer
D. consumer
12. (单选)大数据专业高级2-16.0课程11单元:kafka中，一个topic是逻辑存在的，它可以分成多个（）用来存储具体数据
A. partition
B. hdfs
C. jvm
D. zookeeper
13. (单选)大数据专业高级2-16.0课程11单元:kafka中的message是存放在broker中的，那一个message最大是多少字节？
A. 1000000
B. 1110
C. 2000
D. 2000000
14. (单选)大数据专业高级2-16.0课程11单元:kafka在生产和传输的数据时，元数据信息会具体存放在什么上面？
A. zookeeper
B. channel
C. partition
D. map
15. (单选)大数据专业高级2-16.0课程11单元:Kafka是一个分布式消息系统，它目前是（）公司进行开发和维护的
A. Sun
B. Apache
C. Linkedin
D. Google
16. (单选)大数据专业高级2-16.0课程11单元:kafka中消费者消费数据时，会使用到消费偏移量这个数据，以下哪个组件来存储消费偏移量？
A. flume
B. redis
C. mysql
D. zookeeper
17. (单选)大数据专业高级2-16.0课程11单元:kafka中消息message是生产者发送的数据，那它具体是存放在哪个位置上？
A. zookeeper
B. channel
C. partition
D. map
18. (单选)大数据专业高级2-16.0课程11单元:在kafka中，生产者可以生产数据发送给broker，生产者指的是？
A. Consumer
B. Producer
C. broker
D. topic
19. (单选)大数据专业高级2-16.0课程11单元:kafka中，生产者会将数据发送给broker，那broker指的是什么
A. 服务器
B. 数据库
C. 注册中心
D. 消费者
20. (单选)大数据专业高级2-16.0课程11单元:kafka分区中的数据可以看成是一个队列，它的进出规则是什么
A. 先进后出
B. 先进先出
C. 后进先出
D. 后进不出

三、多选题

1. (多选)大数据专业高级2-16.0课程11单元:kafka为什么传输速度快？
A. 二进制传输
B. 顺序读写磁盘
C. "零"拷贝
D. 发送消息集(message set)
2. (多选)大数据专业高级2-16.0课程11单元:为什么要使用消息系统？
A. 解耦
B. 冗余
C. 扩展性
D. 灵活性 & 峰值处理能力
3. (多选)大数据专业高级2-16.0课程11单元:以下哪些常见会用到kafka
A. 日志收集
B. 消息系统
C. 用户活动跟踪
D. 流式处理
4. (多选)大数据专业高级2-16.0课程11单元:消息系统传输数据有哪些模式？
A. 点对点模式
B. 发布订阅模式
C. 装饰模式
D. mvc模式
5. (多选)大数据专业高级2-16.0课程11单元:以下哪些组件属于消息中间件？
A. mysql
B. oracle
C. rabbitMQ
D. kafka
6. (多选)大数据专业高级2-16.0课程11单元:kafka中，以下说法中，关于partition说法错误的是
A. partition是逻辑存在的，不存放具体数据
B. partition是由topic组成的
C. partition一般是均匀分配在集群中的
D. 每个partition中有可以分为多个segment file
7. (多选)大数据专业高级2-16.0课程11单元:以下关于kafka特点的描述中，说法正确的是
A. 支持Hadoop并行数据加载
B. 通过集群来提供实时的消费
C. 支持通过Kafka服务器和消费机集群来分区消息
D. 高吞吐量
8. (多选)大数据专业高级2-16.0课程11单元:以下选项中，属于kafka的特点的是？
A. 解耦
B. 扩展性
C. 冗余，可靠性
D. 分解峰值压力
9. (多选)大数据专业高级2-16.0课程11单元:kafka中的partition是由多个segment file组成的，而segment file由哪些部分组成？
A. .log结尾的数据文件
B. .data结尾的数据文件
C. .txt结尾的索引文件
D. .index结尾的索引文件
10. (多选)大数据专业高级2-16.0课程11单元:以下哪些语言编写的程序可以成为kafka的客户端？
A. C
B. PHP
C. Python
D. Java

## 四、技能题

要求搭建zookeeper集群

* zk集群，主从节点，心跳机制（选举模式）
* 配置数据文件 myid 1/2/3 对应 server.1/2/3
* 通过 zkCli.sh -server [ip]:[port] 命令检测集群是否配置成功。

我们使用三台虚拟机来搭建zookeeper真实分布式集群,三台机器都必须具备java的运行环境，并且关闭或清空防火墙规则，不想关闭防火墙的话，就需要去配置相应的防火墙规则.

1. 启动三台虚拟机(10分)
2. 每台机器上都安装上zookeeper,如果已经安装可以使用原有的zookeeper（每台5分)共计15分
3. 首先配置一下系统的hosts文件(每台机器正确修改主机名称和host文件各5分)，共计15分，如果修改过了，通过命令查看各个配置.

**提示部分**

编辑vim /etc/sysconfig/network /修改主机名 (提示部分)

[root@localhost ~]*# vim /etc/hosts*

192.168.190.128 zk000

192.168.190.129 zk001

192.168.190.130 zk002

1. 修改每个机器上的zookeeper配置文件正确搭建集群配置每台机器10分，共计30分.

**提示部分:**

tickTime=2000

initLimit=10

syncLimit=5

dataDir=/usr/local/zookeeper/dataDir

dataLogDir=/usr/local/zookeeper/dataLogDir

clientPort=2181

server.1=192.168.190.128:2888:3888 *# 默认server.1为master节点*

server.2=192.168.190.129:2888:3888

server.3=192.168.190.130:2888:3888

1. 启动zookeeper，每台机器正确启动，集群搭建成功每台机器10分，共计30分.

**提示部分**

*zkServer.sh start*