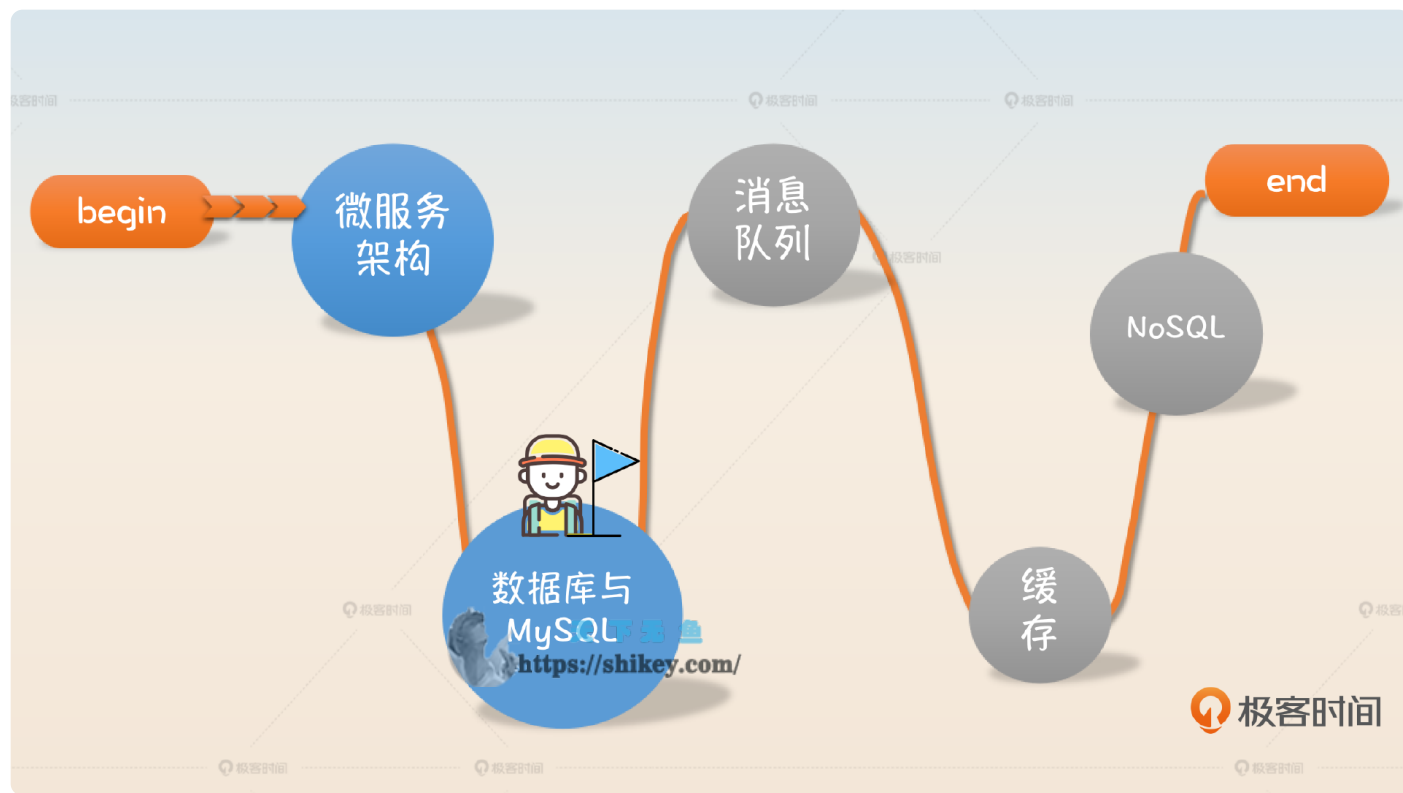


# 模拟面试 | 数据库面试思路一图懂

邓明 · 后端工程师的高阶面经



你好，我是大明，恭喜你学完第二章的内容，又到了要验收成果的时刻了。数据库这一章的内容很多，而且知识点都很重要，所以为了帮助你更好地掌握这部分内容，我们在这里设置了面试题。

你在回答的时候，最好是能够写成一个文档，至少也要口头上说一遍。千万不要仅仅在脑海里回忆一遍。因为在真正面试的时候，脑海中的记忆到嘴里说出的话，还需要一个转换。

## 🔗 10 | 数据库索引：为什么 MySQL 用 B+ 树而不用 B 树？

1. 什么是覆盖索引？
2. 什么是聚簇索引 / 非聚簇索引？
3. 什么是哈希索引？MySQL InnoDB 引擎怎么创建一个哈希索引？
4. 什么回表？如何避免回表？
5. 树的高度和查询性能是什么关系？

6. 什么是索引最左匹配原则?
7. 范围查询、Like 之类的查询怎么影响数据库使用索引?
8. 索引是不是越多越好?
9. 使用索引有什么代价?
10. 如何选择合适的索引列? 组合索引里面怎么确定列的顺序? 状态类的列是否适合作为索引的列?
11. 为什么 MySQL 使用 B+ 树作为索引的数据结构? 为什么不用 B 树? 为什么不用红黑树? 为什么不用二叉平衡树? 为什么不用跳表?
12. NULL 对索引有什么影响?
13. 唯一索引是否允许多个 NULL 值?

## 🔗 11 | SQL 优化: 如何发现 SQL 中的问题?

<https://shikey.com/>

1. 请你解释一下 EXPLAIN 命令。
2. 你有优化过 SQL 吗? 具体是怎么优化的?
3. 你有没有优化过索引? 怎么优化的?
4. 怎么优化 COUNT 查询?
5. 怎么优化 ORDER BY?
6. 怎么优化 LIMIT OFFSET 查询?
7. 为什么要尽量把条件写到 WHERE 而不是写到 HAVING 里面?
8. 怎么给一张表添加新的索引 / 修改表结构? 如果我的数据量很大呢?
9. USE INDEX/FORCE INDEX/IGNORE INDEX 有什么效果?

## 🔗 12 | 数据库锁: 明明有行锁, 怎么突然就加了表锁?

1. 什么是行锁、表锁? 什么时候加表锁? 怎么避免?
2. 什么是乐观锁? 怎么在 MySQL 里面实现一个乐观锁?

3. 什么是意向锁？可以举一个例子吗？
4. 什么是共享锁和排它锁？它们有什么特性？
5. 什么是两阶段加锁？
6. 什么是记录锁、间隙锁和临键锁？
7. RC 级别有间隙锁和临键锁吗？
8. MySQL 是怎么在 RR 级别下解决幻读的？
9. 什么情况下会加临键锁？什么情况下会加间隙锁？什么时候加记录锁？
10. 唯一索引和普通索引会怎么影响锁？
11. 你遇到过什么死锁问题吗？怎么排查的？最终又是怎么解决的？
12. 你有没有优化过锁？怎么优化的？



天下无鱼

<https://shike.com/>

## 🔗 13 | MVCC 协议：MySQL 修改数据时，还能不能读到这条数据？

1. 什么是 MVCC？为什么需要 MVCC？
2. 什么是隔离级别？隔离级别有哪几种？
3. 什么是脏读、不可重复读、幻读？它们与隔离级别的关系是怎样的？
4. 隔离级别是不是越高越好？
5. 你们公司用的是什么隔离级别？为什么使用这个隔离级别？能不能使用别的隔离级别？
6. 你有没有改过隔离级别？为什么改？

## 🔗 14 | 数据库事务：事务提交了，你的数据就一定不会丢吗？

1. 什么是 undo log？为什么需要 undo log？
2. 什么是 redo log？为什么需要 redo log？
3. 什么是 binlog？它有几种模式？用来做什么？
4. 事务是如何执行的？

5. 什么是 ACID? 隔离性和隔离级别是什么关系? 你觉得哪个隔离级别满足这里的隔离性要求?
6. redo log 的刷盘时机有哪些? 该如何选择? 你们公司用的是哪个配置? 为什么用这个配置?
7. binlog 的刷盘时机有哪些? 该如何选择? 你们公司用的是哪个配置? 为什么用这个配置?
8. 我的事务提交了, 就一定不会丢吗? 怎么确保一定不会丢?
9. 什么是 page cache? 为什么不直接写到磁盘?
10. 在分布式环境下, 当服务器告诉我写入成功的时候, 一定写入成功了吗? 如果服务器宕机了可能发生什么?

## 🔗 15 | 数据迁移: 如何在不停机的情况下保证迁移数据的一致性?

1. 你们单库拆分的时候是如何做数据迁移的? 你们修改大表结构的时候是怎么做数据迁移的? 怎么在保持应用不停机的情况下做数据迁移?  
<https://shikey.com/>
2. 什么是双写? 为什么要引入双写?
3. 如果双写的过程中, 有一边写失败了, 怎么办?
4. 你可以用本地事务来保证双写要么都成功, 要么都失败吗? 分布式事务呢?
5. 为什么有一个阶段是双写, 但是以目标表为准? 干嘛不直接切换到单写目标表?
6. 你们有什么容错方案? 比如说如果在迁移过程中出错了, 你们的应用会怎么办?
7. 你们是怎么校验数据的?
8. 增量数据校验你们是怎么做的?
9. 数据迁移你能够做到数据绝对不出错吗?
10. 如果数据出错了你们怎么修复? 怎么避免并发问题?
11. 让你迁移一个 2000 万行的表, 你的方案大概要多久?
12. 你用过 mysqldump/XtraBackup 吗? 它有什么缺点?

## 🔗 16 | 分库分表主键生成: 如何设计一个主键生成算法?

1. 你们分库分表怎么生成主键的？
2. 使用 UUID/ 数据库自增 / 雪花算法有什么优缺点？
3. 雪花算法是如何实现的？
4. 雪花算法是怎么做到全局唯一的？
5. 怎么解决雪花算法的序列号耗尽问题？
6. 怎么解决雪花算法的数据堆积问题？
7. 你有没有优化过主键生成的性能？怎么优化的？效果如何？
8. 你的主键生成的 ID 是严格递增的吗？不是递增有什么问题？
9. 为什么我们一般使用自增主键？
10. 什么是页分裂？有什么缺点？



天下无鱼

<https://shikekey.com/>

## 🔗 17 | 分库分表分页查询：为什么你的分页查询又慢又耗费内存？

1. 你们公司是怎么解决分页查询的？平均查询性能如何？
2. 为什么分页查询那么慢？
3. 全局查询有什么优缺点？对于一个查询 LIMIT X OFFSET Y 来说，如果我命中了三张表，会取来多少数据？
4. 怎么提高分页查询的速度？
5. 什么是二次查询？它的步骤是什么样的？
6. 怎么在二次查询里面计算全局的偏移量？
7. 二次查询有什么优缺点？
8. 代理形态的分库分表中间件有什么优缺点？怎么解决或者改进它的缺点？
9. 使用中间表来进行分库分表，有什么优缺点？怎么设计中间表？
10. 在使用中间表的时候，你怎么保证数据一致性？你能保证强一致吗？如果不能，不一致的时间最差是多久？
11. 你们公司有没有考虑使用别的中间件来解决分页查询？你选择哪一个？为什么？

## 🔗 18 | 分库分表事务：如何同时保证分库分表、ACID 及高性能？

1. 你们公司在分库分表之后，如何解决事务问题？
2. 什么是两阶段提交协议？有什么缺点？
3. 什么是三阶段提交协议？相比两阶段，改进点在哪里？
4. 什么是 XA 事务？
5. 你觉得 XA 事务是否满足 ACID？为什么？
6. 什么是 TCC？
7. 什么是 SAGA？
8. 什么是 AT 事务？
9. 什么是延迟事务？延迟事务失败了怎么办？为什么分库分表中间件喜欢用延迟事务？
10. 你们公司是否允许跨库事务？为什么？有什么场景是必须要使用跨库事务的？

## 🔗 19 | 分库分表无分库分表键查询：按买家分库分表，卖家怎么查？

1. 你们公司的分库分表是怎么分的？一般情况下怎么选分库分表键？
2. 假设说现在我的订单表是按照买家 ID 来分库分表的，现在我卖家要查询，怎么办？
3. 利用中间表来支持无分库分表键查询的时候，怎么设计中间表？
4. 为什么在买家分库分表的时候，按照 4832，但是同样的数据，按照卖家分库分表的时候，就只需要按照 2816？
5. 广播有什么缺点？
6. 可以使用什么中间件来支持复杂查询？你们公司用了什么？

## 🔗 20 | 分库分表容量预估：分库分表时怎么计算所需库、表的数量？

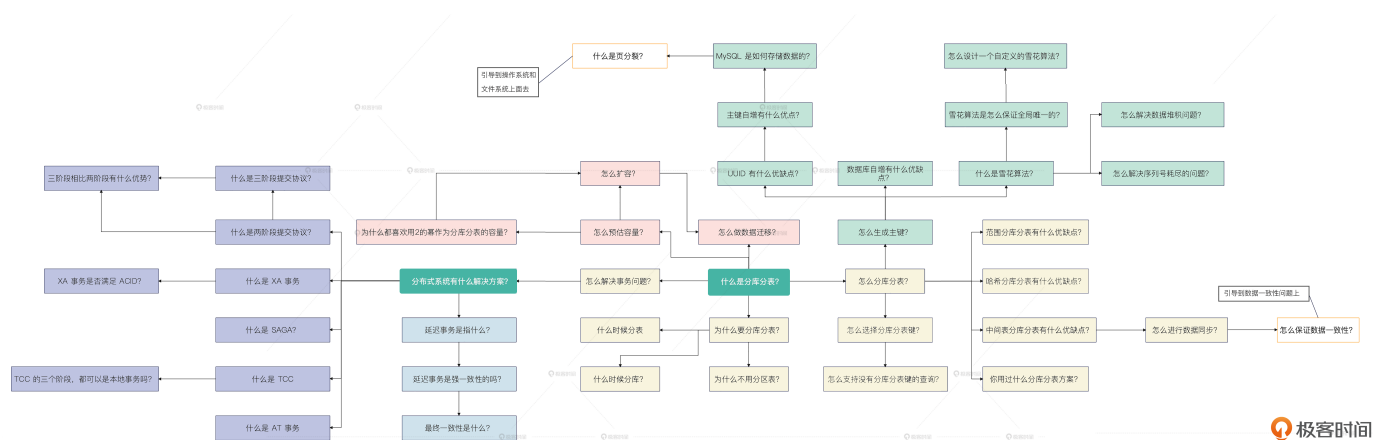
1. 你是怎么估计容量的？考虑了什么因素？
2. 你怎么知道数据未来增长会有多快？
3. 你这容量是预估了几年的数据量？为什么？

4. 你是怎么利用流量录制和重放来验证数据的?
5. 在流量录制之后, 重放之前, 如果数据修改了, 你的数据校验还能正常运行吗?
6. 你公司用的是 HTTPS 协议吗? 使用 HTTPS 协议你怎么录制流量?
7. 为什么大家都喜欢用 2 的幂来作为容量?
8. 怎么扩容? 有哪些步骤?
9. 如果你发现之前分库分表太多了, 能不能缩容? 假如要你缩容, 你怎么办?

## 21 | 数据库综合应用: 怎么保证数据库高可用、高性能?

1. 你们有没有做过数据库优化? 有没有做过 InnoDB 引擎优化?
2. 你调过什么数据库相关的参数? 为什么要调?
3. InnoDB 引擎的 buffer pool 是拿来做什么的? 怎么优化它的性能?
4. buffer pool 是不是越大越好? 过大或者过小都有什么问题? 怎么确定合适的大小?
5. 数据库里面有很多刷盘相关的参数, 你都了解吗? 调过吗? 根据什么来调?
6. 你有没有做过主从分离? 主从延迟是什么? 怎么解决主从延迟?
7. 你们公司的数据库主节点宕机了会发生什么?
8. 什么是查询缓存? 你们公司有没有用查询缓存?

## 数据库面试一图懂



最后我将这节课的内容整理成了一张图片，图片能够更好地展现知识点之间的联系，你可以参考，如果你觉得有用的话，可以把这节课的内容分享给身边的朋友，邀他一起学习，共同进步。好了，我们下一章再见吧！

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

## 全部留言 (2)

最新 精选



**peter**

来自北京

请教老师两个问题：

Q1：实际案例：误删生产数据，怎么回复？

我加了一个群，群里一个人周五(8月4号)说：一个用户的28条数据都让我改了，mysql事务已经提交的数据怎么把找个事务对数据的修改恢复到没提交之前的？

Q2：阿里这种级别的公司，其入口是怎么做的？

用户访问一个网站，肯定是先到这个网站的一个入口，再由此入口转发给后端的服务器。网站最外面的入口，是怎么实现的？现有的外部网关，比如F5、Nginx、LVS等，处理速度都是有限的，对于阿里这种规模的公司，假如只有一个入口，肯定是处理不过来的。那怎么解决这个问题？

作者回复：Q1：找不回来了，你只能说找到历史的备份数据来恢复。

Q2：这种是通过边缘节点接入的，已经超过我们这个专栏的范畴了，我对此研究不多，因为没做过相关的事情。



👍 1



**学无涯**

来自北京

接上面同学的问题，有几个点想要请教老师：

1. 误操作数据后，可能一段时间后才发觉，这时可能有大量binlog，如何快速找到自己误操作的那几条binlog，只能手动一条条binlog的找吗？
2. 找到binlog时，可能已经有新的操作修改了这条数据，这种情况下该怎么办？
3. 老师能否出一节关于误删数据如何补救的标准流程操作



作者回复: 坦白说, 我没遇过这种误删除, 然后要借助历史 binlog 来恢复的场景。一方面来说, 之前待的公司都是软删除, 并没有实际的硬删除。

不过之前遇到过误更新的问题, 但是当时我们是通过别的业务数据来推断出来正确的值, 修复了数据。你也可以考虑借助类似的手段。

2 这种情况, 要是更新的操作已经把数据更新对了, 就不用管了。

