**Kafka 期末复习考试 BD2020**

......................................................................................................................... **第1部分 单选题** 总题数: 25 .........................................................................................................................

1. **【单选题】** (2分) 关于Partition Rebalance的说法，正确的是哪一项？
   1. Partition Rebalance是Kafka Producer将消息进行分区的过程。



* 1. Partition Rebalance是Broker将消息从一个分区转发到另外一个分区的过程。
  2. Partition Rebalance是消费者组中的消费者重新获取分区处理权的过程。

参考答案 **C**

1. **【单选题】** (2分) 关于Kafka Consumer的说法错误的是哪一项？
   1. Kafka Consumer可以同时订阅多个主题的消息。



* 1. Kafka Consumer不能订阅未创建主题的消息。
  2. Kafka Consumer通过调用poll()方法轮询的方式从Broker中接收消息。
  3. 同一个Kafka Consumer对象不能同时属于不同的Consumer Group。

参考答案 **B**

1. **【单选题】** (2分) 在Kafka Broker的配置参数中，包含一个Broker唯一标识符，这个唯一标识符是什么类型的数据？
   1. Float



* 1. Integer
  2. String
  3. Map

参考答案 **B**

1. **【单选题】** (2分) 向Kafka Topic推送消息的是哪一个组件？
   1. 生产者Producer。



* 1. 消费者Consumer。
  2. 分区Partition。
  3. Kafka Broker。

参考答案 **A**

1. **【单选题】** (2分) 关于Kafka中的acks配置参数，说法错误的是哪一项？
2. acks接受三种不同的选项，分别是0，1，all/-1。



1. acks=0 如果设置为零，则生产者将不等待来自服务器的任何确认。
2. acks=1 这意味着leader会将记录写入本地日志，但不会等待所有follower的完全确认。
3. acks=all 这意味着leader将等待所以的同步副本集来确认记录。
4. acks=1的等待时间最长。

参考答案 **E**

1. **【单选题】** (2分) 在一个Kafka集群中创建主题时，如果将主题的复制系数（replication factor) 配置为3，那么这个Kafka集群至少需要几个Broker?
   1. 2



* 1. 1
  2. 3
  3. 4

参考答案 **C**

1. **【单选题】** (2分) 有关于Consumer Group的如下两种说法： 1. 同一个Consumer Group中的每一个Consumer能同时消费所有分区的消息。 2. 不同的Consumer Group能同时消费同一个主题中相同的消息。
   1. 1正确，2正确。



* 1. 1正确，2错误。C. 1错误，2正确。

D. 1错误，2错误。

参考答案 **C**

1. **【单选题】** (2分) 通过jconsole对Kafka进行监控的是什么协议？
   1. JMX



* 1. HTTP
  2. Thrift
  3. RPC参考答案 **A**

1. **【单选题】** (2分) 关于Kafka Consumer中的poll()方法，以下说法中不属于poll()方法的功能是哪一项？
   1. 处理Partition Rebalance。



* 1. 接收分区消息。C. 处理心跳消息。

D. 通知Kafka Producer发送消息。

参考答案 **D**

1. **【单选题】** (2分)

关于消费者提交处理偏移量的交付语义中，下列哪一种交付语义可能会丢失消息？

 A. At least once.

* 1. At most once.



* 1. Exactly once.

参考答案 **B**

1. **【单选题】** (2分) 在开发Kafka Producer应用时， 如果需要配置Producer中的各种参数，需要用到的正确Java类是哪一个？
   1. KafkaProducer



* 1. ProducerRecord
  2. ProducerConfig
  3. RecordMetadata

参考答案 **C**

1. **【单选题】** (2分) 下列选项中列出的脚本命令，哪一项是用于启动Kafka Broker的正确命令？
   1. bin/kafka-start-server.sh config/server.properties



* 1. bin/kafka-start.sh config/server.properties
  2. bin/kafka-server-starts.sh config/server.properties
  3. bin/kafka-server-start.sh config/server.properties

参考答案 **D**

1. **【单选题】** (2分)

ProducerRecord对象用在Kafka Producer发送消息到Broker时，对消息进行初始化。下列ProducerRecord 对象的属性中，哪一项可以用于指定分区？

* 1. Topic



* 1. Partition
  2. Value

参考答案 **B**

1. **【单选题】** (2分) 关于Kafka Producer在向Broker发送消息时的说法正确的是哪一项？
   1. 消息先通过分区器进行分区，然后通过序列化器进行消息的序列化。B. 消息先通过序列化器进行序列化，然后通过分区器进行消息的分区。



C. 消息的分区与序列化是同时进行的。

参考答案 **B**

1. **【单选题】** (2分) 关于Kafka Producer在向Broker发送消息时的序列化器说法正确的是哪一项？
   1. 序列化器将消息转换为byte[]对象。



* 1. 序列化器将消息转换为String对象。
  2. 序列化器将消息转换为Java Object对象。
  3. 序列化器将消息转换为Json对象。

参考答案 **A**

1. **【单选题】** (2分)

如下列出的软件依赖中，哪一项不是运行Kafka的依赖项？

* 1. JRE/JDK



* 1. Python
  2. Zookeeper

参考答案 **B**

1. **【单选题】** (2分) 从服务器收集物理日志文件并放在中心位置（文件服务器或 HDFS）进行处理的过程叫什么？
   1. 流式处理。B. 日志聚合。



* 1. Metrics（指标）。
  2. 消息传递。

参考答案 **B**

1. **【单选题】** (2分)

Kafka Topic中读取消息并进行处理的是哪一个组件？

* 1. 生产者Producer。



* 1. 消费者Consumer。
  2. 分区Partition。D. Kafka Broker。

参考答案 **B**

1. **【单选题】** (2分) 下列关于Kafka Cosumer与Consumer Group的说法中，错误的是哪一项？
   1. Kafka Consumer从Topic中读取订阅的消息。



* 1. Kafka Cosumer将被分到不同的Consumer Group中。
  2. Consumer Group中的每一个Consumer负责读取和处理Topic中部分分区的消息。
  3. 同一个Consumer对像，只能订阅一个主题的消息。

参考答案 **D**

1. **【单选题】** (2分) 关于Kafka Topic的如下说法中，哪一项说法是错误的？
   1. Topic是对数据进行存储的分类。



* 1. Topic中存储的数据可以保存到磁盘中，我们称这种机制为Retention。
  2. 一个Topic只能包含一个Partition。
  3. 发送到Topic中的每一条消息都会赋予一个消息偏移量，消息偏移量是一个整数值。

参考答案 **C**

1. **【单选题】** (2分) 用于从Kafka Producer异步发送消息的正确代码是哪一项？
   1. Producer.send(record, new callback())



* 1. Producer.send(record)
  2. Producer.send(record, “topic”)
  3. Producer.send(record, “key”)

参考答案 **A**

1. **【单选题】** (2分) 在如下的Kafka Topic的配置参数中，哪一项用于配置消息日志段文件的最大值？
   * 1. num.partitions



* + 1. log.retention.ms
    2. log.retention.bytes
    3. log.segment.bytes

参考答案 **D**

1. **【单选题】** (2分) 当我们使用kafka-topics.sh脚本在Kafka中去创建Topic时，用于指定复制系数的正确参数是哪一项？
   * 1. --topic



* + 1. --partitions
    2. --replication-factor
    3. --zookeeper

参考答案 **C**

1. **【单选题】** (2分) 关于Kafka Broker如下两种说法： 1. 一个Kafka集群可以包含多个Broker。 2. 同一个主题分区的所有副本数据可以存储在同一个Broker。 关于以上两种说法，请在以下选项中选出正确的判断。
   * 1. 1正确，2正确。



* + 1. 1正确，2错误。C. 1错误，2正确。

D. 1错误，2错误。

参考答案 **B**

1. **【单选题】** (2分) 下列选项中不是Kafka组件的是哪一项？
   * 1. Producer



* + 1. Consumer
    2. Broker
    3. Spout参考答案 **D**

......................................................................................................................... **第2部分 多选题** 总题数: 4 .........................................................................................................................

1. **【多选题】** (5分) 在下列选项中，选出可能触发Kafka Consumer分区再平衡（Partition Rebalance) 发生的情况。
   * 1. 有消费者加入消费者组。



* + 1. 有消费者离开消费者组。
    2. 消费者所消费的主题分区数发生了变化。



* + 1. 消费者的心跳消息发送到Broker发生了超时。

# 参考答案 A,B,C,D

**27. 【多选题】** (5分) 下列关于Kafka的Topic说法中，正确的是哪些？

1. 一个Topic可以包含多个Partition。



1. 通过增加Topic的Partition数量可以提高消息的处理吞吐量。
2. 通过执行kafka-configs.sh脚本可以减少已经存在Topic的Partition数量。
3. 同一个Topic中，相同Partition的消息串行处理，不同Partition的消息并行处理。

# 参考答案 A,B,D

**28. 【多选题】** (5分)

下列操作或事件中，哪些能导致消费者组中的消费者处理的分区发生变化？

1. 使用kill -9命令停止Kafka Consumer应用。



1. Kafka Consumer连接Broker的网络出现故障，并且长时间没有恢复。
2. Kafka Consumer订阅一个Topic 10天后再创建该主题。
3. 启动一个新的Kafka Consumer应用并加入已存在的消费者组。

# 参考答案 A,B,C,D

1. **【多选题】** (5分)

下列用于消息订阅的代码中，哪些可以用于正确订阅主题t8的消息？

* 1. consumer.subscribe(Collections.singletonList("t8"));

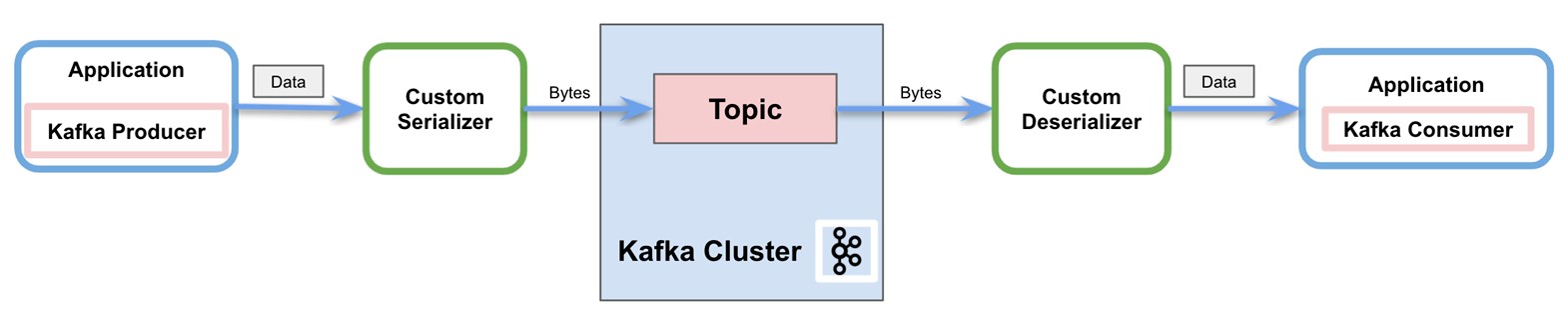


* 1. consumer.subscribe(Arrays.asList("t1", "t8"));
  2. Pattern pattern = Pattern.compile("[a-z][0-9]"); consumer.subscribe(pattern);
  3. consumer.subscribe(Arrays.asList("t10", "t6"));参考答案 **A,B,C**

......................................................................................................................... **第3部分 填空主观题（手动批阅）** 总题数: 3 .........................................................................................................................

1. **【填空题】** (3分) 根据以下的数据处理流程图，消息在生产者端被序列化为 \_\_\_\_\_\_\_\_ 类型，然后发送到Topic，消费者通过订阅主题并接收消息，将 \_\_\_\_\_\_\_\_ 类型数据反序列化为

\_\_\_\_\_\_\_\_\_\_\_。



参考答案 **30.1 byte[]**

**30.2 byte[]**

# 30.3 Java 对象（结构化数据）

1. **【填空题】** (4分) 用户自定义序列化器的类名是CustomizedSerializer, 在以下语句的空白处填入代码片段，将自定义序列化器配

置到Kafka Producer中。 props.put(ProducerConfig.VALUE\_SERIALIZER\_CLASS\_CONFIG, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ );

参考答案 **31.1 CustomizedSerializer.class.getName()**

1. **【填空题】** (3分) 主题mytopic有8个分区，如果将 log.retention.bytes 的值配置为10GB，那么mytopic一共可以保留 \_\_\_\_\_\_\_\_数据。参考答案 **32.1 80GB**

......................................................................................................................... **第4部分 问答题** 总题数: 2 .........................................................................................................................

1. **【问答题】** (10分)

Kafka集群的其中一个Broker运行在192.168.0.56主机上，端口为9092，主题my-topic中的消息key是整数， value是String。 编写消费者代码，订阅主题my-topic中的消息，输出消息的key，value与offset。

参考答案：

public static void main(String[] args) {  
 // 1. 创建消费者配置  
 Properties props = new Properties();  
 // 指定连接的Broker主机名和端口号  
 props.setProperty(ConsumerConfig.*BOOTSTRAP\_SERVERS\_CONFIG*, "192.168.0.56:9092");  
  
 // 指定消费者组的ID  
 props.setProperty(ConsumerConfig.*GROUP\_ID\_CONFIG*, "g1");  
 // 是否自动提交offset(默认自动提交)  
 props.setProperty(ConsumerConfig.*ENABLE\_AUTO\_COMMIT\_CONFIG*, "true");  
 // 提交offset的时间间隔  
 props.setProperty(ConsumerConfig.*AUTO\_COMMIT\_INTERVAL\_MS\_CONFIG*, "1000");  
 // 指定KEY的反序列化类  
 props.setProperty(ConsumerConfig.*KEY\_DESERIALIZER\_CLASS\_CONFIG*, IntegerDeserializer.class.getName());  
 // 指定VALUE的反序列化类  
 props.setProperty(ConsumerConfig.*VALUE\_DESERIALIZER\_CLASS\_CONFIG*, StringDeserializer.class.getName());  
 // 当没有初始偏移量时从哪个位置开始读取  
 props.setProperty(ConsumerConfig.*AUTO\_OFFSET\_RESET\_CONFIG*, "earliest");  
  
 // 2. 创建消费者  
 KafkaConsumer<Integer, String> consumer = new KafkaConsumer<>(props);  
  
 // 3. 订阅主题  
 consumer.subscribe(Collections.*singletonList*("my-topic"));  
  
  
 // 4. 轮询拉取消息 - 可以指定超时时间  
 while (true) {  
 ConsumerRecords<Integer, String> records = consumer.poll(Duration.*ofSeconds*(1).toMillis());  
 // 5. 打印元数据信息  
 for (ConsumerRecord record : records) {  
 int key = (int)record.key();  
 String value = record.value().toString();  
 long offset = record.offset();  
  
 System.*out*.println("key:" + key);  
 System.*out*.println("value:" + value);  
 System.*out*.println("offset:" + offset);  
 }  
 }  
 }

1. **【问答题】** (10分) 请举例说明Kafka能用于大数据处理的场景。

# 参考答案 消息传递、 网站活动跟踪、 Metrics（监控指标）、 日志聚合、 流式处理。