

上海交通大学试卷 (B 卷)

(2022 至 2023 学年 第 1 学期)

班级号 _____ 学号 _____ 姓名 _____

课程名称 _____ 应用系统体系架构 _____ 成绩 _____

在本学期，同学们结合上课内容，对大二时开发的 E-BookStore 系统进行了升级和扩展。在系统设计与开发的过程中，大家碰到了很多问题，请你根据在本课程中学习到的知识，帮助他们分析处理以下问题（每题 5 分，共 100 分）：

所有题目请用中文作答

一、 请你帮助同学们解答在设计和开发服务器端程序时遇到的下列问题：

1. 在应用服务器中都会配置数据库连接池，对数据库连接进行池化管理。由于 E-BookStore 存在大量的用户，因此，有的同学认为连接池应该配置成多大，和并发用户数没有太大的关系，而是和服务器的硬件配置相关。于是，在 8 核的服务器上配置了大小为 20 的连接池，认为这样可以确保数据库访问性能最佳。请问，这种做法正确吗？为什么？
2. 在 MySQL 中，事务中执行的操作都会以预写方式写入 redo log，以备未来事务回滚等操作。请问，写 redo log 是写硬盘操作，会不会导致系统性能下降？为什么？
3. 同学们在设计订单支付服务的 payOrder() 时，在其中调用了两个方法：pay() 从用户账户余额中扣减订单金额；shipping() 生成订单配送信息。同学们希望 pay() 和 shipping() 必需在同一个事务中执行。请问，要怎样设计这三个方法的事务传播属性，以及怎样设计代码才能实现这个目标？

我承诺，我将严格遵守考试纪律。

承诺人：_____

题号	一	二							
得分									
批阅人(流水阅卷教师签名处)									

4. 有的同学认为，即然我们的前端已经是通过 Ajax 异步地发送请求给后端了，那么我们就没有必要像课程作业要求的那样，将下订单功能开发成基于 Kafka 消息队列的方式。请问，你觉得这种想法对不对？为什么？
5. 在多线程编程中，我们经常会看到，在很多代码中用到的是可重入的锁对象。请问，可重入的锁对象是有什么特性的锁对象？我们为什么可能会用到可重入的锁对象？
6. 同学们使用 Redis 缓存来提高系统性能。具体做法是，将所有书籍数据都放入 Redis。在双十一时，当用户生成订单时，先更新 Redis 中书籍的库存量，然后再将库存同步写入数据库。请问，这种使用 Redis 的方式是否合理？为什么？
7. 同学们使用了 JWT 来进行用户认证，在进行认证时，需要在 SSL 协议下传递用户名和密码，进而生成 token。请问，必须使用 SSL 协议传递用户名和密码吗？为什么？

8. 同学们使用搜索引擎 Solr 来建立对图书信息的全文搜索功能。在对搜索到的结果进行排序时，使用了 TF-IDF 算法。有同学认为，该算法本质上就是在计算搜索结果的支持度和置信度。请问，你觉得这种看法对不对？为什么？
9. 同学们将采用微服务架构重构在线书店，将整个在线书店的功能拆分为多个微服务进行单独部署。在部署方案中，还有一个 Gateway 和一个注册中心。请问，注册中心是必需的吗？为什么？
10. 同学们在大二开发的 E-BookStore 系统的基础上，将后端所有的 Controller 都改写为 @RestController，前后端传递的都是 JSON 对象。有的同学认为，这样后端暴露出来的就都是 RESTful Web Services 了。请问，你觉得这种看法正确吗？为什么？

二、 请你帮助同学们解答在配置和使用数据库，并进行系统优化时遇到的下列问题：

1. 大家在设计数据库时，经常会选择遵循第三范式而不是 BCNF。请问，你认为出于怎样的考虑，大家会做出这样的选择？
2. MySQL 的缓存为了避免因为执行了某个很少执行的全表扫描操作而导致大量数据占据缓存无法及时换出的问题，采取了什么样的设计策略来应对？

3. 同学们在设计 MySQL 表时，对于书名字段，应该使用 VARCHAR 还是 CHAR 类型有些分歧。请你告诉同学们，使用 VARCHAR 能带来什么好处和缺点？
4. 同学们决定对 MySQL 数据库执行基于 bin-log 的恢复。Bin-log 是滚动记录的逻辑备份机制，请问，为了防止 bin-log 文件数量过多占据大量存储空间，MySQL 时如何在 bin-log 机制下将全量备份和增量备份机制相结合的？
5. 同学们在对订单表进行分区时，对子分区不是很理解。例如，将订单表分为 4 个区，和分为 2 个区，每个区再分为 2 个子分区相比，后者到底有什么好处？
6. 同学们希望使用 MongoDB 的地理索引(Geo2D)对用户的注册地的位置进行索引，实际上 MySQL 也有类似的索引。请你简述这种索引的工作机制。
7. 同学们打算使用 influxDB 来存储监控到的服务器状态。请你告诉同学们，在 influxDB 中，什么样的数据以 field 存储，什么样的数据以 tag 存储？

8. 同学们构建了 MySQL 集群来提供更好的服务。在集群中，通常是一主多从，写操作只在主节点上完成。但是我们也可以通过设置多主多从来提高数据写入的效率。请问，多主多从会带来哪些负面影响？

9. 同学们打算使用数据湖来存储 E-BookStore 以及其相关系统的所有数据。请你解释一下，为什么在数据湖中存储的都是类似 Parquet 这样的面向列存的数据文件？

10. 同学们打算使用 Hadoop 的 MapReduce 计算范型来执行在线书店的大数据处理功能，例如对订单的统计和查询等。在编写程序时，同学们希望对订单进行过滤，其中订单金额小于 100 元的都过滤掉，只统计大金额的订单金额总额。请问，这种过滤功能应该在哪个类中编写？为什么？