

上海交通大学试卷 (A 卷)

(2021 至 2022 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，同学们设计并开发了一个 BookStore 的升级版，它面向广大的用户，提供书籍在线销售服务。在系统设计与开发的过程中，大家碰到了很多问题，在解决问题的方法上也产生了很多分歧，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. 为了防止单个用户登录系统后长时间不操作导致的安全问题，同学们在后端设计了一个定时器，每当该用户有与网站交互的操作时，就重置定时器。若该用户在定时器时间到达后仍无交互操作，则删除该用户会话信息中的登录状态。请问，这个定时器服务的 Controller 和 Service 类的 Scope 应该如何设置？为什么？
2. 由于在线书店越来越受欢迎，所以按照现有硬件配置，网站长期处于超负荷运行状态，系统性能下降导致用户体验下降。为了提高系统性能，同学们将下订单服务改写为基于异步通信的消息机制实现。请问，你觉得这种策略能够改变系统性能不佳的情况吗？为什么？
3. 同学们在设计订单处理服务的 orderProcess() 时，在其中调用了三个方法：placeOrder() 将订单写入数据库；shipping() 生成订单配送信息；log() 记录日志。同学们希望 placeOrder() 和 shipping() 要么都执行，要么都不执行；同时，log() 是否记录成功不影响订单处理。请问，要怎样设计才能实现这个目标？

我承诺，我将严格遵守考试纪律。

承诺人：_____

题号	一	二	三						
得分									
批阅人(流水阅卷教师签名处)									

4. 在多线程编程中，我们鼓励使用 Immutable Object，即不可变对象，这样就可以避免多线程改写对象属性时造成的状态不一致问题。请问，使用不可变对象时，如果存在更改对象属性的需求，应该如何实现呢？使用不可变对象有什么缺点？
5. 在集群部署方案中，同学们使用了 Redis 缓存来存储 HttpSession 对象，请问，即使不使用 Redis，HttpSession 对象也会存储在内存中，那么为什么要使用 Redis 来存储 HttpSession 对象呢？
6. 同学们在系统中增加了支付功能，调用了包括支付宝和微信支付在内的第三方支付服务，发现它们都是 RESTful 的 Web 服务，请问，它们为什么要设计成 Web 服务？又为什么要设计成 RESTful Web 服务？
7. 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。但是，前端与后端交互时，应该通过单一访问入口来实现，请问，在微服务架构中，如何才能为前端提供单一访问入口？

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. 同学们为了优化 MySQL 的配置，考虑调整 `innodb_buffer_pool` 和 `table_open_cache` 的尺寸。请问，这两个缓存是同一块缓存吗？它们是否有区别？有什么区别？
2. 同学们打算使用第三方工具在 MySQL 上实现快照备份。请问，快照备份是对整个 MySQL 文件系统做物理复制吗？如果是，请说明这样做的好处；如果不是，请说明快照是如何实现的。
3. 因为订单表数量庞大，同学们打算使用分区方式来存储订单表，具体分区方式为使用 RANGE 分区方法按照 `orderid` 分区。请问，你觉得这种分区方式是否合适？为什么？
4. 同学们使用 MongoDB 来存储系统中的一些非结构化数据，例如书的简介、样章和封面图等，对于这些数据，MongoDB 会采用 Sharding 机制做分布式存储。请问，MongoDB 中是如何将一个 Collection 自动 Sharding 后实现分布存储的？
5. 同学们在系统中增加了书评功能，即用户可以对购买的书籍发帖和对他人的书评进行回复，即支持楼上楼的功能。同学们决定使用图数据库存储书评数据，但是有些同学不理解为什么不能使用 MySQL 来存储这些数据。请你告诉这些同学，为什么这类数据更适合用图数据库存储？

6. 同学们在研究是否要使用日志结构合并树类型的数据库来存储订单, 同学们认为既然这种数据库不存在空间放大问题, 那么为什么还存在写放大问题呢? 请你回答这个问题, 帮助同学们解惑。
7. 无论是日志结构合并树类型的数据库, 还是时序数据库, 都包含一个 Write Ahead Log, 请问, 这个 Log 的作用是什么?

三、 请你帮助同学们解答在系统部署和运维, 以及系统优化时遇到的下列问题:

1. 同学们使用集群方式部署在线书店, 并通过 Nginx 实现负载均衡, 为了确保维护会话状态, 采用了 ip hash 的策略。请问, 这种策略的缺点是什么?
2. 同学们决定使用 Docker 来实现在线书店的轻量级虚拟化部署, 在使用 Docker 时, 同学们认为应该使用 named volume 来持久化在线书店的数据。请你告诉同学们 named volume 的实现原理。
3. 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。但是有的同学反对, 认为它是一种读写硬盘操作密集的类型。请问, 你认为 Hadoop 的 MapReduce 是一种频繁读写硬盘的操作吗? 为什么?

4. 同学们准备使用 Spark 来实现订单统计功能,以提升并行计算的效率。有的同学认为,Spark 的 RDD 不能修改,那么在内存中就会创建出很多 RDD,这样很浪费内存。请问,你觉得这种说法对不对?为什么?

5. 同学们在使用 HDFS 存储文件时,设置了 Block 副本数量为 3。请问,当有程序需要读取该 Block 的文件数据时,HDFS 的处理逻辑是如何确定应该访问哪个副本的?

6. 同学们对 HBase 中 ColumnFamily 的作用不是很理解,请你告诉同学们,将数据表分解为 ColumnFamily 的优点是什么?