**《专高2第十二单元》练习手册**

编写人员：

编写时间：

专高二\_高性能架构\_第12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12单元:kafka中消费偏移量需要生产者进行维护
A. 正确
B. 错误
2. (判断)大数据专业高级2-16.0课程12单元:kafka每个partition可以被同一个group中的多个consumer消费
A. 正确
B. 错误
3. (判断)大数据专业高级2-16.0课程12单元:kafka中每个group中消费的offset是存放在zookeeper中的
A. 正确
B. 错误
4. (判断)大数据专业高级2-16.0课程12单元:kafka在传输数据的时候，log文件中的数据直接通过系统内存(内核)直接网络传输，不经过应用(kafka)的内存的数据的交换，这就是“零拷贝”计数
A. 正确
B. 错误
5. (判断)大数据专业高级2-16.0课程12单元:kafka的partition可以在多个不同的broker中备份
A. 正确
B. 错误
6. (判断)大数据专业高级2-16.0课程12单元:kafka producer发送消息的时候，可以指定key，key可以为空
A. 正确
B. 错误
7. (判断)大数据专业高级2-16.0课程12单元:kafka中生产者发送消息，可以指定key，这个key的作用是为消息选择存储分区
A. 正确
B. 错误
8. (判断)大数据专业高级2-16.0课程12单元:kafka中当同一个group中的消费者数量多于partition的数量时，多余的消费者会空闲
A. 正确
B. 错误
9. (判断)大数据专业高级2-16.0课程12单元:kafka中一个partition对应一个broker，一个broker可以管多个partition
A. 正确
B. 错误
10. (判断)大数据专业高级2-16.0课程12单元:kafka中每个consumer都有对应的group
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12单元:kafka作为一个消息中间件，它主要解决了消息\_\_\_\_传输问题
A. 同步
B. 异步
C. 间接
D. 直接
2. (单选)大数据专业高级2-16.0课程12单元:Kafka是由以下哪个组织研发和维护的组件？
A. google
B. apache
C. sun
D. oracle
3. (单选)大数据专业高级2-16.0课程12单元:Kafka最初是由LinkedIn开发编写的，它的底层是使用什么语言编写的？
A. python
B. scala
C. c++
D. c#
4. (单选)大数据专业高级2-16.0课程12单元:以下组件当中，哪个是具有缓存数据的消息系统的组件？
A. storm
B. flume
C. kafka
D. spring
5. (单选)大数据专业高级2-16.0课程12单元:kafka是一个MQ消息中间件，在项目中，以下哪个是它起到的主要作用
A. 缓冲地带
B. 实时计算
C. 离线计算
D. 分布式存储文件
6. (单选)大数据专业高级2-16.0课程12单元:kafka中生产者生产了（key:value）格式的数据，则broker会采用以下哪种策略进行存储？
A. map
B. hash
C. broker
D. 以上都不对
7. (单选)大数据专业高级2-16.0课程12单元:在kafka中，同一个group消费组内，如果有3个消费者，则对于同一个partition，可以有多少个消费者进行消费？
A. 1
B. 2
C. 3
D. 0
8. (单选)大数据专业高级2-16.0课程12单元:Kafka的producer基于（）方式发送消息数据消息
A. Push
B. Pull
C. peer
D. redis
9. (单选)大数据专业高级2-16.0课程12单元:kafka中的消费者进行消费时，会形成一个消费偏移量offset，offset会存储在下面哪个组件？
A. zookeeper
B. hdfs
C. jvm
D. namenode
10. (单选)大数据专业高级2-16.0课程12单元:kafka中生产者生产数据时，数据的生产时间等元数据信息会存储在（）中
A. jvm
B. hdfs
C. zookeeper
D. namenode
11. (单选)大数据专业高级2-16.0课程12单元:Kafka中每个partition都有一个对应的（），它负责所有的读写操作
A. zookeeper
B. leader
C. producer
D. consumer
12. (单选)大数据专业高级2-16.0课程12单元:kafka中，消息会根据Topic来存储，而具体存储数据的是（）
A. partition
B. hdfs
C. jvm
D. zookeeper
13. (单选)大数据专业高级2-16.0课程12单元:kafka中的消息是具体存储在分区之中的，一个消息最大是多少？
A. 1000000字节
B. 1210字节
C. 2000字节
D. 2000000字节
14. (单选)大数据专业高级2-16.0课程12单元:kafka在消费者消费数据时，会从哪个位置读取消费偏移量信息？
A. zookeeper
B. channel
C. partition
D. map
15. (单选)大数据专业高级2-16.0课程12单元:kafka中生产者生产String格式的数据，则broker会采用以下哪种策略进行存储？
A. 轮训
B. 均衡机制
C. broker
D. 以上都不对
16. (单选)大数据专业高级2-16.0课程12单元:kafka中可以缓存生产者生产的数据，如果不进行配置，在默认情况下，该数据可以存储（）天
A. 1
B. 3
C. 5
D. 7
17. (单选)大数据专业高级2-16.0课程12单元:kafka中broker存储数据，具体存放在topic中的（）上
A. zookeeper
B. channel
C. partition
D. map
18. (单选)大数据专业高级2-16.0课程12单元:在kafka中，消费者可以从broker中消费数据，消费者指的是哪个？
A. Consumer
B. Producer
C. broker
D. topic
19. (单选)大数据专业高级2-16.0课程12单元:kafka中，可以配置多个broker用来存储数据，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12单元:kafka中partition是存储具体数据的，它的内部是FIFO的，FIFO指是什么
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9单元:消息系统有多种，它的通信模式有哪些？
A. 点对点
B. 面向对象
C. 发布订阅
D. 面对面
2. (多选)大数据专业高级2-16.0课程9单元:消息系统可以在分布式项目间进行通信，它有哪些特点？
A. 解耦
B. 冗余
C. 可恢复性
D. 异步通信
3. (多选)大数据专业高级2-16.0课程9单元:消息系统有很多种，RabbitMQ、ActiveMQ、Redis等等，相比较而言，kafka的优势在哪里？
A. 高吞吐率
B. 消息持久化
C. 完全分布式
D. 没有缓冲
4. (多选)大数据专业高级2-16.0课程9单元:kafka架构中，有三个主要组件，以下属于kafka架构中的组件的是
A. producer
B. consumer
C. broker
D. agent
5. (多选)大数据专业高级2-16.0课程9单元:Redis是一个nosql数据库，它也可以作为一个消息中间件使用，下面哪些指令能够清空Redis数据库？
A. FLUSHALL
B. SLOWLOG
C. FLUSHDB
D. MONITOR
6. (多选)大数据专业高级2-16.0课程12单元:kafka中，Topic可以由多个partititon组成，以下说法中，关于Topic和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. Topic是物理存在的
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12单元:kafka是一个消息队列，以下关于它的说法正确的是？
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12单元:Kafka是一个分布式发布订阅消息系统，以下哪些是kafka的特点？
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12单元:kafka中的消息是存储在分区（partition）中的，而partition是由多个segment file组成的，其中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12单元:kafka支持多种语言编写的客户端，其中包含以下哪些语言？
A. C++
B. PHP
C. Python
D. Java

## 四、技能题

要求在搭建zookeeper集群，搭建kafka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kafka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kafka,如果已经安装可以使用原有的kafka（每台10分)共计30分
3. 首先配置一下kafka的配置文件(每台机器正确10分)，共计30分，如果修改过了，通过命令查看各个配置.

**提示部分**

1. broker.id=1 #保证每个broker唯一，第一台可以不修改默认为0，后面两台需要修改，如改为2和3
2. num.partitions=3 #分区数量一般与broker保持一致
3. listeners=PLAINTEXT://10.211.55.29:9092 #修改为本机ip
4. zookeeper.connect=10.211.55.29:2181,10.211.55.30:2181,10.211.55.31:2181 #配置三台服务zookeeper连接地址
5. host.name=10.211.55.29 #新增host.name值，分别设为不同的值（3台机器根据自己的ip设置）

6、log.dirs=/var/kafka-logs/ #修改log.dirs目录为之前自定义的目录

1. 启动zookeeper，每台机器正确启动，集群搭建成功每台机器5分，共计15分.
2. 启动kafka,每台机器正确启动，集群搭建成功每台机器5分，共计15分.

**提示部分**

bin/kafka-server-start.sh config/server.properties

1. 演示集群效果

**提示**

1. 集群创建主题：/usr/local/kafka/bin/kafka-topics.sh -create --bootstrap-server 192.168.73.129:9092,192.168.73.130:9092,192.168.73.131:9092 -replication-factor 3 --partitions 3 --topic chjy

2.集群查看主题：/usr/local/kafka/bin/kafka-topics.sh -list --bootstrap-server 192.168.73.129:9092

3.集群生产者：/usr/local/kafka/bin/kafka-console-producer.sh --broker-list 192.168.73.129:9092,192.168.73.130:9092,192.168.73.131:9092 --topic chjy

4.集群消费者：/usr/local/kafka/bin/kafka-console-consumer.sh --bootstrap-server 192.168.73.129:9092,192.168.73.130:9092,192.168.73.131:9092 --topic chjy --from-beginning